



# The quality of systematic reviews and other synthesis in the time of COVID-19

[cambridge.org/hyg](https://cambridge.org/hyg)A. Baumeister<sup>1</sup> , T. Corrin<sup>1</sup>, H. Abid<sup>1</sup>, K. M. Young<sup>2</sup>, D. Ayache<sup>1</sup>  
and L. Waddell<sup>1</sup> 

## Review

**Cite this article:** Baumeister A, Corrin T, Abid H, Young KM, Ayache D, Waddell L (2021). The quality of systematic reviews and other synthesis in the time of COVID-19. *Epidemiology and Infection* **149**, e182, 1–8. <https://doi.org/10.1017/S0950268821001758>

Received: 29 March 2021

Revised: 17 June 2021

Accepted: 19 July 2021

**Key words:**

COVID-19; knowledge synthesis; systematic review; meta-analysis; SARS-CoV-2

**Author for correspondence:**

A. Baumeister,

E-mail: [austyn.baumeister@canada.ca](mailto:austyn.baumeister@canada.ca)

<sup>1</sup>Public Health Risk Sciences Division, National Microbiology Laboratory, Infectious Disease Prevention and Control Branch, Public Health Agency of Canada, Guelph, ON, Canada and <sup>2</sup>Food-borne Disease and Antimicrobial Resistance Surveillance Division, Centre for Food-borne, Environmental and Zoonotic Infectious Diseases, Infectious Disease Prevention and Control Branch, Public Health Agency of Canada, Government of Canada, Guelph, ON, Canada

**Abstract**

COVID-19 research has been produced at an unprecedented rate and managing what is currently known is in part being accomplished through synthesis research. Here we evaluated how the need to rapidly produce syntheses has impacted the quality of the synthesis research. Thus, we sought to identify, evaluate and map the synthesis research on COVID-19 published up to 10 July 2020. A COVID-19 literature database was created using pre-specified COVID-19 search algorithms carried out in eight databases. We identified 863 citations considered to be synthesis research for evaluation in this project. Four-hundred and thirty-nine reviews were fully assessed with A Measurement Tool to Assess systematic Reviews (AMSTAR-2) and rated as very low-quality ( $n = 145$ ), low-quality ( $n = 80$ ), medium-quality ( $n = 208$ ) and high-quality ( $n = 151$ ). The quality of these reviews fell short of what is expected for synthesis research with key domains being left out of the typical methodology. The increase in risk of bias due to non-adherence to systematic review methodology is unknown and prevents the reader from assessing the validity of the review. The responsibility to assure the quality is held by both producers and publishers of synthesis research and our findings indicate there is a need to equip readers with the expertise to evaluate the review conduct before using it for decision-making purposes.

**Introduction**

Since the rapid emergence of the SARS-CoV-2 virus responsible for the COVID-19 pandemic, there has been a need for timely and accurate evidence and evidence summaries to aid in decision-making. This unprecedented pandemic, the novel nature of the virus and the global reach of COVID-19 have resulted in high demand for evidence to inform a large range of public health, healthcare and economic decisions. There has been a higher rate of evidence production during this pandemic than ever before and with that volume, is the challenge of identifying and using quality information [1]. The term infodemic has generally been applied to issues of misinformation for the general public, however has also been used to describe the sheer volume of evidence produced for clinicians and policymakers [1]. Given the volume of research that has been and continues to be produced, reliable syntheses of research are essential.

Synthesis research encompasses a suite of tools for summarizing the primary literature using systematic and reproducible methodologies such as systematic reviews (SR) and meta-analyses (MA), scoping reviews (ScR) and rapid reviews (RR). When appropriately conducted, synthesis research should identify all relevant research, appraise the quality and limitations of the research, and then synthesise information in a way that adheres to a pre-formulated strategy that is designed to minimise bias. Within this category of research, there exists a variance in methods for different types of reviews, and conduct and reporting guidelines are available and widely accepted for each that should be adhered to [2, 3].

The Cochrane review handbook suggests that a review may take a year to 2 years to complete; however, in just a matter of months, there have been hundreds of systematic reviews published on COVID-19 [2]. Given the compressed timelines needed to respond to the current pandemic, we sought to evaluate whether synthesis research was being conducted with the appropriate rigour and objective reporting expected based on the methodology or whether synthesis research methods were being undermined. Several research groups have raised concerns about the contribution of poor quality synthesis to the deluge of the COVID-19 evidence base. Here we evaluated how the need to rapidly produce syntheses has impacted the quality of the synthesis research produced during the early part of the COVID-19 pandemic, by systematically quantifying the type, topic and quality of synthesis research on COVID-19 topics.

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

**CAMBRIDGE**  
UNIVERSITY PRESS

## Methods

### Protocol

An *a priori* protocol was developed and is available in the Supplementary file. The protocol details the methods and tools used in this project, including important definitions, search algorithms and screening strategies.

### Objective

The objective of this study was to evaluate and characterise the synthesis research on COVID-19 to understand its strengths and weaknesses and develop a database where high-quality syntheses on key topics could be easily identified.

### Search strategy

Synthesis research was identified during the daily scan of COVID-19 literature, maintained by the Public Health Agency of Canada since 4 February 2020 and backdated until January 2020. The daily scan retrieves relevant COVID-19 literature from the following databases; PubMed, Scopus, bioRxiv, medRxiv, arXiv, SSRN, Research Square and COVID-19 information centres; Lancet, BMJ, Elsevier, Nature and Wiley. The keywords (COVID-19 OR SARS-CoV-2 OR SARS-Coronavirus-2 OR nCov OR 'novel CoV' OR (novel AND coronavirus)) were adapted for each database and were run daily. No language restrictions were applied to the search. To identify synthesis research within this database, an artificial intelligence classifier built-in DistillerSR (Evidence Partners<sup>®</sup> 2020) automatically classified articles as synthesis research.

As part of the daily COVID-19 literature scan, human reviewers verified the classification of these reviews based on the citation. For this project, synthesis research identified up to and including 10 July 2020 was included. The full protocol can be found in the Supplementary file.

### Inclusion criteria

#### i) Study type

Review articles were included if they identified or described a form of synthesis research (e.g. systematic reviews (SR)/meta-analyses (MA), etc.). All forms of synthesis research were considered including those not labelled using standard nomenclature, but otherwise had a clear methodology and robust and reproducible search strategy.

#### ii) All COVID-19 topics

Synthesis research was included if the review was relevant to the COVID-19 pandemic.

### Exclusion criteria

#### i) Language

Reviews were excluded if they were not written in English or French due to resource constraints.

#### ii) Topic

Reviews were excluded if the review did not contain relevant information for the COVID-19 pandemic.

#### iii) Synthesis quality

Reviews were excluded as very low quality if the review did not meet any of the minimum A Measurement Tool to Assess systematic Reviews (AMSTAR-2) criteria for further evaluation and characterisation:

Did not have an explicit research question where the components of a PICO/PECO (population, intervention/exposure, control, outcome question), where applicable, were well defined.

- Did not explicitly report that there was a protocol registered beforehand or otherwise available that contained a review question, search strategy, inclusion/exclusion criteria, and risk of bias assessment (only for systematic reviews/meta-analyses), synthesis/meta-analysis plan, investigation of heterogeneity plan and justification for deviations from the protocol; or reported methods that were comparable to a protocol and did not explicitly report that methods were created before the review started.
- Did not have a robust or reproducible search strategy (e.g. searched fewer than two databases or did not provide the search algorithm).

### Management of the study

The study was conducted using the web-based systematic review software program, DistillerSR (Evidence Partners<sup>®</sup> 2020). The screening, quality assessment and study characterisation were conducted within DistillerSR using *a priori* developed and pre-tested tools. After a pilot test, each citation was assessed by a single reviewer and a senior reviewer audited a random selection of completed assessments to check for quality and consistency across the review team.

### Study screening

The full texts of articles tagged as synthesis research were assessed for this project based on the inclusion and exclusion criteria by TC, HA, KY, DA, MY, RA, LW and AB. The selection criteria were extensively pre-tested with the team and netted good agreement ( $\kappa > 0.8$ ). For the remaining articles assessed, 20% were verified by a second senior reviewer.

### Study characterisation

The data collection form was pre-tested by reviewers TC, HA, KY, DA, MY, RA, LW and AB. Data extracted to characterise the review included the type of review, whether the review was published or a pre-print, and the basic details of the review (e.g. populations, topic and outcomes available). Results were not extracted as the goal was not to synthesise research findings, but rather to create an encompassing inventory of COVID-19 synthesis research, searchable by topic and overall quality. The screening and data extraction tools can be found in the Supplementary file.

### Quality of synthesis research

The quality of synthesis research was assessed using the AMSTAR-2 quality assessment tool for systematic reviews that include randomised and/or non-randomised studies of healthcare interventions [4]. The AMSTAR-2 tool covers 16 domains to

systematically and comprehensively assess the quality of reviews based on key methodological components of synthesis research. The three questions listed under exclusion criteria were considered the minimum necessary components for any synthesis research. Not meeting any of these criteria meant the review as a synthesis research product was very low quality and was excluded from further data extraction. The remaining 13 questions were used to further assess the quality and adherence to synthesis research methodologies followed in each review. Based on deficiencies in the review, an overall rating of low, medium and high quality was applied as follows: reviews could be downgraded 1.0 point for a full failure in a domain or a 0.5-point downgrade for a partial yes response. The low category encompassed reviews with 3.5 or more downgrades, medium >1.0 and ≤3.0, and high-quality reviews had ≤1.0 downgrades. The AMSTAR-2 tool can be found in the Supplementary file.

## Results

As of 10 July 2020, 863 reviews were identified as synthesis research from the daily scan of COVID-19 literature. Figure 1 shows a flow chart of reviews that were excluded and for which reasons. Of these studies, 499 were published at the time of review and 363 were still in pre-print format, and one review could not be located. Most synthesis research was classified as systematic reviews which included a meta-analysis ( $n = 235$ ), followed by systematic reviews without a meta-analysis ( $n = 233$ ), rapid review with meta-analysis ( $n = 17$ ), rapid reviews without a meta-analysis ( $n = 53$ ), meta-analysis on its own ( $n = 38$ ), scoping reviews ( $n = 33$ ) and umbrella reviews ( $n = 7$ ) (Fig. 2). One hundred and ninety-two reviews were excluded because they were not synthesis research and 35 were excluded because they were review protocols. Nineteen others used non-standard review labels (e.g. clinical reviews, bibliometric analyses, state-of-the-art reviews), but otherwise utilised recognisable synthesis research methods.

In total, 278 articles were excluded by one or more of the exclusion criteria: the review was not a type of synthesis research, was not relevant to the topic, or were published in a language other than French or English. Overall, 584 reviews that were labelled as systematic review and/or a meta-analysis, rapid review, scoping review or an umbrella review, were in English or French, and on a COVID-19 topic were further assessed for eligibility using the minimum criteria for synthesis research. This resulted in the exclusion of 145 reviews, each having at least one failure of the categories; unclear research question (2% of reviews), protocol that was specifically not developed beforehand or methods that were otherwise not laid out well enough to follow (22% of reviews), or lack of robust search methods (8% of reviews). Any single failure or combination of failures was thus deemed to be of very low quality and excluded. Four hundred and thirty-nine reviews were then fully assessed with the AMSTAR-2 tool.

## Results of the quality assessment

There were 439 reviews fully assessed using the AMSTAR-2 tool and further characterised. There were 151 (26%) high-quality reviews, 208 (36%) medium-quality reviews, 80 (14%) low-quality reviews and 145 (25%) were excluded from characterisation as they were very low quality (Table 1). AMSTAR-2 criteria were frequently not met across studies including 81 (18%) that did not identify the types of included study designs. Most reviews at least partially explained what studies were excluded and for

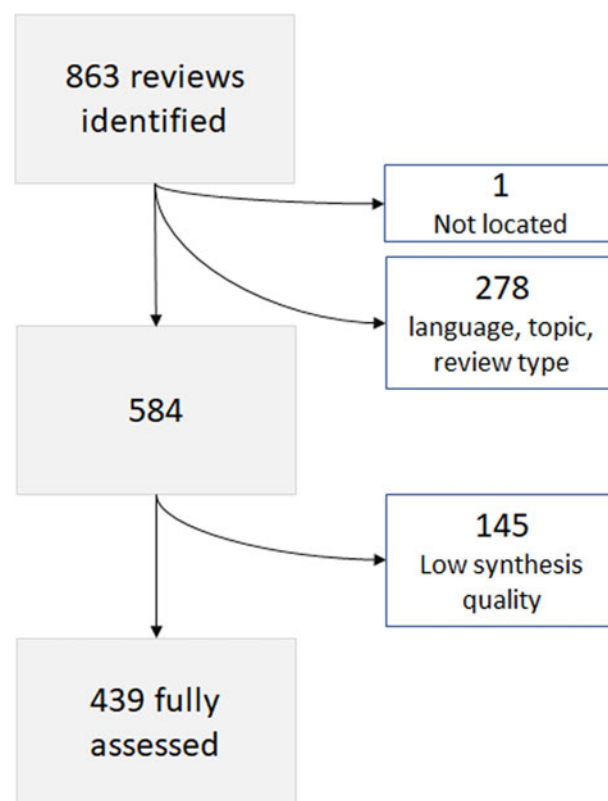


Fig. 1. Flow diagram of the selection of studies.

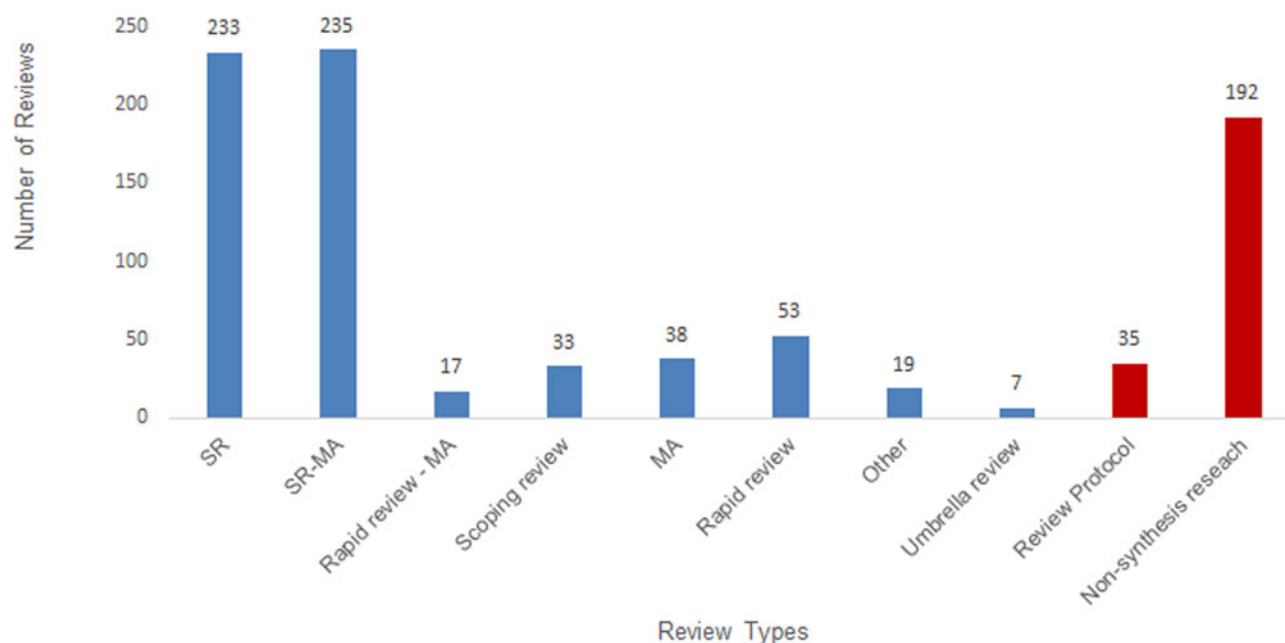
what reasons (79.7%), with 15% failing to report excluded studies, and 5% of reviews listing why each study was excluded. Just over half the reviews (58.3%) described the included studies in good detail, with another 28.7% describing the studies at an adequate level (describing populations, interventions, comparators, outcomes and research designs). Most studies did conduct some risk of bias assessment (69.0%), however a wide range of tools were employed (e.g. Cochrane Risk of Bias (ROB), Newcastle-Ottawa Scale and Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool). Of the 416 studies that conducted risk of bias assessment (195 studies included a meta-analysis), only nine studies (2.2%) reported the Grading of Recommendations, Assessment Development and Evaluations (GRADE) to clarify the strength of associations for particular outcomes. In terms of the reviewing process, the relevance screening and data extraction steps were completed in duplicate (i.e. independently by two reviewers) in only 104 (23.7%) and 118 (26.9%) reviews, respectively.

Few reviews (3.9%) reported on the source of funding for the primary studies included. Overall, 90.4% of reviews included a funding statement or acknowledgement of the funding of their research. The risk of bias results and implications were discussed in less than half of the reviews (48.7%). Heterogeneity and its possible implications were discussed in the results or discussion in approximately 67% of the reviews.

Meta-analysis was conducted in 247 reviews and was considered methodologically appropriate in most studies (94.7%). Just under half the studies did not explore the risk of bias influence on heterogeneity in their meta-analysis (45.3%). Publication bias was assessed in 160 reviews (65%).

In our evaluation, there was no statistical difference between published and unpublished review quality categories (low compared to medium and high).

## Author Reported Review Type



**Fig. 2.** Number of reviews for each author reported review type, red denotes excluded reviews.

**Table 1.** Summary table of responses to the AMSTAR-2 questions of 584 synthesis research articles, 145 were considered to not meet minimum criteria due to a lack of research question, methods or search strategy fully

Response	Yes	Partial yes	No	N/A
Initial screening AMSTAR-2 questions (N = 584 assessed)				
Research question	572	-	12	-
Protocol/adequate methods	118	339	127	-
Search strategy	263	273	48	-
Quality assessment questions (N = 439 assessed)				
Study design selection	358	-	81	-
Inclusion duplicate	335	-	104	-
Extraction duplicate	321	-	118	-
Listed exclusions	24	350	65	-
Sufficient detail	256	126	57	-
RoB assessment	288	15	113	23
Primary study funding	17	-	422	-
Meta-analysis method	234	-	13	192
Meta-analysis RoB	135	-	112	192
RoB discuss	203	-	213	23
Heterogeneity investigation	294	-	145	-
Publication bias	160	-	96	183
Conflict statement	397	-	42	-

N/A, not applicable; -, not a response option.

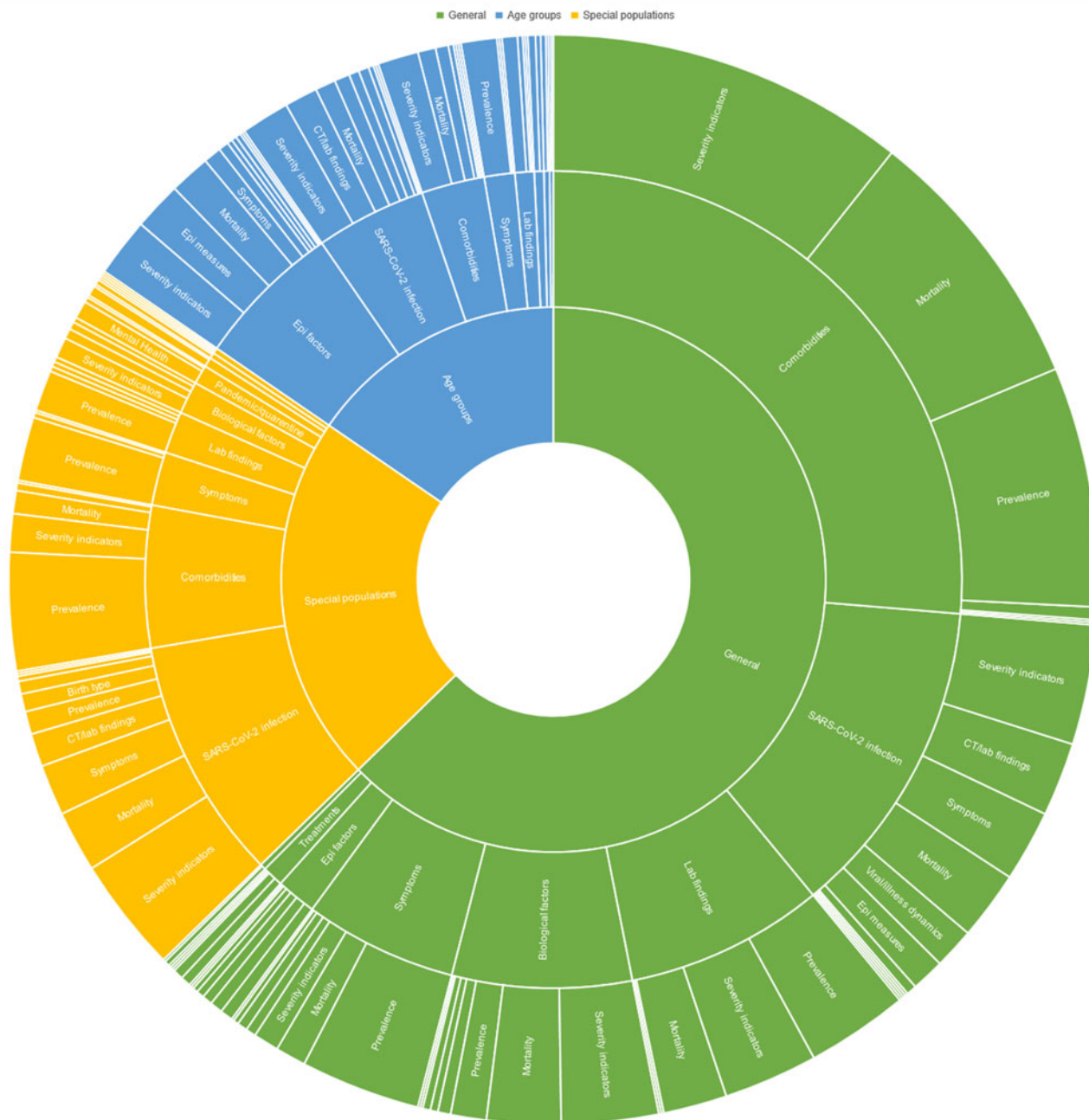
### Topic areas covered by reviews

Topic areas covered in 439 reviews were broadly categorised into three main groups: prognosis and epidemiological parameter studies (70%), studies of interventions (antiviral treatments and non-pharmaceutical interventions) (25%), and studies of guidelines and methods (diagnostic, surgical recommendations and support methods) (5%). Within these categories, populations, risk factors, treatments and outcomes varied greatly. A full list of ungrouped topics is in the Supplementary file.

In the prognosis and epidemiological category ( $n = 287$  reviews), a total of 126 unique combinations of populations, risk factors and outcomes were recorded (1322 total entries) (Fig. 3). Reviews were primarily focused on the general population or did not specify the population (63%) and frequently considered the relationship between comorbidities and severity indicators (40% of reviews that considered comorbidities) or mortality (30% of reviews that considered comorbidities). Special populations were the focus of 22% of the prognosis reviews and included pregnant women, and hospitalised patients and specific age group categories. Reviews with narrow age groups most often looked at epidemiological associations with the severity of COVID-19 and accounted for 15% of reviews.

Nearly all reviews that captured intervention outcomes ( $n = 124$  reviews) focused on antiviral treatment (92%) in either general populations, severe cases or in another special population (e.g. pregnancy, cancer or transplant) (Fig. 4). The next largest slice of reviews was on non-pharmaceutical studies (7%). One hundred and twenty-eight unique combinations were recorded with 466 total entries.

In the guideline and method category ( $n = 68$  reviews), there were 94 total entries with 37 unique combinations of methods, contexts and outcomes (Fig. 5). Reviews were focused on diagnostic



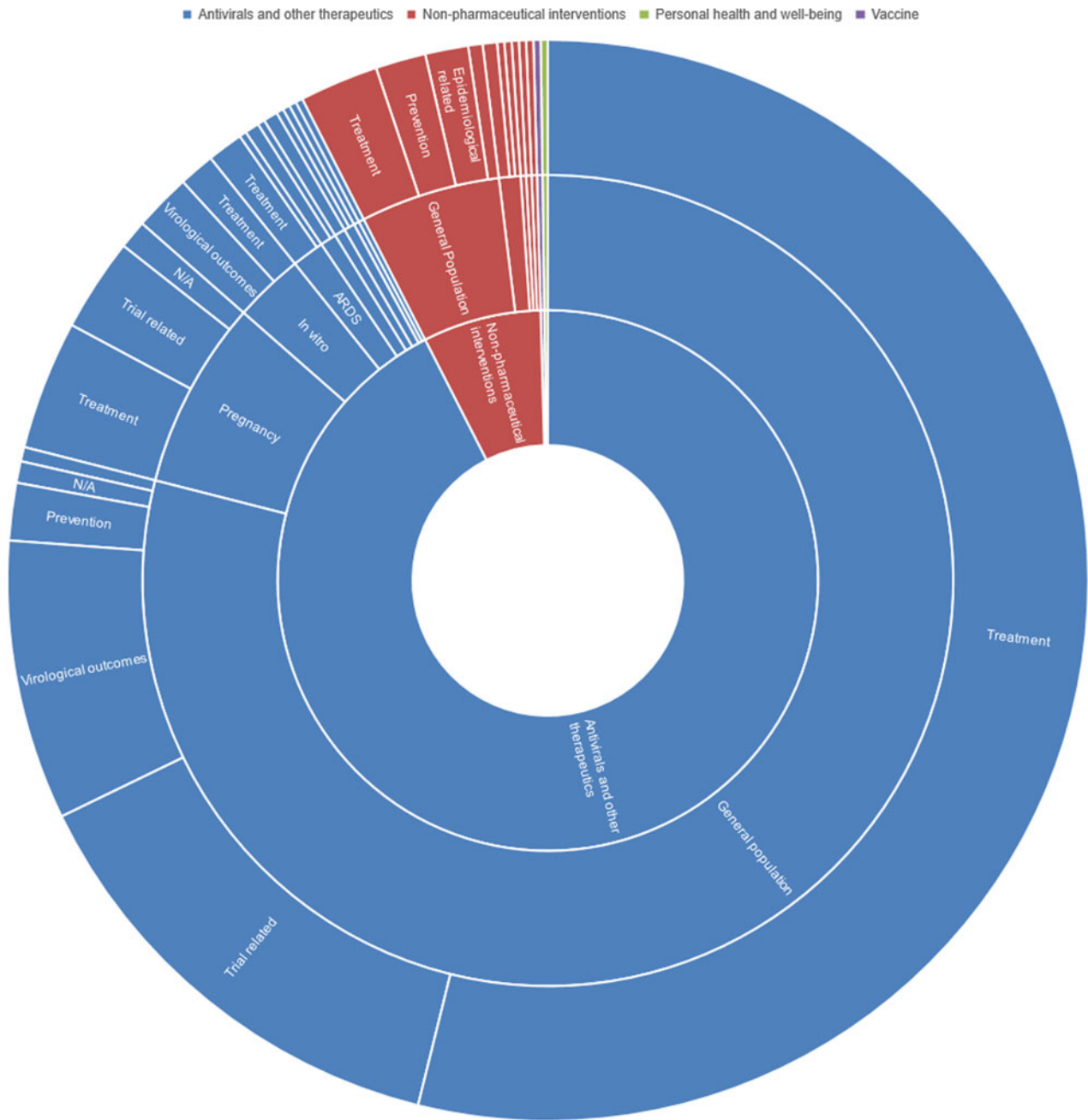
**Fig. 3.** Prognosis and epidemiological parameter topics.

methods in the context of patient care and diagnostic accuracy (44%), followed by guidelines for surgery, support of healthcare workers, patient care and research/management (39%), followed by healthcare solutions (12%), and predictive modelling (3%). Other review topic areas were studied in solitary reviews.

### Discussion

The novel nature of the SARS-CoV-2 virus and its spread has necessitated the building of fundamental knowledge from the ground up. Researchers around the world have worked quickly and tirelessly to produce high-quality primary evidence. Synthesizing these research findings is an important component

of decision-making. However, the speed at which primary and review studies have been published has raised concerns about the quality of research being produced [5]. In this study, we evaluated the quality of the synthesis literature produced between January 2020 and 10 July 2020, the first 5 months of the pandemic. Evaluation of 862 reviews reveals that there are issues with identifying and following a synthesis methodology. When one is chosen, critical steps were often omitted without a clear description of shortcuts taken, why they were taken and the implications to the review results. The poor methodological quality of many reviews means critical assessment of synthesis research is required as a matter of course, given the low barriers to conducting reviews and variable standards for publishing reviews. We did



**Fig. 4.** Intervention topics.

not find a difference between the quality of published compared to pre-print reviews, which indicates that addressing review quality in the published literature should be addressed by researchers and journal editors.

The cause of non-adherence to standard synthesis research methodology during the COVID-19 pandemic is not readily apparent, but there are a few plausible reasons. One possible reason is the compressed timeline of the production and publication of research. The Cochrane handbook, for instance, reports that reviews may take up to 2 years and a recent publication cited the average environmental systematic review took 164 days [2, 6]. Given that these reviews were all produced within 6 months after the discovery of the SARS-CoV-2 virus, they have been

done quickly, often within the span of a couple of weeks to a month. High demands for COVID-19 research may have compounded issues with methodological rigour by the pressure on the peer review process which allowed studies that were not appropriately labelled or did not follow the rigorous methodology expected for synthesis research to be published.

Most systematic or rapid reviews in the first 6 months of the pandemic were focused on summarizing prognostic studies to better understand COVID-19 disease. This is likely a reflection of the novel nature of the SARS-CoV-2 virus, the speed at which it spread around the world and the severity of the disease. There were fewer reviews on epidemiological parameters such as incubation time, length of the infectious period or long-term

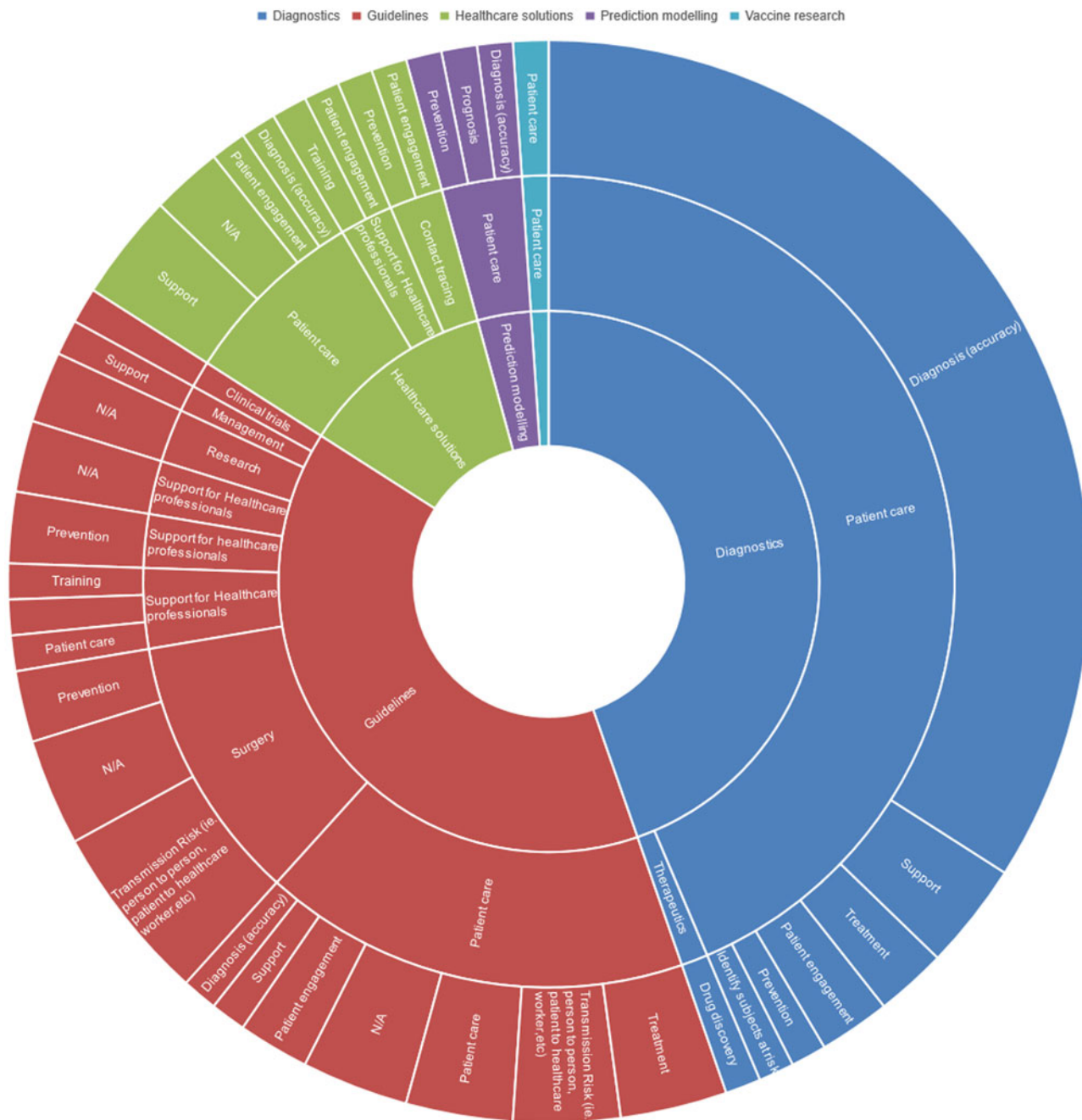


Fig. 5. Guideline topics.

immune response likely due to a lack of primary literature early in the pandemic. Similarly, data regarding treatments reflected the use of repurposed antivirals evaluated through observational studies because randomised controlled trials were not yet completed. Diagnostic accuracy reviews were also common as molecular and serological tests for COVID-19 were developed early on in the pandemic. Whether a topic is represented by one or many reviews, developing a database with synthesis research rated for quality helps to quickly have high-quality syntheses at hand.

Both the Cochrane review guides and the PRISMA reporting guidelines highlight how reviews are to be conducted and reported respectively [4, 5]. After evaluating the COVID-19

synthesis research, it is unclear if some of the consistently omitted details were reporting issues or true methodological errors. Adherence to both conduct and reporting guidelines would have drastically improved the quality of the captured synthesis research. Although this may mean that in practice, reviews take longer and there are fewer of them, adhering to such rigorous guidelines would ultimately improve the utility of individual reviews in the decision-making process. Similarly, we found only nine studies which applied a GRADE assessment to review findings. While not fundamental to the systematic review process, GRADE is a tool to highlight and communicate the certainty of evidence which would be a valuable addition for decision-makers.

Based on our results, synthesis research on COVID-19 needs to be assessed for adherence to methodological rigour and policy-makers should consider only including high-quality reviews in their decision-making process, as these reviews have taken sufficient steps to minimise bias and explain possible sources and implications of heterogeneity.

Across the research–publication pipeline, there are several points in which the burden of quality assurance should be carried out. First, authors bear responsibility for correctly identifying which type of review they conducted and reporting all deviations from the gold-standard methodology for that review type. Second, publishers should be aware of different synthesis research methods and their conduct and reporting guidelines to critically assess the appropriateness of the review label and adherence to the prescribed methodology, so reviews are published with labels that reflect the methodology used. Finally, readers of synthesis research should always critically assess review quality before using the results. Overall, 151 high-quality reviews were identified across all topic areas. This forms a base of quality reviews on a number of COVID-19 topics for reference.

### Limitations

Several limitations have to be considered for this work. The AMSTAR questions give details for how they should be interpreted but individual reviewers' interpretations may have varied despite pre-tests, which is compounded by having only single reviewers conduct both screening and quality assessment, with a second person verifying only a proportion of studies (20%). The impact of non-duplicate reviewing may impact the robustness of results compared to fully verified reviews, although previous research in this area has indicated the impact is likely small given that the tool was clear and extensively pre-tested [8]. Also, the search was only conducted in English and reviews not published in English or French ( $n = 18$ ) were omitted from the project leading to a potential language bias in the reviews evaluated in this project.

### Conclusion

During the current pandemic, there has been a steady flow of synthesis research published on a wide variety of topic areas. Overall, the quality of these reviews fell short of what is expected for

systematic reviews, rapid reviews, scoping reviews and other synthesis research. The influence and impact of omitting key features of a systematic review have been studied during the development of rapid review methodology and it has been shown that omission of the key domains outlined in AMSTAR-2 can lead to bias in the review findings and decreases the utility of the review [7, 8]. The responsibility to assure the quality of published synthesis research is held across both producers and publishers, and it is up to readers to be critical of the review conduct before using the results.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268821001758>

**Acknowledgements.** The corresponding author would like to thank all contributors to the COVID-19 Daily Literature Scan project.

**Financial support.** This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Data availability statement.** Data used in this review are available in the Supplementary file.

### References

1. **World Health Organization** (2020) Novel Coronavirus (2019-nCoV): Situation Report – 13. World Health Organization, 1–7.
2. **Higgins JPT *et al.*** (ed.) (2020) *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.1. Chichester, UK: Cochrane.
3. **Moher D *et al.*** (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* **339**, 332–336. doi: 10.1136/bmj.b2535.
4. **Shea BJ *et al.*** (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal* **358**, 4008.
5. **Yu Y *et al.*** (2020) Assessment of the quality of systematic reviews on COVID-19: a comparative study of previous coronavirus outbreaks. *Journal of Medical Virology* **92**, 883–890.
6. **Haddaway NR and Westgate MJ** (2019) Predicting the time needed for environmental systematic reviews and systematic maps. *Conservation Biology* **33**, 434–443.
7. **Garrity C *et al.*** (2021) Cochrane rapid reviews methods group offers evidence-informed guidance to conduct rapid reviews. *Journal of Clinical Epidemiology* **130**, 136–122.
8. **Pham MT *et al.*** (2016) Implications of applying methodological shortcuts to expedite systematic reviews: three case studies using systematic reviews from agri-food public health. *Research Synthesis Methods* **7**, 433–446.