Annals of Clinical Epidemiology

ORIGINAL ARTICLE

The Methodological Quality Score of COVID-19 Systematic Reviews is Low, Except for Cochrane Reviews: A Meta-epidemiological Study

Yuki Kataoka[1,2,3,4,5], Shiho Oide[3,6], Takashi Ariie[3,7], Yasushi Tsujimoto[3,7,8], Toshi A. Furukawa[9]

ABSTRACT

BACKGROUND

The objective of this study was to investigate the methodological quality of coronavirus disease 2019 (COVID-19) systematic reviews (SRs) indexed in medRxiv and PubMed, compared with Cochrane COVID Reviews.

METHODS

This is a cross-sectional meta-epidemiological study. We searched medRxiv, PubMed, and Cochrane Database of Systematic Reviews for SRs of COVID-19. We evaluated the methodological quality using A MeaSurement Tool to Assess systematic Reviews (AMSTAR) checklists. The maximum AMSTAR score is 11, and minimum is 0. Higher score means better quality.

RESULTS

We included 9 Cochrane reviews as well as randomly selected 100 non-Cochrane reviews in medRxiv and PubMed. Compared with Cochrane reviews (mean 9.33, standard deviation 1.32), the mean AMSTAR scores of the articles in medRxiv were lower (mean difference (MD): −2.85, 98.3% confidence intervals (CI): −0.96 to −4.74), and those in PubMed were also lower (MD: −3.28, 98.3%CI: −1.40 to −5.15), with no difference between the latter two.

CONCLUSIONS

Readers should pay attention to the potentially low methodological quality of SRs related to COVID-19 in both PubMed and medRxiv. Evidence users might be better to search the Cochrane Library rather than medRxiv or PubMed to search SRs related to COVID-19.

KEY WORDS

preprints, peer-review, systematic review, methodological quality, COVID-19

[1] Hospital Care Research Unit, Hyogo Prefectural Amagasaki General Medical Center
[2] Department of Community Medicine in the Graduate School of Medicine, Kyoto University
[3] Systematic Review Workshop Peer Support Group (SRWS-PSG)
[4] Department of Respiratory Medicine, Hyogo Prefectural Amagasaki General Medical Center
[5] Department of Healthcare Epidemiology, Graduate School of Medicine and Public Health, Kyoto University
[6] Department of Gynecology, Women's center
[7] Department of Physical Therapy, School of Health Sciences at Fukuoka, International University of Health and Welfare
[8] Department of Nephrology and Dialysis, Kyoritsu Hospital
[9] Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health

Corresponding author: Toshi A. Furukawa
Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
E-mail: furukawa@kuhp.kyoto-u.ac.j

Received: November 14, 2020
Accepted: December 30, 2020

No.21-06

INTRODUCTION

SARS-Cov-2 virus has caused a once-in-a-century pandemic. As of August 2020, there have been over 800,000 coronavirus disease 2019 (COVID-19) deaths [1]. Recent advances in information and communication technology have led to the publication of many academic articles [2, 3]. Until May 2020, more than a thousand COVID-19 trials were registered in ClinicalTrials.gov [4]. Cochrane started Cochrane COVID Reviews to answer the time-sensitive needs of health decision makers as fast as possible. These reviews are intended to simultaneously assure the scientific quality [5].

While the number of studies is growing, questions are being raised about their methodological quality. A meta-epidemiological study that investigated the quality of randomized controlled studies (RCT) of COVID-19 pointed out the poor methodology [6]. A preliminary meta-epidemiological study including 18 COVID-19 systematic reviews (SRs) published until March 2020 also pointed out the poor methodology [7]. Despite the important role of SRs in clinical decision-making [8], the quality of COVID-19 SRs performed for speedy reporting has not been adequately evaluated. In addition, no studies assessed the quality of SRs published in preprint servers without peer review in comparison with other data sources. Hence, we investigated the methodological quality of COVID-19 reviews indexed in medRxiv, PubMed, and the Cochrane Library.

METHODS

PROTOCOL AND STUDY DESIGN

This is a cross-sectional meta-epidemiological study. We used the reporting guideline of meta-epidemiological study where applicable (**Table 1**) [9]. We published the protocol prior to the conduct of this study [10]. We published the results of the review on the randomized controlled trials separately [11]. After seeing poor quality results from PubMed and medRxiv, we decided to add Cochrane Review as control.

TYPES OF STUDIES INCLUDED

We included SRs, indexed in PubMed, medRxiv, and Cochrane Database of Systematic Reviews (CDSR). We included articles of topics related to the COVID-19 practice, irrespective of publication types. We included Cochrane Reviews that dealt with the COVID-19 pandemic. We included any type of SRs with or without meta-analysis. We included any language. We did not apply language or country restrictions. We excluded study protocols.

The definition of SRs was "a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies." [12]

SEARCH METHODS

We retrieved the abstracts from medRxiv COVID-19 SARS-CoV-2 preprints using the following search terms: "review," "evidence synthesis," "meta-analysis," or "metaanalysis" on 15th June 2020 [13]. We retrieved the abstracts from PubMed using Shokraneh's filter for COVID-19 [14] and PubMed Systematic Reviews Filter [15] on 15th June 2020 (**Table 2**). We retrieved the abstracts from Cochrane Database of Systematic Reviews (CDSR) using search term "COVID-19" on 17th Aug 2020.

STUDY SELECTION

Two of three review authors (YK, SO, and TA) selected abstracts from search results independently. Disagreements were solved through discussion. Two of three review authors (YK, SO, and TA) selected full text articles from selected abstracts independently. Disagreements were solved through discussion. Of the articles indexed in medRxiv and PubMed and meeting the eligibility criteria mentioned in *Types of studies included*, we randomly selected a total of 100 articles from medRxiv and PubMed for inclusion in the present study. The sample size was determined following a previous study [16]. We included all Cochrane reviews.

DATA EXTRACTION AND ASSESSMENT

Methodological quality of systematic reviews

We defined the methodological quality as "to what extent a study was designed, conducted, analyzed, interpreted, and reported to avoid systematic errors" [17].

For calibration training, three review authors (YK, SO, and TA) independently evaluated the methodological quality using A MeaSurement Tool to Assess systematic Reviews (AMSTAR) checklists for five included articles [18]. Disagreements were resolved through discussion. Then one of three review authors (YK, SO, and TA) evaluated other articles. Another author (YK or SO) confirmed the results. We resolved disagreements through discussion. We recorded the individual judgements to evaluate the concordance by kappa statistics.

Table 1 Items for reporting methodological research, adapted from the PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a meta-epidemiologic study.	1
ABSTRACT			
Structured summary	2	Provide a structured summary that includes the background of the topic, goal of the study, data sources, method of data selection, appraisal and synthesis methods, results, limitations, conclusions and implications of key findings.	4–5
INTRODUCTION			
Rationale	3	Describe the rationale for the meta-epidemiological study in the context of what is already known.	6
Objectives	4	Provide an explicit statement of the goal of the meta-epidemiological study and the hypothesis being empirically tested.	6
METHODS			
Protocol	5	Indicate if a protocol exists, if and where it can be accessed (eg, Web address). Registration of a protocol is not mandatory	7
Eligibility criteria	6	Specify study characteristics used as criteria for eligibility with a rationale.	7
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with experts to identify additional studies, Internet searches) and search date.	8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. Search is commonly not driven by a clinical question.	8
Study selection	9	Describe the process for selecting studies for inclusion (ie, how many reviewers selected studies, reviewing in duplicate or by single individuals).	8
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes used for manipulating data or obtaining and confirming data from investigators.	9, 10
Data items	11	List and define all variables for which data were sought and any assumptions and imputations made.	9, 10
Risk of bias in individual studies	12	If risk of bias assessment of individual studies was relevant to the analysis, describe the items used and how this information is to be used during data synthesis.	9, 10
Summary measures	13	State the principal summary measures (eg, ratio of risk ratios, difference in means) and explain its meaning and direction to readers.	11
Synthesis of results	14	Describe the statistical or descriptive methods of synthesis including measures of consistency if relevant. If applicable, describe the development of statistical or simulation modelling based on theoretical background. Describe and justify assumptions and computational approximations. Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.	11
RESULTS			
Study selection	15	Give numbers of studies assessed for eligibility and included in the study, with reasons for exclusions at each stage, ideally with a flow diagram. Present a measure of inter-reviewer agreement (eg, kappa statistic).	12
Study characteristics	16	For each study, present characteristics for which data were extracted and provide the citations. Clinical characteristics may not always be relevant.	12
Risk of bias within studies	17	If risk of bias assessment of individual studies was used in the meta-epidemiological analysis, report risk of bias indicators of each study to allow replication of findings.	12, 13
Results of individual studies	18	Present data elements used in the meta-epidemiological analysis from each study (results of clinical outcomes may not be relevant).	Not applicable
Synthesis of results	19	Present results of statistical analysis done, including measures of precision and measures of consistency. Present validity of assumptions and fit of statistical or simulation modelling, if applicable.	Not applicable
Additional analysis	20	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, meta-regression).	12, 13
DISCUSSION			
Summary of evidence	21	Summarise the main findings and compare them with existing knowledge about the topic. The quality of evidence may not be relevant; however, investigators should describe their certainty in the results to readers.	13, 14
Limitations	22	Discuss limitations at research methodology level (eg, likelihood of reporting or publication bias).	15, 16
Conclusions	23	Provide general interpretation of the results and implications for future research. Provide any plausible impact on clinical practice.	16
FUNDING			
Funding	24	Describe sources of funding for the methodology research and role of funders.	17

From: Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017 Aug;22(4):139–42.

Table 2 Search formula for PubMed

#1 COVID-19	(("Betacoronavirus"[Mesh] OR "Coronavirus Infections"[MH] OR "Spike Glycoprotein, COVID-19 Virus"[NM] OR "COVID-19"[NM] OR "Coronavirus"[MH] OR "Severe Acute Respiratory Syndrome Coronavirus 2"[NM] OR 2019nCoV[ALL] OR Betacoronavirus*[ALL] OR Corona Virus*[ALL] OR Coronavirus*[ALL] OR Coronovirus*[ALL] OR CoV[ALL] OR CoV2[ALL] OR COVID[ALL] OR COVID19[ALL] OR COVID-19[ALL] OR HCoV-19[ALL] OR nCoV[ALL] OR "SARS CoV 2"[ALL] OR SARS2[ALL] OR SARSCoV[ALL] OR SARS-CoV[ALL] OR SARS-CoV-2[ALL] OR Severe Acute Respiratory Syndrome CoV*[ALL]) AND ((2019/11/17[EDAT] : 3000[EDAT]) OR (2019/11/17[PDAT] : 3000[PDAT]))
#2 Systematic Review	(((systematic review[ti] OR systematic literature review[ti] OR systematic scoping review[ti] OR systematic narrative review[ti] OR systematic qualitative review[ti] OR systematic evidence review[ti] OR systematic quantitative review[ti] OR systematic meta-review[ti] OR systematic critical review[ti] OR systematic mixed studies review[ti] OR systematic mapping review[ti] OR systematic cochrane review[ti] OR systematic search and review[ti] OR systematic integrative review[ti]) NOT comment[pt] NOT (protocol[ti] OR protocols[ti])) NOT MEDLINE [subset]) OR (Cochrane Database Syst Rev[ta] AND review[pt]) OR systematic review[pt]
#3	#1 AND #2

We initially intended to use AMSTAR 2, which is a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of healthcare interventions [19]. Because there were few included intervention reviews, we used AMSTAR due to the latter's applicability. In other words, there is no other established tool to quantify the methodological quality of systematic reviews of other than intervention study, we used AMSTAR for the purpose of evaluating prognosis reviews, diagnostic accuracy reviews, scoping reviews, and so on, following previous studies [20, 21]. AMSTAR was developed for assessing the methodological quality of systematic reviews. For each of the 11 items in AMSTAR checklist, we calculated the AMSTAR score by counting the number of "Yes". Higher scores indicate higher quality of the systematic review. The possible maximum score was 11. We added some explanations to AMSTAR to reduce disagreements following a previous study after a calibration training [16]. The details are shown in **Table 3**.

Other variables

The number of included studies was counted at the time of the synthesis of each article. We evaluated types of research questions, the presence of protocol registrations, and the presence of limitation in each abstract. We determined that articles with "rapid" in the title is a "rapid review". We also evaluated the presence of SPIN in the title or abstract conclusion in intervention systematic reviews whose first outcome was non-significant [22]. SPIN was judged present when there were the manipulation of language to potentially mislead readers from the likely truth of the results. One of three review authors (YK, SO, and TA) evaluated these variables. Another author (YK or SO) confirmed the results. We resolved the disagreements through discussion.

Outcomes

Our primary outcome was the total score of AMSTAR.

DATA ANALYSIS

We described summary statistics. We used risk difference (RD) and 95% confidence intervals (CI) to compare binary variables. For quantification of the disagreements of AMSTAR checklists, we calculated kappa statistics between the initial and the final evaluations for the included articles except for the five articles used for calibration. We used ANOVA with Bonferroni correction for the comparison of total AMSTAR score. We conducted sensitivity analysis excluding articles those not intended for meta-analysis. We used Stata ver. 16.1 (StataCorp LLC, College Station, Texas, United States of America). A p-value less than 0.05 was considered statistically significant.

ETHICS

Because this study used only open data, ethics approval was not applicable.

RESULTS

RESULTS OF THE SEARCH

The details of the selection process are shown in **Fig. 1**. We searched a total of 641 abstracts from medRxiv and PubMed. We randomly included 49 articles from medRxiv, and 51 articles from PubMed. We searched a total of 12 abstracts from CDSR and included 9 articles. Detailed citations are publicly available at the study's associated page on the Open Science Framework (https://osf.io/jkyfb/). There were no duplicated articles among medRxiv, PubMed, and Cochrane.

Table 3 Modified AMSTAR checklist

1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review. If the article states that the protocol was created, select "Yes".	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. If the artcle included preprints, select "Yes".	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant. If the article stated limitation in the discussion and weaken the conclusions or used GRADE approach, select "Yes".	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?). If the study did not conduct meta-analysis, select "Not applicable".	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test). For reviews that are inherently difficult to assess publication bias, such as prognostic factor reviews, select "yes" if the article mentioned publication bias as limitation.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable

GRADE: The Grading of Recommendations Assessment, Development and Evaluation
Modified expressions are shown in red.
Cited from: Shea, B.J., Grimshaw, J.M., Wells, G.A. et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 7, 10 (2007). https://doi.org/10.1186/1471-2288-7-10

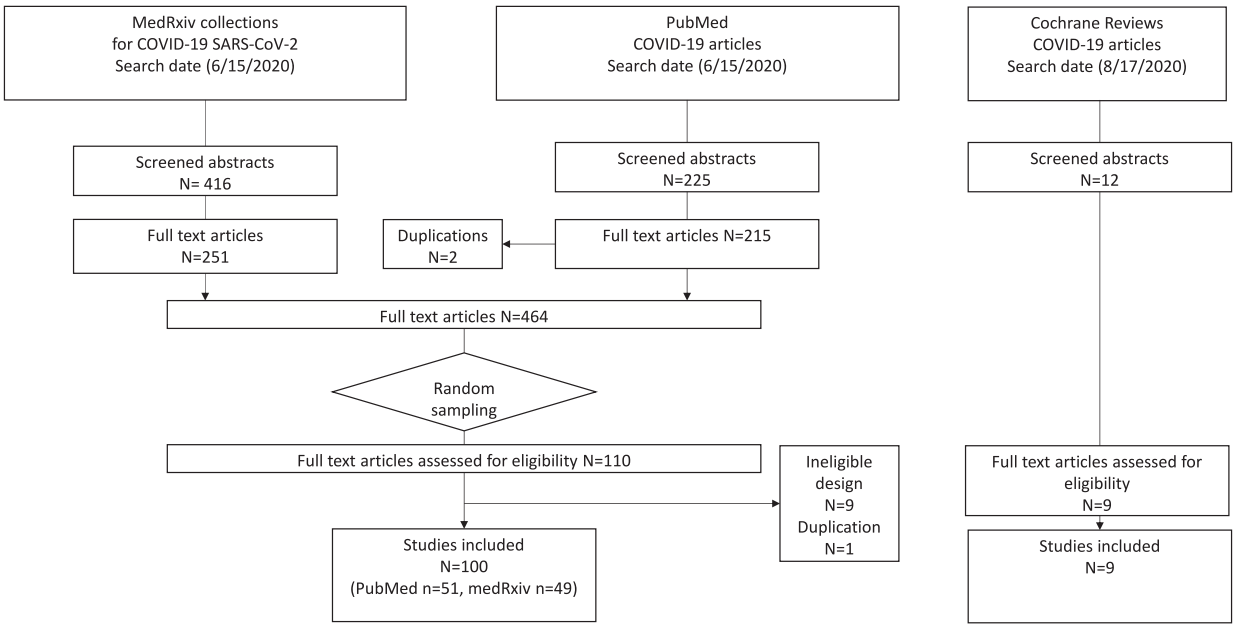

Fig. 1 **Flowchart for the selection**

Table 4 **Characteristics of included systematic reviews**

	medRxiv N = 49	PubMed N = 51	Cochrane N = 9
Number of included articles Median (IQR)	29.5 (17–45.5)	18.5 (11–39)	24 (16–36)
Types of research questions			
DTA	1 (2%)	2 (4%)	2 (22%)
Intervention	9 (18%)	14 (28%)	6 (67%)
Meta-epidemiological	1 (2%)	0 (0%)	
Prevalence, incidence	8 (16%)	16 (31%)	
Prognostic factor	11 (22%)	7 (14%)	
Scoping review, qualitative synthesis	19 (39%)	12 (24%)	1 (11%)
Protocol registration Present	5 (10%)	9 (18%)	3 (33%)
Rapid reviews	4 (8%)	7 (14%)	4 (44%)

IQR: interquartile range, DTA: diagnostic test accuracy

CHARACTERISTICS OF INCLUDED ARTICLES

The characteristics of included articles are shown in **Table 4**. We included 4 rapid reviews from medRxiv, 7 from PubMed, and 4 from Cochrane. The medians [interquartile range (IQR)] of included studies in each article were 29.5 [17–45.5] in medRxiv, 18.5 [11–39] in PubMed, and 24 [16–36] in Cochrane. More than a half of articles did not register their protocols.

QUALITY OF INCLUDED ARTICLES

Table 5 summarizes the characteristics of included articles. The agreement between the initial and the agreed-upon AMSTAR ratings ranged between 0.85 and 1.0. The mean (standard deviation (SD)) scores was 9.33 (1.32) in Cochrane reviews, 6.48 (2.07) of medRxiv, and 6.06 (2.30) of PubMed. Referring to limitations in each abstract were less in medRxiv than Cochrane (RD −56%,

Table 5 The quality of included systematic review articles

	medRxiv N = 49	PubMed N = 51	Cochrane N = 9	Risk difference Cochrane vs. medRxiv % [95%CI]	Risk difference Cochrane vs. PubMed % [95%CI]	kappa
Referring to limitations in each abstract						1
Yes	16 (33%)	12 (24%)	8 (89%)	−56% [−81 to −32]	−64% [−80 to −41]	
No	33 (67%)	37 (73%)	1 (11%)			
Without an abstract	0 (0%)	2 (4%)	0 (0%)			
AMSTAR						
1. Was an 'a priori' design provided?						0.91
Yes	8 (17%)	13 (25%)	8 (89%)	−72% [−95 to −49]	−63% [−88 to −40]	
No	4 (8%)	1 (2%)	0 (0%)			
Can't answer	36 (75%)	38 (73%)	1 (11%)			
2. Was there duplicate study selection and data extraction?						0.96
Yes	41 (85%)	45 (87%)	9 (100%)	−15% [−25 to −5]	−13% [−23 to −4]	
No	1 (2%)	3 (6%)	0 (0%)			
Can't answer	6 (13%)	4 (8%)	0 (0%)			
3. Was a comprehensive literature search performed?						1
Yes	46 (96%)	47 (90%)	7 (78%)	18% [−10 to 46]	13% [−16 to 41]	
No	2 (4%)	5 (10%)	2 (22%)			
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?						0.85
Yes	21 (44%)	24 (46%)	6 (67%)	−23% [−57 to 11]	−21% [−54 to 13]	
No	18 (38%)	14 (27%)	3 (33%)			
Can't answer	8 (17%)	14 (27%)	0 (0%)			
Not applicable	1 (2%)	0 (0%)	0 (0%)			
5. Was a list of studies (included and excluded) provided?						0.96
Yes	5 (10%)	2 (4%)	8 (89%)	−78% [−100 to −56]	−85% [−100 to −64]	
No	43 (90%)	50 (96%)	1 (11%)			
6. Were the characteristics of the included studies provided?						0.85
Yes	47 (98%)	47 (90%)	9 (100%)	−2% [−6 to 2]	−10% [−18 to −2]	
No	1 (2%)	5 (10%)	0 (0%)			
7. Was the scientific quality of the included studies assessed and documented?						0.96
Yes	27 (56%)	22 (42%)	9 (100%)	−44% [−58 to −30]	−58% [−71 to −44]	
No	21 (44%)	30 (58%)	0 (0%)			
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?						0.98
Yes	31 (65%)	24 (46%)	9 (100%)	−35% [−49 to −22]	−54% [−67 to −40]	
No	17 (35%)	28 (54%)	0 (0%)			
9. Were the methods used to combine the findings of studies appropriate?						0.89
Yes	24 (50%)	22 (42%)	3 (33%)	17% [−17 to 51]	9 [−25 to 43]	
No	4 (8%)	2 (4%)	1 (11%)			
Not applicable	20 (42%)	28 (54%)	5 (56%)			
10. Was the likelihood of publication bias assessed?						0.96
Yes	22 (46%)	21 (40%)	7 (78%)	−32% [−63 to −1]	−37% [−68 to −7]	
No	25 (52%)	31 (60%)	2 (22%)			
Not applicable	1 (2%)	0 (0%)	0 (0%)			
11. Was the conflict of interest stated?						0.96
Yes	39 (81%)	48 (92%)	9 (100%)	−19% [−30 to −8]	−8% [−15 to 0]	
No	9 (19%)	4 (8%)	0 (0%)			

95%CI: 95% confidence intervals
We calculated kappa statistics for the initial evaluation and final evaluation.

Table 6 Total AMSTAR score of each type of research question

Types of research questions	Mean (SD)
DTA (n = 5)	8.00 (2.45)
Intervention (n = 29)	7.76 (1.99)
Meta-epidemiological (n = 1)	6
Prevalence, incidence (n = 24)	6.54 (2.08)
Prognostic factor (n = 18)	6.00 (2.30)
Scoping review, qualitative synthesis (n = 32)	5.44 (2.15)

SD: standard deviation

Table 7 The difference of AMSTAR score

	Mean (SD)	Mean difference [98.3%CI]
Cochrane	9.33 (1.32)	Reference
medRxiv	6.48 (2.07)	−2.85 [−0.96 to −4.74]
PubMed	6.06 (2.30)	−3.28 [−1.40 to −5.15]

We used ANOVA with Bonferroni correction.

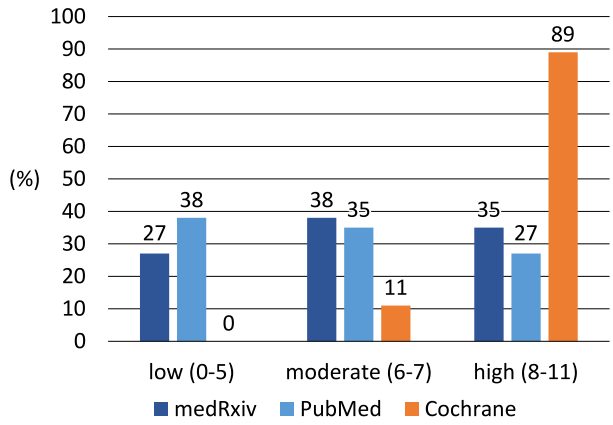

Fig. 2 The summary AMSTAR score by database

95%CI: −81 to −32), and also less in PubMed than Cochrane (RD −64%, 95%CI: −88 to −41). **Table 6** and **Fig. 2** shows the total AMSTAR score. Compared with Cochrane reviews, the mean scores of the reviews in medRxiv were lower (mean difference (MD) −2.85, 98.3%CI: −0.96 to −4.74) and those in PubMed was also lower (MD −3.28, 98.3%CI: −1.40 to −5.15) (**Table 7**). The score difference between articles in medRxiv and those in PubMed were not statistically significant (MD −0.42, 98.3%CI: −1.46 to 0.62). Sensitivity analysis after excluding the articles evaluated as "not applicable" in AMSTAR question 9 showed similar difference (medRxiv vs. Cochrane: MD −2.71, 98.3%CI 0.70 to 4.72, PubMed vs. Cochrane: MD −2.93, 98.3%CI 0.98 to 4.87).

There were six intervention articles whose first presented outcome was not statistically significant. No articles expressed SPIN in both titles and conclusions of abstracts.

DISCUSSION

This is the first meta-epidemiological study evaluating methodological quality of SRs dealing with the COVID-19 pandemic including preprints, peer-reviewed articles, and Cochrane reviews. Compared with Cochrane reviews, the mean scores of AMSTAR were lower in articles from medRxiv and PubMed. The mean scores of medRxiv and PubMed did not differ significantly.

Our findings suggest ordinary peer reviews might not improve the quality of SRs. The mean AMSTAR score difference between medRxiv (6.48) and PubMed (6.06) were not statistically significant. A previous study included SRs published in PubMed from China and US showed their mean AMSTAR score was 6.14 [23]. There's not much difference between our results and regular peer-reviewed SRs. This fact suggests it's not clear whether the quality of COVID-19 SRs will improve in the future with

sufficient time of peer-review. At this point, readers should note that the methodological quality of SRs about COVID-19 in both PubMed and medRxiv may be of poor quality, but this is not the case in the Cochrane COVID reviews. It is important to note that, Cochrane COVID reviews skip some quality control step to provide timely reviews [24]. To summarize so far, it might be better to search the Cochrane Library rather than medRxiv or PubMed to search SRs related to COVID-19.

Strict structured reporting like Cochrane reviews would be more useful than time-limited peer reviews or without peer reviews [25]. Referring to limitations in each abstract were more often in Cochrane than in medRxiv or PubMed. The differences of the quality of the articles we found between Cochrane reviews and others included the presentation of protocols, listing included and excluded studies, considering the quality of included studies at the conclusions, examining publication bias, and presenting the conflict of interests. These domains will be improved by the appropriate use of preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [26, 27]. If the duplicate study selection is difficult due to manpower insufficiency, one solution might be to use crowdsourcing to bring people together through the internet [28].

The proportion of protocol registrations did not differ from those reported in previous studies. Previous studies reported about 20% of SRs registered their protocols [29, 30]. Protocol registration is important to prevent duplicate efforts, prevent outcome reporting bias, and to reduce alpha errors in the results of meta-analyses. PROSPERO, the largest SRs protocol registration site, has some problems including not accepting Scoping reviews and taking more than 30 days to register [31]. Some SRs published their protocols in preprint servers [32, 33]. For speedy and assuring the scientific quality, protocol registration in preprint servers would be useful.

Our study has several limitations. First, AMSTAR is a reliable and valid measurement tool that has been used widely so far [34]. However, it was originally developed for SRs of randomized trials [21], and there is a criticism for calculating the total score [35]. Future research is needed to determine how to assess the quality of SRs which do not target at interventions. Second, we assessed the quality of the articles while not being masked about the published journal names. While the empirical evidence shows that such a bias is unlikely [36], this information bias could have led to overestimating the quality of Cochrane COVID reviews. Third, we were not able to incorporate studies that were published before and after peer review in the current study. Further study to compare studies published in peer-reviewed journals after publication in preprint servers is warranted. Forth, the comparisons with Cochrane COVID reviews were ad-hoc analysis. It should be noted that there is a possibility of alpha errors.

CONCLUSIONS

Readers should pay attention to the potentially low methodological quality of SRs related to COVID-19 in both PubMed and medRxiv. Evidence users might be better to search the Cochrane Library rather than medRxiv or PubMed to search SRs related to COVID-19. The methodological quality of COVID-19 SRs, except for Cochrane COVID reviews, should be improved.

AUTHOR CONTRIBUTIONS

Yuki Kataoka had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: All authors.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Yuki Kataoka.

Critical revision of the manuscript for important intellectual content: All authors.

Supervision: Toshi A. Furukawa.

CONFLICT OF INTEREST DISCLOSURES

Toshi A. Furukawa reports personal fees from Meiji, Mitsubishi-Tanabe, MSD and Pfizer and a grant from Mitsubishi-Tanabe, outside the submitted work; TAF has a patent 2018-177688 pending. Other authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS

We acknowledge Ms. Kyoko Wasai, who assisted retrieving full text articles.

REFERENCES

1. WHO Coronavirus Disease (COVID-19) Dashboard 2020. https://covid19.who.int/ (accessed August 24, 2020).

2. COVID-19: Living systematic map of the evidence. http://eppi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COVID-19Livingsystematicmapoftheevidence/tabid/3765/Default.aspx (accessed August 24, 2020).

3. Welcome to the WHO ICTRP. https://www.who.int/ictrp/en/ (accessed June 11, 2020).
4. Glasziou PP, Sanders S, Hoffmann T. Waste in covid-19 research. *BMJ* 2020 May 12;369:m1847.
5. Cochrane's work on COVID-19 Related Reviews. https://www.cochrane.org/cochranes-work-rapid-reviews-response-covid-19 (accessed August 24, 2020).
6. Alexander PE, Debono VB, Mammen MJ, Iorio A, Aryal K, Deng D, et al. COVID-19 coronavirus research has overall low methodological quality thus far: case in point for chloroquine/hydroxychloroquine. *J Clin Epidemiol* 2020:1–7.
7. Nascimento I, O'Mathúna N, Dónal P, Groote T, Abdulazeem H, Weerasekara I, Marusic A, et al. Coronavirus Disease (COVID-19) Pandemic: An Overview of Systematic Reviews. MedRxiv 2020:2020.04.16.20068213. https://doi.org/10.1101/2020.04.16.20068213.
8. Guyatt G, Rennie D, Meade M, Cook D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2nd ed. McGraw-Hill Professional; 2015.
9. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med* 2017;22:139–42. https://doi.org/10.1136/ebmed-2017-110713.
10. Kataoka Y, Oide S, Ariie T, Tsujimoto Y, Furukawa TA. Quality of COVID-19 research in preprints: a meta-epidemiological study protocol. Protocols.io 2020. https://doi.org/10.17504/protocols.io.bhm8j49w. (accessed December 9, 2020)
11. Kataoka Y, Oide S, Ariie T, Tsujimoto Y, Furukawa TA. COVID-19 randomized controlled trials in medRxiv and PubMed. *Eur J Intern Med* 2020:4627. https://doi.org/10.1016/j.ejim.2020.09.019.
12. Institution of Medicine. Clinical practice guidelines we can trust. 2011.
13. medRxiv COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. https://connect.medrxiv.org/relate/content/181 (accessed July 3, 2020).
14. Shokraneh F. Keeping up with studies on covid-19: systematic search strategies and resources. *BMJ* 2020:m1601. https://doi.org/10.1136/bmj.m1601.
15. Systematic Reviews Filter Strategy. https://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html (accessed August 20, 2020).
16. Tian J, Zhang J, Ge L, Yang K, Song F. The methodological and reporting quality of systematic reviews from China and the USA are similar. *J Clin Epidemiol* 2017;85:50–8. https://doi.org/10.1016/j.jclinepi.2016.12.004.
17. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019.
18. Shea BJ, Grimshaw JM, Wells G a, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10. https://doi.org/10.1186/1471-2288-7-10.
19. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, 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:j4008. https://doi.org/10.1136/bmj.j4008.
20. Arevalo-Rodriguez I, Segura O, Solà I, Bonfill X, Sanchez E, Alonso-Coello P. Diagnostic tools for alzheimer's disease dementia and other dementias: an overview of diagnostic test accuracy (DTA) systematic reviews. *BMC Neurol* 2014;14:183. https://doi.org/10.1186/s12883-014-0183-2.
21. Pieper D, Koensgen N, Breuing J, Ge L, Wegewitz U. How is AMSTAR applied by authors – a call for better reporting. *BMC Med Res Methodol* 2018;18:56. https://doi.org/10.1186/s12874-018-0520-z.
22. Boutron I. Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. *JAMA* 2010;303:2058. https://doi.org/10.1001/jama.2010.651.
23. Tian J, Zhang J, Ge L, Yang K, Song F. The methodological and reporting quality of systematic reviews from China and the USA are similar. *J Clin Epidemiol* 2017;85:50–8. https://doi.org/10.1016/j.jclinepi.2016.12.004.
24. Cochrane Rapid Reviews Methods Group. Cochrane Rapid Review (RR) Definition https://covidreviews.cochrane.org/sites/covidreviews.cochrane.org/files/public/uploads/cochrane_rr_-_guidance-23mar2020-final.pdf (accessed December 6, 2020).
25. Higgins JPT, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, Churchill R. Methodological Expectations of Cochrane Intervention Reviews. Cochrane: London, Version March 2020.
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100. https://doi.org/10.1371/journal.pmed.1000100.
27. Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev* 2007;2. https://doi.org/10.1002/14651858.MR000016.pub3.
28. Semler MW, Stover DG, Copland AP, Hong G, Johnson MJ, Kriss MS, et al. Flash Mob Research. *Chest* 2013;143:1740–4. https://doi.org/10.1378/chest.12-1837.
29. Tsujimoto Y, Tsujimoto H, Kataoka Y, Kimachi M, Shimizu S, Ikenoue T, et al. Majority of systematic reviews published in high-impact journals neglected to register the protocols: a meta-epidemiological study. *J Clin Epidemiol* 2017;0. https://doi.org/10.1016/j.jclinepi.2017.02.008.
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. https://doi.org/10.1016/j.jclinepi.2018.01.003.
31. PROSPERO https://www.crd.york.ac.uk/prospero/ (accessed August 24, 2020).
32. Piechotta V, Chai KL, Valk SJ, Doree C, Monsef I, Wood EM, et al. Convalescent plasma or hyperimmune immunoglobulin for people with COVID-19: a living systematic review. *Cochrane Database Syst Rev* 2020;7. https://doi.org/10.1002/14651858.CD013600.pub2.
33. Khalili M, Karamouzian M, Nasiri N, Javadi S, Mirzazadeh A, Sharifi H. Epidemiological Characteristics of COVID-19: A Systemic Review and Meta-Analysis. *MedRxiv* 2020:2020.04.01.20050138. https://doi.org/10.1101/2020.04.01.20050138.
34. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62:1013–20. https://doi.org/10.1016/j.jclinepi.2008.10.009.
35. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. www.training.cochrane.org/handbook.
36. Berlin JA. Does blinding of readers affect the results of meta-analyses? *Lancet* 1997;350:185–6. https://doi.org/10.1016/S0140-6736(05)62352-5.