

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cross-sectional study of preprints and final journal publications from COVID-19 studies: Discrepancies in results reporting and spin in interpretation
AUTHORS	Bero, Lisa; Lawrence, Rosa; Leslie, Louis; Chiu, Kellia; McDonald, Sally; Page, Matthew; Grundy, Quinn; Parker, Lisa; Boughton, Stephanie; Kirkham, Jamie J.; Featherstone, Robin

VERSION 1 – REVIEW

REVIEWER	Shi, Xiaoting Yale University School of Public Health, Department of Environmental Health Sciences
REVIEW RETURNED	07-Apr-2021

GENERAL COMMENTS	<p>I appreciated the opportunity to review a revised version of this manuscript by Dr. Bero et al. The authors have provided detailed and helpful responses and clarifications to first round of comments (including removing the particular example of toxicity data, providing concrete components of outcomes when making the comparison, re-organizing the tables, etc.). A few additional opportunities exist for improved clarity:</p> <p>First, this is a follow-up question for previous comments #8 and #9, in Page 50, Line 18, authors mentioned that “We extracted data from the medRxiv page and PDF for preprints and the online publication or PDF for journal articles.”. Based on our previous evaluation, the funding, COI and IRB statements reported in the medRxiv page and the ones reported in the pdf versions of the preprint document were not always consistent, could the authors give more details upon which source of information they prioritized? Similarly, in Page 51, Line 9, we previously found that preprints and journal publications of observational studies are not always explicitly defined/specified primary and secondary outcomes. Could the authors possibly provide more clarity on the strategies they used for identification of outcomes?</p> <p>Second, in Page 59, Line 30, in Comparison to other studies section: in addition to the research that the authors have mentioned and compared with, there are opportunities to discuss how the findings are related to other emerging evaluations in this area, including a newly-released preprint on the same/similar topic (https://www.biorxiv.org/content/10.1101/2021.02.20.432090v2), and our recent evaluation of 47 preprint-journal article pairs (doi:10.1001/jamanetworkopen.2021.2110). It might be interesting and helpful for readers to understand the differences of the results in this field.</p> <p>Additional minor comments about some details for perfection:</p>
-------------------------	--

	<p>Page 47, Line 3: could the authors elaborate more on the meaning of “found reporting to be a little higher in journal articles”? for example, does higher reporting mean better quality or more items reported?</p> <p>Page 47, Line 27: Our evaluation of 47 pairs finds 26% had changes in results for primary end point (though not limited to covid-19 studies). Given the limited sample size in the current projects in this field, it’s helpful to enumerate the findings of the relevant studies in this area for readers to better understand the consistency and inconsistency between the results from these studies.</p>
--	---

REVIEWER	Amaral, Olavo Institute of Medical Biochemistry Leopoldo de Meis Federal University of Rio de Janeiro
REVIEW RETURNED	10-Apr-2021

GENERAL COMMENTS	<p>General assessment:</p> <p>This article is a revision of a manuscript previously submitted to the BMJ, in which I also acted as a reviewer. The main improvement in the manuscript has been the sharing of the authors’ dataset as a spreadsheet in the OSF, which is generally well organized and should allow other researchers to build upon the work. The manuscript’s tables have also been expanded to include more qualitative examples, and some of the limitations pointed out by the previous reviews have been acknowledged. The major limitation that remains, in my view, is the fact that the sample size and the lack of inferential statistics limits any kind of generalization beyond the sample. Although I respect the option of the authors to present the data in a descriptive manner only (as this was what was specified in the protocol), this means that they should be much clearer about this limitation when referring to differences between preprints and journal articles.</p> <p>Major points:</p> <ul style="list-style-type: none"> - As pointed out above, in the absence of inferential statistics, the authors should be clear upon the fact that no attempt is being made to generalize comparisons beyond the studied sample of 67 preprint/article pairs. This should definitely be added to the strengths and limitations section of the article summary, for example. Similarly, in the results section, when authors mention that “more preprints than journal publications reported funding source...”, they should be explicit about the numerical differences (which are mostly small) and add that the lack of inferential statistics does not allow one to evaluate how easily these differences could have arisen by chance alone. This should also be mentioned in the discussion, both when describing results in the third paragraph and when discussing limitations of the study in the “strengths and limitations section”. - I also maintain my point that an important finding is that in many cases, the preprint and manuscript seem to contain different primary data (e.g. different number of participants, different length of follow-up, etc.). In these cases, other discrepancies (in statistics or results descriptions, for example) are likely to be expected, as the data itself is different, and cannot be taken as evidence of an effect of peer review or journal publishing. The authors do mention the fact in the discussion (i.e. “... indicates that the preprints were being used to publish preliminary or interim data”), but I think this
-------------------------	--

is important enough that they should try to specify the frequency with which this occurs. They should also mention in the discussion that changes in primary data confounds the evaluation of other discrepancies, as noted above.

Minor points:

Abstract:

- Spin can only be removed if it was present, so it is strange to use all the 67 studies as the denominator for the percentages of “removed spin” and “added spin”.

Strengths and limitations:

- As noted above, the descriptive nature of the study and lack of inferential statistics should be acknowledged as a limitation.

- In line 31, there’s a mention to “reprints” (should be “preprints”)

Introduction:

- Page 7, Line 51: “There has been no systematic assessments” – should either be “have been” or “assessment”.

- Page 8, line 20: “matched pairs preprint and their journal publications” – should be “matched pairs of preprints and their journal publications”.

- Page 8, line 23: “in the abstract of one article, but not its accompanying preprint.” – should be “in its accompanying preprint”.

- Page 8, lines 9-31: Perhaps it might be worth adding quotation marks to “positive without reporting uncertainty” and “positive with reporting of uncertainty” to clearly mark these as categories and facilitate reading.

Methods:

- Page 9, line 27: “also includes preprints records sourced from PubMed”. Should be “preprint records”.

- When referring to OSF pages (e.g. page 9, line 45), it would be clearer if authors referred to specific documents instead of the main page when applicable (e.g. the search strategy link is <https://osf.io/8qfby/>, which does not correspond to the link provided). Similarly, when linking either to specific materials or the whole page (<https://osf.io/5ru8w/>), anything from the “?” onwards refers to the query and is not necessary for the reader).

- Page 9, line 46: “As the register is updated daily, we repeated the search.” – it seems that this refers to the two searches on the two dates mentioned in the next sentences, but this could be made clearer in the sentence (e.g. “we repeated the search twice”).

- As mentioned by other reviewers, it would be useful to visualize examples of spin categories to get a better grasp of what they mean. I understand that these can now be examined in the

dataset, but it might be worth considering whether some examples could also be shown in the methods.

Results:

- Page 15, line 49: If there are 67 pairs, 23 had no changes in outcomes and 23 had changes in outcomes, where are the rest of the studies? Something seems to be missing here.

- Page 16, line 31: "The types of discrepancies were variable, although journal publications consistently included additional statistical analyses and subgroup analyses compared to preprints." What do the authors mean by "consistently". It seems hard to speak of this as a consistent phenomenon, as according to Table 3 additional tests were included in only 7 out of 67 studies.

- Page 16, lines 38-47: This paragraph contains contradictory information: if spin was added between the preprint and journal publication in only 1 study, how can 2 studies have spin in the journal publication and not the preprint?

- Both examples of spin given in the results were found both in preprint and journal publication. Wouldn't it be interesting to include examples found on only one of them as well, in order for the reader to visualize the changes that happened in these cases?

Discussion:

- Page 19, line 35: "A small proportion of medRxiv preprints, 10% during the server's first year, were published as journal publications." It is not clear from the sentence when this assessment was made (i.e. at the end of the year?). The authors could thus word this more clearly.

- Page 19, lines 37-41: "Our sample could be limited to studies that their authors deemed of high enough quality" and "Our sample could be limited to articles that had not been rejected by a journal". I don't see the need for "could be" here: the sample clearly is limited in both of these senses.

- Page 19, lines 44-50: "Under non-pandemic conditions, articles may undergo more revision. For example, peer reviewers may not suggest changes they think are less important, or editors may accept articles when they would have normally requested minor or major revisions." The wording here is confusing – the first sentence speaks of "non-pandemic conditions", but the second seems to be talking about pandemic ones (so it is not really an example of what's said in the first sentence, but rather the contrary).

- Page 20, lines 6-7: "the sample is limited to preprints that could not have benefited from peer review." Although I agree that 7 days is a short time for peer review, it might be too assertive to say that they could not have benefited at all – some changes could still be made in this period.

- Page 20, lines 27-28: "Carnerio et al"- the spelling remains wrong here: please change to Carneiro et al. Also, as previously noted, it's probably worth noting that the sample in that paper was from

	<p>bioRxiv, which might explain the discrepancy in the conflict of interest findings.</p> <p>Tables:</p> <p>Table 1: "Funding statement in preprint provides more detailed" – something seems amiss here (should probably be either "provides more detail" or "provides more detailed information").</p> <p>Table 1: is there any reason for the Conflict of Interest Disclosure statement row to be in bold or is this a mistake?</p> <p>Table 1: "different numbers of patients recruited, but same number randomized; 284 patients included in preprint, 267 in journal publication". I found this a bit confusing: do the numbers refer to patients recruited or to those included in the study/analysis?</p> <p>Table 1: "number do not match". Should be "numbers do not match".</p> <p>Table 2: I think it would be useful to include the number of studies with no change in outcome as a separate row for completeness.</p> <p>Table 2: As there are more outcomes than studies, it would be useful to indicate when the outcomes shown in the right row came from the same studies. Perhaps numbering them as (1a, 1b, etc.) or using different line spacing could help.</p> <p>Table 3: In some descriptions, the direction of the discrepancy is clear (e.g. "journal publication reports outcomes measured over a longer timepoint than preprint), in others they are not (e.g. "differences in numbers of participants or denominators" – in this case, who has more?). I feel that directionality is important in this case and should be described consistently when possible.</p>
--	--

REVIEWER	Carneiro , Clarissa F D Federal University of Rio de Janeiro, Institute of Medical Biochemistry Leopoldo de Meis
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	<p>Dear authors,</p> <p>Thank you for addressing the comments I made previously and for providing clarifications. At this time, I have only minor suggestions to the text that I strongly believe would further improve your manuscript but most of them are not necessarily indispensable.</p> <p>What this study adds</p> <p>Page 4, lines 26-30 – To highlight the two assessments made in this study, I suggest presenting them in separate bullet points. Or maybe break the description of the study itself ("comparison of ...") from its results.</p> <p>Page 4, line 32 – As you did not assess the impact of the discrepancies, I suggest removing this point as a contribution of the study. I believe it speaks more of what is already known about spin or maybe of the potential importance of these assessments.</p> <p>Abstract</p> <p>Page 5, line 24 – A closing parenthesis seems to be missing after "secondary".</p> <p>Page 5, line 54 – Given that you found both publication formats/venues (preprints and journal publication) to be "largely</p>
-------------------------	--

	<p>similar”, it does not make much sense to me that the discrepancies should be the main focus instead of the quality of outcome reporting or presence/absence of spin. I suggest changing “discrepancies” to “results reporting”.</p> <p>Article summary</p> <p>Page 6 – In general, I believe all points should be further summarized as the instructions to authors page mentions “no longer than one sentence” (https://bmjopen.bmj.com/pages/authors/#submission_guidelines). Also, points 3 and 5 are the most important to highlight in my opinion, but I am not sure whether all limitations should be listed here or only a selection of them.</p> <p>Page 6, line 32 – It seems “reprints” should be “preprints”.</p> <p>Methods</p> <p>Page 11, line 20 – It seems the sentence was restructured but the “and” ended up out of place. I think it should read “If they were, the content of the item that differed between the preprint and publication and the details of the discrepancy were recorded.”</p> <p>Discussion</p> <p>Page 18, line 25 - The recommendation/suggestion that preprints with preliminary or interim data should be labelled as such is a valid point, but as it is it could be misinterpreted as a result-derived recommendation. My suggestion: “(...) were being used to publish preliminary or interim data. Preliminary or interim findings should be clearly labeled in preprints, however this study did not assess whether such description was present or not.”</p> <p>Page 19, line 5 – Preprints are likely to be posted without previous peer review in a journal, but most often we cannot be sure about this status. I do not think this distinction changes the position of your study as ‘an indirect investigation of the impact of peer-review’, but may still be worth adding.</p> <p>Page 19, line 41 – Similar to the comment above, I believe a better phrasing for this limitation would be that your sample could be limited to articles that were eventually published in a journal as you cannot be sure whether these preprints have been rejected in previous submissions before or after the preprint posting.</p> <p>Page 20, line 7 – On the same note as the previous comments, I would suggest changing to “(...) preprints that could not have benefited from peer review in the final journal of publication”. Again, I do not believe this would change any of the conclusions from these sensitivity analyses, but it is a warranted clarification to readers.</p> <p>Page 20, line 27 – Please correct Carnerio to Carneiro.</p> <p>Conclusion</p> <p>Page 22, line 10-12 – I do not think the sentence “However, given the urgent need for valid and reliable research on COVID-19 treatment and prevention, even a few important discrepancies could impact decision making.” is a conclusion because the study did not assess the impact of the discrepancies, but only their frequency. I would move it to the beginning of the Discussion or to the Introduction.</p> <p>Page 22, line 15 – As I mentioned in a previous comment, I think the focus of a recommendation should not be about the discrepancies, as they were infrequent. I suggest changing “discrepancies” for “completeness” and adding “presence of” before spin.</p>
--	--

REVIEWER	Schluger, Neil Columbia University Medical Center, Medicine
REVIEW RETURNED	16-Apr-2021

GENERAL COMMENTS	The authors have satisfactorily responded to review concerns, including my prior comments. The manuscript is substantially improved and the limitations are clearly stated. I have no further comments that need to be addressed.
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Xiaoting Shi, Yale University School of Public Health Comments to the Author

Summary: I appreciated the opportunity to review a revised version of this manuscript by Dr. Bero et al. The authors have provided detailed and helpful responses and clarifications to first round of comments (including removing the particular example of toxicity data, providing concrete components of outcomes when making the comparison, re-organizing the tables, etc.). A few additional opportunities exist for improved clarity:

1:

1) First, this is a follow-up question for previous comments #8 and #9, in Page 50, Line 18, authors mentioned that “We extracted data from the medRxiv page and PDF for preprints and the online publication or PDF for journal articles.”. Based on our previous evaluation, the funding, COI and IRB statements reported in the medRxiv page and the ones reported in the pdf versions of the preprint document were not always consistent, could the authors give more details upon which source of information they prioritized?

RESPONSE: We clarified that we checked both the medRxiv page and the PDF, referring to the PDF if information differed (page 10, last para).

#2:

Similarly, in Page 51, Line 9, we previously found that preprints and journal publications of observational studies are not always explicitly defined/specified primary and secondary outcomes. Could the authors possibly provide more clarity on the strategies they used for identification of outcomes?

RESPONSE: We did not select primary or secondary outcomes. As stated in the methods (page 11) we extracted data on all outcomes reported in the preprint or journal article. We recorded the descriptors for the outcomes as used in the preprint or article. We recorded whether an outcome was labeled as primary or not, but extracted data on all outcomes, regardless of how they were labeled.

#3:

Second, in Page 59, Line 30, in Comparison to other studies section: in addition to the research that the authors have mentioned and compared with, there are opportunities to discuss how the findings are related to other emerging evaluations in this area, including a newly-released preprint on the same/similar topic (<https://www.biorxiv.org/content/10.1101/2021.02.20.432090v2>), and our recent evaluation of 47 preprint-journal article pairs (doi:10.1001/jamanetworkopen.2021.2110). It might be interesting and helpful for readers to understand the differences of the results in this field.

RESPONSE: These studies became available several months after we completed our paper.

4

Additional minor comments about some details for perfection:

Page 47, Line 3: could the authors elaborate more on the meaning of “found reporting to be a little higher in journal articles”? for example, does higher reporting mean better quality or more items reported?

RESPONSE: We have clarified to higher quality reporting (page 8, para 1). The direct quote from the source article is: “Peer-reviewed articles had, on average, higher quality of reporting than preprints, although the difference was small, with absolute differences of 5.0% [95% CI 1.4, 8.6] and 4.7% [95% CI 2.4, 7.0] of reported items in the independent samples and paired sample comparison, respectively.” (DOI: <https://doi.org/10.1186/s41073-020-00101-3>)

5

Page 47, Line 27: Our evaluation of 47 pairs finds 26% had changes in results for primary end point (though not limited to covid-19 studies). Given the limited sample size in the current projects in this field, it's helpful to enumerate the findings of the relevant studies in this area for readers to better understand the consistency and inconsistency between the results from these studies.

RESPONSE: To establish the context for our study in the introduction, we have summarized previous studies and provided the references, rather than repeat the findings verbatim.

Reviewer: 2

Dr. Olavo Amaral

Summary: This article is a revision of a manuscript previously submitted to the BMJ, in which I also acted as a reviewer. The main improvement in the manuscript has been the sharing of the authors' dataset as a spreadsheet in the OSF, which is generally well organized and should allow other researchers to build upon the work. The manuscript's tables have also been expanded to include more qualitative examples, and some of the limitations pointed out by the previous reviews have been acknowledged.

#6:

The major limitation that remains, in my view, is the fact that the sample size and the lack of inferential statistics limits any kind of generalization beyond the sample. Although I respect the option of the authors to present the data in a descriptive manner only (as this was what was specified in the protocol), this means that they should be much clearer about this limitation when referring to differences between preprints and journal articles. As pointed out above, in the absence of inferential statistics, the authors should be clear upon the fact that no attempt is being made to generalize comparisons beyond the studied sample of 67 preprint/article pairs. This should definitely be added to the strengths and limitations section of the article summary, for example. Similarly, in the results section, when authors mention that "more preprints than journal publications reported funding source...", they should be explicit about the numerical differences (which are mostly small) and add that the lack of inferential statistics does not allow one to evaluate how easily these differences could have arisen by chance alone. This should also be mentioned in the discussion, both when describing results in the third paragraph and when discussing limitations of the study in the "strengths and limitations section".

RESPONSE: We do not make any conclusions regarding statistical differences. The generalizability of our study is addressed in the limitations (page 21). Conducting a descriptive study is not a limitation as such studies provide valuable qualitative information. Furthermore, the lack of inferential statistics is not a limitation, but is appropriate given our study design. We do not feel it is appropriate to repeat the numerical results in the discussion section.

7

I also maintain my point that an important finding is that in many cases, the preprint and manuscript seem to contain different primary data (e.g. different number of participants, different length of follow-up, etc.). In these cases, other discrepancies (in statistics or results descriptions, for example) are likely to be expected, as the data itself is different, and cannot be taken as evidence of an effect of peer review or journal publishing. The authors do mention the fact in the discussion (i.e. "... indicates that the preprints were being used to publish preliminary or interim data"), but I think this is important enough that they should try to specify the frequency with which this occurs. They should also mention in the discussion that changes in primary data confounds the evaluation of other discrepancies, as noted above.

RESPONSE: In the discussion, we provide possible explanations for the discrepancies we observed (including that preprints are publishing preliminary data). As noted above, this is a descriptive study, so we cannot examine statistical differences in frequencies.

8:

Spin can only be removed if it was present, so it is strange to use all the 67 studies as the denominator for the percentages of “removed spin” and “added spin”.

RESPONSE: As noted on page 16, we measured the presence of spin in each preprint and article publication for the 67 studies. The denominator refers to the number of studies in which spin was added or removed between the preprint and article.

9

As noted above, the descriptive nature of the study and lack of inferential statistics should be acknowledged as a limitation.

RESPONSE: see response to #6.

10

- In line 31, there’s a mention to “reprints” (should be “preprints”)
- Page 7, Line 51: “There has been no systematic assessments” – should either be “have been” or “assessment”.
- Page 8, line 20: “matched pairs preprint and their journal publications” – should be “matched pairs of preprints and their journal publications”.
- Page 8, line 23: “in the abstract of one article, but not its accompanying preprint.” – should be “in its accompanying preprint”.
- Page 8, lines 9-31: Perhaps it might be worth adding quotation marks to “positive without reporting uncertainty” and “positive with reporting of uncertainty” to clearly mark these as categories and facilitate reading.
- Page 9, line 27: “also includes preprints records sourced from PubMed”. Should be “preprint records”.

RESPONSE: These edits were made.

11

When referring to OSF pages (e.g. page 9, line 45), it would be clearer if authors referred to specific documents instead of the main page when applicable (e.g. the search strategy link is <https://osf.io/8qfby/>, which does not correspond to the link provided). Similarly, when linking either to specific materials or the whole page (<https://osf.io/5ru8w/>), anything from the “?” onwards refers to the query and is not necessary for the reader).

RESPONSE: OSF links were updated to the link that takes the reader directly to the document referenced. Links were shortened as suggested.

12

Page 9, line 46: “As the register is updated daily, we repeated the search.” – it seems that this refers to the two searches on the two dates mentioned in the next sentences, but this could be made clearer in the sentence (e.g. “we repeated the search twice”).

RESPONSE: We did the search and repeated it once. We did not repeat it twice. The dates on which the 2 searches were conducted are provided.

13

As mentioned by other reviewers, it would be useful to visualize examples of spin categories to get a better grasp of what they mean. I understand that these can now be examined in the dataset, but it might be worth considering whether some examples could also be shown in the methods.

RESPONSE: To our knowledge, this paper provides the most detailed coding of types of spin that we have seen. Thus, our paper provides more information on the type of spin than most papers. As noted in our previous response, spin is contextual and we, therefore, encourage readers to go to the original preprint or article to review the context.

14

Page 15, line 49: If there are 67 pairs, 23 had no changes in outcomes and 23 had changes in outcomes, where are the rest of the studies? Something seems to be missing here.

RESPONSE: As noted in the methods and results sections, studies had multiple outcomes. Therefore, the summary of changes in outcomes adds to greater than the number of studies. The section the reviewer is referring to described categories that account for the other studies (page 15, last para): “Of the 67 studies, 23 (34%) had no discrepancies in results reporting between preprints and journal publications (Table 2). Twenty-three studies had outcomes that were missing from either the preprint or the journal publication. Fifteen (22%) studies had at least one outcome that was included in the journal publication, but not the preprint; 8 (12%) had at least one outcome that was reported in the preprint only.”

15

Page 16, line 31: “The types of discrepancies were variable, although journal publications consistently included additional statistical analyses and subgroup analyses compared to preprints.” What do the authors mean by “consistently”. It seems hard to speak of this as a consistent phenomenon, as according to Table 3 additional tests were included in only 7 out of 67 studies.

RESPONSE: By consistent, we mean that these were the most common types of discrepancies in these categories, while avoiding the use of the term “frequency” which is incorrectly interpreted by some of the reviewers to mean a statistical difference.

16

Page 16, lines 38-47: This paragraph contains contradictory information: if spin was added between the preprint and journal publication in only 1 study, how can 2 studies have spin in the journal publication and not the preprint?

RESPONSE: As note in Table 4, “Subcategories of spin are not mutually exclusive; a preprint or journal publications could contain multiple subcategories of spin within a category. Preprints and journal publications could contain different subcategories of spin within a category.” The 2 instances of spin in the journal article refers to any category of spin; the addition of spin between a preprint and journal publication refers to a subcategory of spin.

17

Both examples of spin given in the results were found both in preprint and journal publication. Wouldn't it be interesting to include examples found on only one of them as well, in order for the reader to visualize the changes that happened in these cases?

RESPONSE: We chose these examples based on previous feedback from peer reviewers. We encourage readers to go to the original preprint or article to review the context.

18

- Page 19, line 35: “A small proportion of medRxiv preprints, 10% during the server's first year, were published as journal publications.” It is not clear from the sentence when this assessment was made (i.e. at the end of the year?). The authors could thus word this more clearly.

RESPONSE: This was clarified to “at the end of the server's first year” since the dates used were June of 2019 through June of 2020. Also, the percentage was corrected from 10% to 14% to accurately reflect the article. doi: [10.1001/jama.2020.17529](https://doi.org/10.1001/jama.2020.17529)

#19

Page 19, lines 37-41: “Our sample could be limited to studies that their authors deemed of high enough quality” and “Our sample could be limited to articles that had not been rejected by a journal”. I don't see the need for “could be” here: the sample clearly is limited in both of these senses.

RESPONSE: We cannot be certain that are sample was limited for these reasons. An author could just decide not to submit an article or lose their enthusiasm for publication, for example.

#20

Page 19, lines 44-50: “Under non-pandemic conditions, articles may undergo more revision. For example, peer reviewers may not suggest changes they think are less important, or editors may

accept articles when they would have normally requested minor or major revisions.”. The wording here is confusing – the first sentence speaks of “non-pandemic conditions”, but the second seems to be talking about pandemic ones (so it is not really an example of what’s said in the first sentence, but rather the contrary).

RESPONSE: The first sentence was changed to read “Under pandemic conditions, articles may undergo fewer revisions.”

#21

Page 20, lines 6-7: “the sample is limited to preprints that could not have benefited from peer review.” Although I agree that 7 days is a short time for peer review, it might be too assertive to say that they could not have benefited at all – some changes could still be made in this period.

RESPONSE: What this section referred to is slightly different than what the reviewer is referring to. The passage says: “There were minimal changes in the frequency and types of discrepancies between preprints and journal publications when we conducted a sensitivity analysis limiting our sample to studies where the preprints were published before the revision or acceptance date of the journal publication. This suggests that our findings are robust even when the sample is limited to preprints that could not have benefited from peer review.” However, given the ambiguity of the cut off date for the sensitivity analysis, and the possibility that articles may have been reviewed and rejected from other journals, this last sentence was changed to say “when the sample is limited to preprints that likely had not gone through the peer review process.”

22

Page 20, lines 27-28: “Carnerio et al”- the spelling remains wrong here: please change to Carneiro et al. Also, as previously noted, it’s probably worth noting that the sample in that paper was from bioRxiv, which might explain the discrepancy in the conflict of interest findings.

RESPONSE: The spelling was updated to “Carneiro.” The description of this paper in the Introduction (page 8, para 1) notes that the sample is from bioRxiv, so it is not necessary to repeat this.

23

- Table 1: “Funding statement in preprint provides more detailed” – something seems amiss here (should probably be either “provides more detail” or “provides more detailed information”).
- Table 1: is there any reason for the Conflict of Interest Disclosure statement row to be in bold or is this a mistake?
- Table 1: “different numbers of patients recruited, but same number randomized; 284 patients included in preprint, 267 in journal publication”. I found this a bit confusing: do the numbers refer to patients recruited or to those included in the study/analysis?
- Table 1: “number do not match”. Should be “numbers do not match”.

RESPONSE: These edits were made. Re point 2, as suggested by a previous reviewer we are describing the discrepancy verbatim – it is a discrepancy in the number of patients recruited and number randomized (not analyzed).

24

Table 2: I think it would be useful to include the number of studies with no change in outcome as a separate row for completeness.

RESPONSE: In response to previous comments from reviewers regarding how we should modify this table, the number of studies with no discrepancies is provided in the text only. Page 15, first line of “results reporting” section.

25

Table 2: As there are more outcomes than studies, it would be useful to indicate when the outcomes shown in the right row came from the same studies. Perhaps numbering them as (1a, 1b, etc.) or using different line spacing could help.

RESPONSE: The list of outcomes was updated so that numbering indicates unique studies and lettering indicates outcomes from the same study.

26

Table 3: In some descriptions, the direction of the discrepancy is clear (e.g. “journal publication reports outcomes measured over a longer timepoint than preprint), in others they are not (e.g. “differences in numbers of participants or denominators” – in this case, who has more?). I feel that directionality is important in this case and should be described consistently when possible.

RESPONSE: In some cases, the discrepancies are not in a uniform direction.

Reviewer: 3

Dr. Clarissa F D Carneiro , Federal University of Rio de Janeiro Comments to the Author:

Summary: Thank you for addressing the comments I made previously and for providing clarifications. At this time, I have only minor suggestions to the text that I strongly believe would further improve your manuscript but most of them are not necessarily indispensable.

27

What this study adds

Page 4, lines 26-30 – To highlight the two assessments made in this study, I suggest presenting them in separate bullet points. Or maybe break the description of the study itself (“comparison of ...”) from its results.

RESPONSE: This section has been deleted as per editor.

28

Page 4, line 32 – As you did not assess the impact of the discrepancies, I suggest removing this point as a contribution of the study. I believe it speaks more of what is already known about spin or maybe of the potential importance of these assessments.

RESPONSE: This section has been deleted as per editor.

29

Abstract

Page 5, line 24 – A closing parenthesis seems to be missing after “secondary”.

Page 6, line 32 – It seems “reprints” should be “preprints”.

RESPONSE: A closing parentheses was added. “Reprints” was changed to “preprints.”

30

Page 5, line 54 – Given that you found both publication formats/venues (preprints and journal publication) to be “largely similar”, it does not make much sense to me that the discrepancies should be the main focus instead of the quality of outcome reporting or presence/absence of spin. I suggest changing “discrepancies” to “results reporting”.

Page 22, line 15 – As I mentioned in a previous comment, I think the focus of a recommendation should not be about the discrepancies, as they were infrequent. I suggest changing “discrepancies” for “completeness” and adding “presence of” before spin.

RESPONSE: Our protocol was designed to assess and describe discrepancies, not the quality of reporting. Thus, discrepancies remain the focus of our study regardless of our findings.

31

Article summary

Page 6 – In general, I believe all points should be further summarized as the instructions to authors page mentions “no longer than one sentence”

(https://bmjopen.bmj.com/pages/authors/#submission_guidelines). Also, points 3 and 5 are the most important to highlight in my opinion, but I am not sure whether all limitations should be listed here or only a selection of them.

RESPONSE: Edited

32

Methods

Page 11, line 20 – It seems the sentence was restructured but the “and” ended up out of place. I think it should read “If they were, the content of the item that differed between the preprint and publication and the details of the discrepancy were recorded.”

RESPONSE: The original sentence reads as intended: “If they were, and the content of the item differed between the preprint and publication, details of the discrepancy were recorded.”

33

Discussion

Page 18, line 25 - The recommendation/suggestion that preprints with preliminary or interim data should be labelled as such is a valid point, but as it is it could be misinterpreted as a result-derived recommendation. My suggestion: “(...) were being used to publish preliminary or interim data. Preliminary or interim findings should be clearly labeled in preprints, however this study did not assess whether such description was present or not.”

RESPONSE: The discussion section interprets our findings. This is a valid interpretation of our finding that outcomes reported in journal publications were measured over a longer time frame than outcomes reported in preprints.

34

Page 19, line 5 – Preprints are likely to be posted without previous peer review in a journal, but most often we cannot be sure about this status. I do not think this distinction changes the position of your study as ‘an indirect investigation of the impact of peer-review’, but may still be worth adding.

RESPONSE: Our introduction (line 2) describes preprints as non-peer reviewed.

35

Page 19, line 41 – Similar to the comment above, I believe a better phrasing for this limitation would be that your sample could be limited to articles that were eventually published in a journal as you cannot be sure whether these preprints have been rejected in previous submissions before or after the preprint posting.

RESPONSE: As noted in the response to another reviewers, we are offering possible, but not all, reasons that preprints might not be published as journal articles.

36

Page 20, line 7 – On the same note as the previous comments, I would suggest changing to “(...) preprints that could not have benefited from peer review in the final journal of publication”. Again, I do not believe this would change any of the conclusions from these sensitivity analyses, but it is a warranted clarification to readers.

RESPONSE: This sentence was changed to: “when the sample is limited to preprints that likely had not gone through the peer review process.”

37

Page 20, line 27 – Please correct Carnerio to Carneiro.

RESPONSE: “Carnerio” was corrected to “Carneiro.”

38

Conclusion

Page 22, line 10-12 – I do not think the sentence “However, given the urgent need for valid and reliable research on COVID-19 treatment and prevention, even a few important discrepancies could impact decision making.” is a conclusion because the study did not assess the impact of the

discrepancies, but only their frequency. I would move it to the beginning of the Discussion or to the Introduction.

RESPONSE: This study provides descriptive data on the nature of discrepancies (not just their frequency) and some of these could be considered important in different contexts. For examples, failure to publish adverse effects data from a drug study could be considered important to patients. Therefore, we think this is an appropriate summary statement.

Reviewer: 4

Dr. Neil Schluger, Columbia University Medical Center

Summary: The authors have satisfactorily responded to review concerns, including my prior comments. The manuscript is substantially improved and the limitations are clearly stated. I have no further comments that need to be addressed.

VERSION 2 – REVIEW

REVIEWER	Shi, Xiaoting Yale University School of Public Health, Department of Environmental Health Sciences
REVIEW RETURNED	27-May-2021

GENERAL COMMENTS	The authors have provided detailed responses and necessary clarifications of my comments. The manuscript is substantially improved. I have no further comments that need to be addressed.
-------------------------	---

REVIEWER	Amaral, Olavo Institute of Medical Biochemistry Leopoldo de Meis Federal University of Rio de Janeiro
REVIEW RETURNED	13-Jun-2021

GENERAL COMMENTS	<p>General assessment:</p> <p>The manuscript has been improved in many minor points raised by the reviewers. Unfortunately, however, I feel that my main concern has still not been taken into consideration.</p> <p>Major point:</p> <p>- As I mentioned in my previous comments, I have nothing against the authors' decision to provide a merely descriptive account of their findings with no use of statistics. However, they have to be clearer about this point in the discussion, something that has not been addressed.</p> <p>The authors argue in their response to this point that "The generalizability of our study is addressed in the limitations". However, the strengths and limitations section (page 20-21) makes no mention at all about the descriptive nature of the study and the absence of statistical analysis to evaluate generalizability of the findings. This is clearly a limitation in my opinion; although I agree with the authors that the study "can still provide valuable qualitative information", something is clearly lost when no attempt is made to evaluate whether quantitative findings – which make up a large part of the results – can be extrapolated beyond the sample.</p> <p>The authors do mention in this section that "our focus on COVID-19 research may not be representative of other types of research...". However, they fail to acknowledge that, in the</p>
-------------------------	---

absence of any attempt to evaluate inferences on whether findings generalize beyond the sample, these findings may not be “representative” even of the same type of research. In fact, the notion of “representativeness” arguably makes little sense if, as the authors claim, their intention is merely to describe the sample rather than ask whether findings are likely to generalize to a larger population.

On a similar vein, in the last paragraph on page 17, the authors state that “more preprints reported funding source, author conflicts of interest and ethics approval than journal publications. These differences may be due to the screening requirements of medRxiv, the main source of preprints in our sample.” Nowhere do they mention that some of these small quantitative differences could also be very easily explained by chance alone.

To me, if the authors are trying to provide explanations for the quantitative findings (which is the case in this whole paragraph), they seem to be implicitly assuming that there are differences at the population level for which there is an explanation to be found. However, this may or may not be the case and, in the absence of statistics, it is very hard to judge – something that can be easily missed by an inattentive reader if the authors are not more explicit about it.

Minor points:

Methods:

Page 10, line 46: “For each journal publication and preprint, we recorded the number of outcomes reported and, whether outcomes were reported only in the preprint or journal publication, and the outcome descriptor (e.g., mortality, hospitalization, transmission, immunogenicity, harms).” This sentence is a bit truncated: I think it would read better by removing the first “and”.

Page 11, line 3: There are two commas in a sentence here.

Page 12, line 31: “Acknowledge statistically nonsignificant results from the primary outcome but emphasize the beneficial effect of treatment” . For consistency with the other items, this should be “acknowledging” and “emphasizing”

Page 14, line 3: “... all instances of discrepant reporting, even if it occurred once.” This should be “even if they occurred once”.

Page 14 – line 7: “To determine whether preprints that were posted after an article received peer review.” The claim that preprints were posted after peer review is likely but not certain – I would change to “likely posted”.

Results:

Page 15, line 49: “Of the 67 studies, 23 (34%) had no discrepancies in results reporting between preprints and journal publications (Table 2). Twenty-three studies had outcomes that were missing from either the preprint or the journal publication.” I previously raised the point that both categories do not seem to add to the total number of studies, to which the authors responded that the section also describes categories that account for the other studies. However, the remaining categories (“Fifteen (22%) studies had at least one outcome that was included in the journal

	<p>publication, but not the preprint; 8 (12% had at least one outcome that was reported in the preprint only”) are subcategories of the 23 studies with outcomes missing from either the preprint or journal publication, so the response doesn’t seem to hold.</p> <p>My best guess here to explain the apparent conflict is that “no discrepancies in results reporting” includes discrepancies other than missing outcomes: thus, the 20 missing studies are those with no outcomes missing from either version, but with other discrepancies. Nevertheless, the way that these results are worded is clearly not very intuitive, especially as Table 2 (mentioned at the end of the first sentence) is strictly about discrepancies in outcomes. The authors should consider rewording (as they seem to have gotten confused themselves by it when responding to my point).</p> <p>Note that this point also applies to the abstract (Page 4, line 29).</p> <p>Page 16, line 29: “The types of discrepancies were variable, although journal publications consistently included additional statistical analyses and subgroup analyses compared to preprints.” I still find this sentence misleading, considering that additional analyses were present in a minority of preprint-journal pairs. The authors replied that “by consistent, we mean that these were the most common types of discrepancies in these categories”, but this is not what “consistently” is usually meant to imply. If this what they mean, they should consider substituting “consistently” by “most commonly” or something similar.</p> <p>Table 1: Page 30, line 14. “Different numbers of patients recruited, but same number randomized; 284 patients included in preprint, 267 in journal publication”. I found this a bit confusing even after the explanation. If “included” in the second part of the sentence refers to the different numbers of recruited patients, perhaps it could be changed to “recruited”?</p> <p>Table 3: Page 33, line 50. There is an indent mark followed by nothing after it in the last column.</p> <p>Discussion: Page 19, Line 41. “Or, our sample could be limited to articles that had not been rejected by a journal.” Although the authors’ explanation for using “could be” here seems OK for “could be limited to studies that the authors deemed of high enough quality”, I still see no point for its use here. The sample clearly is limited to published articles that were not rejected.</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Xiaoting Shi, Yale University School of Public Health Comments to the Author:

The authors have provided detailed responses and necessary clarifications of my comments. The manuscript is substantially improved. I have no further comments that need to be addressed.

RESPONSE: none needed

Reviewer: 2

Dr. Olavo Amaral

Comments to the Author:

General assessment:

The manuscript has been improved in many minor points raised by the reviewers. Unfortunately, however, I feel that my main concern has still not been taken into consideration.

Major point:

- As I mentioned in my previous comments, I have nothing against the authors' decision to provide a merely descriptive account of their findings with no use of statistics. However, they have to be clearer about this point in the discussion, something that has not been addressed.

The authors argue in their response to this point that "The generalizability of our study is addressed in the limitations". However, the strengths and limitations section (page 20-21) makes no mention at all about the descriptive nature of the study and the absence of statistical analysis to evaluate generalizability of the findings. This is clearly a limitation in my opinion; although I agree with the authors that the study "can still provide valuable qualitative information", something is clearly lost when no attempt is made to evaluate whether quantitative findings – which make up a large part of the results – can be extrapolated beyond the sample.

The authors do mention in this section that "our focus on COVID-19 research may not be representative of other types of research...". However, they fail to acknowledge that, in the absence of any attempt to evaluate inferences on whether findings generalize beyond the sample, these findings may not be "representative" even of the same type of research. In fact, the notion of "representativeness" arguably makes little sense if, as the authors claim, their intention is merely to describe the sample rather than ask whether findings are likely to generalize to a larger population. On a similar vein, in the last paragraph on page 17, the authors state that "more preprints reported funding source, author conflicts of interest and ethics approval than journal publications. These differences may be due to the screening requirements of medRxiv, the main source of preprints in our sample." Nowhere do they mention that some of these small quantitative differences could also be very easily explained by chance alone.

To me, if the authors are trying to provide explanations for the quantitative findings (which is the case in this whole paragraph), they seem to be implicitly assuming that there are differences at the population level for which there is an explanation to be found. However, this may or may not be the case and, in the absence of statistics, it is very hard to judge – something that can be easily missed by an inattentive reader if the authors are not more explicit about it.

RESPONSE: We strongly object to stating that the descriptive nature of our study is a limitation. Strengths and limitations are context-specific and exist in relation to a particular study design; it is not a limitation of study that it did not include elements of a different study design. We note that BMJ Open has an established precedent of publishing descriptive studies that do not indicate the descriptive design nor lack of inferential statistics as a limitation. Here are some recent examples from our authorship team, but there are others:

<https://bmjopen.bmj.com/content/bmjopen/6/4/e010075.full.pdf>

<https://bmjopen.bmj.com/content/10/11/e040541>

<https://bmjopen.bmj.com/content/9/2/e024928>

<https://bmjopen.bmj.com/content/7/6/e016701>

Relying solely on descriptive analysis is, in fact, appropriate given our study design and as we planned in our published protocol. We do not believe that the absence of a formal statistical analysis can be a study limitation when this was not planned.

We also object to removing the text that summarizes our findings. We summarize our main findings based on the empirical results and we do not make any statements about statistical significance, correlation, or causation. Our interpretation of findings aim to provide readers with some understanding of why we may have observed the results we present. As this was an empirical analysis, we can make statements about what we found, which in the context of a major preprint publisher and a global repository of studies used in Cochrane reviews, have practical and policy implications. These implications exist for this sample and the publishing and peer-review processes even in the absence of statistical tests that would allow generalization to a population. Our author team includes researchers with diverse disciplinary backgrounds, including training in a wide variety of experimental, observational, qualitative and social science methods we do not think that our findings will be misunderstood by BMJ Open readers.

Minor points:

Methods:

Page 10, line 46: “For each journal publication and preprint, we recorded the number of outcomes reported and, whether outcomes were reported only in the preprint or journal publication, and the outcome descriptor (e.g., mortality, hospitalization, transmission, immunogenicity, harms).”. This sentence is a bit truncated: I think it would read better by removing the first “and”.

RESPONSE: deleted

Page 11, line 3: There are two commas in a sentence here.

RESPONSE: deleted

Page 12, line 31: “Acknowledge statistically nonsignificant results from the primary outcome but emphasize the beneficial effect of treatment” . For consistency with the other items, this should be “acknowledging” and “emphasizing”

RESPONSE: changed

Page 14, line 3: “... all instances of discrepant reporting, even if it occurred once.” This should be “even if they occurred once”.

RESPONSE: changed

Page 14 – line 7: “To determine whether preprints that were posted after an article received peer review.” The claim that preprints were posted after peer review is likely but not certain – I would change to “likely posted”.

RESPONSE: The articles were definitely posted, not likely. We have edited to read: “To determine whether preprints that were posted after an article had likely received peer review influenced the number of discrepancies...”

Results:

Page 15, line 49: "Of the 67 studies, 23 (34%) had no discrepancies in results reporting between preprints and journal publications (Table 2). Twenty-three studies had outcomes that were missing from either the preprint or the journal publication." I previously raised the point that both categories do not seem to add to the total number of studies, to which the authors responded that the section also describes categories that account for the other studies. However, the remaining categories ("Fifteen (22%) studies had at least one outcome that was included in the journal publication, but not the preprint; 8 (12%) had at least one outcome that was reported in the preprint only") are subcategories of the 23 studies with outcomes missing from either the preprint or journal publication, so the response doesn't seem to hold.

My best guess here to explain the apparent conflict is that "no discrepancies in results reporting" includes discrepancies other than missing outcomes: thus, the 20 missing studies are those with no outcomes missing from either version, but with other discrepancies. Nevertheless, the way that these results are worded is clearly not very intuitive, especially as Table 2 (mentioned at the end of the first sentence) is strictly about discrepancies in outcomes. The authors should consider rewording (as they seem to have gotten confused themselves by it when responding to my point).

Note that this point also applies to the abstract (Page 4, line 29).

RESPONSE: The numbers in this section and the abstract are accurate. As stated in our methods (page 10), we looked at discrepancies in results reporting in 2 ways: 1) number of outcomes reported in preprints and journal publications and, for outcomes reported in both preprints and journal publications, 2) components of results reporting. We could only look at discrepancies in components of outcome reporting for outcomes that were present in both the preprint and final publication. Table 2 shows discrepancies in the number of outcomes reported and table 3 shows discrepancies in the components of outcome reporting. The text in sentence 1 of "Results Reporting" was not properly edited in a previous revision and has been changed to: "Of the 67 studies, 23 (34%) had no discrepancies in the number of outcomes reported between preprints and journal publications (Table 2)."

Page 16, line 29:

"The types of discrepancies were variable, although journal publications consistently included additional statistical analyses and subgroup analyses compared to preprints." I still find this sentence misleading, considering that additional analyses were present in a minority of preprint-journal pairs. The authors replied that "by consistent, we mean that these were the most common types of discrepancies in these categories", but this is not what "consistently" is usually meant to imply. If this what they mean, they should consider substituting "consistently" by "most commonly" or something similar.

RESPONSE: We were accused in previous peer review comments of making a statements about statistical significance by using terms such as "more frequently," so used the term "consistently" instead. We think "more commonly" is an accurate descriptive term and does not suggest a statistical difference. We have now used this term.

Table 1:

Page 30, line 14. "Different numbers of patients recruited, but same number randomized; 284 patients included in preprint, 267 in journal publication". I found this a bit confusing even after the explanation. If "included" in the second part of the sentence refers to the different numbers of recruited patients, perhaps it could be changed to "recruited"?

RESPONSE: changed

Table 3:

Page 33, line 50. There is an indent mark followed by nothing after it in the last column.

RESPONSE: deleted

Discussion:

Page 19, Line 41. "Or, our sample could be limited to articles that had not been rejected by a journal." Although the authors' explanation for using "could be" here seems OK for "could be limited to studies that the authors deemed of high enough quality", I still see no point for its use here. The sample clearly is limited to published articles that were not rejected.

RESPONSE: this was added in response to another peer reviewer and we have not changed

Reviewer: 1

Competing interests of Reviewer: I received fundings from the China Scholarship Council in the past 24 months.

Reviewer: 2

Competing interests of Reviewer: I have acted as an advocate for preprint usage in the life sciences, including as an ambassador for ASAPbio.