






BMJ Open Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation

Lisa Bero ,¹ Rosa Lawrence,² Louis Leslie,² Kellia Chiu,³ Sally McDonald,³ Matthew J Page ,⁴ Quinn Grundy ,⁵ Lisa Parker ,⁶ Stephanie Boughton,⁷ Jamie J Kirkham ,⁸ Robin Featherstone⁷

To cite: Bero L, Lawrence R, Leslie L, *et al*. Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation. *BMJ Open* 2021;**11**:e051821. doi:10.1136/bmjopen-2021-051821

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-051821>).

Received 30 March 2021
Accepted 23 June 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Lisa Bero;
LISA.BERO@CUANSCHUTZ.EDU

ABSTRACT

Objective To compare results reporting and the presence of spin in COVID-19 study preprints with their finalised journal publications.

Design Cross-sectional study.

Setting International medical literature.

Participants Preprints and final journal publications of 67 interventional and observational studies of COVID-19 treatment or prevention from the Cochrane COVID-19 Study Register published between 1 March 2020 and 30 October 2020.

Main outcome measures Study characteristics and discrepancies in (1) results reporting (number of outcomes, outcome descriptor, measure, metric, assessment time point, data reported, reported statistical significance of result, type of statistical analysis, subgroup analyses (if any), whether outcome was identified as primary or secondary) and (2) spin (reporting practices that distort the interpretation of results so they are viewed more favourably).

Results Of 67 included studies, 23 (34%) had no discrepancies in results reporting between preprints and journal publications. Fifteen (22%) studies had at least one outcome that was included in the journal publication, but not the preprint; eight (12%) had at least one outcome that was reported in the preprint only. For outcomes that were reported in both preprints and journals, common discrepancies were differences in numerical values and statistical significance, additional statistical tests and subgroup analyses and longer follow-up times for outcome assessment in journal publications.

At least one instance of spin occurred in both preprints and journals in 23/67 (34%) studies, the preprint only in 5 (7%), and the journal publications only in 2 (3%). Spin was removed between the preprint and journal publication in 5/67 (7%) studies; but added in 1/67 (1%) study.

Conclusions The COVID-19 preprints and their subsequent journal publications were largely similar in reporting of study characteristics, outcomes and spin. All COVID-19 studies published as preprints and journal publications should be critically evaluated for discrepancies and spin.

INTRODUCTION

Preprints have been advocated as a means for rapid sharing and updating of research

Strengths and limitations of this study

- We examine two critical threats to research integrity—components of outcome reporting and the presence of spin—in COVID-19 studies on treatment or prevention published as preprints and journal publications.
- We selected studies from the Cochrane COVID-19 Register rather than conducting a literature search to optimise the identification of COVID-19 clinical research that is useful for systematic reviews.
- We may have identified a different number of discrepancies if we compared later versions of the preprint, rather than the first version, with the journal publication.
- Although clinically important, our focus on COVID-19 research may not be representative of other types of research published as preprints, then journal publications.
- We limited our sample to preprints which authors submitted to journals and that were published.

findings, which could be particularly valuable during a pandemic.¹ Preprints are non-peer-reviewed postings of research articles. Preprints have been a common form of publication in the natural sciences for decades, and more recently in the life sciences. In 2019, *BMJ*, *Yale* and *Cold Spring Harbor Laboratory* launched medRxiv, a preprint server dedicated to clinical and health sciences research.

In April 2020, medRxiv published between 50 and 100 COVID-19-related preprints daily.¹ The accelerated pace of research related to COVID-19 has increased the potential impact and risk of using preprints. Widespread public dissemination of preprints may spread misinformation.² A study comparing 34 preprints and 62 publications about therapies for COVID-19 found that publications had significantly more citations than the preprints (median of 22 vs 5.5 citations; $p =$



0.01), but there were no significant differences for attention and online engagement metrics.³

Most preprint servers conduct some type of screening prior to posting, commonly related to the scope of the article, plagiarism, and compliance with legal and ethical requirements,⁴ but preprints have not been peer reviewed and may not meet the methodological and reporting requirements of a journal. A review of the medRxiv preprint server 1 year after its launch found that 9967 of 11 164 (89%) of submissions passed screening.⁵ It is not clear whether or how preprint servers might screen for quality of results reporting or spin.^{6,7} Spin refers to specific reporting practices that distort the interpretation of results so that results are viewed more favourably.

Preliminary studies suggest that reporting discrepancies may exist between preprints and subsequent publications. However, there has been no systematic assessment of results reporting or spin between preprints and their final journal publications. Carneiro *et al* counted reported items from a checklist meant to cover common points from multiple reporting guidelines and found reporting quality to be marginally higher in journal articles, both in a set of bioRxiv preprints matched to their journal publication (n=56 article/group) and in an unmatched set (n=76 articles/group).⁸ An analysis of preprints from arXiv, a primarily physics/mathematics preprint server, and their journal publications using text comparison algorithms found little difference between preprints and published articles.⁹ However, an analysis of medRxiv and bioRxiv preprints related to COVID-19 pharmacological interventions found that only 24% (23/97) of preprints were published in a journal within 0–98 days (median: 42.0 days). Among these, almost half (11/23, 48%) had modifications in the title or results section, although the nature of these modifications is not described.¹⁰ An analysis of spin in preprints and journal publications for COVID-19 trials found a single difference between two matched pairs of preprints and their journal publications: the discussion of limitations in the abstract. Limitations were discussed in the abstract of one article, but not in its accompanying preprint.¹¹ An analysis of 66 preprint–article pairs of COVID-19 studies found 38% had changes in study results, such as a numeric change in HR or a change in p value, and 29% had changes in abstract conclusions, most commonly from ‘positive without reporting uncertainty’ in the preprint to ‘positive with reporting of uncertainty’ in the article.¹²

The trustworthiness and validity of scientific publications, even after peer review, are weakened by a variety of problems.^{13–14} Selective and incomplete results reporting^{15–16} and spin^{17–18} are two critical threats, especially for clinical studies of treatment or prevention. These reporting practices could be particularly dangerous for users of COVID-19 research as they can inflate the efficacy of interventions and underestimate harms. Given the high prevalence, visibility, and potentially rapid implementation of COVID-19 research published as preprints, this study is the first to compare components of outcome

reporting and the presence of spin in COVID-19 studies on treatment or prevention that are published both as preprints and journal publications.

METHODS

The protocol for this study was registered in the Open Science Framework.¹⁹

Data source and search strategy

We sampled studies from the Cochrane COVID-19 Study Register (<https://COVID-19.cochrane.org/>), a freely available, continually updated, annotated reference collection of human primary studies on COVID-19, including interventional, observational, diagnostic, prognostic, epidemiological and qualitative designs. The register is ‘study based’, meaning references to the same study (eg, press releases, trial registry records, preprints, journal preproofs, journal final publications, retraction notices) are all linked to a single study identifier. References are screened for eligibility to determine if they are primary studies (eg, not opinion pieces or narrative reviews). Data sources for the Cochrane COVID-19 Study Register at the time of the search included ClinicalTrials.gov, the International Clinical Trials Registry Platform, PubMed, medRxiv and Embase.com. The Cochrane register prioritises medRxiv as a preprint source because an internal sensitivity analysis in May 2020 showed that 90% (166/185) of the preprints that were eligible for systematic reviews came from this source. The register also includes preprint records sourced from PubMed.

All studies in the register are classified by study design (interventional, observational, modelling, qualitative, other or unclear) and research aim (prevention, treatment and management, diagnostic/prognostic, epidemiology, health services research, mechanism, transmission, other). Studies may be classified as having multiple research aims. Four searches using the register’s search filters for study reference types (preprints and journal articles) and study characteristics (study type and study aim) were used to retrieve references with a study aim of (a) treatment and management or (b) prevention and classified as interventional or observational (see OSF (Open Science Framework) project for the complete search strategies: (<https://osf.io/8qfby/>)). As the register is updated daily, we repeated the search. The Cochrane COVID-19 Study Register was first searched by RF on 13 October, and updated on 29 October 2020. The results were exported to Excel and duplicates manually identified. The searches identified 297 references for 117 studies, with 67 (21 interventional, 46 observational) that met our inclusion and exclusion criteria for study selection (figure 1).

Inclusion and exclusion criteria for study selection

We included studies of COVID-19 treatment or prevention identified in the search that had both a posted preprint and final journal publication.

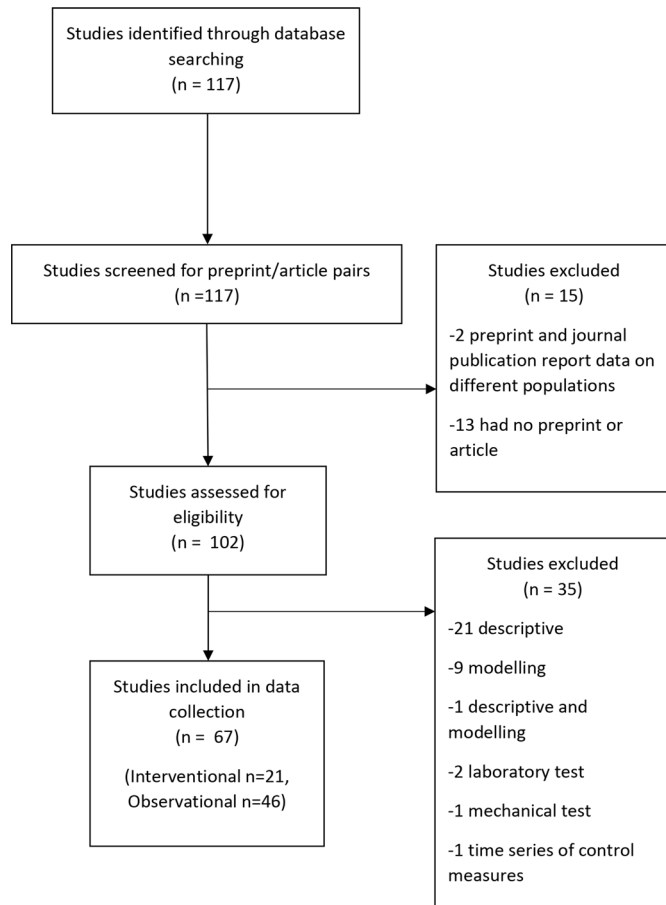


Figure 1 Flowchart of study inclusion.

We included studies with aims of diagnosis/prognosis, epidemiology, health services research, mechanism, transmission and other if they also had an aim coded as (a) treatment and management or (b) prevention. We excluded modelling studies, qualitative studies and studies that reported only descriptive data (eg, demographic characteristics). We screened all records for each included study to identify posted preprints and journal publications from each study. We excluded duplicates and records for protocols, trial registries, commentaries, letters to the editor, news articles and press releases. We excluded records that did not report results and non-English records.

We compared the preprint and journal publication for each included study. In the case of multiple preprints or journal publications reporting study results, we selected the first preprint version and the final journal publication that reported on similar study populations. This was to ensure that the preprint version evaluated in our study had not been altered in response to any comments, which could constitute a form of peer review, and that it was representative of the version most likely to be seen by clinicians, journalists and other research users as new research became available.

Data extraction

Ten investigators (LB, SLB, KC, QG, JJK, LL, RL, SMC, LP and MJP) working independently in pairs extracted

data from the included studies. Discrepancies in data extraction were resolved by consensus. If agreement could not be reached, an investigator who was not part of the coding pair resolved the discrepancies. All extracted data from the included studies were stored in REDCap, a secure web-based application for the collection and management of data.²⁰ We extracted data from both the medRxiv page and PDF for preprints and the online publication or PDF for journal articles, referring to the PDF if information differed. We extracted data on results reporting, presence of spin and study characteristics as described below.

Study characteristics

For each preprint, we recorded the earliest posting date; for each journal publication we extracted the submitted/received, reviewed, revised, accepted and published date(s), where available.

From each journal publication, we extracted: authors, title, funding source, author conflicts of interests, ethics approval, country of study and sample size. For the accompanying preprint, we determined if these study characteristics were also reported. If they were, and the content of the item differed between the preprint and publication, details of the discrepancy were recorded. In addition, we recorded discrepancies between the preprint and journal publication in demographic characteristics of study participants (eg, sex, race/ethnicity, diagnosis), discussion of limitations (regardless of whether there was a labelled limitations section or not), and tables and figures.

Primary outcomes

Our primary outcome measures were (1) discrepancies in results reporting between preprints and journal publications and (2) presence and type of spin in preprints and journal publications.

Results reporting

We collected data on discrepancies in (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. For each journal publication and preprint, we recorded the number of outcomes reported, whether outcomes were reported only in the preprint or journal publication, and the outcome descriptor (eg, mortality, hospitalisation, transmission, immunogenicity, harms).

For outcomes that were reported in both preprints and journal publications, we collected data on components of outcome reporting based on recommendations for clinical study results reporting.^{16,21} We recorded whether there were discrepancies between any components of outcome reporting between journal publications and preprints. We extracted the text relevant to each discrepancy:

- ▶ Measure (eg, PCR test).
- ▶ Metric (eg, mean change from baseline, proportion of people).

- ▶ Time point at which the assessment was made (eg, 1 week after starting treatment).
- ▶ Numerical values reported (eg, effect estimate and measure of precision).
- ▶ Statistical significance of result (as reported).
- ▶ Type of statistical analysis (eg, regression, χ^2 test).
- ▶ Subgroup analyses (if any).
- ▶ Whether outcome was identified as primary or secondary.

Spin

Studies have used a variety of methods to measure spin in randomised controlled trials and observational studies.¹⁷ Based on our previously developed typology of spin derived from a systematic review of spin studies,¹⁷ we developed and pretested a coding tool for spin that can be applied to both interventional and observational studies of treatment or prevention. In the context of research on treatment or prevention of COVID-19, the most meaningful consequences of spin are overinterpretation of efficacy and underestimation of harms. Therefore, our tool emphasises these manifestations of spin. We searched the abstracts and full text of each preprint and journal publication for three primary categories of spin, and accompanying subcategories:

1. Inappropriate interpretation given study design.
 - Claiming causality in non-randomised studies.
 - Interpreting a lack of statistical significance as equivalence.
 - Interpreting a lack of statistical significance of harm measures as safety.
 - Claim of any significant difference despite lack of statistical test.
 - Other.
2. Inappropriate extrapolations or recommendations.
 - Suggestion that the intervention or exposure is more clinically relevant or useful than is justified given the study design.
 - Recommendation made to population groups/contexts outside of those investigated.
 - (Observational) Expressing confidence in an intervention or exposure without suggesting the need for further confirmatory studies.
 - Other.
3. Selectively focusing on positive results or more favourable data presentation.
 - Discussing only significant (non-primary) results to distract from non-significant primary results.
 - Omitting non-significant results from abstract/discussion/conclusion.
 - Claiming significant effects for non-significant results.
 - Acknowledging statistically non-significant results from the primary outcome but emphasising the beneficial effect of treatment.
 - Describing non-significant results as ‘trending towards significance’.

- Mentioning adverse effects in the abstract/discussion/conclusion but minimising their potential effect or importance.
- Misleading description of study design as one that is more robust.
- Use of linguistic spin.
- Other.

Analysis

We report the frequency and types of discrepancies in study characteristics and results reporting between preprints and journal publications. We report the proportion of preprints and journal publications with spin and the types of spin. We iteratively analysed the text descriptions of discrepancies identified; we grouped descriptions into common categories, while still accounting for all instances of discrepant reporting, even if they only occurred once, to demonstrate the range of the phenomenon.

To determine whether preprints that were posted after an article had likely received peer review influenced the number of discrepancies, we conducted a post hoc sensitivity analysis by removing seven studies where the preprint was posted up to 7 days before the revision, acceptance or publication dates of the journal publication.

The OSF project linked to our protocol (<https://osf.io/5ru8w/>) provides our protocol modifications, list of included preprints and journal publications, data dictionary, and dataset.

Patient and public involvement

No patient involvement.

RESULTS

Study characteristics

Of the 67 included studies, 57 were studies of treatment and management, 9 of prevention and 1 of both. The preprints and journal publications were published between 1 March 2020 and 30 October 2020 with a mean time between preprint and journal publication of 65.4 days (range 0–271 days). The topics of the studies varied and included effects of clinical and public health interventions, associations of risk factors with COVID-19 symptoms, and ways to improve implementation of public health measures, such as social distancing. Almost a third of studies (21/67, 31%) were conducted in the USA, followed by Italy and Spain (n=6, 9% each), and China (n=5, 7%). The majority of studies reported public or non-profit funding sources (n=32, 49%) or that no funding was provided (n=24, 36%). Over half the studies also reported that the authors had no conflicts of interest (n=37, 53%).

Discrepancies in study characteristics

Table 1 shows discrepancies in study characteristics reported in preprints and journal publications. The table shows whether each study characteristic was reported

Table 1 Discrepancies in study characteristics (n=67 studies)

| Characteristic | No discrepancies | | | Discrepancies | | |
|---|--|--|--|---|---|---|
| | Characteristic reported in both preprint and journal publication | Characteristic reported in neither preprint or journal publication | Characteristic reported in preprint only | Characteristic reported in journal publication only | Characteristic reported in both preprint and journal publication, but with discrepancies in content | Examples of discrepancies* |
| Title | 47 (70%) | 0 (0%) | 0 (0%) | 0 (0%) | 20 (30%) | <ul style="list-style-type: none"> ▲ Preprint includes study design in the title (n=4) ▲ Journal publication includes study design in the title (n=5) ▲ Change in study design description (n=5) ▲ Change in population description (n=3) ▲ Change in location description in both (n=3) |
| Authors | 49 (73%) | 0 (0%) | 0 (0%) | 0 (0%) | 18 (27%) | <ul style="list-style-type: none"> ▲ Additional author(s) in preprint (n=3) ▲ Additional author(s) in journal publication (n=9) ▲ Change in author order (n=6) ▲ Change in spelling, wording, or order of author first/last names (n=2) |
| Disclosed funding source | 44 (66%) | 3 (4%) | 8 (12%) | 2 (3%) | 10 (15%) | <ul style="list-style-type: none"> ▲ Additional funding sources in journal publication (n=4) ▲ Funding statement in preprint provides more detail (n=1) ▲ Funding statement in journal publication provides more detail (n=2) |
| Conflict of interest disclosure statement | 50 (75%) | 1 (1%) | 5 (8%) | 1 (1%) | 10 (15%) | <ul style="list-style-type: none"> ▲ Additional conflicts reported in journal publication (n=8) ▲ Additional conflicts reported in preprint (n=1) ▲ Additional detail included in journal publication (n=2) |
| Ethics approval | 59 (88%) | 3 (5%) | 1 (1%) | 0 (0%) | 4 (6%) | <ul style="list-style-type: none"> ▲ Preprint contains approval number but journal publication does not (N=1); preprint states approval was waived and journal publication states it was not needed (n=1); preprint contains no information on ethics approval, while journal publication describes the approvals (n=1); preprint states consent was approved prior to sample collection while article states it was approved from next of kin (n=1) |
| Location of study | 63 (94%) | 4 (6%) | 0 (0%) | 0 (0%) | 0 (0%) | |

Continued



Table 1 Continued

| Characteristic | No discrepancies | | Discrepancies | | |
|---------------------------|--|--|--|---|---|
| | Characteristic reported in both preprint and journal publication | Characteristic reported in neither preprint or journal publication | Characteristic reported in preprint only | Characteristic reported in journal publication only | Characteristic reported in both preprint and journal publication, but with discrepancies in content Examples of discrepancies* |
| Number of participants | 61 (91%) | 0 (0%) | 0 (0%) | 0 (0%) | 6 (9%) ▲ Journal publication has larger analytic sample size than preprint (n=2); Journal publication has smaller analytic sample size than preprint (n=1); different numbers of patients recruited, but same number randomized/randomised; 284 patients recruited in preprint, 267 in journal publication (n=1); numbers do not match for any sampling or analysis (n=1); typographical error (n=1) |
| Participant demographics | 38 (58%) | 3 (4%) | 0 (0%) | 1 (1%) | 25 (37%) ▲ Journal publication includes additional demographic categories (n=10) ▲ Preprint includes additional demographic categories (n=4) ▲ Preprint and journal publication report different values for the same demographic characteristics (n=11) ▲ Demographic data reported using different metrics (n=6) |
| Tables and Figures | 18 (27%) | 0 (0%) | 0 (0%) | 0 (0%) | 49 (73%) ▲ Journal publication includes additional tables/figures (n=25) ▲ Preprint includes additional tables/figures (n=10) Additional data in journal publication tables (n=14) Additional data in preprint tables/figures (n=6) ▲ Change in order of tables/figures (n=4) ▲ Change in metrics (eg, mean vs median) (n=15) ▲ Change in labels (n=5) ▲ Numbers reported differed (n=16) |
| Discussion of limitations | 27 (40%) | 7 (11%) | 0 (0%) | 2 (3%) | 31 (46%) ▲ More limitations listed in journal publication than preprint (n=28) ▲ More limitations listed in preprint than journal publication (n=1) |

*Ns do not add to number of discrepancies between preprints and journal publications, as some studies could have more than one discrepancy and not all discrepancies have been included as examples.

or not; if a study characteristic was reported in both the preprint and journal publications, discrepancies in content are described. More preprints than journal publications reported funding source, author conflicts of interest and ethics approval; more journal publications than preprints reported participant demographics and study limitations. In all categories, most discrepancies occurred in the content of items that were reported, rather than in whether the item was present or not. For example, journal publications contained additional information on funding sources, conflicts of interest, demographic characteristics and limitations, as well as more tables and figures compared to preprints (table 1).

Results reporting

Of the 67 studies, 23 (34%) had no discrepancies in the number of outcomes reported between preprints and journal publications (table 2). Twenty-three studies had outcomes that were missing from either the preprint or the journal publication. Overall, 15 (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. The included studies had multiple outcomes. The majority of studies with missing reported outcomes (16/23, 70%) had one outcome missing from either the preprint or journal

Table 2 Discrepancies in Number of Outcomes Reported (N=67 studies)

| Type of discrepancy | Number (%) of studies with at least one outcome that was reported only in the preprint or journal publication (n=67) | Number and description of outcomes across all studies that were reported only in the preprint or journal publication |
|--|--|--|
| Outcome reported in journal publication only | 15 (22%) | N=19 (numbering indicates unique studies, lettering indicates outcomes from the same study) (1a) Treatment-associated toxicities (1b) Adverse reactions (2) Survival at intensive care unit (ICU) discharge (3) Creatine phosphokinase (4) Radiographic scale for acute respiratory distress syndrome (5) Time to negative swab (6) Time to reverse transcription-PCR negativity (7) Clinical outcomes at discharge (8) Ventilator status of those remaining hospitalised at end of follow-up (9a) Secondary composite—cardiovascular complications (9b) Acute renal failure (10) Creatinine phosphokinase (11) Sequential organ failure assessment score (12) Length of stay (13) WHO Clinical Progression Scale (14a) sCD14 levels related to corticoid treatment (14b) Hospital Stay (14c) Onset of symptoms (15) Mechanical ventilation or all-cause mortality at 21 days |
| Outcome reported in preprint only | 8 (12%) | N=17 (numbering indicates unique studies, lettering indicates outcomes from the same study) (1a) Oxygen support need (1b) Invasive mechanical ventilation need (1c) ICU need (1d) Need for inotropics (1e) Naso/oropharyngeal swab viral clearance (2a) Final lymphocyte (cell/mm ³) (2b) Final C reactive protein (CRP) (mg/L) (3a) Negative conversion of SARS-CoV-2 by 28 days (3b) Negative conversion rate at 4-day, 7-day, 10-day, 14-day or 21-day (3c) Changes of CRP values and blood lymphocyte count (3d) Rate of symptoms alleviation within 28-day (3e) Safety endpoints (4) QTc ≥470 ms (5) Cumulative virus clearance rate vs different antiviral regimes in (a) all patients and (b) patients with moderate illness (6) Adverse events (7) Composite cardiovascular and renal failure (8) Nosocomial infections |



publication. However, two studies had five outcomes missing from the journal publication, but reported in the preprint only.^{22–25} As described in [table 2](#), these omissions included important clinical or harm outcomes. For example, one preprint omitted toxicity outcomes that were reported in the journal publication.^{26 27}

[Table 3](#) shows the types of discrepancies in components of results reporting. We report the number of studies that had at least one discrepancy and, because studies have multiple outcomes, the number of discrepancies across all outcomes in the 67 studies. The most frequent types of discrepancies between outcomes reported in both preprints and journal publications were in the numerical values reported, statistical tests performed, subgroup analyses conducted, statistical significance reported and timepoint at which the outcome was assessed ([table 3](#)). The types of discrepancies were variable, although journal publications more commonly included additional statistical analyses and subgroup analyses compared with preprints. Journal publications more frequently reported outcomes measured over a longer time period than preprints.

Spin

At least one instance of spin occurred in the preprint, journal publication, or both in 30 (45%) of the 67 studies. Spin occurred in both preprints and journal publications in 23/67 (34%) studies, the preprint only in 5 (7%) studies, and the journal publications only in 2 (3%) studies ([table 4](#)). Spin, in any category, was removed between the preprint and journal publication in 5/67 (7%) studies; but added between the preprint and journal publication in 1 (1%) study.

[Table 4](#) shows the categories of spin that occurred in preprints and their accompanying journal publications. Overall, 13 of 67 (19%) studies had changes in the type of spin present in the preprint versus the journal publication; 8 (12%) studies had at least one additional type of spin present in the preprint, 2 (3%) studies had at least one additional type of spin present in the journal publication. Inappropriate extrapolation or recommendations was the most frequently occurring type of spin in both preprints and journal publications (11/67, 16% of studies). This type of spin and inappropriate interpretation given the study design occurred more frequently in preprints than journal publications.

An example of inappropriate interpretation was found in both the preprint and journal publication for an open-label non-randomised trial: the study investigated the effect of hydroxychloroquine (and in combination with azithromycin) on SARS-CoV-2 viral load. They found a statistically significant viral load reduction at day 6; however, despite the small sample size and non-randomised study design, they concluded that their findings were ‘so significant’ and recommended that ‘patients with COVID-19 be treated with hydroxychloroquine and azithromycin to cure their infection and to limit the transmission of the virus to other people in order to curb the

spread of COVID-19 in the world’.^{28 29} An example of inappropriate extrapolation or recommendations that occurred in both the preprint and journal publication is a study that recommended specific policy approaches that were not tested in the study: ‘The UK will shortly enter a new phase of the pandemic, in which extensive testing, contact tracing and isolation will be required to keep the spread of COVID-19. For this to succeed, adherence must be improved’.^{30 31} This observational study aimed to identify factors associated with individuals’ adherence to self-isolation and lockdown measures; the authors did not aim to investigate public adherence to testing recommendations or contact tracing, nor test their efficacy.

Sensitivity analysis

The mean time between preprint posting and journal article publication was 65.4 days (range 0–271) (online supplemental table S1). No preprints were posted after the revision, acceptance or publication dates for the accompanying journal publication. One preprint was posted the same date as the publication date. Discrepancies in study characteristics, outcome reporting and spin changed minimally when the analyses were conducted after removing seven studies where the preprint was posted up to 7 days before the revision, acceptance or publication dates of the journal publication (online supplemental table S2–S4).

DISCUSSION

Principal findings

Discrepancies between results reporting in preprints and their accompanying journal publications were frequent, but most often consisted of differences in content rather than a complete lack of reporting. Although infrequent, some outcomes that were not reported would have provided information that is critical for clinical decision-making, such as clinical or harm outcomes that appeared only in the journal publication. The finding that outcomes reported in journal publications were measured over a longer time frame than outcomes reported in preprints indicates that the preprints were being used to publish preliminary or interim data. Preliminary or interim findings should be clearly labelled in preprints.

Although almost half of the preprints and journal publications contained spin, there was no clear difference in the types of spin. Spin is an enduring problem in the medical literature.¹⁷ Our findings suggest that the identification and prevention of spin during journal peer review and editorial processes needs further improvement.

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. When reported in both, journal publications included more detailed information on funding source, conflicts of interest of authors, and demographics of the population studied. Journal publications also included more tables and figures,

Table 3 Discrepancies in components of results reporting for outcomes reported in both preprints and journal publications (N=67 studies; 258 outcomes)

| Type of discrepancy | Number (%) of studies with at least one discrepancy between the preprint and journal publication (n=67) | Number (%) of outcomes across all studies that were discrepant between the preprint and journal publication (n=258) | Descriptive examples* |
|---|---|---|---|
| Outcome measurement | 6 (9%) | 8 (3%) | <ul style="list-style-type: none"> ▶ Journal publication contains more detail on how outcome was measured compared with preprint (n=3) ▶ Journal publication reports an additional or different measurement than the one used for the same outcome in the preprint (eg, preprint reports four adverse events, journal publication reports 12) (n=4) |
| Units of measurement | 3 (4%) | 3 (1%) | <ul style="list-style-type: none"> ▶ For example, journal publication reports events, total and percentage for mortality, preprint reports only percentage; median (IQR) reported in journal publication, mean (SD) in preprint |
| Timepoint assessment was made | 10 (15%) | 24 (9%) | <ul style="list-style-type: none"> ▶ Journal publication reports outcomes measured over a longer timepoint than preprint (n=13) ▶ Journal publication reports additional interim time points compared with preprint (n=3) |
| Numerical values reported | 24 (36%) | 52 (20%) | <ul style="list-style-type: none"> ▶ Differences in number of events or measurement values reported (n=17) ▶ Differences in numbers of participants or denominators (n=5) ▶ More adverse events reported in journal publication than preprint (n=4) |
| Finding of statistical significance | 11 (16%) | 16 (6%) | <ul style="list-style-type: none"> ▶ Different p-value reported with no change in significance (n=3) ▶ Different p-value reported with change in significance; significant result reported in journal publication (n=1) ▶ In multivariate models, journal publication and preprint report different variables as being statistically significant (n=2) |
| Statistical tests performed | 17 (25%) | 31 (12%) | <ul style="list-style-type: none"> ▶ Journal publication contains additional statistical analysis compared with preprint (n=7) ▶ Journal publication uses different statistical adjustments compared with preprint (n=7) ▶ Journal publication and preprint use different statistical tests for same data (n=3) |
| Subgroup analyses conducted | 14 (21%) | 24 (9%) | <ul style="list-style-type: none"> ▶ Journal publication includes subgroup analysis not included in preprint (n=6) ▶ Journal publication finds statistically significant interaction for subgroup, preprint does not (n=1) |
| Identifying the outcome as a primary or secondary outcome | 1 (1%) | 3 (1%) | <ul style="list-style-type: none"> ▶ For example, preprint identifies the primary endpoint as safety; journal publication adds the secondary endpoint of exploration of efficacy |

*Ns do not add to number of reported discrepancies as some studies could have more than one discrepancy and not all discrepancies have been included as examples.

**Table 4** Categories of spin in preprints and Journal publications (n=67 studies)

| Spin categories and subcategories* | No spin N (%) | Occurred in preprint and journal publication N (%) | Occurred in preprint only N (%) | Occurred in journal publication only N (%) |
|---|------------------|--|---------------------------------------|---|
| Any category of spin† | 37 (55%) | 23 (34%) | 5 (7%) | 2 (3%) |
| Category | | | | |
| Inappropriate interpretation given study design‡ | 55 (82%) | 7 (10%) | 4 (6%) | 1 (1%) |
| Subcategory | | | | |
| Claiming causality in non-randomised studies | 62 (93%) | 4 (6%) | 1 (1%) | 0 (0%) |
| Interpreting a lack of statistical significance as equivalence | 66 (99%) | 0 (0%) | 0 (0%) | 1 (1%) |
| Interpreting a lack of statistical significance of harm measures as safety | 65 (97%) | 1 (1.5%) | 0 (0%) | 1 (1.5%) |
| Claim of any significant difference despite lack of statistical test | 67 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Other | 61 (91%) | 2 (3%) | 4 (6%) | 0 (0%) |
| Category | | | | |
| Inappropriate extrapolations or recommendations | 52 (78%) | 13 (19%) | 2 (3%) | 0 (0%) |
| Subcategory | | | | |
| Suggestion that the treatment or test is more clinically relevant or useful than is justified given the study design. | 60 (90%) | 6 (9%) | 1 (1%) | 0 (0%) |
| Recommendations made to population groups/contexts outside of those investigated. | 63 (94%) | 3 (5%) | 1 (1%) | 0 (0%) |
| (Observational) Expressing confidence in a treatment or test without suggesting the need for further confirmatory studies | 66 (99%) | 0 (0%) | 1 (1%) | 0 (0%) |
| (Observational) Making recommendations without stating a randomised controlled clinical should be done to validate the recommendation | 65 (97%) | 2 (3%) | 0 (0%) | 0 (0%) |
| Other | 63 (94%) | 3 (5%) | 1 (1%) | 0 (0%) |
| Category | | | | |
| Selective focusing on positive results or more favourable data presentation | 54 (81%) | 8 (12%) | 2 (3%) | 3 (4%) |
| Subcategory | | | | |
| Discussing only significant (non-primary) results to distract from non-significant (primary) results | 66 (99%) | 0 (0%) | 1 (1%) | 0 (0%) |
| Omitting non-significant results from abstract/discussion/conclusion | 65 (97%) | 1 (1.5%) | 0 (0%) | 1 (1.5%) |
| Claiming significant effects for non-significant results | 67 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Acknowledge statistically non-significant results for the primary outcome but emphasise the beneficial effect of treatment | 66 (99%) | 1 (1%) | 0 (0%) | 0 (0%) |
| Describing non-significant results as 'trending towards significance' | 66 (99%) | 1 (1%) | 0 (0%) | 0 (0%) |
| Mentioning adverse events in the abstract/discussion/conclusion but minimising their potential effect or importance. | 64 (96%) | 2 (3%) | 1 (1%) | 0 (0%) |
| Misleading description of study design as one that is more robust | 67 (100%) | 0 (0%) | 0 (0%) | 0 (0%) |
| No considerations of the limitations of the study | 64 (96%) | 3 (4%) | 0 (0%) | 0 (0%) |
| Use of linguistic spin | 66 (99%) | 0 (0%) | 0 (0%) | 1 (1%) |
| Other | 62 (93%) | 1 (1%) | 2 (3%) | 2 (3%) |

*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.

†This row shows counts of at least one instance of spin in any category. Column category and subcategory counts add to greater than any occurrence of spin because multiple categories and subcategories of spin could occur within a preprint or article publication. Row percents do not add to 100 due to rounding.

‡Row percents may not add to 100 due to rounding.

and more extensive discussion of limitations. Some of these differences may be due to more comprehensive reporting requirements of journals. Other changes, such as more

information on the study population or greater discussion of limitations, may be due to requests for additional information during peer review.

Since preprints are posted without peer review and most journal publications in our sample were likely to be peer reviewed because they were identified from PubMed, our study indirectly investigates the impact of peer review on research articles. Articles may not have been peer reviewed in similar ways. Authors may have made changes in their papers that were independent of peer review. We observed instances where peer review appeared to improve clarity (eg, more detail on measurements)^{32 33} or interpretation (eg, requirement to present risk differences rather than just *n* (%) per treatment group).^{34 35} Empirical evidence on the impact of peer review on manuscript quality is scarce. A study comparing submitted and published manuscripts found that the number of changes was relatively small and, similar to our study, primarily involved adding or clarifying information.¹³ Some of the changes requested by peer reviewers were classified as having a negative impact on reporting, such as the addition of post hoc subgroup analyses, statistical analyses that were not prespecified or optimistic conclusions that did not reflect the trial results. In our sample, additions of subgroup and statistical analyses were common between preprints and journal publications, although we did not determine their appropriateness.

A small proportion of medRxiv preprints, 14% at the end of the server's first year, were published as journal publications.⁵ Therefore, our sample could be limited to studies that their authors deemed of high enough quality to be eligible for submission to a journal. Or, our sample could be limited to articles that had not been rejected by a journal. It is possible that peer review was eliminating publications that were fundamentally unsound, while more quickly processing studies that were sound and useful. Under pandemic conditions, articles may undergo fewer revisions. 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. Thus, in this situation, peer review may mainly be playing the role of determining whether a study should be published in a journal or not.

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 likely had not gone through the peer review process. Given this finding and the observed similarities between preprints and their subsequent journal publications, our results suggest that peer review during the accelerated pace of COVID-19 research publication may not have provided much added value. The urgency related to dissemination of COVID-19 research could have led journals to fast-track publication by abbreviating editorial or peer review processes, resulting in fewer differences between preprints and journal publications.

Comparison to other studies

Our results are consistent with other studies finding small changes in reporting between preprints and journal publications. A number of these studies have been limited by failing to assess the addition or deletion of outcomes and by the use of composite 'scores' that included items related to risk of bias and reporting. In contrast to our study, in a matched sample of preprints and journal publications, Carneiro *et al* found journal publications more likely to have conflict of interest statement than preprints. In a textual analysis using five different algorithms, Klein *et al* found very little difference in text between preprints and articles in a large matched sample.⁹ We also noted preprints and journal publications that were almost identical, or had very minor differences such as corrections of typos. Other studies are limited by comparing unmatched samples of preprints and articles. In a comparison of 13 preprints and 16 articles on COVID-19 that were not reporting on the same studies, Kataoka *et al* found no significant differences in risk of bias or spin in titles and conclusions.¹¹

We found similar changes in numerical results to Oikonomidi *et al* who compared 66 preprint–article pairs for COVID-19 studies and found 25 (38%) of studies had changes.¹² Oikonomidi classified 16 of these changes as 'important' based on (1) an increase or decrease by $\geq 10\%$ of the initial value in any effect estimate and/or (2) a change in the *p* value crossing the threshold of 0.05, for any study outcome. We did not classify changes based on magnitude or threshold *p* values because changes in numerical values may be related to other components of outcome reporting that we observed, such as changes to follow-up times or the use of different statistical tests. Furthermore, deviations from a *p* value of 0.05 do not necessarily indicate changes in scientific or clinical significance. We examined changes in multiple components of outcome reporting that are considered essential, not just the numerical value of the outcome.^{16 21} The diversity of studies included in our sample would make any categorisations of scientific or clinical significance difficult and subjective. For example, studies were observational and experimental and not all studies conducted statistical analysis. The topics of the studies included tests of clinical and public health interventions, associations of risk factors with COVID-19 symptoms and ways to improve implementation of public health measures, such as social distancing.

Strengths and limitations of this study

We selected studies from the Cochrane COVID-19 Register rather than conducting a literature search. However, as the Cochrane COVID-19 Register has been optimised to identify COVID-19 clinical research for systematic reviews, we feel the search was comprehensive for identifying COVID-19 studies related to treatment or prevention that are most likely to have an impact on clinical practice or health policy. As a study-based register, all records related to a study are identified, enabling us to

obtain all preprint and journal publication versions for a single study. Second, we compared the first version of the preprint with the final journal publication. We may have identified a different number of discrepancies if we compared later versions of the preprint with the journal publication. Third, although clinically important, our focus on COVID-19 research may not be representative of other types of research published as preprints, then journal publications. This study should be replicated in a sample of non-COVID-related interventional and observational clinical studies. Future research could also include assessment of outcome reporting components and spin in preprints that have not been published in journals. Fourth, although we compared non-peer-reviewed preprints to their accompanying journal publications, we did not directly assess the effects of peer review. Finally, coders were not blinded to the source or authors of preprints and journal publications, as this was not feasible and there is no evidence that it would alter the decisions made.

CONCLUSIONS

The COVID-19 preprints and their subsequent journal publications were largely similar in reporting of study characteristics, outcomes and spin in interpretation. 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. All COVID-19 studies, whether published as preprints or journal publications, should be critically evaluated for discrepancies in outcome reporting or spin, such as failure to report data on harms or overly optimistic conclusions.

Author affiliations

¹General Internal Medicine/Public Health/Center for Bioethics and Humanities, University of Colorado—Anschutz Medical Campus, Denver, Colorado, USA

²Center for Bioethics and Humanities, University of Colorado - Anschutz Medical Center, Denver, Colorado, USA

³Charles Perkins Centre and School of Pharmacy, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

⁴School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

⁵Faculty of Nursing, University of Sydney, Toronto, Ontario, Canada

⁶Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia

⁷Editorial and Methods Department, Cochrane, London, UK

⁸Biostatistics, Manchester University, Manchester, UK

Twitter Quinn Grundy @QuinnGrundy

Contributors LB conceived the project, drafted the protocol, acquired data, conducted analysis, interpreted data and drafted the paper. RL edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. LL edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. KC edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. SM edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. MP edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. QG edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. LP edited the protocol, acquired data, conducted analysis, interpreted data and revised the paper. SB edited the protocol, acquired data, conducted analysis, interpreted data

and revised the paper. JJK edited the protocol, acquired data, conducted analysis, interpreted data, and revised the paper. RF edited the protocol, conducted the search, conducted analysis, interpreted data and revised the paper. All authors (LB, RL, LL, KC, SM, MJ, QG, LP, SB, JJK and RF) have approved the final manuscript. LB served as guarantor for all aspects of the work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests RF is a Cochrane employee and part of the development team for the Cochrane COVID-19 Study Register. No other authors declare any other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval This study analyses publicly available information and is exempt from ethics review.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data from this study are available in OSF project file (<https://osf.io/5ru8w/>).

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

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

Author note Data access: LB had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

ORCID iDs

Lisa Bero <http://orcid.org/0000-0003-1893-6651>

Matthew J Page <http://orcid.org/0000-0002-4242-7526>

Quinn Grundy <http://orcid.org/0000-0002-7640-8614>

Lisa Parker <http://orcid.org/0000-0001-8635-6953>

Jamie J Kirkham <http://orcid.org/0000-0003-2579-9325>

REFERENCES

- All that's fit to preprint. *Nat Biotechnol* 2020;38:507
- Flanagin A, Fontanarosa PB, Bauchner H. Preprints involving medical Research-Do the benefits outweigh the challenges? *JAMA* 2020;324:1840.
- Jung YE, Sun Y, Schluger NW. Effect and reach of medical articles posted on preprint servers during the COVID-19 pandemic. *JAMA Intern Med* 2020.
- Kirkham JJ, Penfold NC, Murphy F, et al. Systematic examination of preprint platforms for use in the medical and biomedical sciences setting. *BMJ Open* 2020;10:e041849.
- Krumholz HM, Bloom T, Sever R, et al. Submissions and Downloads of Preprints in the First Year of medRxiv. *JAMA* 2020;324:1903-5.
- Krumholz HM, Bloom T, Ross J. Preprints can fill a void in times of rapidly changing science, 2020. STAT. Available: <https://www.statnews.com/2020/01/31/preprints-fill-void-rapidly-changing-science/> [Accessed 17 Dec 2020].
- Bloom T. Shepherding preprints through a pandemic. *BMJ* 2020;371:m4703.
- Carneiro CFD, Queiroz VGS, Moulin TC, et al. Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Res Integr Peer Rev* 2020;5:16.
- Klein M, Broadwell P, Farb SE. Comparing published scientific Journal articles to their pre-print versions. Association for Computing Machinery. *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, New York, NY, USA, 2016:153-62.
- Nicolalde B, Añazco D, Mushtaq M, et al. Citations and publication rate of preprints on pharmacological interventions for COVID-19: the good, the bad and, the ugly. *Res Sq* 2020;version 2.

- 11 Kataoka Y, Oide S, Arie T, *et al.* COVID-19 randomized controlled trials in medRxiv and PubMed. *Eur J Intern Med* 2020;81:97–9.
- 12 Oikonomidi T, Boutron I, Pierre O, *et al.* Changes in evidence for studies assessing interventions for COVID-19 reported in preprints: meta-research study. *BMC Med* 2020;18:402.
- 13 Hopewell S, Collins GS, Boutron I, *et al.* Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014;349:g4145.
- 14 Lazarus C, Haneef R, Ravaud P, *et al.* Peer reviewers identified spin in manuscripts of nonrandomized studies assessing therapeutic interventions, but their impact on spin in abstract conclusions was limited. *J Clin Epidemiol* 2016;77:44–51.
- 15 Chan A-W, Hróbjartsson A, Haahr MT, *et al.* Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457.
- 16 Mathieu S, Boutron I, Moher D, *et al.* Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977–84.
- 17 Chiu K, Grundy Q, Bero L. 'Spin' in published biomedical literature: a methodological systematic review. *PLoS Biol* 2017;15:e2002173.
- 18 Boutron I, Ravaud P. Misrepresentation and distortion of research in biomedical literature. *Proc Natl Acad Sci U S A* 2018;115:2613–9.
- 19 Bero L, Lawrence R, Leslie L. Comparison of preprints with peer-reviewed publications on COVID-19: discrepancies in results reporting and conclusions, 2020. Available: <https://osf.io/j62eu> [Accessed 17 Dec 2020].
- 20 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
- 21 Zarin DA, Tse T, Williams RJ, *et al.* The ClinicalTrials.gov results database—update and key issues. *N Engl J Med* 2011;364:852–60.
- 22 Borba MGS, Val FFA, Sampaio VS, *et al.* Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: a randomized clinical trial. *JAMA Netw Open* 2020;3:e208857.
- 23 Borba MGS, FFA V, Sampaio VS, *et al.* Chloroquine diphosphate in two different dosages as adjunctive therapy of hospitalized patients with severe respiratory syndrome in the context of coronavirus (SARS-CoV-2) infection: preliminary safety results of a randomized, double-blinded, phase IIb clinical trial (CloroCovid-19 study). *MedRxiv Prepr* 2020.
- 24 Tang W, Cao Z, Han M, *et al.* Hydroxychloroquine in patients with COVID-19: an open-label, randomized, controlled trial. *MedRxiv Prepr* 2020.
- 25 Tang W, Cao Z, Han M, *et al.* Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. *BMJ* 2020;369:m1849.
- 26 Weber AG, Chau AS, Egeblad M, *et al.* Nebulized in-line endotracheal dornase alfa and albuterol administered to mechanically ventilated COVID-19 patients: a case series. *MedRxiv Prepr* 15 May 2020.
- 27 Weber AG, Chau AS, Egeblad M, *et al.* Nebulized in-line endotracheal dornase alfa and albuterol administered to mechanically ventilated COVID-19 patients: a case series. *Mol Med* 2020;26:91.
- 28 Gautret P, Lagier J-C, Parola P, *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *MedRxiv* 2020.
- 29 Gautret P, Lagier J-C, Parola P, *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020;56:105949.
- 30 Smith LE, Amlôt R, Lambert H, *et al.* Factors associated with adherence to self-isolation and lockdown measures in the UK: a cross-sectional survey. *MedRxiv Prepr* 2020.
- 31 Smith LE, Amlôt R, Lambert H, *et al.* Factors associated with adherence to self-isolation and lockdown measures in the UK: a cross-sectional survey. *Public Health* 2020;187:41–52.
- 32 Chorin E, Wadhvani L, Magnani S, *et al.* QT interval prolongation and torsade de pointes in patients with COVID-19 treated with hydroxychloroquine/azithromycin. *Heart Rhythm* 2020;17:1425–33.
- 33 Chorin E, Wadhvani L, Magnani S, *et al.* QT interval prolongation and torsade de pointes in patients with COVID-19 treated with Hydroxychloroquine/Azithromycin. *MedRxiv Prepr* 2020.
- 34 Agarwal A, Mukherjee A, Kumar G, *et al.* Convalescent plasma in the management of moderate COVID-19 in India: an open-label parallel-arm phase II multicentre randomized controlled trial (PLACID trial). *MedRxiv Prepr* 2020.
- 35 Agarwal A, Mukherjee A, Kumar G, *et al.* Convalescent plasma in the management of moderate covid-19 in adults in India: open label phase II multicentre randomised controlled trial (PLACID trial). *BMJ* 2020;371:m3939.

Supplemental Files**Table S1: Timing of preprint to journal publication (days)****Table S2: Sensitivity Analysis of Discrepancies in Study Characteristics****Table S3: Sensitivity Analysis of Discrepancies in Outcome Reporting****Table S4: Sensitivity Analysis of Categories of Spin in Preprints and Journal Publications**

Table S1: Timing of preprint to journal publication (days)

| | Days from preprint to published, mean (range) |
|--|--|
| All Studies (n=67) | 65.4 (0 - 271) |
| Subgroup: Preprint posted before submission to journal (n=32) | 87.1 (10 - 271) |
| Subgroup: Preprint posted after submission to journal (n=27) | 52.2 (0 - 120) |

Table S2: Sensitivity Analysis of Discrepancies in Study Characteristics (n=60) ^a

| | No Discrepancies | | Discrepancies | | |
|----------------------------------|---------------------------|------------------------------|--|------------------------------------|---|
| | Reported in Both, No. (%) | Reported in Neither, No. (%) | Reported in Both With Discrepancies, No. (%) | Reported in Preprint Only, No. (%) | Reported in Journal Publication Only, No. (%) |
| Title | 44 (73) | 0 (0) | 16 (27) | 0 (0) | 0 (0) |
| Authors | 43 (72) | 0 (0) | 17 (28) | 0 (0) | 0 (0) |
| Disclosed Funding Source | 39 (65) | 3 (5) | 10 (17) | 6 (10) | 2 (3) |
| COI Disclosure Statement | 45 (75) | 1 (2) | 9 (15) | 4 (7) | 1 (2) |
| Ethics Approval | 54 (90) | 2 (3) | 4 (7) | 0 (0) | 0 (0) |
| Location of Study | 56 (93) | 4 (7) | 0 (0) | 0 (0) | 0 (0) |
| Number of Participants | 54 (90) | 0 (0) | 6 (10) | 0 (0) | 0 (0) |
| Participant Demographics | 34 (57) | 3 (5) | 22 (37) | 0 (0) | 1 (2) |
| Tables and Figures | 15 (25) | 0 (0) | 45 (75) | 0 (0) | 0 (0) |
| Discussion of Limitations | 23 (38) | 6 (10) | 30 (50) | 0 (0) | 1 (2) |

^a Studies that had a preprint posted on-or-after the date of revision, acceptance, or publication were removed. This removed 1 study. Due to differences in journal reporting of these dates, there was overlap in those studies and no comparison in others. Therefore, we expanded the studies removed to include those with preprints posted 1-7 days before the date of revision, acceptance, or publication, thus removing 7 studies from the sensitivity analysis.

Table S3: Sensitivity Analysis of Discrepancies in Outcome Reporting (n=60) ^a

| | Number (%) of studies with at least 1 discrepancy n=60 | Number (%) of Outcomes n=242 |
|--|---|---|
| Outcome in journal publication only | 14 (23) | 18 (7) |
| Outcome in preprint only | 7 (12) | 16 (7) |
| Outcome measurement | 5 (8) | 7 (3) |
| Units of measurement | 3 (5) | 3 (1) |
| Timepoint assessment was made | 10 (17) | 24 (10) |
| Numerical values reported | 23 (38) | 49 (20) |
| Finding of statistical significance | 11 (18) | 16 (7) |
| Statistical tests performed | 16 (27) | 30 (12) |
| Subgroup analyses conducted | 13 (22) | 23 (10) |
| Identifying the outcome as a primary or secondary outcome | 1 (2) | 3 (1) |

^a Studies that had a preprint posted on-or-after the date of revision, acceptance, or publication were removed. This removed 1 study. Due to differences in journal reporting of these dates, there was overlap in those studies and no comparison in others. Therefore, we expanded the studies removed to include those with preprints posted 1-7 days before the date of revision, acceptance, or publication, thus removing 7 studies from the sensitivity analysis.

Table S4: Sensitivity Analysis of Categories of Spin in Preprints and Journal Publications (n=60)^a

| | Neither, No. (%) | Both, No. (%) | Preprint Only, No. (%) | Journal Publication Only, No. (%) |
|---|------------------|----------------|------------------------|-----------------------------------|
| Inappropriate interpretation given study design | 49 (82) | 6 (10) | 4 (7) | 1 (2) |
| Claiming causality in non-randomized studies | 56 (93) | 3 (5) | 1 (2) | 0 (0) |
| Interpreting a lack of statistical significance as equivalence | 59 (98) | 0 (0) | 0 (0) | 1 (2) |
| Interpreting a lack of statistical significance of harm measures as safety | 58 (97) | 1 (2) | 0 (0) | 1 (2) |
| Claim of any significant difference despite lack of statistical test | 60 (100) | 0 (0) | 0 (0) | 0 (0) |
| Other | 54 (90) | 2 (3) | 4 (7) | 0 (0) |
| | | | | |
| Inappropriate extrapolations or recommendations | 46 (77) | 12 (20) | 2 (3) | 0 (0) |
| Suggestion that the treatment or test is more clinically relevant or useful than is justified given the study design. | 54 (90) | 5 (8) | 1 (2) | 0 (0) |
| Recommendations made to population groups / contexts outside of those investigated. | 56 (93) | 3 (5) | 1 (2) | 0 (0) |
| (Observational) Expressing confidence in a treatment or test without suggesting the need for further confirmatory studies | 59 (98) | 0 (0) | 1 (2) | 0 (0) |
| (Observational) Making recommendations without stating an RCT should be done to validate the recommendation | 59 (98) | 1 (2) | 0 (0) | 0 (0) |
| Other | 56 (93) | 3 (5) | 1 (2) | 0 (0) |
| | | | | |
| Selective focusing on positive results or more favorable data presentation | 48 (80) | 7 (12) | 2 (3) | 3 (5) |
| Discussing only significant (non-primary) results to distract from non-significant (primary) results | 59 (98) | 0 (0) | 1 (2) | 0 (0) |
| Omitting non-significant results from Abstract/Discussion/Conclusion | 58 (97) | 1 (2) | 0 (0) | 1 (2) |
| Claiming significant effects for non-significant results | 60 (100) | 0 (0) | 0 (0) | 0 (0) |
| Acknowledge statistically nonsignificant results for the primary outcome but emphasize the beneficial effect of treatment | 59 (98) | 1 (2) | 0 (0) | 0 (0) |

| | | | | |
|--|----------|-------|-------|-------|
| Describing non-significant results as "trending towards significance" | 59 (98) | 1 (2) | 0 (0) | 0 (0) |
| Mentioning adverse events in the abstract/discussion/conclusion but minimizing their potential effect or importance. | 58 (97) | 1 (2) | 1 (2) | 0 (0) |
| Misleading description of study design as one that is more robust | 60 (100) | 0 (0) | 0 (0) | 0 (0) |
| No considerations of the limitations of the study | 58 (97) | 2 (3) | 0 (0) | 0 (0) |
| Use of linguistic spin | 59 (98) | 0 (0) | 0 (0) | 1 (2) |
| Other | 55 (92) | 1 (2) | 2 (3) | 2 (3) |

^a Studies that had a preprint posted on-or-after the date of revision, acceptance, or publication were removed. This removed 1 study. Due to differences in journal reporting of these dates, there was overlap in those studies and no comparison in others. Therefore, we expanded the studies removed to include those with preprints posted 1-7 days before the date of revision, acceptance, or publication, thus removing 7 studies from the analysis