



Consistency of covid-19 trial preprints with published reports and impact for decision making: retrospective review

Dena Zeraatkar ,¹ Tyler Pitre,² Gareth Leung,³ Ellen Cusano,⁴ Arnav Agarwal,⁵ Faran Khalid,² Zaira Escamilla,² Matthew Adam Cooper,⁶ Maryam Ghadimi,² Ying Wang,⁷ Francisca Verdugo-Paiva ,^{8,9} Gabriel Rada ,⁸ Elena Kum,¹ Anila Qasim,¹ Jessica Julia Bartoszko ,² Reed Alexander Cunningham Siemieniuk ,¹ Chirag Patel,¹⁰ Gordon Guyatt,² Romina Brignardello-Petersen^{1,11}

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjmed-2022-000309>).

For numbered affiliations see end of article.

Correspondence to: Dr Dena Zeraatkar, Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; dena.zeraa@gmail.com

Cite this as: *BMJ MED* 2022;0:doi:10.1136/bmjmed-2022-000309

Received: 5 July 2022

Accepted: 30 August 2022

ABSTRACT

OBJECTIVE To assess the trustworthiness (ie, complete and consistent reporting of key methods and results between preprint and published trial reports) and impact (ie, effects of preprints on meta-analytic estimates and the certainty of evidence) of preprint trial reports during the covid-19 pandemic.

DESIGN Retrospective review.

DATA SOURCES World Health Organization covid-19 database and the Living Overview of the Evidence (L-OVE) covid-19 platform by the Epistemonikos Foundation (up to 3 August 2021).

MAIN OUTCOME MEASURES Comparison of characteristics of covid-19 trials with and without preprints, estimates of time to publication of covid-19 preprints, and description of differences in reporting of key methods and results between preprints and their later publications. For the effects of eight treatments on mortality and mechanical

ventilation, the study comprised meta-analyses including preprints and excluding preprints at one, three, and six months after the first trial addressing the treatment became available either as a preprint or publication (120 meta-analyses in total, 60 of which included preprints and 60 of which excluded preprints) and assessed the certainty of evidence using the GRADE framework.

RESULTS Of 356 trials included in the study, 101 were only available as preprints, 181 as journal publications, and 74 as preprints first and subsequently published in journals. The median time to publication of preprints was about six months. Key methods and results showed few important differences between trial preprints and their subsequent published reports. Apart from two (3.3%) of 60 comparisons, point estimates were consistent between meta-analyses including preprints versus those excluding preprints as to whether they indicated benefit, no appreciable effect, or harm. For nine (15%) of 60 comparisons, the rating of the certainty of evidence was different when preprints were included versus being excluded—the certainty of evidence including preprints was higher in four comparisons and lower in five comparisons.

CONCLUSION No compelling evidence indicates that preprints provide results that are inconsistent with published papers. Preprints remain the only source of findings of many trials for several months—an unsuitable length of time in a health emergency that is not conducive to treating patients with timely evidence. The inclusion of preprints could affect the results of meta-analyses and the certainty of evidence. Evidence users should be encouraged to consider data from preprints.

Introduction

During the covid-19 pandemic, the scientific community adopted preprint servers, which allow investigators to disseminate research findings before publication in peer reviewed journals. Authors of seminal covid-19 trials—for example, representing massive international collaborations such as RECOVERY^{1–4} and SOLIDARITY⁵—reported their results in preprints before subsequent publication in journals.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Clinicians and decision makers need rapidly available, credible information on the comparative effectiveness of treatments and prophylaxis for covid-19
- ⇒ During the covid-19 pandemic, the scientific community adopted preprint servers, which allow the rapid dissemination of research findings before publication in peer reviewed journals
- ⇒ The medical community, however, has been cautious about adopting preprints owing to concerns that they could lead to the dissemination of erroneous provisional findings

WHAT THIS STUDY ADDS

- ⇒ After a review of covid-19 trial preprints and published reports, no compelling evidence indicated any important discrepancies between preprints and published reports
- ⇒ The inclusion of preprints could affect the results of meta-analyses and the certainty (quality) of evidence
- ⇒ Generalisability of these results is limited to covid-19; furthermore, preprints that are subsequently published in journals might be the most rigorous and might not represent all trial preprints—particularly those that remain unpublished

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

- ⇒ Evidence users—including systematic reviewers, guideline developers, and clinicians—are encouraged to consider evidence from preprint trials in contexts in which decisions are being made rapidly and evidence is being produced faster than can be peer reviewed and published
- ⇒ Scepticism might still be warranted when suspicion arises regarding falsified data (for which criteria is provided in this article)

Growing interest in preprints predates the covid-19 pandemic.^{6–7} Researchers and evidence users have raised concerns that the traditional publication model is slow, peer review might not always improve the quality of manuscripts, journals impede dissemination owing to paywalls and high publication fees and encourage publication bias by prioritising significant or anomalous findings—issues that preprints may avoid.^{8–14} Despite these concerns, and the potential of preprints to resolve them, preprints might result in the dissemination of provisional findings that contain important errors that, presumably, published papers do not—therefore, the medical community has been cautious regarding their adoption.^{15–16}

Authors of systematic reviews, guideline developers, and other decision makers face a trade-off when considering preprints: on the one hand, including preprints could reduce the credibility of evidence syntheses and risk serious errors if important differences appear in published reports; on the other, including preprints might increase the precision of estimates, allow timely dissemination of research, and minimise the effects of publication bias.

Knowledge of the extent to which preprints might accelerate the dissemination of findings, the frequency and nature of discrepancies between preprints and subsequent published reports, and the impact that preprints might have on meta-analytic estimates could inform the trade-off that evidence users face. Our study capitalises on our living systematic reviews and network meta-analyses (SRNMAs) of drug treatments, antiviral antibodies and cellular treatments, and prophylaxis for covid-19—an initiative launched in July 2020 that provides real time summaries of the comparative effectiveness of treatments and prophylaxis for covid-19.^{17–19} These living SRNMAs informed linked guidelines for covid-19 treatments and prophylaxis.²⁰ We use these reviews to assess the degree of discrepancies between covid-19 trial preprints and their later publications and to assess the effects of considering evidence from preprints on meta-analytic estimates, certainty (quality) of evidence, and decision making.

Methods

We submitted a protocol for this study for publication on 7 September 2021. Because the protocol was still under review at the time the study was completed, we withdrew the protocol for publication and present it in online supplement 1.

Search

Our study uses the search strategy of our living SRNMA that includes daily searches in the World Health Organization covid-19 database—a comprehensive multilingual source of global published and preprint literature on covid-19 (<https://search.bvsalud.org/global-literature-on-novel-corona->

[virus-2019-ncov/](https://search.bvsalud.org/global-literature-on-novel-corona-virus-2019-ncov/)). Before it was merged with the WHO covid-19 database on 9 October 2020, we searched the US Centres for Disease Control and Prevention's (CDC) covid-19 research articles downloadable database. A validated machine learning model facilitates efficient identification of randomised trials.²¹

Our search was supplemented by ongoing surveillance of living evidence retrieval services, including the Living Overview of the Evidence (L-OVE) covid-19 platform by the Epistemonikos Foundation (<https://app.iloveevidence.com/loves/5e6fdb9669c00e4ac072701d>) and the Systematic and Living Map on Covid-19 Evidence by the Norwegian Institute of Public Health (<https://www.fhi.no/en/qk/systematic-reviews-hta/map/>). Using the above sources, we monitor for retraction notices. Online supplement 2 includes additional details of our search strategy. This study included trials identified up to 3 August 2021.

Study selection

As part of the living SRNMA, pairs of reviewers, after calibration exercises to ensure sufficient agreement, worked independently and in duplicate to screen titles and abstracts of search records and subsequently the full texts of records determined as potentially eligible at the title and abstract screening stage. Reviewers also linked preprint reports with their subsequent publications based on trial registration numbers, the names of investigators, recruiting centres and countries, dates of recruitment, and baseline patient characteristics. When links between preprints and subsequent publications were unclear, we contacted trial authors for confirmation. Reviewers resolved discrepancies by discussion or, when necessary, by adjudication with a third party reviewer.

Eligible preprint and peer reviewed articles reported trials that randomised patients with suspected, probable, or confirmed covid-19 to drug treatments, antiviral antibodies and cellular treatments, placebo, or standard care, or reported trials that randomised healthy participants exposed or unexposed to covid-19 to prophylactic drugs, standard care, or placebo. We did not apply any restrictions on severity of illness, setting, or language of publication but excluded trials reporting on nutritional interventions, traditional Chinese herbal medicines without standardisation in formulations and dosing across batches, and non-drug supportive care interventions.

We did not perform a sample size calculation because we included all eligible trial reports identified through our living SRNMAs up to 3 August 2021. While the parallel living SRNMA performed ongoing daily searches, we pragmatically limited our search because it was no longer feasible to continue to collect additional data from preprints beyond this timepoint.

Data collection

As part of the living SRNMA, for each eligible trial, pairs of reviewers, after training and calibration exercises, independently extracted trial characteristics, methods, and results using a standardised, pilot tested data extraction form. To assess risk of bias, reviewers, after training and calibration exercises, used a revision of the Cochrane tool for assessing risk of bias in randomised trials (RoB 2.0)²² (online supplement 3). Reviewers resolved discrepancies by discussion and, when necessary, by adjudication with a third party.

For the current study, pairs of trained and calibrated reviewers, working independently and in duplicate and using a pilot tested data collection form, collected data on differences in key methods and results between preprint and published trial reports. We prioritised collecting information on key methods and results that might affect the interpretation of trials and decision making by evidence users. For key methods, we focused on aspects of the methods that could affect risk-of-bias judgments, which included description of the randomisation process and allocation concealment, blinding of patients and healthcare providers, extent of and handling of missing outcome data, blinding of outcome assessors and adjudicators, and prespecification of outcomes and analyses. Key results included the number of participants analysed and number of events in each trial arm for dichotomous outcomes and number of participants analysed, means or medians and measures of variability for continuous outcomes. We focused on the same outcomes as our living SRNMA and linked guidelines that were identified as being important or critical for decision making by the review authors and authors of the parallel guidelines, including patient partners: mortality, mechanical ventilation, adverse events leading to discontinuation, admission to hospital, viral clearance, hospital length of stay, length of stay in intensive care, duration of mechanical ventilation, time to symptom resolution or clinical improvement, days free from mechanical ventilation, and time to viral clearance.^{17–20} For preprints with more than one version, we extracted data from the first version of the preprint, which is the least likely to have been modified in response to peer review.

Because risk of bias might vary across outcomes, for this analysis we presented risk-of-bias judgments corresponding to the following hierarchy of outcomes for therapy trials: mortality, mechanical ventilation, duration of hospital stay, time to symptom resolution or clinical improvement, and virological outcomes. For prophylaxis trials, we used the following hierarchy: mortality, laboratory confirmed and suspected covid-19 infection, and laboratory confirmed covid-19 infection. These hierarchies represent the relative importance of outcomes based on rankings made by the linked WHO guideline panel.²⁰

Data synthesis and analysis

We compared the characteristics and risk of bias of trials with preprints, trials with publications, and trials first posted as a preprint and subsequently published by calculating differences in proportions, associated confidence intervals, and z tests to test for differences in independent proportions. To compare the number of participants in trials with preprints, trials with publications, and trials first posted as a preprint and subsequently published, we performed Mann Whitney U tests.

We calculated the median time from a trial being posted on a preprint server to its eventual publication in a journal and used Kaplan-Meier curves and log-rank tests to assess whether the following factors were predictive of time to publication of trial preprints: source of funding, number of centres and participants, early termination for benefit, intensity of care (inpatient v outpatient), and severity (mild/moderate v severe/critical covid-19), significant primary or secondary outcomes (based on cut-off thresholds defined by the authors or, when no cut-off thresholds were defined, based on a cut-off threshold of $P < 0.05$ or confidence intervals not including the null), and risk of bias (trials rated at low v high risk of bias).

Among trial preprints that were subsequently published in a peer reviewed journal, we described the number and types of discrepancies in key methods and results between preprint and published trial reports. For discrepancies in the reporting of key methods, we reported the number and percentage of the changes between preprints and publications that affected risk-of-bias judgments—changes that we considered to be critical. We also compared the number of preprint and published trials that have been retracted.

For trials that reported on interventions that have been addressed by the linked WHO living guideline²⁰ up to 3 August 2021 (ie, corticosteroids, remdesivir, lopinavir-ritonavir, hydroxychloroquine, ivermectin, interleukin 6 receptor blockers, and convalescent plasma for treatment and hydroxychloroquine for prophylaxis) and the two most commonly reported outcomes (ie, mortality, mechanical ventilation), we conducted pairwise frequentist random effects meta-analyses with the restricted maximum likelihood estimator including versus excluding evidence from preprints at one, three, and six months after the first trial of the drug of interest was made public, either via preprint or publication. The choice of timepoints was informed by timeframes within which guideline developers needed to issue recommendations.²⁰ We also conducted an analysis including versus excluding evidence from preprints at 3 August 2021—the longest timepoint at which we collected data. For hydroxychloroquine for prophylaxis, because mechanical ventilation was not an outcome of interest for prophylaxis trials, we reported only on mortality.

To facilitate interpretation, we calculated absolute effects. To calculate absolute effects, for drug treatments, we used mortality data from the CDC and data on ventilation from the International Severe Acute Respiratory and Emerging Infection covid-19 database.^{23–25} For prophylaxis, we used the event rate among all participants randomised to standard care or placebo to calculate the baseline risk.

We compared the direction of effect between meta-analyses including preprints and meta-analyses excluding preprints. We considered the direction of effect to be different if one point estimate suggested no effect and another suggested a benefit or harm or if one point estimate suggested benefit and another suggested harm. For treatment, we considered an effect to be beneficial if the point estimate indicated a reduction in risk of mortality of 1% or greater or a reduction in risk of mechanical ventilation of 2% or greater for treatment. For prophylaxis, we considered an effect to be beneficial if the point estimate indicated a reduction in risk of mortality of 0.5% or greater. For treatment, we considered an effect to be harmful if the point estimate indicated an increase in risk of mortality of 1% or greater or an increase in risk of mechanical ventilation of 2% or greater. For prophylaxis, we considered an effect to be harmful if the point estimate indicated an increase in risk of mortality of 0.5% or greater for prophylaxis. Otherwise, we inferred that there was no important effect. Our thresholds for beneficial and harmful effects were informed by surveys of the coauthors in the parallel living SRNMA.^{17–19}

We used the GRADE approach to assess the certainty of evidence, considering risk of bias (limitations in trial design leading to systematic underestimation or over-estimation of treatment effects), inconsistency (heterogeneity in results reported across trials), indirectness (differences between the question asked in trials and the question of interest), imprecision (width of confidence intervals), and publication bias (propensity for studies with significant results, notable results, or results that support a particular hypothesis to be published, published faster, or published in journals with higher visibility). We also assessed whether meta-analyses including preprints versus excluding preprint reports led to differences in ratings of the overall certainty of evidence, judgments related to specific GRADE domains, and whether differences in ratings were likely to affect decision making (ie, evidence rated as high/moderate v low/very low).²⁶ We used a minimally contextualised approach to make judgments about imprecision.²⁷ This approach considers whether confidence intervals include the null effect and thus does not consider whether plausible effects, captured by confidence intervals, include both important and trivial effects. We considered any effect on mortality and mechanical ventilation to be important. Thresholds of 1% risk difference

for mortality and 2% risk difference for mechanical ventilation informed judgments of minimal or no treatment effect. For prophylaxis and mortality, we used a 0.5% risk difference.²⁷ We performed all statistical analyses in R (version 4.0.3, R Foundation for Statistical Computing), using the meta, forestplot, survival, and survminer packages.

Patient and public involvement

Patients were involved in outcome selection, interpretation of results, and the generation of parallel recommendations, as part of the parallel SRNMA and guidelines.²⁰ Patients were not involved in the present secondary study.

Results

Trial characteristics

As of 3 August 2021, we identified 356 eligible trials, of which 101 were only available as preprints, 181 only available as journal publications, and 74 first available as preprints and subsequently published as journal articles. Online supplement 4 presents additional details on the results of the search and table 1 presents trial characteristics.

Most trials were registered, completed at the time of reporting, addressed drug treatments, enrolled fewer than 250 participants, reported one or more outcomes that were statistically significant, and were funded by governments or institutions. Nearly two thirds of trials were at high risk of bias, primarily because of their open label design.

Compared with published trials without preprints, trials only available as preprints and trials first available as preprints and then subsequently published were more likely to be registered; trials only available as preprints were more likely to report on interim results, describe drug treatments compared with antiviral antibodies and cellular treatments or prophylaxis, and to have received industry funding; and trials first posted as preprints and subsequently published were more likely to have received government funding.

Predictors of publication and time to publication

During the 1.5 year span of this study, of 175 preprints, 74 (42.3%) were subsequently published in peer reviewed journals. Table 2 presents the proportion of preprints published up to one year. The median time to publication of preprints was 5.9 months. At one year, a third of preprints remained unpublished.

Table 3 presents predictors for the publication of preprints. Preprints that received government funding, reported on inpatients, or reported on patients with severe disease were published faster than preprints that did not receive government funding, reported on outpatients, or reported on patients with mild or moderate disease.

Table 1 | Trial characteristics. Data are number (%) of trials unless statement otherwise

Characteristic	All trials (n=356)	Published trials without preprints (n=181)	Trials only available as preprints (n=101)	Difference between trials as preprints and published trials without preprints (%; 95% CI)	P value	Trials first posted as preprints and subsequently published (n=74)	Difference between trials first posted as preprints and subsequently published and published trials without preprints (%; 95% CI)	P value
Trial registered	319 (89.6)	149 (82.3)	97 (96.0)	13.72 (7.0 to 20.5)	<0.001	73 (98.6)	16.3 (10.2 to 22.5)	<0.001
Study status								
Completed	268 (75.3)	142 (78.5)	75 (74.3)	-4.2 (-14.6 to 6.2)	0.44	51 (68.9)	-9.5 (-21.7 to 2.6)	0.12
Interim results	20 (5.6)	5 (2.8)	10 (9.9)	7.1 (0.8 to 13.4)	0.03	5 (6.8)	4.0 (-2.2 to 10.2)	0.21
Terminated early for benefit	2 (0.6)	0	2 (2.0)	2.0 (-0.7 to 4.7)	0.15	0	0 (0%)	1.00
Terminated early owing to feasibility	66 (18.5)	34 (18.8)	14 (13.9)	-4.9 (-13.7 to 3.9)	0.28	18 (24.3)	5.5 (-5.8 to 16.9)	0.34
Type of treatment								
Drug treatment	293 (82.3)	157 (86.7)	77 (76.2%)	-10.5 (-20.2 to -0.8)	0.03	59 (79.7)	-7.0 (-17.4 to 3.4)	0.19
Prophylaxis	17 (4.8)	7 (3.9)	8 (7.9)	4.0 (-1.9 to 10.0)	0.18	2 (2.7)	-1.2 (-5.8 to 3.5)	0.64
Antiviral antibodies and cellular treatments	46 (12.9)	17 (9.4)	16 (15.8)	6.5 (-1.8 to 14.7)	0.13	13 (17.6)	8.2 (-1.5 to 17.8)	0.10
Inpatient*	252 (74.3)	129 (72.3)	65 (69.9)	-4.3 (-15.6 to 7.1)	0.47	58 (80.6)	6.4 (-4.8 to 17.6)	0.27
Severe/critical disease*	88 (26.0)	46 (26.4)	26 (28.0)	1.5 (-9.7 to 12.8)	0.80	16 (22.2)	-4.2 (-15.8 to 7.4)	0.49
No of centres								
Single centre	151 (42.4)	86 (47.5)	39 (38.6)	-8.9 (-20.9 to 3.1)	0.15	26 (35.1)	-12.4 (-25.5 to 0.7)	0.06
Multicenter	177 (49.7)	85 (47.0)	48 (47.5)	0.6 (-12.0 to 12.7)	0.93	44 (59.5)	12.5 (-0.8 to 25.8)	0.07
Median (IQR) number of participants	106 (60-266)	101 (60-208)	163 (64-412)	—	0.01	100 (60 to 249)	—	0.54
Primary outcome—significant	151 (42.4)	74 (40.9)	49 (48.5)	7.6 (-4.5 to 19.7)	0.22	28 (37.8)	-3.1 (-16.2 to 10.1)	0.66
Any secondary outcome(s)—significant	186 (52.2)	90 (49.7)	53 (52.5)	2.8 (-9.4 to 14.9)	0.67	43 (58.1)	8.4 (-5.0 to 21.8)	0.22
Reasons for high risk of bias								
Randomisation	122 (34.3)	67 (37.0)	34 (33.7)	-3.4 (-15.0 to 8.2)	0.58	19 (25.7)	-11.3 (-23.5 to 0.9)	0.07
Deviations from intended intervention	214 (60.1)	111 (61.3)	56 (55.5)	-5.9 (-17.9 to 6.1)	0.34	47 (63.5)	2.2 (-10.9 to 15.3)	0.76
Missing outcome data	19 (5.3)	11 (6.1)	5 (5.0)	-1.1 (-6.6 to 4.4)	0.70	3 (4.1)	-2.0 (-7.7 to 3.7)	0.50
Measurement of the outcome	7 (2)	6 (3.3)	1 (1.0)	-2.3 (-5.6 to 0.9)	0.16	0	-3.3 (-5.9 to -0.7)	0.01
Selective reporting	22 (6.2)	9 (5.0)	8 (7.9)	3.0 (-3.2 to 9.1)	0.35	5 (6.8)	1.8 (-4.8 to 8.3)	0.60
Overall high risk of bias	228 (64)	117 (64.6)	61 (60.4)	-4.2 (-16.1 to 7.6)	0.49	50 (67.6)	2.9 (-9.8 to 15.7)	0.67
Funding†								
Industry	79 (22.2)	32 (17.7)	30 (29.7)	12.0 (1.5 to 22.5)	0.02	17 (23)	5.3 (-5.8 to 16.4)	0.35
Government	126 (35.4)	47 (26.0)	37 (36.6)	10.7 (-0.7 to 22.0)	0.07	42 (56.8)	30.8 (17.8 to 43.8)	<0.001
Institution	128 (36.0)	69 (38.1)	32 (31.7)	-6.4 (-17.9 to 5.1)	0.28	27 (36.5)	-1.6 (-14.7 to 11.4)	0.82
Not-for-profit	51 (14.3)	24 (13.3)	12 (11.9)	-1.4 (-9.4 to 6.6)	0.75	15 (20.3)	7.0 (-3.4 to 17.4)	0.19
Not reported	23 (6.5)	12 (6.6)	10 (9.9)	3.3 (-3.6 to 10.1)	0.36	1 (1.4)	-5.3 (-9.8 to -0.8)	0.02
No funding	49 (13.8)	33 (18.2)	10 (9.9)	-8.3 (-16.4 to -0.2)	0.04	6 (8.1)	-1.1 (-11.1 to 8.5)	0.02

CI=confidence interval; IQR=interquartile range.

*Estimates only include trials addressing treatment.

†Trials might be classified in more than one category.

Table 2 | Time to publication of covid-19 trial preprints

Time	Proportion of preprints published (No/total No (%))	Cumulative proportion (%) of preprints published
0 to 2 weeks	3/174 (1.7)	1.70
2 weeks to 1 month	5/169 (3.0)	4.70
1 month to 2 months	12/162 (7.4)	12.10
2 months to 3 months	19/133 (14.3)	25.70
3 months to 6 months	27/98 (27.6)	50.10
6 months to 1 year	9/42 (21.4)	65.90
Median (95%CI) time to publication	5.9 months (5.1 to 10.3)	

CI=confidence interval.

Differences between preprint and published trial reports

Forty two (56.8%) trials had one or more discrepancies in the reporting of key methods and results between the preprint and the later published trial report. We identified a median of 1 (interquartile range 0-2) discrepancy per pair of preprint and publication reports. Online supplement 5 describes these discrepancies.

Thirty (40.5%) trials had one or more discrepancies in the reporting of key methods. The most common discrepancy in the reporting of key methods was the description of allocation concealment, which occurred in eight trials. For four of these eight trials, our judgment of risk of bias for the randomisation domain changed from "probably high" to "low" owing to additional details reported in the published report. [Box 1](#) presents an example.

Other differences in the reporting of key methods were the publication reporting one or more additional statistics important for meta-analysis (eg, interquartile ranges or standard deviations) that were not previously reported in the preprint (n=6; 8.1%), the preprint reporting on interim results and the publication on completed trial results (n=4; 5.4%), and the publication including a protocol or statistical analysis plan as a supplementary that was not previously included with the preprint (n=3; 4.1%). The overall trial rating of risk of bias, however, changed only for one trial based on additional information provided in the published report.

Thirty one (41.9%) trials had one or more differences in the reporting of key results between preprints and publications. The most common discrepancy in the reporting of key results were changes in outcome data between preprints and publications, which was seen in 20 (27.0%) trials—although most of these discrepancies might likely be attributed to events accumulating in trials from the time when the preprint was posted to when the trial was published. [Table 4](#) presents an example.

Despite discrepancies in outcome data being common, results were similar between preprints and publications both in magnitude and precision. [Figure 1](#) shows differences in results on

mortality and mechanical ventilation between preprints and publications. Among all preprints with differences in outcomes, differences in relative effects did not exceed 15%, except for one trial with very few events that included just one additional event in the publication.²⁸ Other differences between preprints and publications in key results included the publication reporting at least one additional key outcome that was not included in the preprint (n=11; 14.9%).

Retractions

We identified four retracted trials.^{29–36} Two trials reported on (hydroxy)chloroquine,^{29–32} two on favipiravir,^{31–34} and two on ivermectin.^{29 30 35 36} One of the trials was retracted when the authors noticed an error in their analysis^{35 36} and the remainder were retracted owing to concerns about data fabrication or falsification (eg, inconsistencies between the eligibility criteria and patients included in the trial, discrepancies between when the trial was reported to have been conducted and when patients were recruited, inconsistencies between the dataset and the results reported in the preprint, and inconsistencies between the distribution of baseline variables and the described randomisation procedure). We compared the number of retractions between preprints and journal publications. One of the retracted trials was posted as a preprint^{29 30} and the remainder were published in peer reviewed journals.

Meta-analyses including v excluding preprint reports

[Tables 5 and 6](#) presents results of meta-analyses including and excluding data from unpublished preprints for the comparison of placebo or standard care with several treatments for covid-19, with the outcomes of mortality and mechanical ventilation. Treatments included corticosteroids, remdesivir, lopinavir-ritonavir, hydroxychloroquine, ivermectin, interleukin 6 receptor blockers, and convalescent plasma. Results were recorded at one, three, and six months after the first trial addressing the intervention was made public, either as a preprint or a publication, and at the longest point of follow-up of the trials (up to 3 August 2021). Online supplement 6 presents a more detailed table.

In total, we performed and assessed the certainty of evidence of 120 meta-analyses, 60 of which included preprints and 60 of which excluded preprints. Online supplement 7 presents forest plots for meta-analyses.

Because of insufficient data, we could not perform meta-analyses for six comparisons without preprints: mortality and mechanical ventilation, at one month, for ivermectin versus placebo or standard care and for interleukin 6 receptor blockers versus placebo or standard care; and mortality and mechanical ventilation, at three months, for ivermectin versus placebo or standard care.

Table 3 | Predictors of time to publication of covid-19 trial preprints

	Median (95% CI) time to publication (days)	P value
Funding		
Industry funding (n=47)	171 (145 to NA)	0.72
No industry funding (n=127)	219 (177 to NA)	
Government funding (n=78)	155 (120 to 192)	0.02
No government funding (n=96)	309 (183 to NA)	
Institutional funding (n=58)	136 (90 to NA)	0.14
No institutional funding (n=116)	192 (166 to NA)	
Not-for-profit funding (n=26)	110 (79 to NA)	0.07
No not-for-profit funding (n=148)	188 (161 to NA)	
No of centres		
Single centre (n=65)	Undefined (130 to undefined)	0.13
Multicentre (n=90)	161 (136 to 219)	
No of participants		
More than median (n=88)	142 (120 to NA)	0.14
Less than median (n=85)	219 (171 to NA)	
Intensity of care		
Inpatient (n=122)	110 (95 to 144)	<0.0001
Outpatient (n=38)	Undefined	
Severity		
Mild or moderate disease (n=79)	182 (159 to 263)	0.00001
Severe or critical disease (n=42) ³⁷	101 (89 to 128)	
Early termination for benefit	NA	NA
Primary outcome significance		
Significant (n=77)	243 (155 to undefined)	0.12
Not significant (n=91)	161 (127 to 219)	
Any secondary outcomes statistically significant		
Significant (n=96)	187 (120 to NA)	0.66
Not significant (n=64)	152 (120 to NA)	
Risk of bias		
Low (n=52)	155 (124 to 187)	0.17
High (n=102)	122 (98 to 161)	

CI= confidence interval; NA=not available. Upper bounds of confidence intervals could often not be estimated owing to insufficient follow-up of preprints.

Differences in estimates from meta-analyses including v excluding preprints

Except for two (3.3%) cases, all meta-analyses including and excluding results from unpublished preprints produced point estimates that were consistent as to whether they indicated benefit, no appreciable effect, or harm. The meta-analysis of corticosteroids at onemonth suggested a reduction in risk of mechanical ventilation when preprints were excluded (43 fewer per 1000 people (95% confidence interval 59.24 fewer to 22.12 fewer); moderate

BOX 1 | EXAMPLE OF TRIAL REPORTING ADDITIONAL INFORMATION ON ALLOCATION CONCEALMENT IN ITS PUBLISHED REPORT

The PANAMO trial, which was initially available as a preprint on the Social Science Research Network and later published in *Lancet Rheumatology*, provided additional details on allocation concealment in the publication. The publication described central randomisation with an online tool and the development of the randomisation list by a third party—all of which were not reported in the preprint.^{49 50} This addition resulted in a change in the rating of the risk of bias owing to randomisation, from "probably high risk of bias" to "definitely low risk of bias."

PREPRINT

"Patients were randomly assigned in a 1:1 ratio to receive IFX-1, at a dose of 800 mg intravenously, for a maximum of seven doses, plus best supportive care, or best supportive care only ... Randomisation was performed with an online tool within the eCRF (electronic case report form) and was stratified by study site."

PUBLICATION

"Patients were randomly assigned in a 1:1 ratio to IFX-1 plus best supportive care (the IFX-1 group) or to best supportive care only (the control group). Randomisation was done by investigators centrally with an online tool within the electronic case report form and was stratified by study site. The tool used a randomised variable block length of either two or 4. The randomisation list was only available to contract research organisation (Metronomia) staff involved in the production of the randomisation list and set-up of the online randomisation tool."

certainty) and no appreciable effect when preprints were included (1.2 more per 1000 people (60.3 fewer to 131.1 more); very low certainty). The meta-analysis without preprints included one trial with 5418 participants and the meta-analysis with preprints included two trials with 5472 participants.

The meta-analysis of ivermectin at six months suggested no appreciable effect on risk of mechanical ventilation when preprints were excluded (2.3 fewer per 1000 people (95% confidence interval 52.2 fewer to 83.5 more); low certainty) and a reduction in risk of mechanical ventilation when preprints were included (26.7 fewer per 1000 people (74.2 fewer to 75.4 more); very low certainty). The meta-analysis without preprints included seven trials with 1826 participants and the meta-analysis with preprints included nine trials with 4000 participants. Four of 60 meta-analyses had results that were significant with preprints and not significant without preprints, or vice versa.

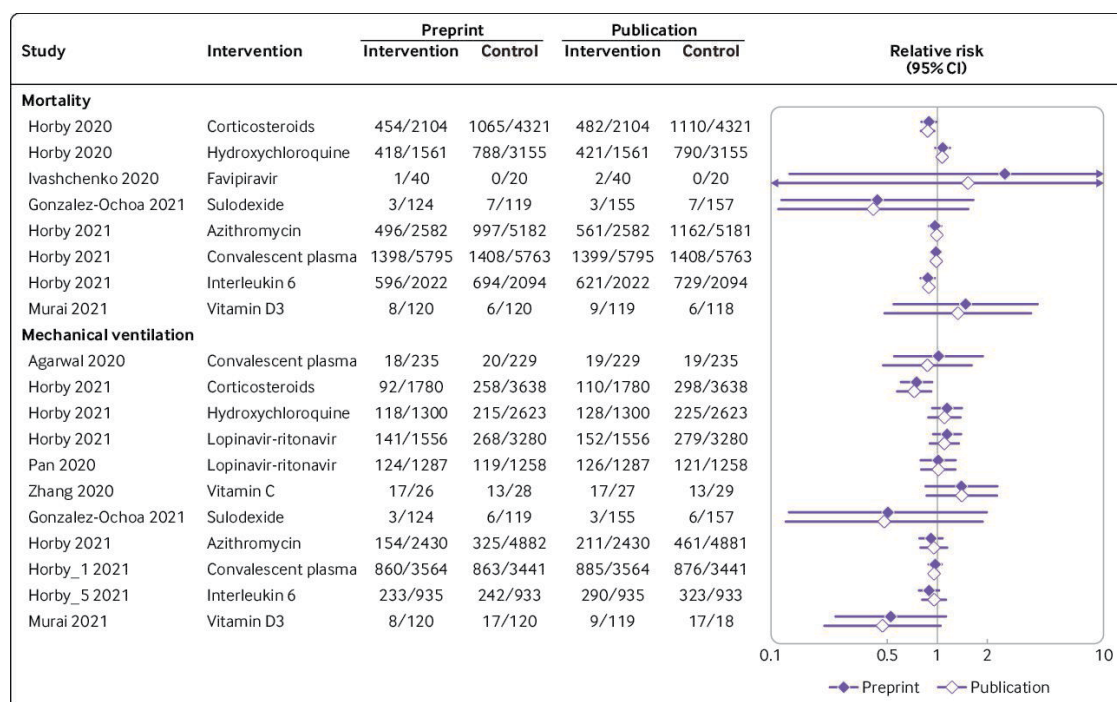


Figure 1 | Differences in results on mortality and mechanical ventilation between preprints and publications of covid-19 trials.^{1-5 28} Data are events/total number of participants

Differences in ratings of certainty of evidence from meta-analyses including v excluding preprints

We judged nine (15%) of 60 meta-analyses to have different ratings of the certainty of evidence when preprints were included versus when preprints were excluded. For four of these nine cases, we rated the certainty of evidence for the meta-analyses including preprints to be higher than the evidence excluding preprints. For five of these nine cases, we rated the certainty of meta-analyses excluding preprints to be higher than the meta-analyses including preprints. In six of these cases, differences in ratings of the certainty of evidence could have affected decision making (ie, evidence including preprints is rated as high or moderate whereas evidence excluding preprints is rated as low or very low, or vice versa).

Differences in ratings of GRADE domains from meta-analyses including v excluding preprints

Risk of bias

Between meta-analyses including preprints and meta-analyses excluding preprints, judgments

related to the GRADE risk-of-bias domain differed only for one meta-analysis (remdesivir v standard care or placebo for mechanical ventilation at six months). We judged the meta-analysis excluding preprints to not have any concerns related to risk of bias and downgraded the meta-analysis including preprints due to serious risk of bias.

Imprecision

Between meta-analyses including preprints and meta-analyses excluding preprints, judgments related to the GRADE imprecision domain differed for 13 of 60 meta-analyses. We judged nine meta-analyses excluding preprints to have more serious concerns related to imprecision than their counterparts including preprints (ie, additional data from preprints narrowed confidence intervals). Furthermore, we judged four meta-analyses excluding preprints to have less serious concerns related to imprecision than meta-analyses including preprints (ie, additional data from preprints increased statistical heterogeneity and hence imprecision).

Table 4 | Example of trial* reporting different outcome data between preprint⁵¹ and publication,² by outcome and treatment group

Outcome	Preprint ⁵¹	Publication ²
Mortality	Corticosteroids: 454/2104 (21.6) Standard care: 1065/4321 (24.6)	Corticosteroids: 482/2104 (22.9) Standard care: 1110/4321 (25.7)
Mechanical ventilation	Corticosteroids: 92/1780 (5.2) Standard care: 258/3638 (7.1)	Corticosteroids: 110/1780 (6.2) Standard care: 298/3638 (8.2)

Data are events/total number (%) of participants in each treatment group.

*The RECOVERY publication on dexamethasone treatment for covid-19 reported different results for mortality and mechanical ventilation—likely not because of an error in the preprint but because of events accumulating in the trial from when the preprint was posted to when it was published.^{2 51}

Table 5 | Results of meta-analyses excluding and including results from preprints, by mortality outcome

Drug treatment duration	Meta-analyses excluding unpublished preprints			Meta-analyses including all preprints		
	No of studies (participants)	Risk difference with standard care/ placebo (95% CI; per 1000 people)	GRADE assessment for certainty of evidence	No of studies (participants)	Risk difference with standard care/ placebo (95% CI; per 1000 people)	GRADE assessment for certainty of evidence
Mortality						
Corticosteroids						
1 month	1 (6425)	14.3 fewer (24.7 fewer to 2.6 fewer)	Moderate	2 (6489)	14.3 fewer (24.7 fewer to 2.6 fewer)	Moderate
3 months	5 (7667)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate	6 (7731)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate
6 months	5 (7667)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate	6 (7731)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate
Current	10 (7959)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate	10 (7959)	13 fewer (22.1 fewer to 3.9 fewer)	Moderate
Remdesivir						
1 month	2 (1298)	27.3 fewer (53.3 fewer to 6.5 more)	Moderate	2 (1298)	27.3 fewer (53.3 fewer to 6.5 more)	Moderate
3 months	2 (1298)	27.3 fewer (53.3 fewer to 6.5 more)	Moderate	2 (1298)	27.3 fewer (53.3 fewer to 6.5 more)	Moderate
6 months	3 (1882)	28.6 fewer (53.3 fewer to 5.2 more)	Moderate	4 (7333)	13 fewer (35.1 fewer to 14.3 more)	Low
Current	5 (7415)	11.7 fewer (32.5 fewer to 14.3 more)	Low	6 (8247)	10.4 fewer (27.3 fewer to 9.1 more)	Moderate
Lopinavir-ritonavir						
1 month	1 (199)	29.9 fewer (71.5 fewer to 39 more)	Very low	2 (250)	29.9 fewer (71.5 fewer to 39 more)	Very low
3 months	2 (250)	29.9 fewer (71.5 fewer to 39 more)	Very low	2 (250)	29.9 fewer (71.5 fewer to 39 more)	Very low
6 months	2 (250)	29.9 fewer (71.5 fewer to 39 more)	Very low	2 (250)	29.9 fewer (71.5 fewer to 39 more)	Very low
Current	7 (9427)	5.20 more (6.5 fewer to 18.2 more)	Very low	7 (9427)	5.20 more (6.5 fewer to 18.2 more)	Very low
(Hydroxy)chloroquine						
1 month	1 (30)	NA	NA	1 (30)	NA	NA
3 months	1 (30)	NA	NA	2 (180)	NA	NA
6 months	5 (1287)	20.8 more (54.6 fewer to 174.2 more)	Very low	9 (6135)	10.4 more (2.60 fewer to 24.7 more)	Low
Current	19 (10634)	11.7 more (0 fewer to 24.7 more)	Low	23 (10997)	9.10 more (2.60 fewer to 22.1 more)	Low
Ivermectin						
1 month	0 (0)	NA	NA	1 (180)	106.6 fewer (122.2 fewer to 65 fewer)	Very low
3 months	0 (0)	NA	NA	4 (517)	87.1 fewer (118.3 fewer to 22.1 more)	Very low
6 months	1 (398)	87.1 fewer (128.7 fewer to 916.5 more)	Very low	6 (1169)	85.8 fewer (115.7 fewer to 0 fewer)	Low
Current	5 (1220)	36.4 fewer (93.6 fewer to 110.5 more)	Very low	9 (1879)	63.7 fewer (100.1 fewer to 16.9 more)	Very low
Interleukin 6 receptor blockers						
1 month	0 (0)	NA	NA	1 (97)	20.8 fewer (70.2 fewer to 66.3 more)	Very low
3 months	2 (26)	91 fewer (124.8 fewer to 165.1 more)	Very low	4 (435)	15.6 fewer (54.6 fewer to 41.6 more)	Very low
6 months	6 (1292)	23.4 fewer (42.9 fewer to 0 fewer)	Low	7 (5408)	22.1 fewer (31.2 fewer to 13 fewer)	Moderate
Current	8 (5457)	22.1 fewer (31.2 fewer to 13 fewer)	Moderate	11 (6303)	22.1 fewer (31.2 fewer to 13 fewer)	Moderate
Convalescent plasma						
1 month	1 (101)	45.5 fewer (92.3 fewer to 59.8 more)	Very low	2 (187)	52 fewer (87.1 fewer to 13 more)	Very low
3 months	1 (101)	45.5 fewer (92.3 fewer to 59.8 more)	Very low	4 (428)	57.2 fewer (88.4 fewer to 3.90 fewer)	Very low
6 months	3 (898)	6.5 fewer (41.6 fewer to 42.9 more)	Very low	7 (1185)	22.1 fewer (48.1 fewer to 14.3 more)	Very low
Current	9 (12962)	2.60 fewer (10.4 fewer to 6.5 more)	Moderate	14 (16073)	2.60 fewer (9.10 fewer to 3.90 more)	Moderate
(Hydroxy)chloroquine (prophylaxis)						
1 month	1 (744)	NA	NA	1 (744)	NA	NA
3 months	1 (744)	NA	NA	2 (3151)	0.8 fewer (2.3 fewer to 3.7 more)	High
6 months	4 (8569)	0.8 fewer (2 fewer to 3.7 more)	High	4 (8569)	0.8 fewer (2 fewer to 3.7 more)	High
Current	4 (8569)	0.8 fewer (2 fewer to 3.7 more)	High	4 (8569)	0.8 fewer (2 fewer to 3.7 more)	High

GRADE=grading of recommendations assessment, development, and evaluation; NA=not available.

Discussion

Main findings

Our study presents a detailed assessment of the degree of discrepancies between covid-19 trial preprints and their later publications and the impact of trial preprints meta-analytic estimates, the certainty of evidence, and decision making. We show that preprints remain the only source of findings of many trials for several months. Half of all preprints, for example, remain unpublished at six months and a third at one year—a length of time that might be unacceptable in a health

emergency or to patients who might expect that their care is guided by the most recent and best available evidence. Preprints can importantly accelerate the time to dissemination of trial findings.

We did not find compelling evidence of important differences between preprint and published reports of trials—although preprint reports of trials that are subsequently published in journals might not be representative of all trial preprints. Further, we found retractions to occur for both preprints and publications, suggesting that publication in a peer reviewed

Table 6 | Results of meta-analyses excluding and including results from preprints, by mechanical ventilation outcome

Drug treatment duration	Meta-analyses excluding unpublished preprints			Meta-analyses including all preprints		
	No of studies (participants)	Risk difference with standard care/ placebo (95% CI; per 1000 people)	GRADE assessment for certainty of evidence	No of studies (participants)	Risk difference with standard care/ placebo (95% CI; per 1000 people)	GRADE assessment for certainty of evidence
Mechanical Ventilation						
Corticosteroids						
1 month	1 (5418)	43 fewer (59.24 fewer to 22.12 fewer)	Moderate	2 (5472)	1.2 more (60.3 fewer to 131.1 more)	Very low
3 months	5 (6324)	32.56 fewer (44.16 fewer to 19.8 fewer)	Moderate	6 (6378)	17.4 fewer (29 fewer to 3.5 fewer)	Moderate
6 months	5 (6324)	32.56 fewer (44.16 fewer to 19.8 fewer)	Moderate	6 (6378)	17.4 fewer (29 fewer to 3.5 fewer)	Moderate
Current	9 (6576)	27.92 fewer (39.52 fewer to 15.16 fewer)	Moderate	9 (6576)	13.9 fewer (25.5 fewer to 1.2 fewer)	Moderate
Remdesivir						
1 month	2 (1001)	47.6 fewer (65 fewer to 24.4 fewer)	High	2 (1001)	47.6 fewer (65 fewer to 24.4 fewer)	High
3 months	2 (1001)	47.6 fewer (65 fewer to 24.4 fewer)	High	2 (1001)	47.6 fewer (65 fewer to 24.4 fewer)	High
6 months	3 (1585)	51 fewer (67.3 fewer to 30.2 fewer)	High	4 (6549)	39.4 fewer (68.4 fewer to 8.1 more)	Low
Current	5 (6619)	32.5 fewer (62.6 fewer to 13.9 more)	Low	6 (7451)	27.8 fewer (52.2 fewer to 4.6 more)	Low
Lopinavir-ritonavir						
1 month	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low
3 months	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low
6 months	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low	1 (198)	30.2 fewer (71.9 fewer to 48.7 more)	Very low
Current	5 (8474)	16.2 more (2.3 fewer to 30.2 more)	Low	5 (8474)	16.2 more (2.3 fewer to 30.2 more)	Low
(Hydroxy)chloroquine (treatment)						
1 month	2 (642)	16.2 more (45.2 fewer to 129.9 more)	Very low	4 (4616)	17.4 more (7.0 fewer to 45.2 more)	Low
3 months	4 (4693)	19.7 more (4.6 fewer to 46.4 more)	Low	6 (6430)	17.4 more (3.5 fewer to 40.6 more)	Low
6 months	7 (6877)	15.1 more (4.6 fewer to 37.1 more)	Low	9 (7417)	12.8 more (4.6 fewer to 33.6 more)	Low
Current	12 (8053)	26.7 more (5.8 more to 53.4 more)	Moderate	12 (8053)	26.7 more (5.8 more to 53.4 more)	Moderate
Ivermectin						
1 month	1 (45)	NA	NA	2 (90)	60.3 more (107.9 fewer to 496.5 more)	Very low
3 months	1 (45)	NA	NA	4 (354)	69.6 fewer (109 fewer to 169.4 more)	Very low
6 months	4 (642)	2.3 fewer (52.2 fewer to 83.5 more)	Low	7 (951)	26.7 fewer (74.2 fewer to 75.4 more)	Very low
Current	8 (1464)	7.0 fewer (48.7 fewer to 61.5 more)	Low	9 (1616)	7.0 fewer (48.7 fewer to 61.5 more)	Low
Interleukin six receptor blockers						
1 month	0 (0)	NA	NA	1 (273)	27.8 fewer (54.5 fewer to 10.4 more)	Low
3 months	3 (495)	37.1 fewer (66.1 fewer to 10.4 more)	Low	5 (1145)	33.6 fewer (52.2 fewer to 8.1 fewer)	Moderate
6 months	7 (1826)	30.2 fewer (42.9 fewer to 16.2 fewer)	Moderate	9 (4000)	20.9 fewer (31.3 fewer to 8.1 fewer)	Moderate
Current	10 (4170)	19.7 fewer (30.2 fewer to 8.1 fewer)	Moderate	12 (4560)	19.7 fewer (30.2 fewer to 9.3 fewer)	Moderate
Convalescent plasma						
1 month	1 (464)	9.3 more (47.6 fewer to 104.4 more)	Very low	3 (705)	19.7 fewer (70.8 fewer to 90.5 more)	Very low
3 months	3 (827)	16.2 more (22 fewer to 70.8 more)	Very low	6 (1108)	4.6 more (29 fewer to 48.7 more)	Very low
6 months	4 (987)	12.8 more (24.4 fewer to 62.6 more)	Very low	6 (1108)	4.6 more (29 fewer to 48.7 more)	Very low
Current	8 (8252)	2.3 fewer (11.6 fewer to seven more)	Moderate	9 (8333)	2.3 fewer (11.6 fewer to 5.8 more)	Moderate

GRADE=grading of recommendations assessment, development, and evaluation; NA=not available.

journal alone does not indicate the trustworthiness of a trial report.

We also found that in most cases, meta-analyses including versus excluding evidence from preprints yielded consistent results. In a minority of circumstances, however, including preprints improved the certainty (quality) of evidence. At six months after trial data first became available, when preprints were excluded, we found low certainty evidence indicating that interleukin 6 receptor blockers might reduce mortality—downgraded because of risk of bias and imprecision. But when preprints were included, we found moderate certainty evidence indicating that interleukin 6 receptor blockers probably reduce mortality—downgraded because of risk of bias. For interleukin 6 receptor blockers, consideration of evidence from preprints could have importantly

accelerated the time to incorporation of this treatment as part of standard care.²⁰

In a minority of circumstances, meta-analyses including preprints also reduced the certainty of evidence. At one month after trial data first became available, for example, when preprints were excluded, we found moderate certainty evidence that corticosteroids probably reduce mechanical ventilation, downgraded because of serious risk of bias; we also found very low certainty evidence when we included preprints, downgraded because of serious risk of bias and very serious imprecision. For corticosteroids, including preprints increased imprecision.

Implications

Our findings have implications for evidence users, such as clinicians, who are concerned with the

quality of preprints and for systematic reviewers and guideline developers deciding whether to consider preprint reports in systematic reviews and guideline recommendations. Our results support the overall usefulness of preprints, and preprints as a venue through which the dissemination of trial findings might be accelerated. While we found that preprints could both increase or reduce the certainty of evidence, we encourage systematic reviewers and guideline developers to consider evidence from preprints, appraise preprint reports, and only consider excluding preprints in situations where there are data fabrication or integrity issues.

Because at the point of submission to peer reviewed journals trials have already been completed and major methodological decisions (such as whether to collect data on an outcome or whether to blind investigators) cannot be changed, peer review probably looks at only the transparency of trial reports and interpretation of results. We encourage systematic reviewers and guideline developers to consider trial evidence from preprints, especially in circumstances in which decisions are being made rapidly and evidence is being produced faster than can be peer reviewed and published.

We caution, however, that preprints (and publications) might describe untrustworthy trials with fabricated or falsified data. Evidence users might consider scrutinising both preprint and published trials for anomalies that suggest fabrication and falsification (examples of which are reported in box 2).^{37 38} While such methods require subjective judgments and cannot be used to definitively identify untrustworthy trials, they could be useful to identify trials that are at high risk of such issues. Evidence users might subsequently investigate such trials further or systematic reviewers might consider sensitivity analyses excluding such trials from meta-analyses. We direct readers to other sources that describe these methods.³⁹ These methods could also be useful to journal editors and peer reviewers to apply when considering trials for publication.

Review authors who are concerned about publication bias or are wanting to make decisions within timeframes that might not be conducive to peer review and publication might also consider conducting prospective meta-analyses—meta-analyses that are conducted using an inception cohort of registered trials and incorporate unpublished data from investigators.⁴⁰ During the covid-19 pandemic, investigators have conducted such prospective meta-analyses to review the effectiveness of corticosteroids and interleukin 6 receptor blockers for covid-19.^{41 42}

Relation of study's findings to previous work

Our study presents data on the contribution of preprints to the body of evidence of covid-19 treatments and prophylaxis. Three studies have reported on differences between covid-19 preprint and

BOX 2 | METHODS TO ASSESS FOR FABRICATION AND FALSIFICATION IN CLINICAL TRIALS³⁹

- ⇒ Judge whether the reported recruitment speed is feasible given local disease patterns, trial eligibility criteria, and capacity of the recruiting centres^{52 53}
- ⇒ Review the trial registration or protocol and assess the consistency between the registration and the manuscript in aspects of the trial that cannot be modified after completion (eg, blinding status)
- ⇒ Review profiles of the investigators or institutions involved, for a history of research misconduct
- ⇒ Review baseline patient characteristics and test whether reported distributions are consistent with randomisation^{37 38 54}
- ⇒ Review primary data, when available, for duplicate records or inconsistencies between the data and reported statistics in the trial manuscript
- ⇒ Review primary data, when available, and assess whether correlations between variables are plausible⁵²

published study reports and citations and Altmetric attention metrics.^{43–45} One study looked at publication characteristics and dissemination of covid-19 preprints, another looked at outcome reporting and spin in interpretation of results, and another looked at risk of bias and spin. These studies were, however, restricted to only publications up to August and October 2020—which is not representative of the current landscape of covid-19 research and which does not include the majority of evidence being currently used to guide covid-19 care, including critical trials of the effects of corticosteroids and interleukin 6 receptor blockers.^{1 2} These studies did not compare the effects of including preprints on meta-analytic estimates and the certainty of the body of evidence, which is particularly important because evidence users use the totality of the body of evidence, rather than single studies, to make treatment decisions and recommendations.⁴³ One study has reviewed publication rates and citations for covid-19 research but it primarily deals with basic science research and so its findings might not apply to covid-19 clinical trials.⁴⁶

The covid-19 pandemic highlighted the need for rapid dissemination of research and incited increased interest in preprint servers, which yielded an substantial amount of research that was published on preprint servers and which made this study possible. We are not aware of studies looking at the trustworthiness or impact of preprints in other areas—although such research in other areas would

also be useful. Our results are also aligned with previous assessments before covid-19 that study interpretations and other study details do not change importantly between preprints and their later publications in high impact journals.⁴⁷

Strengths and limitations of the study

The strengths of this study include the comprehensive search for and inclusion of preprint and published covid-19 trial reports and rigorous data collection. The generalisability of our results is, however, limited to covid-19. Journals have expedited the publication of covid-19 research and have been publishing more prolifically on covid-19 than on other areas, which could reduce opportunity for revisions between preprints and their subsequent publications and might mean that time to and predictors of publication might differ than in other research areas. For these reasons, our estimated time to publication of covid-19 preprints is different from estimates made from before covid-19.⁴⁸

Although the WHO covid-19 database is a comprehensive source of published and preprint literature, it does not include all preprint servers—but preprint servers not covered by our search strategy include other subjects and are unlikely to include covid-19 trials.

Preprints of trials that are subsequently published in journals could represent the most rigorous or transparently reported preprints, which are not representative of all trial preprints. To minimise this limitation, we compared the characteristics of published and unpublished trial preprints and we did not identify any important differences. This comparison suggested that published and unpublished trial reports in our sample are comparable.

Furthermore, published trial reports might still contain errors, posting trial reports as preprints could allow more errors to be identified before final publication, and even trials without discrepancies between the preprint and later published report might have important limitations that reduce their trustworthiness.

Our assessment of differences in key results between preprints and publications was limited to the outcomes that were included in our living SRNMA. While these outcomes were identified as being important or critical to decision making by coauthors of the living SRNMA and the parallel guideline, they do not include adverse events. Differences in such outcomes could exist between preprints and publications.⁴⁴ Further, other aspects of the reporting of results (eg, baseline characteristics of patients) could be different between preprints and publications.

We report on the number of publications and preprints that were retracted. Preprints, however, could be less likely to be retracted because they might draw less attention and because preprint

servers could be less likely than journals to have formal policies on research integrity. Further, we found too few retractions to be able to draw confident conclusions.

Our assessment of the impact of preprints focused only on the impact of preprints on meta-analytic estimates, the certainty of evidence, and decision making and did not consider other aspects of impact, such as number of citations or Altmetrics.

We used the GRADE approach to assess the certainty of evidence.²⁶ While the GRADE framework provides a transparent and systematic framework of all factors that might bear on the certainty of evidence, its application is subjective.

Our assessment of the contribution of trial preprint reports to meta-analytic estimates and their effect on the certainty of evidence was undertaken in the context of pairwise meta-analysis and the minimally contextualised approach for assessing the certainty of evidence, whereas parallel guideline recommendations have been based on network meta-analyses and the fully contextualised approach.^{20 27} Despite no compelling reasons that our results will not be generalisable to network meta-analyses, more differences in judgments will be related to the certainty of evidence in a fully contextualised framework where judgments are more dependent on the magnitude and precision of estimates.

We limited our assessment of the impact of meta-analyses including versus excluding preprint reports on meta-analytic estimates to only interventions that have been addressed by the WHO living guideline at the time of analysis.²⁰ While the effects of including or excluding preprints in meta-analyses might vary across interventions, our analysis looks at the interventions that have had sufficient interest and research to instigate guideline recommendations. Interventions informed by fewer trials may be more sensitive to including or excluding evidence from preprints.

Our estimate of the time to publication of preprint reports might have been over-estimated if some preprint authors did not attempt to subsequently publish in peer reviewed journals—although evidence shows that most authors of covid-19 preprints intend to publish their findings.⁴³ Time to publication might also be under-estimated if preprints are made public later in the submission process.

Where discrepancies exist between preprints and publications, systematic reviewers and guideline developers will also need to consider whether the results of preprints or publications are more trustworthy and should be incorporated in the meta-analysis. In such situations, systematic reviewers and guideline developers might assume that changes between preprints and publications are due to errors or inaccuracies in the reporting of the preprint that were later corrected during peer review. Future research should investigate whether including

results from preprints or final publications affects overall findings and decisions.

Finally, although we describe discrepancies in the reporting of key methods and results, we did not assess differences in the discussion or conclusion sections of trial reports.

Conclusions

We found no compelling evidence indicating important discrepancies between preprint and published trial reports. We show that including preprints might affect the results of meta-analyses and the certainty of evidence, and might encourage evidence users to consider data from preprints in contexts in which decisions are being made rapidly and evidence is being produced faster than can be peer reviewed and published. Scepticism might still be warranted when suspicion arises regarding falsified data (for which we provide criteria).

AUTHOR AFFILIATIONS

- ¹Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
²McMaster University, Hamilton, ON, Canada
³University of Ottawa, Ottawa, ON, Canada
⁴Internal Medicine Residency Program, University of Calgary Cumming School of Medicine, Calgary, AB, Canada
⁵Department of Medicine, University of Toronto, Toronto, ON, Canada
⁶Department of Medicine, University of Alberta Faculty of Medicine and Dentistry, Edmonton, AB, Canada
⁷Department of Pharmacy, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
⁸Epistemonikos Foundation, Santiago, Chile
⁹UC Evidence Centre, Cochrane Chile Associated Centre, Pontificia Universidad Católica de Chile, Santiago, Chile
¹⁰Biomedical Informatics, Harvard Medical School, Boston, MA, USA
¹¹Faculty of Dentistry, University of Chile, Santiago, Chile

Twitter Dena Zeraatkar @dena.zera

Contributors DZ, RACS, and RB-P conceptualised the study. JJB, FV-P, GR, and AQ screened studies for eligibility. TP, GL, EC, AA, FK, ZE, MAC, MG, YW, EK, JJB, and AQ collected data. DZ drafted the first version of the manuscript. All authors reviewed the manuscript and provided important intellectual input. DZ is study guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Transparency: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Funding This study was supported by the Canadian Institutes of Health Research (CIHR-IRSC:057900132) and Coronavirus Rapid Research Funding Opportunity (OV2170359). DZ is funded by a Banting Postdoctoral Fellowship. The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: support from the Canadian Institutes of Health Research for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available at <https://osf.io/gadxb/>.

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

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

ORCID iDs

Dena Zeraatkar <http://orcid.org/0000-0003-4287-0541>
 Francisca Verdugo-Paiva <http://orcid.org/0000-0003-0199-9744>
 Gabriel Rada <http://orcid.org/0000-0003-2435-0710>
 Jessica Julia Bartoszko <http://orcid.org/0000-0001-6743-3538>
 Reed Alexander Cunningham Siemieniuk <http://orcid.org/0000-0002-3725-3031>

REFERENCES

- Horby P, Lim WS, Emberson J. Effect of dexamethasone in hospitalized patients with COVID-19: preliminary report. *medRxiv* 2020. doi:10.1056/NEJMoA2021436
- Horby P, Lim WS, Emberson JR. Dexamethasone in hospitalized patients with Covid-19 - preliminary report. *N Engl J Med* 2020. doi:10.1056/NEJMoA2021436
- Horby P, Mafham M, Linsell L. Effect of hydroxychloroquine in hospitalized patients with COVID-19: preliminary results from a multi-centre, randomized, controlled trial. *medRxiv* 2020. doi:10.1056/NEJMoA2022926
- RECOVERY Collaborative Group, Horby P, Mafham M, *et al.* Effect of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med* 2020;383:2030–40. doi:10.1056/NEJMoA2022926
- WHO Solidarity Trial Consortium, Pan H, Peto R, *et al.* Repurposed antiviral drugs for Covid-19 - interim WHO solidarity trial results. *N Engl J Med* 2021;384:497–511. doi:10.1056/NEJMoA2023184
- Maslove DM. Medical preprints-a debate worth having. *JAMA* 2018;319:443–4. doi:10.1001/jama.2017.17566
- Lauer MS, Krumholz HM, Topol EJ. Time for a prepublication culture in clinical research? *Lancet* 2015;386:2447–9. doi:10.1016/S0140-6736(15)01177-0
- Walker R, Rocha da Silva P. Emerging trends in peer review-a survey. *Front Neurosci* 2015;9:169. doi:10.3389/fnins.2015.00169
- Chan A-W, Song F, Vickers A, *et al.* Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257–66. doi:10.1016/S0140-6736(13)62296-5
- Franco A, Malhotra N, Simonovits G. Social science. publication bias in the social sciences: unlocking the file drawer. *Science* 2014;345:1502–5. doi:10.1126/science.1255484
- Schmucker C, Schell LK, Portalupi S, *et al.* Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS One* 2014;9:e114023. doi:10.1371/journal.pone.0114023
- Scherer RW, Meerpohl JJ, Pfeifer N, *et al.* Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2018;11:MR000005. doi:10.1002/14651858.MR000005.pub4
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the food and drug administration: review of publication and presentation. *PLoS Med* 2008;5:e217. doi:10.1371/journal.pmed.0050217
- Carneiro CF, 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. doi:10.1186/s41073-020-00101-3
- van Schalkwyk MCI, Hird TR, Maani N, *et al.* The perils of preprints. *BMJ* 2020;370:m3111. doi:10.1136/bmj.m3111
- Flanagin A, Fontanarosa PB, Bauchner H. Preprints involving medical research-do the benefits outweigh the challenges? *JAMA* 2020;324:1840–3. doi:10.1001/jama.2020.20674
- Siemieniuk RA, Bartoszko JJ, Zeraatkar D, *et al.* Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ* 2020;370:m2980. doi:10.1136/bmj.m2980

- 18 Siemieniuk RA, Bartoszko JJ, Díaz Martínez JP, *et al.* Antibody and cellular therapies for treatment of covid-19: a living systematic review and network meta-analysis. *BMJ* 2021;374:n2231. doi:10.1136/bmj.n2231
- 19 Bartoszko JJ, Siemieniuk RAC, Kum E, *et al.* Prophylaxis against covid-19: living systematic review and network meta-analysis. *BMJ* 2021;373:n949. doi:10.1136/bmj.n949
- 20 Agarwal A, Rochwerf B, Lamontagne F, *et al.* A living who guideline on drugs for covid-19. *BMJ* 2020;370:m3379. doi:10.1136/bmj.m3379
- 21 Marshall LJ, Noel-Storr A, Kuiper J, *et al.* Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. *Res Synth Methods* 2018;9:602–14. doi:10.1002/jrsm.1287
- 22 Sterne JAC, Savović J, Page MJ, *et al.* Rob 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898
- 23 Centers for Disease Control and Prevention. COVIDView. A Weekly surveillance summary of U.S COVID-19 activity; 2020. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>
- 24 Centers for Disease Control and Prevention. Daily updates of totals by week and state: provisional death counts for coronavirus disease 2019 (COVID-19); 2020. <https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm>
- 25 ISARIC (International Severe Acute Respiratory and Emerging Infections Consortium). COVID-19 report: 08 June 2020 medRxiv; 2020.
- 26 Guyatt GH, Oxman AD, Vist GE, *et al.* Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6. doi:10.1136/bmj.39489.470347.AD
- 27 Hultcrantz M, Rind D, Akl EA, *et al.* The grade working group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 2017;87:4–13. doi:10.1016/j.jclinepi.2017.05.006
- 28 Ivashchenko AA, Dmitriev KA, Vostokova NV, *et al.* AVIFAVIR for treatment of patients with moderate coronavirus disease 2019 (COVID-19): interim results of a phase II/III multicenter randomized clinical trial. *Clin Infect Dis* 2021;73:531–4. doi:10.1093/cid/ciaa1176
- 29 Elgazzar A, Hany B, Youssef SA. Efficacy and safety of ivermectin for treatment and prophylaxis of COVID-19 pandemic. *Research Square* 2021. doi:10.21203/rs.3.rs-100956/v4
- 30 Brown N. A bug and a dilemma, 2021. Available: <https://steamtraen.blogspot.com/2021/07/Some-problems-with-the-data-from-a-Covid-study.html>
- 31 Dabbous HM, Abd-Elsalam S, El-Sayed MH, *et al.* Efficacy of favipiravir in COVID-19 treatment: a multi-center randomized study. *Arch Virol* 2021;166:949–54. doi:10.1007/s00705-021-04956-9
- 32 Dabbous HM, Abd-Elsalam S, El-Sayed MH, *et al.* Retraction note to: efficacy of favipiravir in COVID-19 treatment: a multi-center randomized study. *Arch Virol* 2022;167:277. doi:10.1007/s00705-021-05307-4
- 33 Dabbous HM, El-Sayed MH, El Assal G, *et al.* Retraction note: safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: a randomised controlled trial. *Sci Rep* 2021;11:18983. doi:10.1038/s41598-021-98683-5
- 34 Dabbous HM, El-Sayed MH, El Assal G, *et al.* Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: a randomised controlled trial. *Sci Rep* 2021;11:7282. doi:10.1038/s41598-021-85227-0
- 35 Samaha AA, Mouawia H, Fawaz M, *et al.* Effects of a single dose of ivermectin on viral and clinical outcomes in asymptomatic SARS-CoV-2 infected subjects: a pilot clinical trial in Lebanon. *Viruses* 2021;13:989. doi:10.3390/v13060989
- 36 Samaha AA, Mouawia H, Fawaz M, *et al.* Effects of a single dose of ivermectin on viral and clinical outcomes in asymptomatic SARS-CoV-2 infected subjects: a pilot clinical trial in Lebanon. *Viruses* 2021;13:989. doi:10.3390/v13060989
- 37 Carlisle JB. The analysis of 168 randomised controlled trials to test data integrity. *Anaesthesia* 2012;67:521–37. doi:10.1111/j.1365-2044.2012.07128.x
- 38 Carlisle JB. Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia* 2017;72:944–52. doi:10.1111/anae.13938
- 39 Bordewijk EM, Li W, van Eekelen R, *et al.* Methods to assess research misconduct in health-related research: a scoping review. *J Clin Epidemiol* 2021;136:189–202. doi:10.1016/j.jclinepi.2021.05.012
- 40 Tierney JF, Fisher DJ, Vale CL, *et al.* A framework for prospective, adaptive meta-analysis (FAME) of aggregate data from randomised trials. *PLoS Med* 2021;18:e1003629. doi:10.1371/journal.pmed.1003629
- 41 WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, *et al.* Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. *JAMA* 2020;324:1330–41. doi:10.1001/jama.2020.17023
- 42 WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Shankar-Hari M, Vale CL, *et al.* Association between administration of IL-6 antagonists and mortality among patients hospitalized for COVID-19: a meta-analysis. *JAMA* 2021;326:499–518. doi:10.1001/jama.2021.11330
- 43 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. doi:10.1186/s12916-020-01880-8
- 44 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
- 45 Kataoka Y, Oide S, Arië T, *et al.* COVID-19 randomized controlled trials in medRxiv and PubMed. *Eur J Intern Med* 2020;81:97–9. doi:10.1016/j.ejim.2020.09.019
- 46 Añazco D, Nicolalde B, Espinosa I, *et al.* Publication rate and citation counts for preprints released during the COVID-19 pandemic: the good, the bad and the ugly. *PeerJ* 2021;9:e10927. doi:10.7717/peerj.10927
- 47 Shi X, Ross JS, Amancharla N, *et al.* Assessment of concordance and discordance among clinical studies posted as preprints and subsequently published in high-impact journals. *JAMA Netw Open* 2021;4:e212110. doi:10.1001/jamanetworkopen.2021.2110
- 48 Ross JS, Mocanu M, Lampropoulos JF, *et al.* Time to publication among completed clinical trials. *JAMA Intern Med* 2013;173:825–8. doi:10.1001/jamainternmed.2013.136
- 49 Vlaar APJ, de Bruin S, Busch M, *et al.* Anti-C5a antibody IFX-1 (vilobelimab) treatment versus best supportive care for patients with severe COVID-19 (PANAMO): an exploratory, open-label, phase 2 randomised controlled trial. *Lancet Rheumatol* 2020;2:e764–73. doi:10.1016/S2665-9913(20)30341-6
- 50 Vlaar Alexander PJ, Bruin de, Marcus J, *et al.* Anti-C5a antibody (IFX-1) treatment of severe COVID-19: an exploratory phase 2 randomized controlled trial. *SSRN Journal*. doi:10.2139/ssrn.3658226
- 51 Horby P, Lim WS, Emberson J. Effect of dexamethasone in hospitalized patients with COVID-19 – preliminary report. *medRxiv* 2020. doi:10.1101/2020.06.22.20137273
- 52 Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. *Clin Trials* 2013;10:783–806. doi:10.1177/1740774513494504
- 53 van den Bor RM, Vaessen PWJ, Oosterman BJ, *et al.* A computationally simple central monitoring procedure, effectively applied to empirical trial data with known fraud. *J Clin Epidemiol* 2017;87:59–69. doi:10.1016/j.jclinepi.2017.03.018
- 54 Carlisle JB, Dexter F, Pandit JJ, *et al.* Calculating the probability of random sampling for continuous variables in submitted or published randomised controlled trials. *Anaesthesia* 2015;70:848–58. doi:10.1111/anae.13126

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjmed-2022-000309>).