Statistical review of *A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19* by B. Cao *et al*

April 8, 2020

Simon Gates, PhD - Professor of Biostatistics and Clinical Trials, Cancer Research UK Clinical Trials Unit, University of Birmingham, UK

Richard Emsley, PhD - NIHR Research Professor & Professor of Medical Statistics and Trials Methodology, Department of Biostatistics and Health Informatics, King's College London, UK

Darren Dahly, PhD - Principal Statistician, HRB Clinical Research Facility - Cork. HRB Trials Methodology Research Network. Senior Lecturer in Patient Focused Research Methods, University College Cork School of Public Health

The following review has been prepared in collaboration with members of the MRC-NIHR Trials Methodology Research Partnership [1]. The reviewers named above, and other, unnamed discussants of the paper, are all qualified statisticians with experience in clinical trials. Our objective is to provide a rapid review of publications, preprints and protocols from clinical trials of COVID-19 treatments, independent of journal specific review processes. We aim to provide timely, constructive, focused, clear advice aimed at improving both the research outputs under review, as well as future studies. Given our collective expertise (clinical trial statistics) our reviews focus on the designs of the trials and other statistical content (methods, presentation and accuracy of results, inferences).

DOI:10.5281/zenodo.3746512

Study Summary

Here we review *A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19*, by B.Cao *et al*, published in the New England Journal of Medicine on 18 March 2020 [3]. This was one of the first randomised controlled trials of treatments for COVID-19 to be published, and the first in a high-profile medical journal.

The paper reports a randomised trial of 199 hospitalised patients with a positive test for SARS-CoV-2; pneumonia confirmed by chest imaging; and an oxygen saturation (Sao2) of 94% or less breathing ambient air, or Pao2:Fio2 ratio less than or equal to 300 mg Hg. It was conducted at a single hospital in Wuhan, China in January and February of 2020. Patients were randomised to standard care plus twice daily oral lopinavir–ritonavir (400 mg and 100 mg; 99

patients) or standard care alone (100 patients). The primary outcome was the time to clinical improvement, where clinical status was measured on a 7-category ordinal scale, ranging from *no longer hospitalized with resumption of normal activities* to *death*, and improvement was then defined as moving down two or more categories (towards better outcomes) on this scale.

In the intention to treat analysis, the hazard ratio for this outcome in a was 1.31, favouring lopinavir-ritonavir (95% confidence interval, 0.95 to 1.85; P = 0.09). Day 28 mortality was 19/99 in the lopinavir-ritonavir group (19.2%) and 25/100 (25%) in the control group (risk difference = −5.8%, 95% CI, −17.3 to 5.7). On the basis of these findings, the authors concluded that there was that "no benefit was observed… beyond standard care." However, in a modified intention to treat analysis for which 3 patients were omitted from the lopinavir-ritonavir group (because they died within 24 hours of randomisation and did not receive the study intervention), the hazard ratio for the primary outcome was 1.39 (95% CI, 1.00 to 1.91) which equated to a 1 day reduction in median time to clinical improvement. Further, the authors reported a number of secondary outcomes that also favoured lopinavir-ritonavir, including length of ICU stay (−5 days, 95% CI −9 to 0); duration of hospitalization (-1 day, 95% CI, -3 to 0), and the proportion of patients who had clinically improved by day 14 (risk difference of 15.5%, 95% CI 2.2 to 28.8).

We sincerely thank the authors for their contribution to our collective understanding of COVID-19, and for their commitment to the timely dissemination of research results.

Major comments

The authors chose a complicated primary outcome measure that might be better analysed with alternative statistical methods.

The primary outcome had a complex definition, which was the time from randomisation to an improvement of two categories on a 7-level ordinal scale. It is likely that a 2-category improvement will be different for different starting points on the scale. For example, an improvement from category 6 (*hospitalized, requiring ECMO, invasive mechanical ventilation, or both*) to 4 (*hospitalized, requiring supplemental oxygen*) is likely not to represent the same improvement in health as a change from 3 (*hospitalized, not requiring supplemental oxygen*) to 1 (*not hospitalized with resumption of normal activities*). Moreover, dichotomisation of the ordinal measure into "improvement" and "non-improvement" does not use the information efficiently. It would be preferable to analyse the original 7-category outcome using ordinal regression models. The time to event outcome was probably used because the speed of clinical improvement is an important outcome; an outcome could be beneficial by causing people to recover more quickly, even if the proportion recovering was not increased overall.

Recommendations:

For future studies

- Outcome measures need careful specification; core outcome sets are in development and may give useful guidance.

For this study

- Analyse the 7-category outcome using ordinal regression models, which would use the information more efficiently, though it may be difficult to incorporate both the ordinal outcome and time to recovery into the same analysis.

The overall interpretation was overly cautious, given that the study was not well powered to detect clinically important effects.

The estimated effect of lopinavir–ritonavir on the time to clinical improvement was a hazard ratio of 1.31 (95% confidence interval 0.95 to 1.85, P = 0.09), while mortality was 19.2% in the lopinavir-ritonavir arm versus 25.0% in standard care arm (a risk difference −5.8%,, 95% CI −17.3 to 5.7). The conclusions in the paper were that “no benefit was observed with lopinavir–ritonavir treatment beyond standard care” (Abstract), and that “lopinavir–ritonavir treatment did not significantly accelerate clinical improvement, reduce mortality, or diminish throat viral RNA detectability in patients with serious Covid-19...” (end of Discussion). The authors appropriately called for further trials, noting that their data could be used to justify such efforts.

Their focus on a lack of benefit was apparently driven by the lack of statistically significant differences in the primary outcome and mortality, which in turn encouraged the widespread interpretation, evident in media reports, that lopinavir–ritonavir was not more effective than standard care. However, in our expert opinion, this is an over-interpretation; it would be more appropriate to conclude that the results were compatible with a wide range of treatment effects, from substantial benefit to small or moderate harm, likely due to the relatively small sample size (199 patients; see additional notes below), and thus we cannot conclude that the treatment was either effective *or* ineffective. Furthermore, results for several important secondary outcomes were consistent with clinical benefit, while a Bayesian re-analysis of the data showed a substantially higher probability of benefit than harm [4].

As noted above, this ambiguity in the conclusions was largely due to the relatively small sample size. The planned sample size of n= 160 was based on being able to detect a reduction in the median time to clinical improvement of 8 days, with 80% power. Given that the median time to clinical improvement found in the trial was 16 days, this was an unrealistically large effect size to base the trial’s design on. Although the study recruited more patients than originally planned (199 rather than 160), this wouldn’t have been enough of an increase to appreciably improve the power of the study to detect minimally important effects. The results therefore need to be interpreted in this light - that the trial was unlikely to find clinically meaningful effects even if they existed.

Recommendations:

For future studies

- Avoid over-interpretation, especially dichotomisation of results into “positive” and “negative” based on significance tests. This is specifically warned against by the American Statistical Association’s guidelines for the use of p-values and significance tests.[5]
- Ensure that the trial is designed to have high power to detect minimally important effects.
- Plan your trial so that individual patient data can be pooled with data from other studies of the treatment under investigation.

For this study

- Interpret the results in the light of the small sample size and resulting low power to find realistic effect sizes.

For the reader

- Do not make the error of interpreting lack of statistical significance as proof of an ineffective treatment.

Lack of blinding

The trial was not blinded. This was for pragmatic reasons, as it would have been difficult or impossible to produce a matching placebo on the necessary timescale. Most of the outcomes used are reasonably objective, so there may not be a high risk that the results were affected by knowledge of the treatment allocations. However, we cannot be completely certain about this, and additional treatments could have been influenced, consciously or unconsciously, by clinicians’ knowledge of which patients had received which treatment.

For future studies

- Placebo control and full blinding or participants, clinicians, and people ascertaining outcomes would be ideal. This may be difficult to achieve in the context of trials that are set up quickly to run during the pandemic.

For this study

- The lack of blinding means that the possibility of bias cannot be excluded, though the risk does not seem high.

For the reader

- Study participants (or their assessors) may behave differently if they are aware of the treatment given. This may matter for outcomes that are subjective (such as the clinical assessment categories) and for which subconscious bias may creep in, though is unlikely to make an appreciable impact on death or stay in ICU. Where possible, people in the trial should be unaware of the treatment given in order to avoid this criticism.

Minor points

- There is a slight lack of clarity about the outcome measures and which were prespecified. The lists of outcomes in the Methods and Results do not match precisely, and some outcomes are reported in the results which do not appear in the methods, or are described in a different way in the Methods section. For example, time to deterioration and duration of ICU stay are not mentioned in the Methods, and clinical improvement is not described as a binary variable.

- The reasons for extending recruitment beyond the originally planned sample size, and for termination of the trial after 199 patients had been recruited, are not clearly explained and remain somewhat opaque.

- Subgroup analyses were not handled appropriately. They were analysed and presented in separate strata, rather than using interaction terms to directly evaluate whether there was a difference in treatment effect between the subgroups. Also, the effects of time from onset of symptoms to randomisation would be more appropriately analysed as a continuous variable, rather than dichotomising it at 12 days.

- The analysis did not take advantage of prognostic covariates measured prior to randomization. Using multivariable models to adjust for these covariates would have led to more precise estimates of the treatment effects [6] and more appropriate conditional estimates of effects for non-linear models.

- The main analysis reported in the Abstract is different from the intention to treat analysis reported in the main text of the paper, and does not correspond to any of the reported analyses. The Abstract reports a hazard ratio of 1.24 (95% confidence interval 0.90, 1.72) for the primary outcome, whereas the Results section reports a hazard ratio of 1.31 (95% CI 0.95, 1.85). The result presented in the Abstract is presumably an error.

Open Data

A data sharing statement is included (available on the NEJM website). It specified that data will be available one year after publication. No justification for the delay is provided.

Open Analysis Code

Not provided. Not mentioned in the data sharing statement.

Pre-registered study design

Yes.

PubPeer

There may be comments on the PubPeer page for the published version of this paper.
https://pubpeer.com/publications/F1FD9E33F410969AA7E8CAF860F6C2

References

1. MRC-NIHR Trials Methodology Research Partnership.

   https://www.methodologyhubs.mrc.ac.uk/about/tmrp/

2. Creative Commons Attribution 4.0 International License.

   https://creativecommons.org/licenses/by/4.0/

3. Cao, B *et al.* A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19.

   NEJM 2020; DOI: 10.1056/NEJMoa2001282

4. https://discourse.datamethods.org/t/first-rct-of-lopinavir-ritonavir-on-sars-cov2/3105/24

5. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and

   Purpose. *American Statistician* 70: 129-133 (2016).

6. Kahan, B. C., Jairath, V., Doré, C. J. & Morris, T. P. The risks and rewards of covariate

   adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. Trials 15, 139

   (2014)

7. Schulz, K. F., Altman, D. G., Moher, D. & for the CONSORT Group. CONSORT 2010

   Statement: updated guidelines for reporting parallel group randomised trials. BMJ 340,

   c332–c332 (2010)

CONSORT CHECKLIST

To support the review, we completed the CONSORT [6] checklist below. Material taken directly from the paper (or trial registry) is in *italics.* Our additional comments are in **bold**.

Title and abstract

1a Identification as a randomised trial in the title

No. Title describes the study as a "trial" but does not mention randomisation.

1b Abstract: Structured summary of trial design, methods, results, and conclusions.

Title: Identification of the study as randomised	No
Authors: Contact details for the corresponding author	Yes
Trial design: Description of the trial design (eg, parallel, cluster, non-inferiority)	Yes
Methods	
Participants: Eligibility criteria for participants and the settings where the data were collected	Yes
Interventions: Interventions intended for each group	Yes
Objective: Specific objective or hypothesis	No
Outcome: Clearly defined primary outcome for this report	Yes
Randomisation: How participants were allocated to interventions	Yes
Blinding (masking): Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	No blinding used
Results	
Numbers randomised: Number of participants randomised to each group	Yes
Recruitment: Trial status	No
Numbers analysed: Number of participants analysed in each group	Yes
Outcome: For the primary outcome, a result for each group and the estimated effect size and its precision	Yes
Harms: Important adverse events or side-effects	Yes
Conclusions: General interpretation of the results	Yes
Trial registration: Registration number and name of trial register	Yes
Funding: Source of funding	Yes

Introduction

Background and objectives

2a Scientific background and explanation of rationale

Yes. This is sufficiently described in the Introduction.

2b Specific objectives or hypotheses

“To evaluate the efficacy and safety of oral lopinavir–ritonavir for SARS-CoV-2 infection,,,”

Methods

Trial design

3a Description of trial design (such as parallel, factorial) including allocation ratio

“Open-label, individually randomized, controlled trial”

3b Important changes to methods after trial commencement (such as eligibility criteria), with reasons

Only change noted was to the sample size, which was increased during recruitment (see below).

Participants

4a Eligibility criteria for participants

“Male and nonpregnant female patients 18 years of age or older were eligible if they had a diagnostic specimen that was positive on RT-PCR, had pneumonia confirmed by chest imaging, and had an oxygen saturation (Sao2) of 94% or less while they were breathing ambient air or a ratio of the partial pressure of oxygen (Pao2) to the fraction of inspired oxygen (Fio2) (Pao2:Fio2) at or below 300 mg Hg.”
“Exclusion criteria included a physician decision that involve -ment in the trial was not in the patient’s best in -terest, presence of any condition that would not allow the protocol to be followed safely, known allergy or hypersensitivity to lopinavir–ritonavir, known severe liver disease (e.g., cirrhosis, with an alanine aminotransferase level >5× the upper limit of the normal range or an aspartate amino -transferase level >5× the upper limit of the normal range), use of

medications that are contra -indicated with lopinavir–ritonavir and that could not be replaced or stopped during the trial period"

4b Settings and locations where the data were collected

"Conducted from January 18, 2020, through February 3, 2020 (the date of enrollment of the last patient), at Jin Yin-Tan Hospital, Wuhan, Hubei Province, China"

Interventions

5 The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

Intervention arm: *"lopinavir–ritonavir (400 mg and 100 mg, orally; freely provided by the national health authority) twice daily, plus standard care,... for 14 days"*
Standard care arm: *"as necessary, supplemental oxygen, noninvasive and invasive ventilation, antibiotic agents, vasopressor support, renal-replacement therapy, and extracorporeal membrane oxygenation (ECMO)."*
Delivery of treatments: *"In the lopinavir–ritonavir group, 5 patients did not receive any doses of lopinavir– ritonavir: 3 because of early death within 24 hours after randomization and 2 others because the attending physician refused to prescribe lopinavir– ritonavir after randomization."*

Outcomes

6a Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed

Primary outcome: time from randomisation to clinical improvement, defined as an increase of two points on a 7-category ordinal scale:
1, not hospitalized with resumption of normal activities;
2, not hospitalized, but unable to resume normal activities;
3, hospitalized, not requiring supplemental oxygen;
4, hospitalized, requiring supplemental oxygen;
5, hospitalized, requiring nasal high-flow oxygen therapy, noninvasive mechanical ventilation, or both;
6, hospitalized, requiring ECMO, invasive mechanical ventilation, or both;
7, death

Secondary outcomes:

1. **clinical status as assessed with the seven-category ordinal scale on days 7 and 14,;**
2. **mortality at day 28;**
3. **duration of mechanical ventilation;**
4. **duration of hospitalization in survivors;**

5. **time (in days) from treatment initiation to death;**
6. **proportions with viral RNA detection over time and viral RNA titer area-under-the-curve (AUC) measurements.**

Safety outcomes

1. **adverse events that occurred during treatment;**
2. **serious adverse events;**
3. **premature discontinuation of treatment.**

6b Any changes to trial outcomes after the trial commenced, with reasons

None reported.

Sample size

7a How sample size was determined

"The original total sample size was set at 160, since it would provide the trial with 80% power to detect a difference, at a two-sided signifi -cance level of α = 0.05, of 8 days in the median time to clinical improvement between the two groups, assuming that the median time in the standard-care group was 20 days and that 75% of the patients would reach clinical improvement."

7b When applicable, explanation of any interim analyses and stopping guidelines

"The planned enrollment of 160 patients in the trial occurred quickly, and the assessment at that point was that the trial was underpowered; thus, a decision was made to continue enroll -ment by investigators. Subsequently, when another agent (remdesivir) became available for clinical trials, we decided to suspend enrollment in this trial."

Randomisation

Sequence generation

8a Method used to generate the random allocation sequence

Computer-generated (SAS).

8b Type of randomisation; details of any restriction (such as blocking and block size)

Individual randomisation. Stratification; *"no oxygen support or oxygen support with nasal duct or mask, or high-flow oxygen, noninvasive ventilation, or invasive ventilation including ECMO."* **Not completely clear how many strata were used.**
Blocking of randomisation: *"permuted block (four patients per block) randomization sequence,"*

Allocation concealment mechanism

9 Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

"we performed allocation concealment with an interactive Web-based response system until randomization was finished on the system through a computer or phone."
That sentence is not completely clear; maybe a translation issue?

Implementation

10 Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

Randomisation prepared by a statistician not involved in the trial.

Blinding

No blinding was used.

11a If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how

11b If relevant, description of the similarity of interventions

Statistical methods

12a Statistical methods used to compare groups for primary and secondary outcomes

Primary outcome: Cox proportional hazards model. Data right-censored at day 28 if failure to reach clinical improvement or death before day 28.
No specific methods mentioned for analysis of secondary outcomes.

12b Methods for additional analyses, such as subgroup analyses and adjusted analyses

"Modified intention to treat analysis" performed, which omitted the three patients in the intervention group that died within 24 hours and did not receive any of the drug.

Two post hoc subgroup analyses were performed: 1. National Early Warning Score 2 (NEWS2) of 5 or below versus greater than 5; 2. randomization up to 12 days after onset of illness versus more than 12 days.

No methods described for subgroup analyses.

Results

Participant flow (a diagram is strongly recommended)

13a For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

13b For each group, losses and exclusions after randomisation, together with reasons

Flow chart included; 5 patients in intervention arm did not receive intervention, reasons documented.

Recruitment

14a Dates defining the periods of recruitment and follow-up

"Conducted from January 18, 2020, through February 3, 2020 (the date of enrollment of the last patient), at Jin Yin-Tan Hospital, Wuhan, Hubei Province, China"

14b Why the trial ended or was stopped

"when another agent (remdesivir) became available for clinical trials, we decided to suspend enrollment in this trial."

Baseline data

15 A table showing baseline demographic and clinical characteristics for each group

Baseline data shown in tables 1 and 2.

Numbers analysed

16 For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups

Specified in descriptions of analyses; all patients included in intention to treat analyses.

Outcomes and estimation

17a For each primary and secondary outcome, results for each group, and the

estimated effect size and its precision (such as 95% confidence interval)

Effect sizes and 95% confidence intervals given for all outcomes.

17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended

Time to clinical improvement presented as both hazard ratio (with Kaplan-Meier plot) and difference in medians. Binary outcomes presented only as risk differences.

A number of non-prespecified outcomes were reported, such as days of oxygen support, time to deterioration, and percentage with clinical improvement at days 7, 14, and 28.

Ancillary analyses

18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory

Subgroups analysed separately rather than using interaction tests or another preferable method.

Harms

19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms42)

Adverse events and serious adverse events were reported.

Discussion

Limitations

20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

Lack of blinding discussed as a limitation.

Generalisability

21 Generalisability (external validity, applicability) of the trial findings

Mentions that high mortality suggests a severely ill population was recruited.

Interpretation

22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

Some of the interpretations in various parts of the paper may have encouraged the widespread interpretation that this trial showed that the treatment was not beneficial. In the Abstract (conclusions) the authors state: *"No benefit was observed… beyond standard care."* **In the Results (primary outcome) they provide a statement that suggests more strongly that there was no difference;** "*Patients assigned to lopinavir–ritonavir did not have a time to clinical improvement different from that of patients assigned to standard care alone."* **At the beginning of the Discussion that statement suggests more strongly still that the study found that there was no difference:** *"This randomized trial found that lopinavir-ritonavir treatment added to standard supportive care was not associated with clinical improvement or mortality in seriously ill patients with Covid-19 differnet from that associated with standard care alone."*

Other information

Registration

23 Registration number and name of trial registry

Provided

Protocol

24 Where the full trial protocol can be accessed, if available

Protocol available from the NEJM website.

Funding

25 Sources of funding and other support (such as supply of drugs), role of funders

Sources of funding declared.