Statistical review of *Chloroquine diphosphate in two different dosages as adjunctive therapy of hospitalized patients with severe respiratory syndrome in the context of coronavirus (SARS-CoV-2) infection: Preliminary safety results of a randomized, double-blinded, phase IIb clinical trial (CloroCovid-19 Study)*

April 20, 2020

Dr. Andrew Althouse, PhD - Assistant Professor of Medicine, University of Pittsburgh, Division of General Internal Medicine, Center for Research on Health Care (CRHC), Center for Clinical Trials and Data Coordination (CCDC)

Beatriz Goulao, MSc - Statistician, Health Services Research Unit, University of Aberdeen

Mike Bradburn, MSc - Senior medical statistician, Clinical Trials Research Unit, University of Sheffield

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). This review reflects the expert opinions of the named authors, and does not imply endorsement by the MRC-NIHR Trials Methodology Research Partnership, its wider membership, or any other organization.

DOI:10.5281/zenodo.3757821

Study Summary

Here we review *Chloroquine diphosphate in two different dosages as adjunctive therapy of hospitalized patients with severe respiratory syndrome in the context of coronavirus (SARS-CoV-2) infection: Preliminary safety results of a randomized, double-blinded, phase IIb clinical trial (CloroCovid-19 Study* by Borba *et al*. When we started our review, the report presented data collected up to day-6 post-randomization and was posted on MedRxiv here: https://www.medrxiv.org/content/10.1101/2020.04.07.20056424v1.full.pdf. During the process of

our review, the authors updated the preprint to report data up to 13 days post-randomization (https://www.medrxiv.org/content/10.1101/2020.04.07.20056424v2).

The paper reports an interim analysis of a randomized controlled trial carried out in Manaus, Brazil. The trial aimed to recruit 440 adult patients with potential COVID-19 infection based on their presentation (but not requiring laboratory confirmation, to avoid any delay in treatment). Patients were randomized to receive a low-dose chloroquine regimen (450mg for 5 days, twice daily only on the first day, or total dose of 2.7g) or a high-dose regimen (600mg twice daily for 10 days, or total dose of 12g) using a 1:1 allocation ratio.

Rather than comparing outcomes in the randomized treatment arms to one another, the analysis plan was to compare the results from each arm against an expected 20% mortality at 28 days, following what they observed in a previous patient cohort from Wuhan, China. The sample size of 220 per arm (n = 440) was set to provide 80% power to detect a 50% reduction in the risk of mortality in either arm. However, after enrollment of 81 patients, the randomization of patients to the high-dose arm was halted following the recommendation by the independent data safety and monitoring board (DSMB), who cited a greater number of deaths in that arm, as well as concerns about cardiotoxicity due to increased QTc and/or ventricular tachycardia.

When we started our review, day-6 mortality was slightly greater in the high-dose arm (7/41 patients) versus the low-dose arm (4/40 patients). When the updated preprint with day-13 outcomes was released, 13 of 41 patients in the high-dose arm had died, versus 6/40 patients in the low-dose arm. The study will carry on enrolling patients for low-dose chloroquine (as a single arm) and will eventually compare the outcome observed in that arm against the 20% mortality based on the Wuhan data.

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. We also appreciate their thoughtful discussion of the limitations of their work - the authors have clearly been forthcoming about these. Our detailed review follows.

Major comments

The planned statistical analysis is inconsistent with the typical purpose of a randomized controlled trial.

Randomizing patients to a high-dose regimen versus a low-dose regimen would be most commonly used to test the safety and efficacy of the two dosing regimens against one another. However, the investigators have not planned a direct comparison of the randomized arms, but have instead decided to compare the results from each arm (separately) against a “historical

control" derived from two case series of patients from Wuhan where mortality was 20%. Effectively, the analysis plan treats the study as two separate "single arm" trials, rather than analyzing the data as a two-arm parallel comparative-effectiveness trial. While randomization ensures that participant health status or other characteristics do not influence the chloroquine dose given, it cannot ensure that the severity of disease in these patients in Manaus, or their standard-of-care, was comparable to that of the external control group from Wuhan (more on this below). This makes the comparison difficult to interpret. It is unusual to employ randomization to create comparable groups and then eschew that entirely in the main analysis.

Using "historical controls" to draw conclusions about efficacy creates substantial challenges for interpreting the study's results.

Following from the above, without a control group of comparable patients receiving therapy other than chloroquine, the trial can not provide information about its efficacy versus alternative treatment strategies, including standard-of-care. We acknowledge the ethical challenges of a placebo-controlled trial in a pandemic situation for a potentially lethal disease, and the authors discuss this in their manuscript, noting that "use of placebo in Brazil in severe cases of COVID-19 infections is not considered ethically acceptable by national regulatory health agencies, especially due to the compassionate use of chloroquine – and because early reports seem to indicate its effectiveness *in vitro* and *in vivo*".

However, it must also be acknowledged that comparisons against a historical control suffer from many potential biases. The use of a historical control may be sufficient in specific settings where the disease is well-understood and the treated patients are known to be very similar to the historical controls. This is clearly not the case here. Are the patients in the historical control group similar to the patients included in this trial with respect to prognostic factors such as age and comorbidity? Did clinicians in Manaus have the same understanding of COVID-19 than those in Wuhan at the start of the outbreak? Are they receiving other competing therapies that could affect their outcomes (patients in this study were also on ceftriaxone and azithromycin)? How does the local context, such as potentially overwhelmed intensive care units in some regions, affect the mortality rate? The importance of these contextual differences are clearly reflected by the variability in COVID-19 mortality rates around the world. Indeed, the two studies themselves report different mortality figures (14% and 28%), and while their average is a best guess of the true rate, it highlights the difficulties of extrapolating mortality rates from one setting to another.

Concurrent controls are thus essential for detecting what are likely to be modest treatment effects, and without them we are left with inferences that rest on substantial, untestable assumptions. This is why the inclusion of a standard-of-care arm (e.g. a "placebo arm") is ethically sound in circumstances where the effectiveness of a new treatment is unproven and its risks are unknown.

Recommendations:

For future studies

- Ideally, include a randomized comparison of chloroquine against the current standard-of-care.
- Alternatively, if randomizing patients to two chloroquine-based treatment strategies, design the trial with planned contrasts of the randomized arms against one another rather than comparing each independently against a "historical control". However, doing this without establishing assay sensitivity will still lead to conclusions of limited value.

For the reader

- Treat this interim report as very modest evidence of harm from high-dose chloroquine in comparison to low-dose chloroquine, and consider other existing data on chloroquine and related agents.

The authors did not provide enough detail about the decision for early stopping.

We at first found it surprising that the DSMB recommended discontinuation of the high-dose arm based on the day-6 data (reported April 13th), as well as the authors' unequivocal conclusion that the higher dose was unsafe. However, in the updated preprint (April 17) with data from day-13, the mortality figures are more strongly stacked against high-dose arm, and the decision makes more sense in that light. The DSMB is, of course, charged with protecting patients and may have valid reasons for an abundance of caution. Without any apparent mortality benefit (though the study was not designed to demonstrate such benefits at this stage of enrollment), the observed difference in deaths, along with the early hints of cardiotoxicity (two patients with ventricular tachycardia in the high-dose arm), were clearly enough for the DSMB to suggest action. However, the DSMB's decision-making process is barely mentioned in the preprint, which only says that DSMB recommended halting randomization due to concerns about safety in the high-dose arm. It would be helpful if the authors could further report what exact data the DSMB had access to when the trial was stopped, along with any other pertinent details. While we appreciate the importance of promptly disseminating results, new and ongoing studies will use these earlier study reports as guidance. This reinforces the importance of communicating exactly how and why decisions were made. We thus encourage the study team to include more details about what information was used to base the decision on, and/or the DSMB to write its own explanation for the recommendation.

Finally, given that the study stopped based on a relatively small number of events without "statistically significant" evidence of harm, it is worth noting that the high-dose arm included more elderly patients, more high-risk patients (defined by qSOFA >= 2) and more patients with heart disease. Ordinarily we advise against consideration of baseline balance in randomized trials because the standard errors, when properly calculated, make an allowance for any random differences between the groups. However, when the evidence to conclude harm was

based on so few events, and formal statistical analysis provides minimal or weak evidence of "harm" - it may be worth asking if that small number of events could be driven by the small surplus of patients with heart problems in the high-dose arm, given that the small number of cardiotoxicity events was the likely reason for stopping.

Recommendations:
For future studies

- Clear report the rationale for any changes made during the course of the trial, whether due to a recommendation from the DSMB, or any other reason.

For this study

- Provide more detailed rationale for the DSMB's decision to halt enrollment into the high-dose arm

Minor points

- Even though the stated aim of the paper was to comprehensively evaluate the safety and efficacy of two different chloroquine doses in patients with established severe COVID-19, the authors weren't able to confirm patients' COVID-19 diagnosis until after they were randomized (due to delays in testing). For that reason, even though 81 patients were enrolled in the trial, only 62 had confirmed COVID-19. The argument for including participants ahead of formal diagnosis is well justified and reflective of "real world" clinical practice (e.g. treatment may need to be started before infection can be confirmed by lab testing) but it may impact the generalizability of the results to COVID-19 patients. The title and abstract could be modified to say "with suspected coronavirus…" to make it more clear that not all enrolled patients had "established" COVID-19.

- The authors presented p-values in Table 1 comparing the baseline characteristics of the randomized treatment arms. This practice goes against expert guidance, as explained in the CONSORT explanation and elaboration document (http://www.consort-statement.org/checklists/view/32--consort-2010/510-baseline-data). Even in cases like this one where early stopping is based on very small number of events, and there may be a need to carefully examine the "balance" of important prognostic factors at baseline (e.g. noting that the 5 patients in the high-dose arm were over 75 years versus 0 in the low-dose arm, or that 5 patients in the high-dose arm had heart disease versus 0 in the low-dose arm) this wouldn't be based on the reported p-values.

- There were several other inconsistencies in the reporting, obscuring exactly what was done. In some places, p-values are presented but the tests that led to those p-values are not clear. The methods note that, "For qualitative variables, Chi-square tests and Fisher's exact test were performed", but this is not specific enough to understand what was done for each outcome. The reader should not have to guess this information. Further, several outcomes presented under

methods were not analysed for the publication, although in some cases that is due to the ongoing nature of the study (e.g. day 28 follow-up). It is also unclear whether any participants recovered or were censored for a different reason during the analysed follow-up.

- There were outcomes noted in the April 13th version of the pre-print that are missing entirely from the April 17th update. Further, the April 13 preprint included a Table with some adverse events that were incompletely reported, adding confusion regarding the overall outcomes. This has been corrected in the update posted on April 17th, but we note it here for completeness.

Open Data

No.

Open Analysis Code

No.

Pre-registered study design

No.

PubPeer

There may be comments on the PubPeer page for the published version of this paper.
https://pubpeer.com/publications/58DB1B0295C425BFDC0DAB54A36DC0

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. Schulz KF et al CONSORT 2010 Statement: updated guidelines for reporting parallel group

   randomised trials. BMJ 2010;340:c332

CONSORT CHECKLIST

To support the review, we completed the CONSORT checklist [10] 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: **Yes**

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

Title: Identification of the study as randomised	Yes
Authors: Contact details for the corresponding author	No
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	Yes
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	Yes
Results	
Numbers randomised: Number of participants randomised to each group	Yes
Recruitment: Trial status	Yes
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	Sort of?
Harms: Important adverse events or side-effects	Yes, but poorly
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: **Present**

2b Specific objectives or hypotheses: **Present**

Methods

Trial design

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

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

Participants

4a Eligibility criteria for participants: **Present**

4b Settings and locations where the data were collected: **Present**

Interventions

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

Outcomes

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

6b Any changes to trial outcomes after the trial commenced, with reasons: **Present, but secondary outcomes seem inconsistent with trial registration; could be related to early stopping?**

Sample size

7a How sample size was determined: **Present, but poorly done (see review)**

7b When applicable, explanation of any interim analyses and stopping guidelines: **Minimal description despite trial stopping early**

Randomisation

Sequence generation

8a Method used to generate the random allocation sequence: **Present**

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

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: **Present (not ideally done - sealed envelopes - but it is reported)**

Implementation

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

Blinding

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

11b If relevant, description of the similarity of interventions: **Present**

Statistical methods

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

Poorly reported

12b Methods for additional analyses, such as subgroup analyses and adjusted analyses: **Poorly reported**

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: **Present**

13b For each group, losses and exclusions after randomisation, together with reasons: **Poorly reported**

Recruitment

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

14b Why the trial ended or was stopped: **Present, but would like more details about DSMB decision**

Baseline data

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

Numbers analysed

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

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): **Present, but oddly displayed (tied to odd analysis plan - a randomized trial that doesn't plan to compare the randomized groups for a primary outcome of mortality, but then stops early due to safety concerns based on comparing the randomized groups)**

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

Ancillary analyses

18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory: **Present, but oddly displayed (see review)**

Harms

19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms42): **Present, but oddly displayed (see review)**

Discussion

Limitations

20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses: **Fairly good**

Generalisability

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

Interpretation

22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence: **okay for most part, but surprising how strongly they recommend against high-dose chloroquine given that these data are fairly weak**

Other information

Registration

23 Registration number and name of trial registry: **Present**

Protocol

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

Funding

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