Statistical review of *Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial*

April 3, 2020

Professor Christina Yap, PhD - Professor of Clinical Trials Biostatistics, ICR Clinical Trials and Statistics Unit, The Institute of Cancer Research

Dr Jack Wilkinson, PhD - Research Fellow, Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester

Dr 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.3739134

Study Summary

Here we review *Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial* by Chen et al [3]. This paper reports a two-arm parallel trial that took place in Renmin Hospital of Wuhan University (China) between Feb 4th to the 29th (2020), where they randomized 31 patients to standard care, and another 31 patients to receive standard care plus five days of hydroxychloroquine treatment (400 mg/d). All enrolled patients were adults (age ≥ 18 years) with RT-PCR confirmed SARS-CoV-2, diagnosed pneumonia (by chest CT), and a SaO2/SPO2 ratio > 93% or a PaO2/FIO2 ratio > 300 mmHg (mild illness). The outcomes were time to clinical recovery, where recovery was defined as a return of body temperature and/or cough (measured 3 times daily) to normal levels (and maintained for more

than 72 h); and pulmonary recovery, which was the change in radiological assessment of the pneumonia (*stable*, *worsening*, *improving*) between enrollment and day + 6.

The authors reported that the mean recovery time for fever was 2.2 days (SD 0.4) in the hydroxychloroquine arm, vs 3.2 days (SD 1.3) in the control arm ($p = 0.0008$); and that the mean remission time for cough was 2.0 days (SD 0.2) in the hydroxychloroquine arm, vs 3.1 days (SD 1.5) in the control arm ($p = 0.0016$). They also reported that radiologically assessed pneumonia improved in 17/31 (54.8%) patients in the control arm, vs. 25/31 (80.6%) in the hydroxychloroquine arm ($p = 0.048$). They also noted that there were four patients that progressed to severe illness, and that they were all in the control arm.

Based on these findings, they concluded that “Despite our small number of cases, the potential of CQ in the treatment of COVID-19 has been partially confirmed. Considering that there is no better option at present, it is a promising practice to apply HCQ to COVID-19 under reasonable management. However, Large-scale clinical and basic research is still needed to clarify its specific mechanism and to continuously optimize the treatment plan.”

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

Major comments

The choice of endpoints and how they were analyzed was suboptimal, and poor reporting further confused what was actually done.

The authors note that the primary outcome was time to clinical recovery, with observation up to day + 5 (or the occurrence of an adverse event). Time to clinical recovery was then measured two different ways, each reported as a separate outcome: recovery time for fever (measured by study staff using a variety of methods) and remission time for cough (based on patient self-report, where *no cough* or *slight cough* were considered remission). Each of these was assessed 3 times a day; remission/recovery had to be maintained for at least 72 hours. We had a number of questions/comments on how these outcomes were analyzed:

- There were a number of patients in each arm that weren’t coughing and/or didn’t have a fever at enrollment. It is not clear how these patients contributed to the analysis (were they excluded, or were their outcomes recorded as 0 days to recovery?).

- The last day of observation was day + 5, meaning that some patients would have had cough/fever last beyond the period of observation, meaning that their data were right-censored.

- It was not clear if the 72 hours of recovery maintenance was still assessed when it would have extended past day + 5.

- The data were analysed with t-tests, suggesting a normal model for the outcome. However, the data couldn't have been generated by a normal model, given the floor and ceiling effects at day +1 and day +5, respectively. We also suspect there could have been substantial skew. In other words, it seems unlikely that the distribution of outcomes would have been adequately described by a mean and standard deviation. A statistical model for time-to-event data would have likely been a more appropriate choice.

- The authors should consider covariate-adjusted estimates of treatment effects [4] using the appropriate multivariable statistical model (ordered logistic regression for pulmonary recovery; Cox proportional hazards models for time-to-event outcomes). This would result in more appropriate conditional measures of treatment effects, as well as more efficient estimates, with narrower confidence intervals, given the fixed sample size.

- There was not enough detail to allow us to replicate the reported analysis. We tried different versions of the t-test, under various configurations (equal/unequal variances and/or sample sizes), but could not replicate the reported p-values (though they were qualitatively similar enough).

Recommendations:
For future studies
- Properly report all methods, following CONSORT [5].
- Whenever possible, please consult or collaborate with an experienced trial statistician, especially at the design phase.
- Pre-register key prognostic covariates and include multivariable models that adjust for those covariates in your statistical analysis plan.

For this study
- Report additional post-hoc analyses following the advice above.
- Clarify reporting of the statistical methods and results, providing enough detail that the analyses can be accurately replicated.

For the reader
- To correctly interpret the results, it is important that the data were properly analysed and that the methods and results are properly reported. The concerns we raised above suggest that the results should be interpreted with caution.

The study might not have been properly blinded.

The authors noted that “Neither the research performers nor the patients were aware of the treatment assignments.” However, this was the only mention of blinding, and there was no additional information in the paper about any methods used to ensure blinding throughout the study (though the trial registration does mention the use of a starch pill in the control group). Further, there was no mention of allocation concealment. This lack of vital information about the study is particularly concerning, given the subjectivity in some of the outcomes (e.g. changes in radiologically assessed pneumonia; patient reported coughing).

Recommendations:
For future studies

- Ensure you use robust procedures to ensure allocation concealment [6], and subsequent blinding (when possible).
- Properly report those procedures, following CONSORT [5].

For this study

- Clarify what procedures were used, if any, to ensure allocation concealment and blinding.
- If there was no allocation concealment and/or blinding, interpret the results in light of these limitations.

For the reader

- It is well understood that a lack of blinding can bias subjective outcome assessments [7], thus these results should be viewed cautiously.

There were substantial differences between what was reported in the paper, and what was described in the trial registration.

The paper reports a two arm parallel trial of hydroxychloroquine vs. control where time to clinical recovery was the primary outcome and n = 62. The trial registry however describes a three arm parallel trial with 2 different doses of hydroxychloroquine vs control, where viral clearance and t-cell recovery were the outcomes, where the planned sample size was 300 patients. There was no explanation for the discrepancies.

These discrepancies are not without consequence. To support the frequentist approach to statistical inference underpinning this clinical trial, it is well understood that key aspects of the design (including the comparators, outcomes, and the sample size) should be fixed in advance and pre-registered (adaptive and/or Bayesian designs notwithstanding, though these also take a substantial amount of planning). When these details aren’t pre-registered, the reader can no longer identify data-driven decisions that can invalidate statistical inferences (e.g. by inflating the probability of a false positive result). For example, how did the authors of this study arrive at a sample size of 62? Did they make this decision on the basis of analyses of the accruing data? Did the authors come up with their definitions for recovery after seeing the data? Were different

analyses tried, but not reported? In our expert opinion, these questions substantially limit our ability to accept the trial's results at face value.

Recommendations:
For future studies

- Always report deviations from the trial registration and/or pre-registered trial protocol.

For this study

- Clarify the reasons for these deviations in the trial report.

For the reader

- Until the discrepancies above are resolved, the results should be interpreted with caution.

Minor points

- The paper notes that the randomization was stratified by site, but this does not appear to be a multi-site study. Further, they wound up with an equal sample size in each arm, but there was no mention of restricting the randomization to force an equal sample size in each arm (the probability of this happening by chance is about 10%, so not implausible).

- Table 1 combines information on baseline covariates and the estimates of the treatments' effects. These should be presented separately.

- Estimates of treatment effects should be presented as differences between study arms. For example, instead of providing the mean and SD of an outcome for each arm, as in table 1, please explicitly report the differences in means and their standard errors or confidence intervals.

- Ethical approval appears to have been obtained for the version of the study outlined in the trial registration, rather than for the substantially amended version described in the preprint.

Open Data

No.

Open Analysis Code

No.

Pre-registered study design

No.

PubPeer

There are additional comments on the PubPeer page for the published version of this preprint.
https://pubpeer.com/publications/71E74AD5896DCF99981A47917DE097

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. Chen, Z. *et al. Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial*. http://medrxiv.org/lookup/doi/10.1101/2020.03.22.20040758 (2020) doi:10.1101/2020.03.22.20040758.
4. 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).
5. 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).
6. Vickers, A. J. How to randomize. *J. Soc. Integr. Oncol.* **4**, 194–198 (2006).
7. Day, S. J. Statistics Notes: Blinding in clinical trials and other studies. *BMJ* **321**, 504–504 (2000).

CONSORT CHECKLIST

To support the review, we completed the CONSORT checklist [5] 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

Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial

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	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	No
Interventions: Interventions intended for each group	Yes
Objective: Specific objective or hypothesis	Yes
Outcome: Clearly defined primary outcome for this report	No
Randomisation: How participants were allocated to interventions	No
Blinding (masking): Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	No
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	No
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	No

Introduction

Background and objectives

2a Scientific background and explanation of rationale

Yes

2b Specific objectives or hypotheses

This study aims to evaluate the efficacy of hydroxychloroquine (HCQ) in the treatment of patients with COVID-19. [abstract]

As one of the clinical research registration units in China, we aimed to investigate the efficiency of HCQ in patients with COVID-19 in this study. [introduction]

Methods

Trial design

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

All participants were randomized in a parallel-group trial [abstract]

1:1 allocation

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

In the registry, this study was listed as a 3 arm parallel trial with 100 patients in each arm, where the arms were 2 different doses of hydroxychloroquine vs a control. No mention of the change in the paper.

Participants

4a Eligibility criteria for participants

For this trial, the selection criteria: 1. Age ≥ 18 years; 2. Laboratory (RT-PCR) positive of SARS-CoV-2; 3. Chest CT with pneumonia; 4. SaO2/SPO2 ratio > 93% or PaO2/FIO2 ratio > 300 mmHg under the condition in the hospital room (mild illness); 5. Willing to receive a random

assignment to any designated treatment group and not participating in another study at the same time.

The exclusion criteria: 1. Severe and critical illness patients or participating in the trial does not meet the patient's maximum benefit or does not meet any criteria for safe follow-up in the protocol after a doctor's evaluation; 2. Retinopathy and other retinal diseases; 3. Conduction block and other arrhythmias; 4. Severe liver disease (e.g., Child-Pugh score ≥ C or AST> twice the upper limit); 5. Pregnant or breastfeeding; 6. Severe renal failure [estimated glomerular filtration rate (eGFR) ≤ 30 mL/min/1.73m2] or receiving renal replacement therapy; 7. Possibility of being transferred to another hospital within 72 h; 8. Received any trial treatment for COVID-19 within 30 days before this research.

From the registry:
Inclusion: patients with novel coronavirus pneumonia who agreed to participate in this trial and signed the informed consent form.
Exclusion: The investigator considers that the subject has other conditions that make him/her unsuitable to participate in the clinical trial or other special circumstances.

4b Settings and locations where the data were collected

The clinical research protocol was reviewed and approved by the Ethics Committee in Renmin Hospital of Wuhan University (Wuhan, China).

From February 4, 2020, to February 28, 2020, 142 patients with confirmed COVID-19 were admitted. Diagnosis and classification of COVID-19 were based on the criteria of the China National Health Commission.

Interventions

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

62 patients who met the trial criteria were randomly assigned in a to two groups, all received the standard treatment (oxygen therapy, antiviral agents, antibacterial agents, and immunoglobulin, with or without corticosteroids), patients in the HCQ treatment group received additional oral HCQ (hydroxychloroquine sulfate tablets, Shanghai Pharma) 400 mg/d (200 mg/bid) between days 1 and 5 (Figure 1), patients in the control group with the standard treatment only.

Outcomes

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

5 days after enrollment or severe adverse reactions appeared was the observation endpoint.

Changes in time to clinical recovery (TTCR) and clinical characteristics of patients were evaluated after administration. TTCR is defined as the return of body temperature and cough relief, maintained for more than 72 h. Normalization and mitigation criteria included the following: a. Body temperature ≤ 36.6 °C on the surface, ≤ 37.2 °C under the armpit and mouth or ≤ 37.8 °C in the rectum and tympanic membrane; b. Cough from patients' reports, slight or no cough was in the asymptomatic range. Body temperature, cough check three times daily to calculate the average level.

For radiological changes, the chest CT results in one day before (Day 0) and one day after (Day 6) the study for evaluation.

Pulmonary recovery is defined as three levels: exacerbated, unchanged, and improved, moderately improved when less than 50 % of pneumonia were absorbed, and more than 50 % means significantly improved.

From the registry:

- *The time when the nucleic acid of the novel coronavirus turns negative*
- *T cell recovery time*

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

Different outcomes appear in the registry compared to the paper ('the time when the nucleic acid of the novel coronavirus turns negative' and 'T cell recovery time', vs 'time to clinical recovery' and 'pulmonary recovery'). There is no acknowledgement or explanation for the change in the paper.

Sample size

7a How sample size was determined

There was no information provided about how they arrived at a sample size of 62 patients.

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

Not applicable.

Randomisation

Sequence generation

8a Method used to generate the random allocation sequence

Randomization was performed through a computer-generated list stratified by site.

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

No additional information.

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

Treatments were assigned after confirming the correctness of the admission criteria.

Implementation

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

No additional information.

Blinding

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

Neither the research performers nor the patients were aware of the treatment assignments.

No additional information.

11b If relevant, description of the similarity of interventions

No mention of a placebo to mimic administration of hydroxychloroquine in the paper. In the registry, there is mention of a starch pill.

Statistical methods

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

Data were described as the mean (standard deviation, SD), n (%), the t-test or χ^2 test was used to compare the differences between the two groups. A two-sided p-value of less than 0.05 was considered statistically significant. Statistical analyses were performed using Graphpad Prism, version 6.0.

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

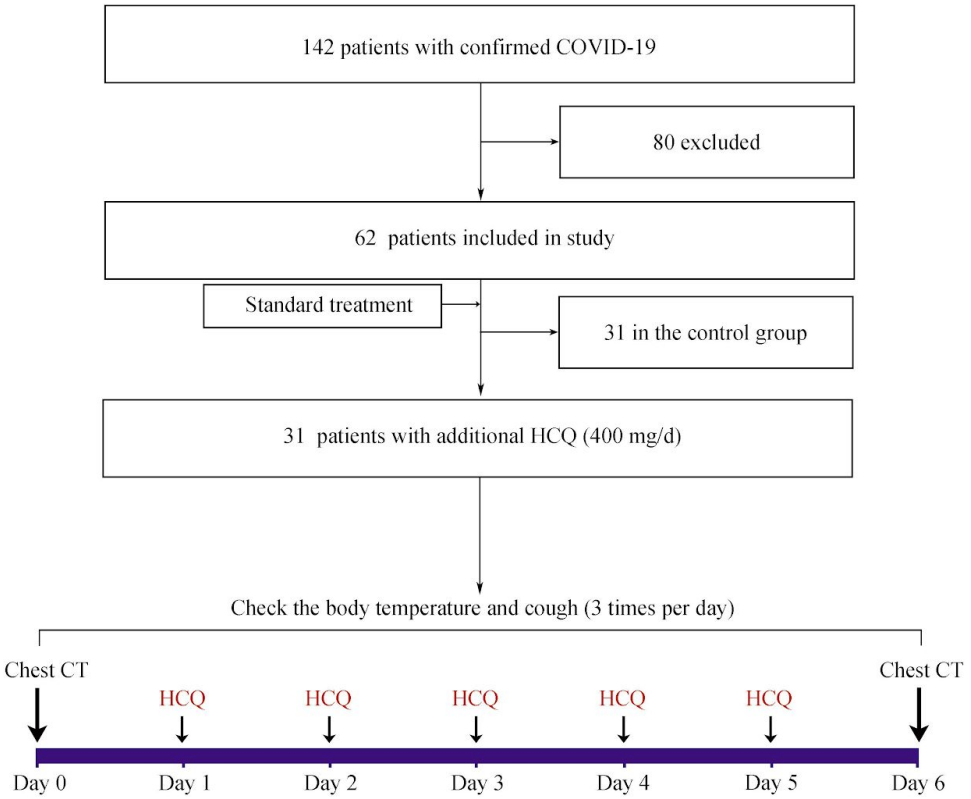
No additional analyses described.

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



Recruitment

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

From February 4, 2020, to February 28, 2020, 142 patients with confirmed COVID-19 were admitted. Diagnosis and classification of COVID-19 were based on the criteria of the China National Health Commission.

From the registry*: 2020-01-31 To 2020-02-29*

14b Why the trial ended or was stopped

Registry entry still listed as recruiting (last updated Feb 12, 2020)

Baseline data

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

As shown in Table 1, For all patients, the age was 44.7 (15.3) years old, 46.8% (29 of 62) were male and 53.2% (33 of 62) were female. Patients were randomly assigned into two groups. There was no significant difference in the age and sex distribution between the two groups of patients...

For fever, 17 patients in the control group and 22 patients in the HCQ treatment group had a fever in day 0…

For cough, 15 patients in the control group and 22 patients in the HCQ treatment group had a cough in day 0...

Characteristics	All	Control	HCQ	P value
Cases, n	62	31	31	
Age, mean (SD)	44.7 (15.3)	45.2 (14.7)	44.1 (16.1)	0.8809
Sex, n (%)				0.7991
Male	29 (46.8%)	15 (48.3%)	14 (45.2%)	
Female	33 (53.2%)	16 (51.7%)	17 (54.9%)	
Fever, day (SD)[a]	2.6 (1.0)	3.2 (1.3)	2.2 (0.4)	0.0008
Cough, day (SD)[b]	2.4 (1.1)	3.1 (1.5)	2.0 (0.2)	0.0016
Progressed to severe illness	4 (6.5 %)	4 (12.9 %)	0	
Adverse effects	2 (3.2 %)	0	2 (6.4 %)	

Table 1: Characteristics of patients in this trial.

[a]22 patients in the HCQ treatment group,17 patients in the control group with a fever one day before the intervention. [b]22 patients in the HCQ treatment group,15 patients in the control group with a cough one day before the intervention. Abbreviations: SD, standard deviation; HCQ, hydroxychloroquine; CT, computed tomography.

Numbers analysed

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

62 patients were identified as having COVID-19 and enrolled in this study, none quit (Figure 1).

The analysis was intention-to-treat as far as we can tell, though we can't be certain that participants who did not recover by day 5 were included in analyses.

It is not clear how patients without a cough or fever on day 0 were analysed.

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)

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

...but there are significant differences in TTCR between the two groups. For fever, 17 patients in the control group and 22 patients in the HCQ treatment group had a fever in day 0. Compared with the control group [3.2 (1.3) days], the body temperature recovery time was significantly shortened in the HCQ treatment group [2.2 (0.4) days]. For cough, 15 patients in the control group and 22 patients in the HCQ treatment group had a cough in day 0, The cough remission time was significantly reduced in the HCQ treatment group. Notably, a total of 4 of the 62 patients progressed to severe illness, all of which occurred in the control group not receiving HCQ treatment.

From table 1 (above)

Fever, day (SD)[a]	2.6 (1.0)	3.2 (1.3)	2.2 (0.4)	0.0008
Cough, day (SD)[b]	2.4 (1.1)	3.1 (1.5)	2.0 (0.2)	0.0016
Progressed to severe illness	4 (6.5 %)	4 (12.9 %)	0	

To further explore the effect of HCQ on pneumonia, we compared and analyzed the chest CT of patients on day 0 and day 6. In our study, pneumonia was improved in 67.7% (42/62) of patients, with 29.0% moderately absorbed and 38.7% significantly improved. Surprisingly, a larger proportion of patients with improved pneumonia in the HCQ treatment group (80.6%, 25 of 31) compared with the control group (54.8%, 17 of 31). Besides, 61.3% of patients in the HCQ treatment group had a significant pneumonia absorption.

Ancillary analyses

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

No additional analyses reported.

Harms

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

For adverse effects, it should be noted that there were two patients with mild adverse reactions in the HCQ treatment group, one patient developed a rash, and one patient experienced a headache, none severe side effects appeared among them.

Discussion

Limitations

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

No discussion of limitations.

Generalisability

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

No discussion of generalizability.

Interpretation

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

Among patients with COVID-19, the use of HCQ could significantly shorten TTCR and promote the absorption of pneumonia. [abstract]

Despite our small number of cases, the potential of HCQ in the treatment of COVID-19 has been partially confirmed. Considering that there is no better option at present, it is a promising practice to apply HCQ to COVID-19 under reasonable management. However, Large-scale clinical and basic research is still needed to clarify its specific mechanism and to continuously optimize the treatment plan. [conclusion]

Other information

Registration

23 Registration number and name of trial registry

This trial for SARS-CoV-2 has already been registered in the Chinese Clinical Trial Registry (ChiCTR), the unique identifier: ChiCTR2000029559

http://www.chictr.org.cn/showprojen.aspx?proj=48880

Protocol

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

We have not been able to identify a full protocol.

Funding

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

Funding: This study was supported by the Epidemiological Study of COVID-19 Pneumonia to Science and Technology Department of Hubei Province (2020FCA005).