Review

Use of hydroxychloroquine and chloroquine in COVID-19: How good is the quality of randomized controlled trials?

Faizan Mazhar[a], Muhammad Abdul Hadi[b], Chia Siang Kow[c],
Albaraa Mohammed N. Marran[d], Hamid A. Merchant[e], Syed Shahzad Hasan[e,*]

[a] Department of Biomedical and Clinical Sciences L. Sacco, "Luigi Sacco" University Hospital, Università di Milano, Milan, Italy
[b] School of Pharmacy, Institute of Clinical Sciences, University of Birmingham, Birmingham, United Kingdom
[c] School of Postgraduate Studies, International Medical University, Kuala Lumpur, Malaysia
[d] Department of Pharmaceutical Care, Prince Mohammed Bin Nasser Hospital, Jizan, Saudi Arabia
[e] School of Applied Sciences, University of Huddersfield, Huddersfield, United Kingdom

ABSTRACT

Objectives: We critically evaluated the quality of evidence and quality of harm reporting in clinical trials that evaluated the effectiveness of hydroxychloroquine (HCQ) or chloroquine (CQ) for the treatment of coronavirus disease 2019 (COVID-19).
Study design and setting: Scientific databases were systematically searched to identify relevant trials of HCQ/CQ for the treatment of COVID-19 published up to 10 September 2020. The Cochrane risk-of-bias tools for randomized trials and non-randomized trials of interventions were used to assess risk of bias in the included studies. A 10-item Consolidated Standards of Reporting Trials (CONSORT) harm extension was used to assess quality of harm reporting in the included trials.
Results: Sixteen trials, including fourteen randomized trials and two non-randomized trials, met the inclusion criteria. The results from the included trials were conflicting and lacked effect estimates adjusted for baseline disease severity or comorbidities in many cases, and most of the trials recruited a fairly small cohort of patients. None of the clinical trials met the CONSORT criteria in full for reporting harm data in clinical trials. None of the 16 trials had an overall 'low' risk of bias, while four of the trials had a 'high', 'critical', or 'serious' risk of bias. Biases observed in these trials arise from the randomization process, potential deviation from intended interventions, outcome measurements, selective reporting, confounding, participant selection, and/or classification of interventions.
Conclusion: In general, the quality of currently available evidence for the effectiveness of CQ/HCQ in patients with COVID-19 is suboptimal. The importance of a properly designed and reported clinical trial cannot be overemphasized amid the COVID-19 pandemic, and its dismissal could lead to poorer clinical and policy decisions, resulting in wastage of already stretched invaluable health care resources.

What is new?
Key findings
- Currently, published controlled clinical trials that evaluated the effectiveness and safety of hydroxychloroquine (HCQ) or chloroquine (CQ) in the treatment of coronavirus disease 2019 (COVID-19) have not only several methodological limitations but also the findings were significantly confounded by the lack of reporting and/or statistical adjustment of baseline disease severity.

- None of the 16 controlled trials included in this review met the Consolidated Standards of Reporting Trials (CONSORT) criteria in full for reporting harm data in clinical trials. Published trials were associated with a moderate to high risk of bias, which raises further questions on the validity of the trials' findings.
- Given the questionable quality of evidence available, it is not possible to draw a meaningful conclusion on the effectiveness and the safety of HCQ/CQ for the treatment of patients with COVID-19.

What does this add to what was known?
- The methodological and reporting quality of trials is of fundamental importance to inform evidence-based medicine.

* Corresponding author at: University of Huddersfield, Huddersfield, United Kingdom.
E-mail address: s.hasan@hud.ac.uk (S.S. Hasan).

This article highlights the methodological inadequacies in clinical trials of HCQ/CQ to assess their effectiveness in patients with COVID-19.

What are the implications and what should change now?

- Well-designed, well-conducted, and transparently reported clinical trials are important for clinicians and public health organizations to make a careful risk–benefit assessment of interventions for patients with COVID-19. The need to swiftly find a cure for a pandemic such as COVID-19 should not be an excuse to lower the quality threshold for the design and reporting of clinical trials.

Introduction

Since its outbreak in December 2019, coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has affected more than fourty million individuals across the world (World Health Organization, 2020; Dong et al., 2020). The full spectrum of clinical manifestations of COVID-19 ranges from asymptomatic or mild, self-limiting respiratory tract illness to severe progressive pneumonia, multi-organ failure, and death (Huang et al., 2020a). The reported case fatality rate of COVID-19 is highly variable, ranging from about 0.06% to 19%, depending on countries, settings, and age groups (Anon, 2020a). The case fatality rate is higher for hospitalized COVID-19 patients, with population data from the UK suggesting a case fatality rate of 26% (Docherty et al., 2020).

Many candidate drugs have been proposed for the treatment of COVID-19, but the antimalarial agents chloroquine (CQ) and hydroxychloroquine (HCQ) have attracted much attention (Anon, 2020b). *In vitro* studies have suggested direct antiviral properties of these antimalarial agents through the inhibition of pH-dependent steps of viral replication, while other researchers have suggested anti-inflammatory effects mediated by inhibition of the production of tumour necrosis factor alpha and interleukin 6, thus blocking the cascade of events leading to acute respiratory distress syndrome (Savarino et al., 2003). The use of antimalarials first attracted media attention in February 2020 after a news briefing by the Chinese government revealed that, according to several Chinese studies, CQ and HCQ seemed "to have apparent efficacy and acceptable safety for the treatment of COVID-19" (Hasan et al., 2020). A second boost of attention on antimalarials came after the publication of a non-randomized study—with considerable methodological limitations—claiming that a combination of HCQ and azithromycin achieved more rapid SARS-CoV-2 clearance in respiratory secretions of 20 patients (Gautret et al., 2020), which was followed by the drug combination being touted numerous times by the president of the United States as a potential cure for patients with COVID-19 in the media (Anon, 2020c).

Evidence-based medicine is one of the cornerstones of high-quality clinical care. It has been well established that the best evidence comes from well-designed and well-conducted randomized controlled trials. The promising signals from *in vitro* studies or uncontrolled data must be rigorously confirmed or refuted in high-quality randomized controlled trials. Ideally, efficacy-based trials, including proof-of-mechanism studies, should precede larger pragmatic effectiveness trials (Ford and Norrie, 2016). However, the development of robust evidence through well-designed and well-conducted trials can be challenging, particularly during a pandemic (Knottnerus and Tugwell, 2020), and thus there may be a temptation to lower the 'quality threshold' and overlook the limitations associated with study design either in the wider interest of public health or to claim a 'breakthrough'. However, such temptation must be resisted because falsely adopting

ineffective and potentially unsafe interventions on the basis of studies with methodological flaws may only cause harm without a noteworthy benefit. This may eventually have a negative impact not only on the design of other clinical trials but also on the course to find truly effective and safe interventions.

In addition to robust trial design, transparent and accurate reporting of trial data is equally important, especially reporting of harm data related to interventions. Optimal collection and reporting of adverse events (AEs) during any clinical trial should not be overlooked in order for clinicians to make a comprehensive risk–benefit assessment for their patients, who may have other underlying conditions that contraindicate the use of a particular drug, either relatively or absolutely. Studies have examined the methods for AE collection and presentation and highlighted inadequacies and inconsistencies in AE reporting in various published clinical trials (Ioannidis and Contopoulos-Ioannidis, 1998; Edwards et al., 1999; Cornelius et al., 2013; Hadi et al., 2017). In 2004, the Consolidated Standards of Reporting Trials (CONSORT) Group produced an extension to its guidelines for reporting trial results to include the reporting of harms, but these guidelines are poorly implemented in practice (Ioannidis et al., 2004).

The number of clinical trials assessing various treatment strategies for COVID-19 continues to increase; however, although some are good, the quality of most of these trials remains questionable. It is, therefore, more important than ever to critically assess the quality of emerging evidence from clinical trials amid the COVID-19 pandemic. This review aims to systematically summarize and critically evaluate the quality of evidence from all clinical trials of CQ or HCQ for the treatment of COVID-19 and to evaluate the quality of AE assessment and reporting in these trials.

Review methods

Search strategy and selection criteria

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was followed for the study design, search protocol, screening, and reporting. Articles were searched with a predefined search strategy and eligibility criteria. A systematic search was performed in electronic databases— PubMed, EMBASE, medRxiv (a preprint repository), Bibliovid, Google Scholar, and Dimensions—to retrieve eligible articles published between 1 December 2019 and 10 September 2020. We also searched for other eligible studies by screening the reference lists of relevant articles as well as unpublished studies in the ClinicalTrials.gov database. The search strategy included all MeSH terms and free keywords found for COVID-19, SARS-CoV-2, and CQ/HCQ. The following search terms were also used: "anti-malarial", "antimalarials", "chloroquine* OR CQ OR Aralen", "hydroxychloroquine OR HCQ OR Plaquenil", and "COVID-19 OR 2019-nCoV OR SARS-COV-2 OR Wuhan virus OR coronavirus".

Two authors (SSH and FM) independently searched the electronic databases and selected the articles against the eligibility criteria. Discrepancies between them in the selection of articles for inclusion were resolved by discussion with a third author to achieve a consensus. The study was included if (1) it was a randomized or non-randomized controlled trial and (2) it reported the effects of CQ and/or HCQ compared with placebo and/or active comparator treatment (s) in COVID-19 patients. Studies were excluded if they were (1) observational studies, animal studies, reviews, case reports, or *in vitro* studies and (2) duplicate publications.

Risk of bias assessment

Two authors (SSH and FM) assessed the risk of bias in the studies included in the systematic review. Version 2 of the

Cochrane risk-of-bias tool for randomized trials (RoB 2) (Higgins et al., 2011), which is a standardized method for assessing potential bias in reports of randomized interventions, was used to assess the risk of bias in the included randomized trials. RoB 2 is structured into a fixed set of domains of bias, focusing on different aspects of trial design, conduct, and reporting. A proposed judgement about the risk of bias arising from each domain is generated by an algorithm, where judgement can be 'low' or 'high' risk of bias or can express 'some concerns'. For non-randomized trials, the bias was assessed by the Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool (Sterne et al., 2016). Similarly, ROBINS-I, which is structured into a fixed set of domains of bias, includes signalling questions that inform the risk of bias judgements, and on the basis of answers to the questions, judgements for each bias domain, and the overall risk of bias can be classified as 'low', 'moderate', 'serious', or 'critical' risk of bias.

Quality of reporting of harm data

For this systematic review, we considered harms to be a continuum of all adverse treatment effects, including tolerability issues at the lower end and safety concerns at the upper end, a definition consistent with the 2004 CONSORT harm recommendations (Ioannidis et al., 2004). Data on harm reporting were assessed with use of a 10-item checklist, 'CONSORT Extension for Harms' (Ioannidis et al., 2004). The 10-item checklist was adapted and modified as there are multiple items of interest within a single CONSORT harm recommendation, and thus scoring the multiple items within a single recommendation would have been difficult and misleading. Therefore, where appropriate, we split the single CONSORT harm extension items into two or three items, resulting in a 19-item checklist. Each item of the 19-item checklist was scored individually and weighted with equal importance in line with CONSORT harm recommendations. Each item carries a score of 1 if it was adequately reported or 0 if it was inadequately reported or not reported at all. The total harm reporting score (THRS) was calculated by summation of all the individual scores, with maximum and minimum scores of 19 and 0, respectively. All included trials were coded by one author (FM) using the descriptors from the CONSORT Extension for Harms and subsequently cross-verified by a second author (SSH).

Data extraction, analysis, and reporting

Extracted data from all included studies were compiled into an electronic summary table. The following pertinent information was extracted: all-cause mortality, the requirement for mechanical ventilation, virological clearance, radiological results, admission to the intensive care unit, confirmation of COVID-19 status, development of new symptoms, serious AEs, and total AEs. Further parameters of interest included the number of patients, number of controls, mean age, sex distribution, baseline disease severity, treatment options, treatment dosage, and treatment duration.

We summarized the quantitative findings of individual studies and used a descriptive approach for reporting of harms. The percentage of trials fulfilling each CONSORT Extension for Harms recommendation and the number of recommendations fulfilled by each trial were tabulated descriptively. In addition, after the calculation of the THRS for all included trials, we determine the median THRS along with the interquartile range.

Results

Results of the search

A total of 1320 records were identified from the literature search. After the initial screening, the full text was retrieved for 74 potentially eligible abstracts. Implementation of the eligibility criteria resulted in 16 clinical trials being included (Abd-Elsalam et al., 2020; Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Gautret et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020; Esper et al., 2020). Data were extracted from these 16 clinical trials (Figure 1).
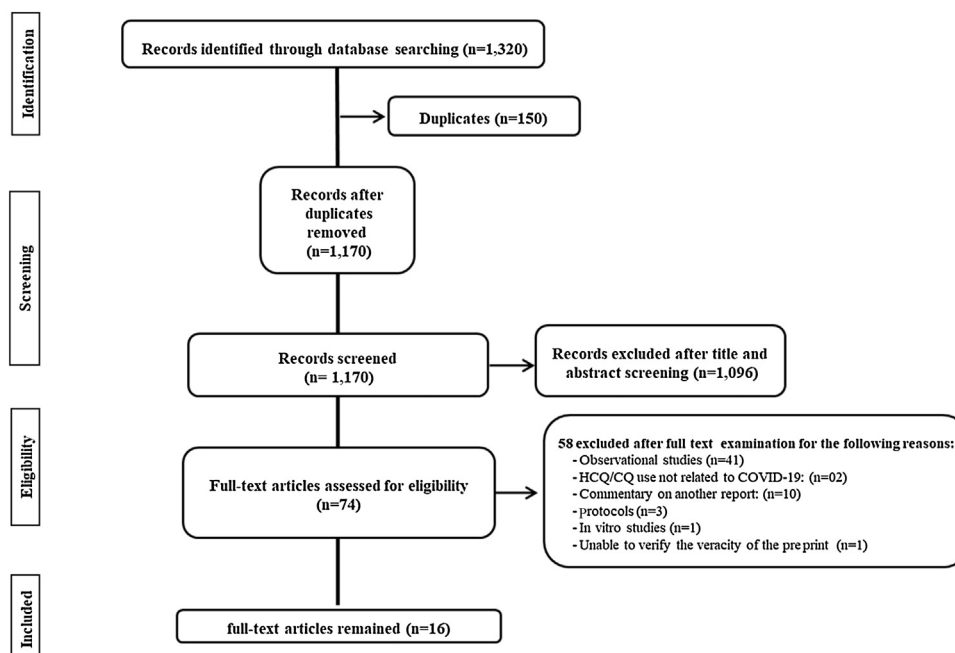
Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the study selection process. COVID-19, coronavirus disease 2019; CQ, chloroquine; HCQ, hydroxychloroquine.

Study characteristics

Sixteen clinical trials were included in this systematic review (Abd-Elsalam et al., 2020; Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Gautret et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020; Esper et al., 2020): two non-randomized controlled trials (Gautret et al., 2020; Esper et al., 2020), two double-blind randomized controlled trials (Boulware et al., 2020; Skipper et al., 2020), and 12 open-label randomized controlled trials (Abd-Elsalam et al., 2020; Borba et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Tang et al., 2020). The sample size in these 16 trials ranged from 22 to 4716 participants. The characteristics of these trials are summarized in Table 1.

Most trials (n = 9) (Gautret et al. (2020), Chen et al. (2020a), Chen et al. (2020b), Esper et al. (2020), Borba et al. (2020), Boulware et al. (2020), Horby et al. (2020), Abd-Elsalam et al. (2020), and Mitjà et al., 2020) provided no information on baseline disease severity, although it may not be significant for the trials by Esper et al. (2020), Boulware et al. (2020) and Mitjà et al., 2020 since they evaluated HCQ as postexposure prophylaxis. Tang et al. (2020) defined disease severity on the basis of the Chinese guidelines for the management of COVID-19. Chen et al. (2020c) categorized patients into different disease severity levels on the basis of chest radiographic findings. Mitjà et al., 2020 enrolled patients with mild disease presenting with symptoms including fever, acute cough, shortness of breath, sudden olfactory or gustatory loss, or influenza-like illness. Skipper et al. (2020) evaluated the severity of disease on the basis of a 10-point visual analogue scale. Furtado et al. (2020) included patients with severe disease based on the use of oxygen supplementation with a flow rate of more than 4 L/min, the use of a high-flow nasal cannula, the use of non-invasive positive-pressure ventilation, and the use of mechanical ventilation. Although Cavalcanti et al. (2020) specified that they enrolled participants with mild-to-moderate disease, no information was given on such a definition or the breakdown of the proportions of participants with either mild or moderate disease. Huang et al. (2020b) recruited patients with moderate-to-severe disease but did not give information on such a definition, although they did provide the proportions of participants with moderate or severe disease, respectively. The trials by Tang et al. (2020), Huang et al. (2020b) and Chen et al. (2020c) were found to have baseline differences in disease severity between the treatment arm and the comparator arm. None of the trials statistically adjusted disease severity at the baseline.

Outcome measures

Four trials (Gautret et al., 2020; Chen et al., 2020a; Tang et al., 2020; Chen et al., 2020c) of orally administered HCQ used viral clearance as the outcome measure. With a daily oral dose of 400 mg of HCQ for 5 days, Chen et al. (2020a) reported a negative conversion rate on day 7 in a pharyngeal swab of 86.7% in the HCQ group (n = 15), compared with 93.0% for the control group (n = 15). Tang et al. (2020), who used a loading plus maintenance dose regimen of HCQ, reported a 28-day negative conversion rate of 85.4% in the HCQ arm (n = 75), compared with 81.3% in the standard of care arm (n = 75). Chen et al. (2020c) used a loading plus maintenance dose regimen of HCQ, and reported a 14-day negative conversion rate of 81.0% for the HCQ group (n = 21) and 75.0% for the standard of care group (n = 12). Gautret et al. (2020) use a daily oral dose of 600 mg of HCQ, and reported a higher viral clearance

rate of 70% in the HCQ arm (n = 20) compared with 12.5% in the control arm (n = 16) at 6 days.

Hospital admission as the outcome measure was used in three trials (Mitjà et al., 2020; Skipper et al., 2020; Esper et al., 2020). Esper et al. 2020 administered HCQ with azithromycin (HCQ at a dose of 800 mg on the first day and 400 mg for another 6 days and azithromycin at a dose of 500 mg once daily for 5 days; n = 412) for patients with suspected COVID-19, and reported a lower hospitalization rate of 1.9% compared with 5.4% in the control group, who refused the trial drug (n = 224). Mitjà et al., 2020 in their trial involving patients with mild symptoms of COVID-19 reported hospitalization rates of 5.9% in the HCQ arm (800 mg on day 1, followed by 400 mg once daily for another 6 days; n = 136) and 7.1% in the control arm (n = 157). Skipper et al. (2020), who enrolled non-hospitalized adults with suspected or confirmed COVID-19, reported hospitalization rates of 1.8% for the HCQ group (800 mg once, followed by 600 mg in 6–8 h, then 600 mg daily for another 4 days; n = 212) and 4.7% for the placebo group (n = 211).

Death as the outcome was reported in four trials (Cavalcanti et al., 2020; Horby et al., 2020; Abd-Elsalam et al., 2020; Furtado et al., 2020). Cavalcanti et al. (2020) reported 15-day mortality of 1.7% with HCQ plus azithromycin (HCQ 400 mg twice daily plus azithromycin 500 mg once daily for 7 days), 3.1% with HCQ alone (HCQ 400 mg twice daily for 7 days), and 2.9% with control treatment. Horby et al. (2020) reported that patients randomized to receive HCQ (n = 1561) had 28-day all-cause mortality of 26.8%, while patients who randomized to receive usual care (n = 3155) had 28-day all-cause mortality of 25.0%. In the trial by Abd-Elsalam et al. (2020), death at 28 days occurred in 6.1% of patients in the HCQ group (400 mg twice daily on day 1, followed by 200 mg twice daily for another 14 days; n = 97) and 5.1% of patients in the control group (n = 97). Furtado et al. (2020) randomized patients to receive either HCQ (400 mg twice daily for 10 days; n = 183) or HCQ plus azithromycin (HCQ 400 mg twice daily plus azithromycin 500 mg once daily for 10 days; n = 214), and reported that death at 29 days occurred in 40% of patients in the HCQ group and 42% of patients in the HCQ plus azithromycin group.

The other outcome measure in HCQ trials was radiological lung clearance, where Chen et al. (2020b) used a dosing regimen of 400 mg daily and reported an improvement in radiological results on day 6 in 80.6% of patients in the HCQ arm (n = 31) versus 54.8% of patients in the control arm (n = 31).

Boulware et al. (2020) reported the incidence of either laboratory-confirmed COVID-19 or illness compatible with COVID-19 within 14 days of administration of either HCQ (800 mg once, followed by 600 mg in 6–8 h, then 600 mg daily for an additional 4 days) or placebo for patients who had been exposed to individuals with confirmed COVID-19, and reported a lower incidence of 11.8% in the HCQ arm (n = 414) compared with 14.3% in the placebo arm (n = 407). Mitjà et al., 2020 compared the incidence of laboratory-confirmed symptomatic COVID-19 within 14 days of administration of either HCQ (800 mg on day 1, followed by 400 mg once daily for 6 days; n = 1,116) or treatment with usual care (n = 1198) for healthy contacts of COVID-19 index patients, and observed that the incidence was lower in the HCQ arm (5.7%) than in the usual care arm (6.2%).

For CQ trials, Huang et al. (2020b) compared CQ in a regimen of 500 mg orally twice daily for 10 days (n = 10) with lopinavir/ritonavir in a regimen of 400 mg/100 mg for 10 days (n = 12), and observed that all patients in the CQ arm achieved virological clearance on day 14, while 11 of 12 patients in the lopinavir/ritonavir arm achieved virological clearance on day 14. In terms of lung clearance rate based on computed tomography findings, 60% of patients in the CQ group achieved radiological lung clearance by day 9, compared with 25% in the lopinavir/ritonavir group. Borba et al. (2020) compared high (600 mg twice daily for 10 days; n = 41)

Table 1
Summary of findings in included trials.

Study authors	Study design	No of patients	Treatment regimen	Patient characteristics: Age (years) (mean ± SD)	Male (%)	Common comorbid conditions	Baseline disease severity	Outcomes: Viral clearance (n/N)	Lung clearance (n/N)	Patients discharged (n/N)	Death or transfer to ICU or hospitalization (n/N)	Confirmed or probable COVID-19 (n/N)	Development of new symptoms (n/N)	Adverse events[c] (n/N)	Treatment discontinuation[b] (n/N)	Summary of findings
Boulware et al. (2020)	Double-blind placebo-controlled RCT	414	Orally administered HCQ 800 mg once, followed by 600 mg in 6–8 h, then 600 mg daily for an additional 4 days	41 (33–51)[a]	47.3	Hypertension Asthma Diabetes	NR	NR	NR	NR	None	49/414 (11.8%)	57/414 (13.8%)	Any adverse event =140/349 (40.1%); nausea = 80/349 (32.9%); diarrhoea =81/349 (23.2%); neurologic reaction = 19/349 (5.4%); visual changes = 3/349 (0.9%)	17/414 (4.1%)	Inconclusive: no statistically significant benefit
		407	Placebo	40 (32–50)[a]	49.4	Hypertension Asthma Diabetes	NR	NR	NR	NR	None	58/407 (14.3%)	59/407 (14.5%)	Any adverse event = 59/351 (16.8%); nausea = 27/351 (7.7%); diarrhoea = 15/351 (4.3%); neurologic reaction = 13/351 (3.7%); visual changes = none	8/407 (2.0%)	
Gautret et al. (2020)	Open-label non-randomized controlled trial	20	Orally administered HCQ 200 mg, three times per day for 10 days	51.2 ± 18.7	45.0	NR	NR	14/20 (70.0%; 6 dropped out) at day 6	NR	NR	ICU admission = 3/20 (15.0%); death = 1/20 (5.0%)	NR	NR	Nausea reported in one patient where the patient ceased HCQ treatment and was excluded from the analysis	One patient discontinued treatment because of nausea and was excluded from the analysis	Suggestive of benefits: increased negative conversion rate with HCQ but with a low degree of certainty because of critical risk of bias
		16	Control	37.3 ± 24.0	37.5	NR	NR	2/16 (12.5%) at day 6	NR	NR	NR	NR	NR	NR	NR	
Chen et al. (2020a)	Open-label RCT	15	Orally administered HCQ 400 mg daily for 5 days plus conventional treatments	48	NR	Hypertension Diabetes COPD	NR	13/15 (86.7%) at day 7	5/15 (33.3%) on day 3	NR	None	NR	NR	Any adverse event = 4/15 (26.7%); diarrhoea = 2/15 (13.3%); abnormal liver function = 1/15 (6.6%)	NR	Inconclusive: no statistically significant benefit
		15	Control	48	NR	Hypertension Diabetes COPD	NR	14/15 (93.3%) at day 7	7/15 (46.7%) on day 3	NR	None	NR	NR	Any adverse event = 3/15 (20.0%); abnormal liver function and anaemia = 1/15 (6.6%)	NR	
Tang et al. (2020)	Open-label RCT	75	Orally administered HCQ loading dose of 1200 mg daily for 3 days followed by a maintenance dose of 800 mg/day for remaining days (total treatment duration 2 weeks or 3 weeks for patients mild/ moderate or severe disease, respectively)	48.0 ± 14.1	56.0	Diabetes Hypertension	Mild = 15/ 75 (20.0%); moderate = 59/75 (78.7%); severe = 1/75 (1.3%)	Frequency was not reported; probability of viral clearance estimated by Kaplan–Meier method = 85.4%	NR	NR	NR	NR	NR	Any adverse event = 21/70 (30.0%); serious adverse event = 2/70 (2.9%); diarrhoea =7/70 (10.0%)	1/75 (1.3%) due to blurred vision	Inconclusive: no statistically significant benefit
		75	SOC as per Chinese guidelines	44.1 ± 15.0	53.3	Diabetes Hypertension	Mild = 7/ 75 (9.3%); moderate	Frequency was not reported;	NR	NR	NR	NR	NR	Any adverse event = 7/80 (8.8%); serious adverse	None	

Table 1 (*Continued*)

Study authors	Study design	No of patients	Treatment regimen	Patient characteristics			Baseline disease severity	Outcomes[b]						Adverse events[c] (n/N)	Treatment discontinuation[b] (n/N)	Summary of findings
				Age (years) (mean ± SD)	Male (%)	Common comorbid conditions		Viral clearance (n/N)	Lung clearance (n/N)	Patients discharged (n/N)	Death or transfer to ICU or hospitalization (n/N)	Confirmed or probable COVID-19 (n/N)	Development of new symptoms (n/N)			
							= 67/75 (89.3%); severe = 1/75 (1.3%)	probability of viral clearance estimated by Kaplan–Meier method = 81.3%						event = none; diarrhoea = none		
Chen et al. (2020b)	Open-label RCT	31	Orally administered HCQ 200 mg twice daily	44.1 ± 16.1	45.2	NR	NR	NR	25/31 (80.6%) at day 6	NR	NR	NR	NR	Mild adverse event = 2/31 (6.5%); serious adverse event = none	None	Suggestive of benefits: more rapid lung clearance but with a moderate degree of certainty because of some concerns regarding the risk of bias
		31	Control	45.2 ± 14.7	48.3	NR	NR	NR	17/31 (54.8%) at day 6	NR	NR	NR	NR	None	None	
Esper et al. (2020)	Open-label non-randomized controlled trial via telemedicne	412	Orally administered HCQ 800 mg on the first day and 400 mg for another 6 days and AZT 500 mg once daily for 5 days	63.6 ± 14.9	36.4	Diabetes	NR	NR	NR	NR	Hospitalization = 8/412 (1.9%); death = 2/412 (0.49%) but non COVID-19 related	NR	NR	Serious adverse event = none; dizziness = 8/412 (1.9%); diarrhoea = 68/412 (16.5%); nausea = 31/412 (7.5%) vomiting = 5/412 (1.2%) visual disturbance = 1/412 (0.2%); allergy = 4/412 (1.0%)	NR	Suggestive of benefits: reduced need for hospitalization in suspected COVID-19 patients but with a low degree of certainty because of serious risk of bias
		224	Control (refused HCQ + AZT)	61.0 ± 16	38.4	Diabetes	NR	NR	NR	NR	Hospitalization = 12/224 (5.4%)	NR	NR	NR	NR	
Huang et al. (2020b)	Open-label parallel RCT	10	Orally administered CQ 500 mg orally twice daily for 10 days	41.5 (33.8–50.0)[a]	30.0	NR	Severe disease = 3/10 (30.0%)	At day 10 = 9/10 (90%); at day 14 = 10/10 (100%)	At day 10 = 2/10 (20%); at day 14 = 10/10 (100%)	At day 14 = 10/10 (100%)	None	NR	NR	Any adverse event = 9/10 (90.0%); serious adverse event = none; vomiting = 5/10 (50.0%); diarrhoea = 5/10 (50.0%); nausea = 4/10 (40.0%)	None	Inconclusive: no statistically significant benefit
		12	Orally administered lopinavir/ ritonavir 400 mg/ 100 mg for 10 days	53.0 (41.8–63.5)[a]	50.0	NR	Severe disease = 5/12 (41.7%)	At day 10 = 9/12 (75.0%); at day 14 = 11/12 (91.7%)	At day 10 = 2/12 (8.3%); at day 14 = 9/12 (75.0%)	At day 14 = 6/12 (50%)	None	NR	NR	Any adverse event = 10/12 (83.3%); serious adverse event = none; vomiting = 1/12 (8.3%); diarrhoea = 8/12 (66.7%); nausea = 5/12 (41.7%)	NR	
Borba et al. (2020)	Double-blind parallel RCT	41	Orally administered CQ (4 × 150 mg tablets, twice daily for 10 days; total dose 12 g)	54.7 ± 13.7	75.0	Hypertension kidney disease	NR	NR	NR	NR	Death = 16/41 (39.0%)	NR	NR	Ventricular tachycardia = 2/37 (5.4%); prolonged QT interval = 7/37 (18.9%)	NR	Suggested: safety concerns with high dose of CQ
		40	Orally administered CQ (3 × 150 mg tablets and 1 placebo tablet)	47.4 ± 13.3	74.5	Hypertension	NR	NR	NR	NR	Death = 6/40 (15.0%)	NR	NR	Severe rhabdomyolysis = 1/36 (2.8%); ventricular tachycardia = 0/36	NR	

Study	Design	N	Intervention	Age (mean ± SD)		Comorbidities	Disease severity		Outcome			Adverse events	Conclusion
			twice daily on day 0, 3 × 150 mg tablets plus 1 placebo tablet once a day followed by 4 placebo tablets from day 1 to day 4, then 4 placebo tablets twice daily from day 5 to day 9; total dose 2.7 g)									(5.4%); prolonged QT interval = 4/36 (11.1%)	
Cavalcanti et al. (2020)	Open label RCT	217	Orally administered HCQ 400 mg twice daily plus AZT 500 mg once daily for 7 days	49.6 ± 14.2	56.7	Heart failure HypertensionCOPD Asthma ObesityDiabetes AIDSChronic kidney disease Current or former smoking Cancer	Mild-to-moderate disease	NR	At day 15 = 184/217 (84.8%)	At day 15 death = 3/217 (1.4%)	NR	Any adverse event = 94/239 (39.3%); serious adverse event = 5/239 (2.1%); prolonged QT interval = 17/116 (14.7%); arrhythmia = 3/239 (1.3%); nausea = 6/239 (2.5%); abnormal liver function = 26/239 (10.9%)	Suggested: no clinical improvement after treatment with HCQ alone or in combination with AZT
		221	Orally administered HCQ 400 mg twice daily for 7 days	51.3 ± 14.5	64.3	Heart failure HypertensionCOPD AsthmaObesity Diabetes AIDSChronic kidney disease Current or former smoking Cancer	Mild-to-moderate disease	NR	At day 15 = 185/221 (83.7%)	At day 15 death = 7/221 (3.2%)	NR	Any adverse event = 67/199 (33.7%); serious adverse event = 2/199 (1.0%); prolonged QT interval = 13/89 (14.6%); arrhythmia = 3/199 (1.5%); nausea = 9/199 (4.5%); abnormal liver function = 17/199 (8.5%)	
		227	SOC	49.9 ± 15.1	54.2	Heart failure HypertensionCOPD AsthmaObesity Diabetes AIDSchronic kidney disease Current or former smoking Cancer	Mild-to-moderate disease	NR	At day 15 = 195/227 (85.9%)	At day 15 death = 6/227 (2.6%)	NR	Any adverse event = 49/227 (21.6%); serious adverse event = 2/227 (0.9%); prolonged QT interval = 1/64 (1.6%); arrhythmia = 1/227 (0.4%); nausea = 2/227 (0.9%); abnormal liver function = 8/227 (3.5%)	
Mitjà et al., 2020a	Open-label RCT	136	Orally administered HCQ 800 mg once on day 1, followed by 400 mg daily for 6 days	41.6 ± 12.4	27.9	Cardiovascular disease Respiratory disease Metabolic disease Nervous system disease	NR	NR	128/136 (94.1%)	Hospitalization = 8/136 (5.9%); death = none	NR	Any adverse event = 121/169 (72.0%); gastrointestinal disorders = 148/169 (88.1%); infections and infestations = 9/169 (5.4%); general disorders = 30/169 (17.9%); nervous system disorder = 63/169 (37.5%); ear and labyrinth disorders = 5/169 (3.0%)	Inconclusive: no statistically significant benefit
		157	Control	41.7 ± 12.6	34.4	Cardiovascular disease Respiratory disease Metabolic disease Nervous system disease	NR	NR	143/155 (92.3%)	Hospitalization = 11/155 (7.1%); death = none	NR	Any adverse event = 16/184 (8.7%); gastrointestinal disorders = 7/184 (3.8%); infections and infestations = 12/184 (6.6%);	

Table 1 (*Continued*)

Study authors	Study design	No of patients	Treatment regimen	Patient characteristics				Outcomes[b]						Adverse events[c] (n/N)	Treatment discontinuation[b] (n/N)	Summary of findings
				Age (years) (mean ± SD)	Male (%)	Common comorbid conditions	Baseline disease severity	Viral clearance (n/N)	Lung clearance (n/N)	Patients discharged (n/N)	Death or transfer to ICU or hospitalization (n/N)	Confirmed or probable COVID-19 (n/N)	Development of new symptoms (n/N)			
														general disorders = 1/184 (0.5%); nervous system disorder = 3/184 (1.6%); ear and labyrinth disorders = 0/184 (0%)		
Skipper et al. (2020)	Double-blind placebo-controlled RCT	212	Orally administered HCQ 800 mg once, followed by 600 mg in 6–8 h, then 600 mg daily for 4 days (5 days in total)	41 (33–49)[a]	42.0	Hypertension Diabetes Asthma	Mean symptom severity score = 4.2	NR	NR	NR	Hospitalization = 4/212 (1.8%); death = 1/212 (0.4%)	73/212 (34.4%)	NR	Any adverse event = 92/212 (43.4%); serious adverse event = none; upset stomach/nausea = 66/212 (31.1%); rash = 6/212 (2.8%); changes in vision = 4/212 (1.9%)	NR	Inconclusive: no statistically significant benefit
		211	Control	39 (31–50)[a]	45.4	Hypertension Diabetes Asthma	Mean symptom severity score = 4.1	NR	NR	NR	Hospitalization = 10/211 (4.7%); death = 1/211 (0.5%)	72/211 (34.1%)	NR	Any adverse event = 46/211 (21.8%); serious adverse event = none; upset stomach/nausea = 26/211 (12.3%); rash = 2/211 (1.0%); changes in vision = 5/211 (2.4%)	NR	
Chen et al. (2020c)	Open-label RCT	21	Orally administered HCQ 400 mg twice on day 1, followed by 200 mg twice daily for an additional 6 days	33.0 ± 12.0	52.4	NR	Mild = 19/21 (90.5%); moderate = 2/21 (9.5%)	At day 14 = 17/21 (81.0%)	NR	NR	NR	NR	NR	Serious adverse event = none	NR	Inconclusive: no statistically significant benefit
		12	SOC	32.8 ± 8.3	66.7	NR	Mild = 10/12 (83.3%); moderate = 2/12 (16.7%)	At day 14 = 9/12 (75.0%)	NR	NR	NR	NR	NR	Serious adverse event = none	NR	
Horby et al. (2020)	Open-label RCT	1561	Orally administered HCQ 800 mg at 0 and 6 h followed by 400 mg at 12 h after the initial dose and then every 12 h for the next 9 days or until discharge	65.2 ± 15.2	61.6	Diabetes Heart disease Lung disease Tuberculosis HIV positive Severe liver disease Severe kidney impairment	NR	NR	NR	At day 28 = 941/1561 (60.3%)	Death at day 28 = 418/1561 (26.8%)	NR	NR	Serious adverse event = 1/698 (0.14%); torsades de pointes = 1/698 (0.14%)	NR	Inconclusive: no statistically significant benefit
		3155	SOC	65.4 ± 15.4	62.6	Diabetes Heart disease Lung disease Tuberculosis HIV positive Severe liver disease Severe kidney impairment	NR	NR	NR	At day 28 = 1,982/3155 (62.8%)	Death at day 28 = 788/3155 (25.0%)	NR	NR	NR	NR	
Abd-Elsalam	Open-label RCT	97	Orally administered HCQ 400 mg	40.4 ± 18.7	57.7	Liver diseases Renal impairment	NR	NR	NR	NR	ICU admission = 11/97 (11.3%);	NR	NR	NR	NR	Inconclusive: no statistically

et al. (2020)			twice daily on day 1, followed by 200 mg twice daily							death = 6/97 (6.1%)				significant benefit
		97	Control	41.09 ± 20.07	59.8	Liver diseases Renal impairment	NR	NR	NR	NR	ICU admission = 13/97 (13.4%); death = 5/97 (5.1%)	NR	NR	NR	NR
Furtado et al. (2020)	Open-label RCT	214	Orally administered HCQ 400 mg twice daily plus AZT 500 mg once daily for 10 days	59.4 (49.3–70.0)[a]	65.4	Hypertension DiabetesHeart failurePrevious strokeP revious myocardial infarction COPDActive cancerC hronic kidney failure	NR	NR	NR	NR	Death at day 29 = 90/214 (42.1%)	NR	NR	Serious adverse events = 102/241 (42.3%); corrected QT interval prolongation = 47/241 (19.5%); gastrointestinal intolerance = 61/241 (25.3%); ventricular arrhythmias = 8/241 (3.3%); acute kidney failure = 147/241 (61.0%); death due to acute kidney failure = 2/241 (0.8%)	NR	Inconclusive: no statistically significant benefit with addition of AZT
		183	Orally administered HCQ 400 mg twice daily for 10 days	60.2 (52.0–70.1)[a]	66.7	Hypertension DiabetesHeart failurePrevious strokeP revious myocardial infarction COPDActive cancerC hronic kidney failure	NR	NR	NR	NR	Death at day 29 = 73/183 (39.9%)	NR	NR	Serious adverse events = 75/198 (37.9%); corrected QT interval prolongation = 42/198 (21.2%); gastrointestinal intolerance = 48/198 (24.2%); ventricular arrhythmias = 5/198 (2.5%); acute kidney failure = 103/198 (52.0%); death due to acute kidney failure = 3/198 (1.5%)	NR	
Mitjà et al., 2020b)	Open-label RCT	1116	Orally administered HCQ 800 mg once, followed by 400 mg daily for 6 days	48.6 ± 18.7	27.1	Cardiovascular disease Respiratory disease Metabolic disease Nervous system disease	NR	NR	NR	NR	NR	64/1116 (5.7%)	NR	Any adverse event = 671/1,197 (51.6%); palpitations = 5/1,197 (0.4%); gastrointestinal disorder = 510/1,197 (42.6%); nervous system disorder = 260/1,197 (21.7%); general disorder = 103/1,197 (8.6%)	NR	Inconclusive: no statistically significant benefit
		1198	Control	48.7 ± 19.3	26.9	Cardiovascular disease Respiratory disease Metabolic disease Nervous system disease	NR	NR	NR	NR	NR	74/1198 (6.2%)	NR	Any adverse event = 77/1,300 (5.9%); palpitations = 1/1,300 (0.1%); gastrointestinal disorder = 33/1,300 (2.5%); nervous system disorder = 32/1300 (2.5%); general disorder = 10/1300 (0.8%)	NR	

AIDS, acquired immunodeficiency syndrome: AZT, azithromycin; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; CQ, chloroquine; HCQ, hydroxychloroquine; HIV, human immunodeficiency virus; ICU, intensive care unit; NR, not reported; RCT, randomized controlled trial; SD, standard deviation; SOC: standard of care.

[a] Reported as the median (interquartile range).

[b] Intention-to-treat population.

[c] Safety population.

and low (450 mg twice daily on day 1 and once daily for 4 days; *n* = 40) doses of CQ, and reported higher mortality (17.5%) with the high-dose regimen than with the low-dose regimen (9.7%).

Reporting of harms

Tang et al. (2020) in their HCQ trial found a higher rate of any AEs among HCQ recipients (30.0%, *n* = 21/70) compared with patients who received standard of care treatment (8.8%, *n* = 7/80). Similarly, Boulware et al. (2020) reported a higher rate of any AEs in HCQ recipients (40.1%, *n* = 140/414) relative to placebo recipients (16.8%, *n* = 59/407). In both trials, the most common AE among the HCQ recipients was diarrhoea (10.0%, *n* = 7/70, in the trial by Tang et al. (2020) and 19.6%, *n* = 81/414, in the trial by Boulware et al. (2020)). This was also reported in a larger trial by Esper et al. (2020), who also reported diarrhoea (16.5%, *n* = 68/412) as the most common AE among 412 patients who received HCQ.

In both trials by Mitjà et al., 2020a, 2020, there was also a higher rate of any AEs in the HCQ arm compared with the control/placebo arm (72.0%, *n* = 121/169, versus 8.7%, *n* = 16/184, in the trial by Mitjà et al., 2020a and 51.6%, *n* = 671/1,197, versus 5.9%, *n* = 77/1300, in the trial by Mitjà et al., 2020b). The most frequent AEs reported among participants given HCQ in both trials by Mitjà et al., 2020a, 2020 were related to the gastrointestinal system (diarrhoea, nausea, and abdominal pain) without individual presentation of the AEs. Similarly, Skipper et al. (2020) reported more frequent occurrences of any AEs in the HCQ group (43.4%, *n* = 92/212, versus 21.8%, *n* = 46/211), with the most frequent AEs reported being upset stomach and nausea (31.1%, *n* = 66/212).

Although more AEs were also reported in the trial by Cavalcanti et al. (2020) in patients who received HCQ plus azithromycin (*n* = 94/239, 39.3%) or HCQ alone (*n* = 67/199, 33.7%) than in those who received azithromycin alone (*n* = 9/50, 18.0%) or neither of the trial drugs (*n* = 40/177, 22.6%), a corrected QT interval greater than 480 ms within 7 days was the most frequent AE observed in patients who received HCQ (14.7%, *n* = 17/116, in the HCQ plus azithromycin group and 14.6%, *n* = 13/89, in the HCQ-alone group). Chen et al. (2020b) reported two patients (3.2%) with mild adverse reactions in the HCQ group (*n* = 31), where one patient developed a rash and one patient experienced a headache, while no patients experienced an AE in the control group.

Furtado et al. (2020) did not compare the proportion of patients with any AEs between HCQ recipients and HCQ plus azithromycin recipients, although they reported a higher proportion of serious AEs among HCQ plus azithromycin recipients (42%, *n* = 102/241) relative to HCQ recipients (38%, *n* = 75/198). Similarly, Chen et al. (2020c) did not compare the proportion of patients with AEs between the two study arms, although they reported headache as the most frequent grade 1 and grade 2 HCQ-related AEs.

Huang et al. (2020b) in their CQ trial observed that almost all of the patients (*n* = 9/10) experienced CQ-related AEs, with the most common AEs being vomiting (*n* = 5) and diarrhoea (*n* = 5). Borba et al. (2020), who compared a high-dose regimen and a low-dose regimen of CQ, reported that a higher proportion of patients (in the safety population) who received high-dose CQ (18.9%, *n* = 7/37) experienced QT prolongation compared with patients who received low-dose CQ (11.1%, *n* = 4/36).

Among the trials (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020; Esper et al., 2020) which presented harm data (*n* = 12), all trials presented harm data in both text and table form except the trial by Chen et al. (2020c), which presented harm data in text form only. Nevertheless, only four (Boulware et al., 2020; Cavalcanti et al., 2020; Mitjà et al., 2020; Skipper et al., 2020) of the 12 trials described the scale or any

criteria used to measure the severity of AEs. In all but four trials (Boulware et al., 2020; Furtado et al., 2020; Mitjà et al., 2020; Skipper et al., 2020), safety data were presented as frequencies only, without a statistical comparison of the occurrence of the AE between the investigational arm and the control arm. All but the two trials by Esper et al. (2020) and Skipper et al. (2020) attributed AEs to HCQ. Nine trials (Cavalcanti et al., 2020; C. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020) reported both frequent and serious AEs, while the other three trials (Esper et al., 2020; Borba et al., 2020; Boulware et al., 2020) reported only AEs selected by the investigators.

Out of a maximum score of 19, the median THRS was 6.5 (interquartile range 5). Among the total of 19 CONSORT items, all but five trials (Borba et al., 2020; Cavalcanti et al., 2020; Furtado et al., 2020; Mitjà et al., 2020; Skipper et al., 2020) reported less than 50% of the items (THRS range 1–9). Only the two trials by Borba et al. (2020) and Cavalcanti et al. (2020) reported more than 60% of the items (THRS 13).

Adherence to CONSORT recommendations

The number and percentages of the randomized controlled trials fulfilling each of the CONSORT harm recommendations are presented in Table 2. Scoring for each recommendation can be found in Table S1. Nine trials (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; Chen et al., 2020; Furtado et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020) mentioned AEs in the title or abstract (CONSORT recommendation 1). However, only five trials (Esper et al., 2020; Borba et al., 2020; Cavalcanti et al., 2020; Abd-Elsalam et al., 2020; Furtado et al., 2020) provided information on AEs in the introductory section (CONSORT recommendation 2). Although only five trials (Boulware et al., 2020; Cavalcanti et al., 2020; Horby et al., 2020; Mitjà et al., 2020; Skipper et al., 2020) used a validated scale to measure the severity of AEs, half of the trials (*n* = 8) (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; Chen et al., 2020; Furtado et al., 2020; Mitjà et al., 2020; Tang et al., 2020; Esper et al., 2020) defined the AEs (CONSORT recommendations 3) in their report. Only seven trials (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; Chen et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020) described how AE-related data were collected (CONSORT recommendation 4 (4a)), but more than half of the trials (*n* = 9) (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Esper et al., 2020) described when AE data were collected (CONSORT recommendation 4 (4b)). Only two trials (Borba et al., 2020; Cavalcanti et al., 2020) described methods of presenting and/or analysing AEs (CONSORT recommendation 5). Fewer than half of the trials (*n* = 7) (Borba et al., 2020; Boulware et al., 2020; Chen et al., 2020; Huang et al., 2020; Mitjà et al., 2020; Skipper et al., 2020; Tang et al., 2020) described the number of withdrawals due to AEs in each arm (CONSORT recommendation 6 (6a)). More than half of the trials (*n* = 10) (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Furtado et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020) provided denominators for AEs (CONSORT recommendation 7 (7a)). Most of the trials presented results for each arm separately (*n* = 13) (Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020; Esper et al., 2020) and presented a balanced discussion on both the safety and the efficacy of the drug (*n* = 11) (Abd-Elsalam et al., 2020; Borba et al., 2020; Cavalcanti et al., 2020; Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Huang

Table 2
Items adequately reported against the quality of reporting criteria (Consolidated Standards of Reporting Trials (CONSORT) Extension for Harm).

Recommendations of 2004 CONSORT harm extension	Quality of reporting criteria	Number of trials
1. If the study collected data on harms and benefits, the title of the abstract should state so	Adverse events mentioned in the title	0 (0%)
	Adverse events mentioned in the abstract	9 (56%)
2. If the trial addresses both harms and benefits, the introduction should state so	Information on adverse events mentioned in the introduction	5 (31%)
3. List addressed adverse events with definitions for each (with attention, when relevant, to grading, expected versus unexpected events, reference to standardized and validated definitions, and description of new definitions)	3a. If article mentioned the use of a validated instrument to report adverse event severity	5 (31%)
	3b. If article mentioned definition of adverse event	8 (50%)
4. Clarify how harm-related information was collected (mode of data collection, timing, attribution methods, intensity of ascertainment, and harm-related monitoring and stopping rules, if pertinent)	4a. Description of how harm data were collected (e.g. diaries, phone interviews, face-to-face interviews)	7 (44%)
	4b. Description of when adverse event data were collected	9 (56%)
	4c. Whether or not adverse events were attributed to trial drug (e.g. how adverse events were attributed to drugs)	2 (13%)
5. Describe plans for presenting and analysing information on harms (including coding, handling of recurrent events, specification of timing issues, handling of continuous measures, and any statistical analyses)	5. Description of methods for presenting and/or analysing adverse events	2 (13%)
6. Describe for each arm the participant withdrawals that are due to harms and the experience with the allocated treatment	6a. If the article reported number of withdraws caused by adverse events in each arm	7 (44%)
	6b. Description of adverse events leading to withdrawals	5 (31%)
	6c. Description of adverse events leading to death	2 (13%)
7. Provide the denominators for analyses on harms	7a. If the article provided denominators for adverse events	10 (63%)
	7b. If the article provided definitions used for analysis set (intention to treat, per protocol, safety data available, unclear)	3 (19%)
8. Present the absolute risk of each adverse event (specifying type, grade, and seriousness per arm), and present appropriate metrics for recurrent events, continuous variables, and scale variables, whenever pertinent	8a. Results presented separately for each arm	13 (81%)
	8b. Separate reporting of severe adverse events s (grade >2 or serious adverse events)	11 (69%)
	8c. Provided both number of adverse events and number of patients with adverse events	2 (13%)
9. Describe any subgroup analyses and exploratory analyses for harms	–	2 (13%)
10. Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms	10a. If the discussion was balanced with regard to efficacy and adverse events	9 (56%)
	10b. Limitations of the study specifically in relation to adverse events discussed	2 (13%)

et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020) (CONSORT recommendations 8 (8a) and 10 (10a)). The trials by Esper et al. 2020 and Cavalcanti et al. (2020) were the only trials that described subgroup analyses and exploratory analyses for harms (CONSORT recommendation 9).
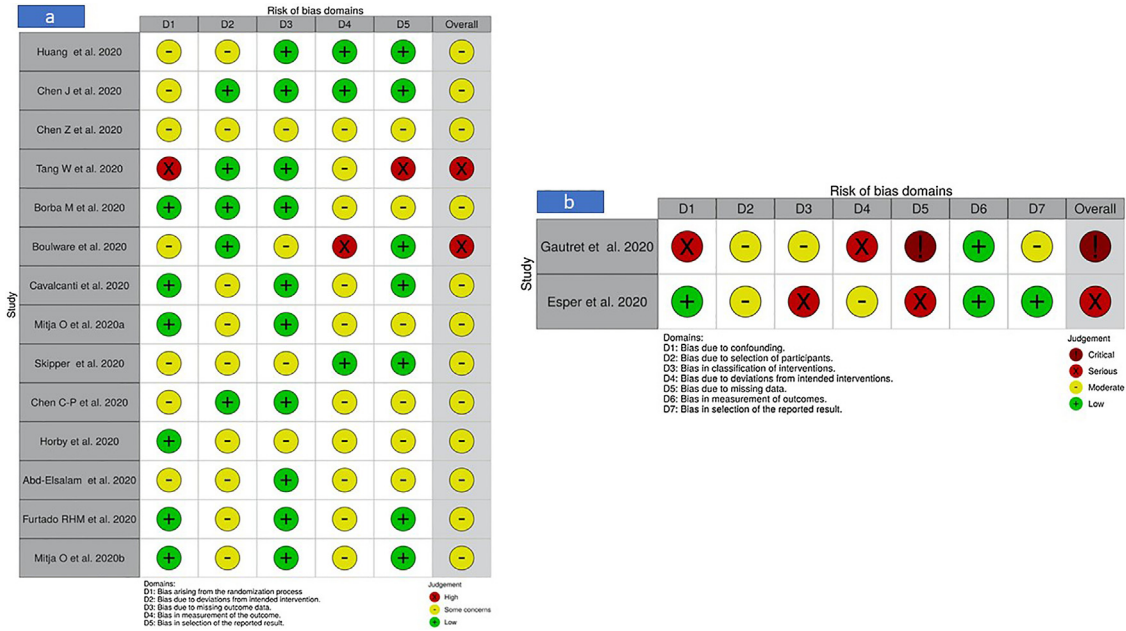
Risk of bias

The risk of bias analysis using the RoB 2 and ROBINS-I frameworks for the 16 trials included in this review is summarized in Figure 2. Surprisingly, none of the 16 clinical trials scored an overall 'low risk', four trials scored a high (Tang et al., 2020; Boulware et al., 2020), critical (Gautret et al., 2020), or serious (Esper et al., 2020) risk, and the remaining trials (Abd-Elsalam et al., 2020; Borba et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020) were classified into a moderate risk.

Among the 14 randomized controlled trials (Abd-Elsalam et al., 2020; Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020) included in this review which had their risks of bias assessed with RoB 2, all of the trials had at least some concerns of risk of bias in at least one of the bias domains in RoB 2. The most significant risk of bias across all 14 randomized controlled trials was bias in the measurement of the outcome, in which 11 of the trials (Abd-Elsalam et al., 2020; Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Mitjà et al., 2020a, 2020b; Tang et al., 2020) had at least some concerns on this domain of bias. Such concerns were due to the possibility that assessments of outcomes, either efficacy outcomes or outcomes of occurrence of AEs, would be affected by

knowledge of the intervention assignment by the outcome assessors owing to the open-label study design. The trial by Boulware et al. (2020) had a particularly high risk of bias in the measurement of the outcome because 24% of those randomized to receive the study treatment were excluded in the analysis on the outcome of occurrence of AEs. Although similar proportions of participants were excluded from analysis in the two groups on the outcome of the occurrence of AEs, the reasons for exclusion differ between the two arms and are likely to be related to the outcome. Also, the trial by Boulware et al. (2020) had some concerns of bias due to missing outcome data since there were critical differences between interventions in the proportion of participants with missing data (4.1% of patients in HCQ-treated group, whereas 2.0% of patients in the placebo group discontinued treatment).

Another significant risk of bias across the included randomized trials was bias due to deviations from the intended intervention, in which nine of the trials (Abd-Elsalam et al., 2020; Cavalcanti et al., 2020; Chen et al., 2020; Furtado et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020) had some concerns on this domain of bias. These trials provided no information on co-interventions of interest such as the use of systemic corticosteroids and antivirals among participants, and thus signalled some concerns regrading the existence of bias due to the possibility of unequal use of co-interventions of interest between the study groups.

The trial by Tang et al. (2020) had an overall high risk of bias, with a particularly high risk of bias arising from the randomization process; it was an open-label study and the authors did not assess baseline differences between intervention groups. Also, a high risk of bias was noted in the trial by Tang et al. (2020) in the selection of the reported results as there were unplanned analyses of the changes in C-reactive protein values and blood lymphocyte count, and there were unplanned post hoc subgroup analyses. The trial by Chen et al. (2020b) exhibited some concerns regarding the risk of

Figure 2. Summary of risk of bias: (a) RoB 2 for randomized trials; (b) Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool for non-randomized trials.

bias in all the domains assessed in RoB 2. Similarly, the trials by Horby et al. (2020) and Abd-Elsalam et al. (2020) had some concerns regarding the risk of bias in all the domains assessed except the domain of bias arising from the randomization process in the former and the domain of bias due to missing outcome data in the latter. The trial by Chen et al. (2020a) was noted to have the least risk of bias, with some concerns regarding the randomization process since no information was provided about the concealment of the allocation sequence, while all the other domains had a low risk of bias.

The trial by Borba et al. (2020) had some concerns regarding the risk of bias in the selection of the reported results because of the early interruption of the high-dose treatment, the unmasking of treatment allocation in some participants, and missing data on radiological findings. BoBbBoth the trials by Mitjà et al., 2020 and Skipper et al. (2020)had some concerns regarding the risk of bias due to deviations from the intended intervention. The trial by Mitjà et al., 2020 also had some concerns regrading the risk of bias in the selection of the reported results because of the unavailability of the study protocol and statistical analysis plan. The trial by Skipper et al. (2020) also had some concerns regarding the risk of bias arising from the randomization process because 20% of the participants assigned to the intervention and control arms were also randomized in a separate trial. Moreover, with nearly 5% of participants with missing data for death as the outcome and with more than 10% of participants with missing data for the primary outcome in the trial by Skipper et al. (2020), there appear to be some concerns regarding the risk of bias due to missing outcome data. Although the trial by Chen et al. (2020c) had a low risk of bias due to deviations from the intended intervention, there were some concerns regarding the risk of bias arising from the randomization process because of differences in the proportion of participants with mild disease severity and moderate disease severity between the two study arms at the baseline and some concerns regarding the risk of bias in the selection of the reported results because of unavailability of the study protocol and statistical analysis plan.

The risk of bias for non-randomized controlled trials (Gautret et al., 2020; Esper et al., 2020) was assessed with the ROBIN-I tool. Across the seven domains of the risk of bias in the ROBIN-I tool, the trial by Gautret et al. (2020) had at least a moderate risk of bias in

all domains except bias in the measurement of outcomes (low risk). In particular, the trial by Gautret et al. (2020) had a critical risk of bias due to missing data, in which there were critical differences between the two intervention arms in the proportion of participants with missing data (30% of HCQ-treated patients were lost to follow-up, while no patients (0%) were lost to follow-up in the control group), and a serious risk of bias due to confounding because of the baseline differences in participants' characteristics between the treatment arm and the control arm. Also, the trial had a serious risk of bias due to deviations from the intended interventions since the use of important co-intervention (azithromycin) was not balanced across the intervention groups. For the trial by Esper et al. (2020), a serious risk of bias in the classification of interventions and a serious risk of bias due to missing data were observed. The remaining domains for the trial by Esper et al. (2020) had either low or moderate risk of bias.

Discussion

This systematic review was aimed to critically assess and summarize the quality of published clinical trials evaluating the effectiveness of CQ or HCQ in the treatment of COVID-19. Overall, the 16 trials (Abd-Elsalam et al., 2020; Borba et al., 2020; Boulware et al., 2020; Cavalcanti et al., 2020; C. Chen et al., 2020; J. Chen et al., 2020; Z. Chen et al., 2020; Furtado et al., 2020; Gautret et al., 2020; Horby et al., 2020; Huang et al., 2020; Mitjà et al., 2020a, 2020b; Skipper et al., 2020; Tang et al., 2020; Esper et al., 2020) included in our review recruited 10,873 participants (5036 patients in the HCQ/CQ group and 5837 patients in the comparator group). Potentially a meta-analysis of these trials would have definitively answered the effectiveness and safety questions regarding CQ or HCQ in the treatment of COVID-19, but given the methodological inadequacies in these trials, a meta-analysis is likely to produce misleading outcomes.

Since the publicity of HCQ/CQ in the press as a potential cure for COVID-19 on the basis of promising preliminary clinical experiences (Liu et al., 2020) and *in vitro* investigations (Gao et al., 2020), several trials were designed across the world to testify the effectiveness and safety of these antimalarial drugs for the treatment of COVID-19. However, we found that most of these

trials recruited a fairly small number of participants and hence lack statistical power and significance. On critical review of the trial design and reporting, we found that the findings from these trials were far from conclusive, in which inferences were contradictory and findings were significantly confounded by covariates such as comorbidities and/or use of co-interventions, along with lack of adjustment for disease severity at the baseline. Indeed, some trials provided no information on the baseline severity for the patients recruited. Moreover, all trials were associated with moderate to high risk of bias; the examples included risks of bias arising from the randomization process, potential deviation from intended interventions, outcome measurements, selective reporting, confounding, participant selection, and/or classification of interventions. Also, none of the trials included in this review met the CONSORT criteria in full for reporting harms in the trials, with few trials presenting no harm data. Although HCQ/CQ has a history of safe use, AEs such as toxic retinopathy and corrected QT prolongation especially may have a fatal consequence and, therefore, inadequate reporting on AEs in these trials can be misleading and pose serious risks to public safety, particularly during more widespread use during the COVID-19 pandemic. This is not surprising given that amid the current pandemic, researchers may place stronger emphasis on the benefits than the risks in an attempt to save lives.

It is, however, acknowledged that there are genuine methodological challenges in designing studies for COVID-19 pandemic crises and there is a great sense of urgency for the containment of the COVID-19 pandemic. A recent editorial by Knottnerus and Tugwell (2020) critically summarized these issues. However, the quality of trial design and reporting is imperative for the adoption of trial findings in clinical practice as poorly designed or poorly reported trials can only lead to contradictory findings and confusion among front-line clinicians in their decision-making in the management of patients with COVID-19. Clinicians are eagerly looking for definitive answers as to what works and what does not, and they may not have sufficient time amid the COVID-19 emergency to critically appraise every single trial. Complete and accurate reporting is, therefore, invaluable to inform policy-makers and guide clinical decisions. Furthermore, the responsibility to ensure a greater balance between reporting of both benefits and harms lies with the authors and the journals publishing those trials. Although we acknowledge the limited availability of space in journals often leads to selective reporting of outcomes, it should not be an excuse since this can be easily overcome by the reporting of supplementary data. We hope that this review will encourage clinical researchers to better design, conduct, and report trials to uphold the principles of evidence-based medicine even amid a global health emergency such as COVID-19.

Strengths and limitations

The strengths of our review lie in the comprehensive literature search. We used multiple databases, such as PubMed and EMBASE, a clinical trial registry, and a COVID-19 specific database (Dimensions) to search for relevant clinical trials. Furthermore, we used standardized quality reporting tools and methods used in the synthesis of evidence, in which for research synthesis we followed standard PRISMA guidelines in searching for, selection of, inclusion of, and exclusion of studies, as well as for data extraction. We evaluated the harm reporting and risk of bias in the included trials using standardized CONSORT harm recommendations and the RoB 2 and ROBINS-I tools, respectively.

Although we used all the possible terms, free-text terms, and the MeSH terms to search for relevant HCQ/CQ clinical trials, the sensitivity of our strategy is still unknown. In addition, as with most studies examining the design and reporting methods, it is challenging particularly when the reporting of these aspects is incomplete. An example of this is the reporting of baseline disease activity in the HCQ/CQ trials, where it can be difficult to determine the methodology used as it was often not defined.

Implications of this research

There are implications of this work in designing and conducting randomized controlled trials for the treatment of COVID-19. This review suggests that basic requirements for designing and conducting randomized controlled trials should never be compromised, and a standardized protocol must be used and followed. It is also important that confounders should always be identified and adjusted. To uphold public trust in medical practice amid the COVID-19 pandemic, randomized controlled trials should be designed and reported more exhaustively, particularly when the effects of the specific treatment that could be potentially life-saving are being reported. The authors should always report all AEs experienced by the patients during a trial with optimal quality standards (CONSORT harm recommendations), where all AEs (instead of a mere selection) are explicitly described and provide details for patients who dropped out because of AEs. Definitions of baseline disease activity should always be provided (and appropriately adjusted if there are differences) to determine if a particular treatment is effective in the mild, moderate, severe, or critical stage of the disease.

Conclusion

Given the quality of evidence available, it is not possible to draw a meaningful conclusion on the effectiveness and the safety of CQ or HCQ for the treatment of patients with COVID-19. The quality of evidence should be carefully considered in the making of clinical and policy decisions, particularly during a pandemic. The importance of designing and reporting trials properly cannot be overemphasized for the synthesis of clinical evidence, and its dismissal in its entirety or partially amid pandemic crises could not only lead to a waste of invaluable health care resources but may also risk precious lives.

Conflict of interest

None declared.

Funding source

The authors of this review received no funding.

Ethics approval

Not required.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2020.09.1470.

References

Abd-Elsalam S, Esmail ES, Khalaf M, et al. Hydroxychloroquine in the treatment of COVID-19: a multicenter randomized controlled study. Am J Trop Med Hyg 2020;, doi:http://dx.doi.org/10.4269/ajtmh.20-0873.

Global Covid-19 Case Fatality Rates [Internet] UK: Centre for Evidence-Based Medicine [Accessed 26 May 2020]. Available from: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/.

What If Hydroxychloroquine Doesn't Work? What if it does? Right Now, We Don't Know [Internet] Statnews. [Accessed 26 May 2020]. Available from: https://www.statnews.com/2020/03/27/we-dont-know-hydroxychloroquine/.

What Do We Know About Hydroxychloroquine? [Internet] Euronews. [Accessed 26 May 2020]. Available from: https://www.euronews.com/2020/05/19/what-do-we-know-about-hydroxychloroquine-euronews-answers.

Borba MGS, Val FFA, Sampaio VS, et al. Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: a randomized clinical trial. JAMA Netw Open 2020;3(4)e208857.

Boulware DR, Pullen MF, Bangdiwala AS, et al. A randomized trial of hydroxychloroquine as postexposure prophylaxis for Covid-19. N Engl J Med 2020;383:517–25, doi:http://dx.doi.org/10.1056/NEJMoa2016638.

Cavalcanti AB, Zampieri FG, Rosa RG, et al. Hydroxychloroquine with or without azithromycin in mild-to-moderate Covid-19. N Engl J Med 2020;, doi:http://dx.doi.org/10.1056/NEJMoa2019014.

Chen J, Liu D, Liu L, et al. A pilot study of hydroxychloroquine in treatment of patients with moderate COVID-19. Zhejiang Da Xue Xue Bao Yi Xue Ban 2020a;49:215–9.

Chen Z, Hu J, Zhang Z, et al. Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial. medRxiv 2020b; 03.22.20040758.

Chen C, Lin Y, Chen T, et al. A multicenter, randomized, open-label, controlled trial to evaluate the efficacy and tolerability of hydroxychloroquine and a retrospective study in adult patients with mild to moderate coronavirus disease 2019 (COVID-19). medRxiv 2020c; Preprint, 2020.07.08.20148841.

Cornelius VR, Sauzet O, Williams JE, et al. Adverse event reporting in randomised controlled trials of neuropathic pain: considerations for future practice. Pain 2013;154:213–20.

Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020;369:m1985.

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4, doi:http://dx.doi.org/10.1016/S1473-3099(20)30120-1.

Edwards JE, McQuay HJ, Moore RA, et al. Reporting of adverse effects in clinical trials should be Improved. J Pain Symptom Manage 1999;18:427–37.

Esper R.B., da Silva R.S., Teiichi F., et al. Empirical treatment with hydroxychloroquine and azithromycin for suspected cases of COVID-19 followed-up by telemedicine. Prevent Senior Institute SP, Brazil, ed. São Paulo, 25. [Accessed 19 October 2020]. Available from https://static.poder360.com.br/2020/04/2020.04.15-journal-manuscript-final.pdf.

Ford I, Norrie J. Pragmatic trials. N Engl J Med. 2016;375:454–63.

Furtado RHM, Berwanger O, Fonseca HA, et al. Azithromycin in addition to standard of care versus standard of care alone in the treatment of patients admitted to the hospital with severe COVID-19 in Brazil (COALITION II): a randomised clinical trial. Lancet 2020;396:959–67, doi:http://dx.doi.org/10.1016/S0140-6736(20)31862-6.

Gao J, Tian Z, Yang X. Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends 2020;14(1):72–3.

Gautret P, Lagier JC, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents 2020;56:105949.

Hadi MA, McHugh GA, Conaghan PG. Quality of reporting of harms in randomised controlled trials of pharmacological interventions for rheumatoid arthritis: a systematic review. Evid Based Med 2017;22(5):170–7.

Hasan SS, Kow CS, Merchant H. Is it worth the wait? Should chloroquine or hydroxychloroquine be allowed for immediate use in CoViD-19?. Br J Pharm 2020;5(1), doi:http://dx.doi.org/10.5920/bjpharm.745.

Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

Horby P, Mafham M, Linsell L, et al. Effect of hydroxychloroquine in hospitalized patients with COVID-19: preliminary results from a multi-centre, randomized, controlled trial. medRxiv 2020; Preprint, 2020.07.15.20151852.

Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020a;395:497–506.

Huang M, Tang T, Pang P, et al. Treating COVID-19 with chloroquine. J Mol Cell Biol 2020b;12(4):322–5.

Ioannidis JP, Contopoulos-Ioannidis DG. Reporting of safety data from randomised trials. Lancet 1998;352:1752–3.

Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141 (10):781–8.

Knottnerus JA, Tugwell P. Methodological challenges in studying the COVID-19 pandemic crisis. J Clin Epidemiol 2020;121:A5–7.

Liu J, Cao R, Xu M, et al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov 2020;6:16.

Mitjà O, Corbacho-Monné M, Ubals M, et al. Hydroxychloroquine for early treatment of adults with mild Covid-19: a randomized-controlled trial. Clin Infect Dis 2020a;, doi:http://dx.doi.org/10.1093/cid/ciaa1009.

Mitjà O, Ubals M, Corbacho M, et al. A cluster-randomized trial of hydroxychloroquine as prevention of Covid-19 transmission and disease. medRxiv 2020b; Preprint, 2020.07.20.20157651.

Savarino A, Boelaert JR, Cassone A, Majori G, Cauda R. Effects of chloroquine on viral infections: an old drug against today's diseases?. Lancet Infect Dis 2003;3:722–7.

Skipper CP, Pastick KA, Engen NW, et al. Hydroxychloroquine in nonhospitalized adults with early COVID-19: a randomized trial. Ann Intern Med 2020;, doi:http://dx.doi.org/10.7326/M20-4207 M20-4207.

Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.

Tang W, Cao Z, Han M, et al. Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. BMJ 2020;369:m1849.

World Health Organization. Director-General's Remarks at the Media Briefing on 2019-nCoV on 11 February 2020. [Internet]. [Accessed 3 April 2020]. Available from:. World Health Organization; 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020.