Supplementary Methods

Participants

The patients enrolled in this study were with a confirmed diagnosis of Coronavirus disease 2019 (COVID-19) in accordance with the Diagnosis and Treatment Protocol for COVID-19 (sixth and seventh edition) issued by National Health Commission in China.

Trial design

We conducted two multicenter, open-label, randomized trials at six sites (Peking University Third Hospital, Peking University People's Hospital, Peking University First Hospital, Haidian Section of Peking University Third Hospital, Tongji Hospital Affiliated to Tongji Medical college HUST (Wuhan), The First Affiliated Hospital of Nanchang University) in China from February 19, 2020 to March 3, 2020.

Eligible patients randomly assigned in a 1:1 ratio to orally receive hydroxychloroquine (600 mg b.i.d. on the first day followed by 200 mg b.i.d. in the last four days) plus standard care or chloroquine (500 mg b.i.d. for the first three days followed by 250 mg b.i.d. in the last two days) plus standard care (SC) for 5 days. Vital signs, clinical symptoms, and adverse events were captured on day 0 to day 10, day 14, day 28, and on the day of discharge/death. Several nasopharyngeal or oropharyngeal swab samples were collected for testing Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on day 0, day 5 (±2) and additionally upon doctors' requests. Other laboratory tests such as interleukins, ferritin, TNF-α and routine blood test were scheduled on day 1, day 3, day 5, day 10 and the day of discharge (Appendix Figure. 1-2).

The real-world study (RWS) database consisted of data from all the hospitalized patients in the hospitals where the randomized controlled trial (RCT) was carried out but not under hydroxychloroquine and chloroquine treatments. Clinical data were recorded on paper-based report forms and then double-entered into the electronic database and validated.

Inclusion and exclusion criteria of clinical study

Patients who met the following inclusion criteria were included in the study of moderate COVID-19: male and nonpregnant female; SARS-CoV-2 tested positive by a quantitative reverse-transcriptase polymerase-chain-reaction (qRT-PCR) assay or IgM / IgG Enzyme-Linked ImmunoSorbent Assay (ELISA); age ≥ 18 years, with confirmed moderate COVID-19 which was defined as having at least one manifestation as follows: 1) mild clinical symptoms and no imaging findings of pneumonia; 2) fever and respiratory symptoms, imaging findings of pneumonia. Patients who met the following inclusion criteria were included in the study of severe COVID-19: male and nonpregnant female; SARS-CoV-2 tested positive by qRT-PCR assay or IgM / IgG ELISA; age ≥ 18 and ≤ 75 years; with confirmed severe COVID-19 which was defined as having at least one manifestation as follows: 1) respiratory distress, respiratory frequency (RR) ≥30 /min; 2) oxygen saturation (SaO2) of 93% or less while they were in resting state; 3) a ratio of the partial pressure of oxygen(PaO2) to the fraction of inspired oxygen (FiO2)(PaO2:FiO2) at or below 300 mmHg; 4) respiratory failure requiring mechanical ventilation.

Patients with the following conditions were excluded from our studies: a known allergy to 4-aminoquinolines drugs or analogs; any comorbidities including hematological diseases, mental illnesses, retinopathy, maculopathy, severe liver diseases, hypokalemia, hypomagnesemia, pancreatitis, severe renal impairment, and glucose-6-phosphate dehydrogenase deficiency; a history of acute myocardial infarction or other conditions were assessed ineligible (The complete eligibility criterion are provided in protocols).

Matching Method of RCT and RWS

Patients with RWS were hospitalized at the same time as patients with RCT who met the inclusion criteria. To account for potential confounders, we calculated propensity scores using logistic regression, which estimated the probability of treatment with chloroquine and hydroxychloroquine, given 5 baseline characteristics (age, gender, course of disease, baseline of absolute value of Lymphocyte and C-reactive protein). What needs to be explained is the matching method of disease course, the RCT group

was selected from symptom onset time to enrollment time and the RWS group was symptom onset time to hospitalization time. We then matched patients treated with Chloroquine or Hydroxychloroquine in RCT and conventionally treated patients in a 1:1 ratio according to propensity score, thus creating a distinct matched analysis set. Matching was performed with the use of the nearest-neighbor matching algorithm (caliper width=0.1). We assessed the balance of covariates that was achieved from matching by evaluating standardized differences between groups. We considered covariates with a standardized difference (SMD) of less than 20% to be acceptable balanced.

Outcome Measures

The primary outcome in the study of moderate COVID-19 was time to clinical recovery. The clinical recovery was defined as a total relief of the following COVID-19 related signs and sustained for more than 72 hours: (1) body temperature was back to normal (underarm temperature ≤36.9℃, oral temperature ≤37.2℃), (2) respiratory rate was ≤24 /min, (3) oxygen saturation was > 94%, (4) remission of cough.

The primary outcome in the study of severe COVID-19 was time to clinical improvement, the clinical improvement was defined as an improvement of one level using a 6-level ordinal scale (Appendix Table 1).

The secondary outcomes in the study of moderate COVID-19 included: (1) 28-day mortality, (2) time to virological clearance, (3) frequency of dyspnea progression, (4) time to the body temperature returned to normal, (5) time to the cough improved to mild or absent, (6) time to dyspnea reduced or absent, (7) duration of oxygen therapy or non-invasive ventilation.

The secondary outcomes in the study of severe COVID-19 included: (1) the percentage of patients reporting each severity rating on a 6-level ordinal scale on day 7, day 14 and day 28, (2) 28-day mortality, (3) duration of invasive mechanical ventilation (IMV), (4) duration of oxygen therapy, (5) length of hospital stay (from randomization), and (6) time to virological clearance.

Other outcomes of the studies of moderate and severe COVID-19 included results of the routine blood test, and levels of procalcitonin, c-reactive protein, ferritin, Interleukin (IL)-1, IL-2, IL-6, IL-8, IL-10, and TNF-α. Outcomes of the studies of moderate and severe COVID-19 consisted of the length of hospital stay in moderate patients, virological clearance rate on day 5 and CT results assessed by CT scores on day 10 (Bernheim et al., 2020; Fujimoto et al., 2012).

Safety outcomes of the studies of moderate and severe COVID-19 included the incidence and severity of adverse events, abnormalities of clinical symptoms, and abnormalities of laboratory results. Adverse events were classified following the National Cancer Institute Common Terminology Criteria for Adverse Events, version 5.0.

CT scan scoring system

All CT images were reviewed by two physicians with approximately 10 years' experience each (SL and SYL) independently, using a modified scoring system. The lung was divided into six zones (upper, middle, and lower on both sides) by the level of the tracheal carina and the level of the inferior pulmonary vein bilaterally on CT. The observers recognized ground-glass opacity (GGO), consolidation and crazy-paving pattern following Fleischner Society definitions. Bronchiectasis, cavity, pleural effusion, etc., were not included in CT reading and analysis because of low incidence. The reviewers evaluated the extent of the targeted patterns and overall affected lung parenchyma for each zone, using Likert scale (0=absent; 1=1–25%; 2=26–50%; 3=51–75%; 4=76–100%). Thus, GGO score, consolidation score, and overall lung involvement score were sum of 6 zones ranging from 0–24. For crazy-paving pattern, it was only coded as absent or present (0 or 1) for each zone and therefore ranging from 0-6.

Statistical Analysis

The full analysis set (FAS) consisted of patients successfully enrolled in our clinical trials and received at least one dose of medication (hydroxychloroquine or

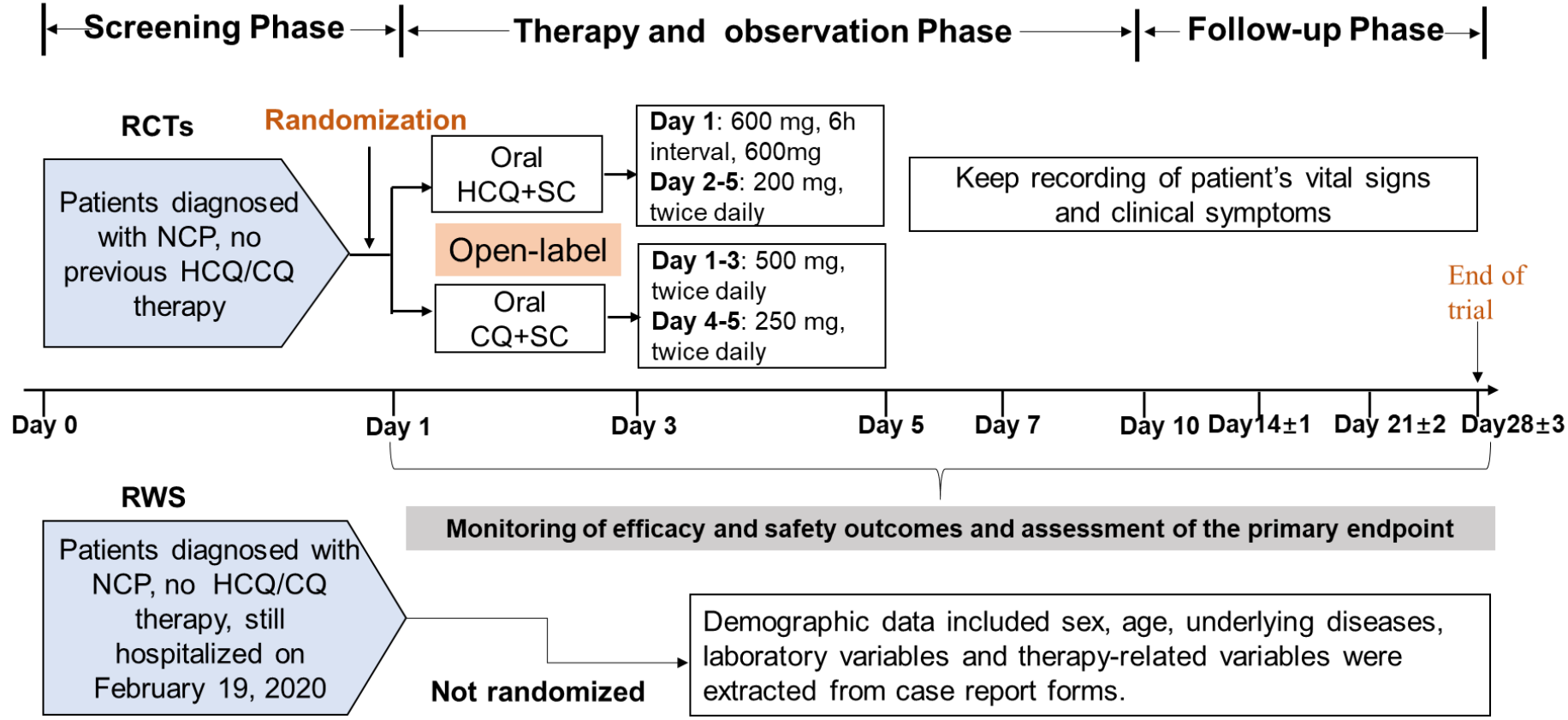
chloroquine). The efficacy analysis set (ES) only excluded patients who did not meet inclusion criteria and were mistakenly included in FAS at the beginning. The safety analysis set (SS) did not exclude patients who did not meet inclusion criteria and were mistakenly included in FAS initially.

Kaplan-Meier method was used to estimate the median of TTCR, TTCI and other time to events, and their 95% confidence intervals (CIs) were estimated by the Brookmeyer-Crowley method. Hazard ratios (HRs) with 95% CIs were calculated by the Cox proportional-hazards regression model. The log-rank test was used to compare the differences between hydroxychloroquine and chloroquine.

Continuous data were described by means and standard deviation. Independent t-test and Kruskal-Wallis rank sum test were carried out to compare the differences of continuous data between hydroxychloroquine and chloroquine groups depending on the distribution of data. Nominal data were described by counts and percentages. Chi-square test was used for comparisons of nominal data between groups.

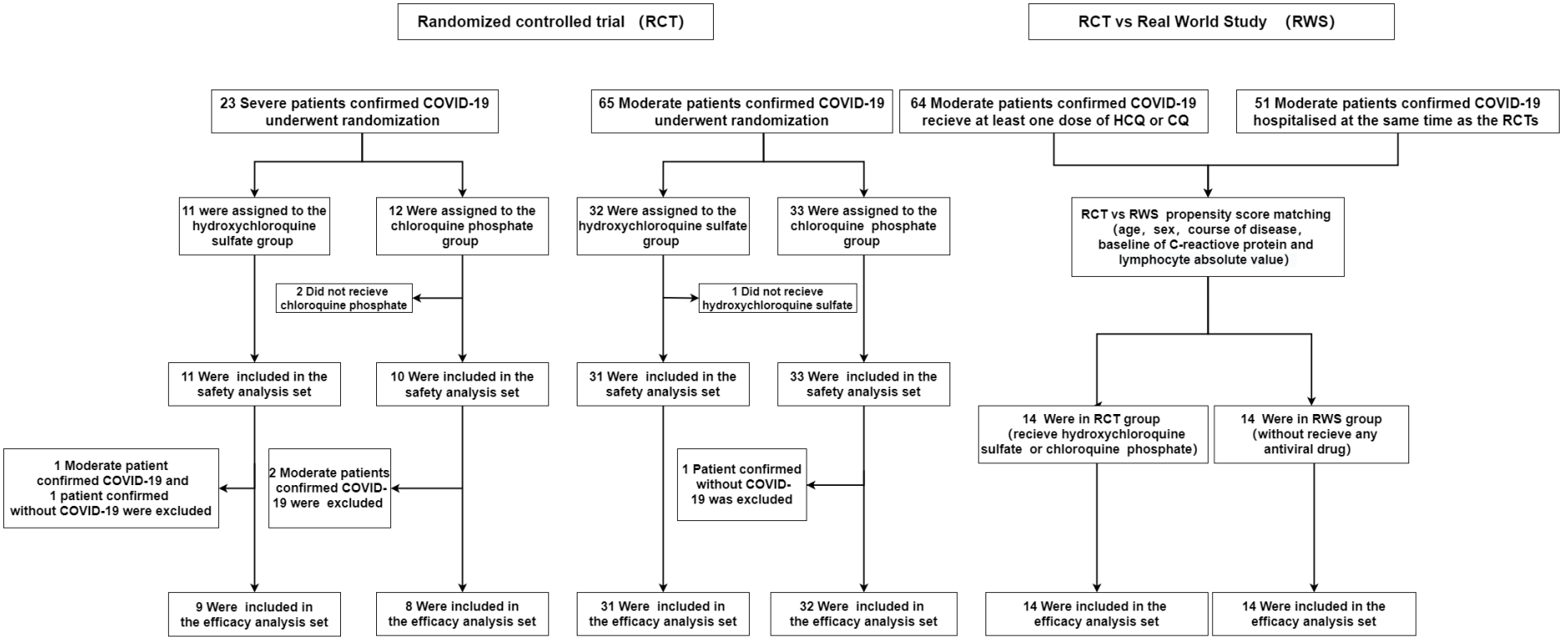
We matched the moderate patients in the hydroxychloroquine and chloroquine groups with patients in the RWS database in a 1: 1 ratio using the propensity score matching methods. All endpoints and indicators were statistically analyzed in the same method as RCT. Statistical analyses were conducted with SAS software, version 9.4 (SAS Institute).

Appendix Figure 1. Trial design



RCT：Randomized control trial; RWS: Real world study; HCQ: Hydroxychloroquine; CQ: chloroquine; SC: standard care based on National guidance

Appendix Figure 2. Patient disposition



Appendix Table 1. Clinical status scale

Grade	Content
1	Discharged
2	Hospitalized, not requiring oxygen therapy
3	Hospitalized and requiring oxygen therapy (excluding items in 4-5)
4	Hospitalized and requiring noninvasive ventilation, high flow oxygen therapy or both
5	Hospitalized and requiring ECMO, mechanical ventilation or both
6	Death

References

1. Bernheim A, Mei X, Huang M, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology* 2020; **295**(3): 200463.
2. Fujimoto K, Taniguchi H, Johkoh T, et al. Acute exacerbation of idiopathic pulmonary fibrosis: high-resolution CT scores predict mortality. *Eur Radiol* 2012; **22**(1): 83-92.