Clinical Design for Phase II/III Clinical Trials for Testing Therapeutic Interventions in COVID-19 Patients

Shesh Rai (shesh.rai@louisville.edu)  
University of Louisville  
https://orcid.org/0000-0002-8377-353X

Chen Qian  
University of Louisville

Jianmin Pan  
University of Louisville

Anand Seth  
SK Patent Associates

Deo Kumar Srivastava  
Saint Jude Children's Research Hospital

Aruni Bhatnagar  
University of Louisville

Research article

Keywords: composite outcomes, COVID-19, efficacy, hospitalization, interim analysis, intubation, power, mortality, sample size, toxicity monitoring

DOI: https://doi.org/10.21203/rs.3.rs-30558/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background

Researchers around the world are urgently conducting clinical trials to develop new treatments for reducing mortality and morbidity related to COVID-19. However, due to unknown features of the disease and complexity of the patient population, traditional trial designs may not be optimal in such patients. We propose two independent clinical trials designs based on careful grouping of the expected characteristics of patient population. This could serve as a useful guide for researchers designing COVID-19 related Phase II/III trials.

Methods

Using the commonly utilized World Health Organization ordinal scale on patient status, we classify patients into three risk groups. In this approach, patients in Stages 3, 4 and 5 are categorized as the intermediate-risk group while patients in Stages 6 and 7 are categorized as the high-risk group. To ensure that an intervention, if deemed efficacious, is promptly made available to vulnerable patients, we propose a group sequential design with two interim analyses along with a final analysis and a toxicity monitoring rule for the intermediate-risk group. For the high-risk group, we propose a group sequential design with two interim analyses without toxicity monitoring.

Results

Based on different response rates, effect sizes, and power, required sample size and toxicity boundaries are calculated for each scenario. Sample size requirements for the designs with interim analyses are only marginally greater than the ones without. In addition, for both the intermediate-risk group and the high-risk group, conducting two interim analyses have identical required sample size compared with just one interim analysis. Additional issues that could potentially impact the trial are discussed.

Conclusions

We recommend using composite endpoints, with binary outcome for those in Stages 3, 4 and 5 with a power of 90% to detect an improvement of 20% in response rate, and 30 days mortality rate outcome for those in Stages 6 and 7 with a power of 90% to detect 15% (effect size) reduced mortality rate, in the COVID-19 trial design. For the intermediate-risk group, two interim analyses for efficacy evaluation along with toxicity monitoring are encouraged. For the high-risk group, two interim analyses without toxicity monitoring is advised.

Background

The ongoing COVID-19 (SARS-COV-2 infection) crisis is an unprecedented public health challenge as there are no clinically-proven interventions with substantial evidence that can effectively manage the infection. To meet this challenge, researchers around the world have been working diligently on developing new treatment plans or drugs. Several clinical interventions including those that involve the use of convalescent plasma, a combination of existing drugs, or repurposing drugs, such as Remdesivir, have either entered the clinical trial phase or completed small size studies (a partial list of drugs/therapies used for COVID-19 treatment is given in the appendix). However, in general, many clinical trials have failed and continue to fail due to various reasons, including lack or inappropriate control group and/or rigorous statistical designs. Key reasons for the failure of the approaches attempted so far are due to the uniqueness of this patient population compared with clinical trials for other patient populations and speed at which such trials must be
conducted. With patients showing up with a variety of characteristics and fast changing status, it is difficult to recruit and conduct an appropriate trial that could best show the effectiveness of an intervention. Many factors, such as patient status, age, gender, race, co-morbidity, etc., can affect the design or the outcome of the trial and therefore they must be taken into consideration as stratification factors in designing the study. Consideration of a wide set of such factors makes it challenging to develop a design that minimizes the imbalance in treatment allocation with respect to stratification factors while ensuring that the number of strata remain manageable.

The purpose of this article is to propose effective statistical designs for COVID-19 clinical trials. Two parallel clinical trials design with respect to different patient risk groups are described. Issues and limitations are discussed. Required sample size in each arm under different scenarios along with toxicity boundaries are calculated and presented in a tabular form for ease of implementation and to inform clinical trial design considerations.

Methods

A flowchart that illustrates the overall design of both such trials is shown in Fig. 1.

World Health Organization Ordinal Scale

To best describe the clinical status of the patient, we adopted the World Health Organization’s ordinal scale. Similar scale such as the 7-category ordinal scale was used in a previous trial (1). This scale has proven to be an effective way of describing the severity of illness as well as for assessing clinical outcomes in hospitalized patients. This 7-category ordinal scale has been used recently by Wang (2) to categorize outcomes in patients hospitalized with seasonal influenza infection. The authors found the scale to be a useful in capturing a broad range of clinical states as well as tracking patient’s status change. Although the ordinal scale is useful for patient classification, it is difficult to design a trial based on every stage. Therefore, we used a composite endpoint that combines similar groups together. The details of this World Health Organization ordinal scale are given in Table 1. Stage 0 is not included here since we are not interested in the uninfected population.

| Stage | Condition                                      |
|-------|-----------------------------------------------|
| 8     | Death                                         |
| 7     | Ventilation with Additional Organ Support (ECMO*) |
| 6     | Intubation and Mechanical Ventilation         |
| 5     | Non-invasive Ventilation or High-flow Oxygen   |
| 4     | Oxygen by Mask or Nasal Prongs                |
| 3     | Hospitalized; No Oxygen Therapy               |
| 2     | Limitation of Activities                      |
| 1     | No Limitation of Activities                   |

*ECMO: extracorporeal membrane oxygenation

Composite Endpoints
A composite endpoint is a single measure of effect, based on a combination of individual components endpoints. Composite endpoints have high utility in evaluating the efficacy of therapeutic interventions that could individually or concurrently alter several different symptoms or outcomes. For example, in Type II diabetics, a drug may affect HbA1C (hemoglobin A1C), body weight, and systolic blood pressure (3). Often, the frequency of events in individual components of a composite endpoint may be low, so several components are combined to assess the overall efficacy of an intervention. However, each component of a composite endpoint should be clinically meaningful. Ideally, all component should be weighted equally, but this is rarely possible, therefore the relative importance of the components may have to be determined by the frequency of occurrence of the component outcomes. For instance, in cardiovascular trials, death, myocardial infarction (MI), stroke, coronary revascularization and hospitalization for angina are commonly combined, although fatal and non-fatal events are not be treated as the same. In a recent study, patients and clinical trial authors, when asked to assign “spending weights” to five events - death, myocardial infarction, stroke, coronary revascularization and hospitalization for angina, assigned different weights to each of these components (4).

In trials where death is a possible outcome, it is often included as a part of a composite outcome to capture the overall efficacy of the treatment. In this regard, the statistical theory of competing risk provides support to the notion of including mortality as a component of a composite outcome (5). In a review of 14 journals between January 2000 to January 2007, of the 1231 cardiovascular trials, 37% used composite endpoints, and 98% of these trials included mortality as a component (6).

For COVID-19 patients, it is nearly impossible to come up with a composite endpoint that could include all stages. Therefore, further classification is needed. Based on the World Health Organization ordinal scale, patients are separated into three group: High-Risk Group (Stages 6 and 7), Intermediate-Risk Group (Stages 3, 4, and 5), Low-Risk Group (Stages 1 and 2). Patients in intermediate-risk group are treated in a similar way while the high-risk group patients need some innovative and aggressive treatment. Two independent clinical trials with two different designs are proposed for the intermediate-risk group and the high-risk group. Note that if a patient in the intermediate-risk group could not recover and progresses to the high-risk group, then that patient could be eligible for the high-risk group trial.

In our study design considerations, we looked at the 15th day to determine the patient’s status because 14-days period is the mean number of patient recovery, or as a complete cycle of treatment, as shown in Cao (1). Other useful values (median days) adopted from the same manuscript are:

- **Time to Clinical Improvement** — 16 Days
- **Intensive Care Unit (ICU) Length of Stay** — 10 Days
- **Duration of Invasive Mechanical Ventilation (IMV)** — 5 Days
- **Days on Oxygen Support** — 13 Days
- **Length of Hospitalization** — 15 Days.

The “time to clinical improvement” is defined by Cao (1) as the time from randomization to either an improvement of two points on a 7-category ordinal scale or discharge from the hospital, whichever came first.

**Design for Intermediate -Risk Group:**

**Outcome Variables**
Since a larger number of patients are expected in the intermediate-risk group, it is feasible to use binary endpoints (success or failure).

We define $Y$ as the primary outcome variable. Let $Y = 1$ indicate the success outcome if the patient is discharged from the hospital by the 15th day. Let $Y = 0$ indicate the failure if the patient is not discharged from the hospital by the 15th day or dead. Then $Y = 1$ is the success with probability $P$ and $Y = 0$ is the failure with probability $1-P$. Accordingly, we calculated results based on the improvement of response rate from 40% in the standard arm to various rates (50%, 55%, 60%, 65%, 70%, 75%, 80%) in the treatment arm.

Some secondary outcome variables might also be considered. For example, the change in viral load or biomarkers of inflammation such as ferritin or IL-6, time to reduced viral load, or the number of event-free days in the hospital (event-free survival).

Stratification

For ethical reason, group sequential designs are recommended in the current setting. Since many factors could impact the outcome, stratified randomization is more suitable. In previous work Srivastava (7) found that, with several factors appearing to affect the primary outcome of interest with their true distributions being unknown, or the possibility of causing heterogeneous treatment response among individuals in a group with unknown effect size, stratified randomization approach offered consistently better results if the effect size can be assumed to be similar within each stratum.

Factors, such as age, race, sex, co-morbidity and viral load, which might impact the primary outcomes should be addressed by stratification. However, choosing the right factor for stratification is critically important. Many different issues need to be considered when choosing stratification factors. Based on the current clinical experience showing a strong dependence of COVID-19 outcomes on age, sex and diabetes, obesity and hypertension (8, 9), we consider four such factors for the intermediate-risk group: patient stage, at least one cardiovascular disease risk factor among obesity, hypertension and diabetes (Yes/No), age (< 60 and ≥ 60 years), and gender (Male/Female).

For the intermediate risk group, we further classified patients in the three stages, those in Stages 3 and 4 and those in Stage 5 (essentially classifying patients into those who are not in ICU vs. those who are in ICU) and grouping them into two groups. This is suggested to minimize the number of strata for randomization while ensuring that the patients within each stratum are relatively homogenous. All these factors are readily identifiable; however, for defining metabolic syndrome status, it may be necessary to include other factors that are representative of a patient’s health condition. Alternatively, composite risk scores such as the Framingham, Reynolds, or GRACE risk scores may be used. Data to calculate cardiovascular risk score and/or obesity may be readily available, as patients are usually weighed and their blood pressure, cholesterol status, and diabetes are known upon admission to most hospitals. Although age is usually included in risk factor score, it could also be considered as a separate variable when the risk score cannot be calculated. Assuming that no risk scores are available, in our recommended design, we define two age groups: less than 60 years of age and greater than or equal to 60 years of age. It is generally known that patients in the intermediate-risk group are mostly elderly. Patients less than 50 years of age only count a very small percent of the patients admitted to a hospital. Therefore, the cutoff line at 60 years of age is selected to have balanced strata. Moreover, in view of studies showing that the recovery rate for males is lower than of females (10), gender should also be considered. With these four factors for stratification, there would be a total of 16 strata in the design, which makes the trial design somewhat manageable. Race is not explicitly considered, as there is no indication yet of race-dependent variations in outcome, independent of pre-existing disease burden.

Interim Analysis
For the intermediate-risk group, two interim analyses are recommended. Results are presented for no interim analysis, one interim analysis and two interim analyses. Because the virus is life threatening, it is important to ascertain the efficacy of the intervention as early as possible and make the drug available to this patient population as soon as possible. Without interim analysis, researchers would know the outcome of the trials only after all patients have been enrolled. If one choose to perform one interim analysis, when 50% of patients are enrolled, then using G-rho spending function with rho equals 2.3, one would stop the trial at the interim evaluation if the p-value of the test for comparing the two groups is less than 0.01 (11, 12). Otherwise, the trial should continue, and the final analysis will be conducted, and the efficacy of the treatment should be declared only if when the p-value is less than 0.046. However, due to the insidious nature of this infection, waiting until 50% of patients enrolled to find out the result may still not be aggressive enough. Therefore, to fast track the process and make sure that the drug can be made available to those who need it urgently, we recommend two interim analyses, with first interim analysis to be performed when one-third of total patient population has been enrolled and evaluated (p = 0.002), second being performed when two-third of the patients are enrolled and evaluated (p = 0.014), and the final analysis when all patients are enrolled (p = 0.046). Rho equals three is used in the G-rho spending function (11, 12). The choice of Rho was based on the consideration that we need to make the drug available to the patients quickly but we need to make sure that the trial is stopped early only if we have strong evidence the drug is effective and this is the reason why we chose the p-values cut-offs at interim evaluations to be somewhat conservative (making sure that there is strong evidence in favor of the drug and avoid false positive findings). To explain with an example, assume the overall sample size is 237 in which 79 belongs to the standard care arm and the rest 158 belongs to the treatment arm. At the first interim analysis, we have 53 patients (one third of 158). If p < 0.002, then there is strong evidence to declare that the intervention is working, and the trial should stop right away. With this design, researchers can find out early whether the intervention works, or stop if it is causing unacceptable harm to patients by monitoring toxicities. Considering some unforeseen reasons, the sample size should be increased by approximately 5% with resulting total n = 249.

**Group Ratio**

We calculate here sample sizes for both 1:1 and 1:2 randomizations for the intermediate-risk group. However, patients enrolled in treatment arm may be the same or twice the number of patients enrolled in the standard treatment arm. The choice of group ratio depends on the efficacy of the intervention in the pilot studies. If it is a new intervention that has not been approved by the FDA, then 1:1 randomization with block size of 4 is recommended in consideration of patient’ safety. If it is an approved procedure or drug with some preliminary data on efficacy with known toxicity profile, then 1:2 randomization with block size of 6 is recommended to ensure that if the drug is effective more patients get the advantage of being treated on the more efficacious arm.

**Toxicity Monitoring**

Toxicity monitoring is challenging, but necessary. For the intermediate-risk group, a toxicity rate of 25% is recommended due to the urgent need of drugs for treatment. In other words, if intervention provides even minimal improvement, then it should still be considered as there is currently no known drug that is 100% effective against COVID-19. Also note that patient may die during the treatment due to other causes. We recommend monitoring for those toxicities (Common Terminology Criteria for Adverse Events (CTCAE) Grades > 2) that are related, possibly related and probably related to the drug, evaluated by the Data and Safety Monitoring Board (DSMB) (13).

**Design for High-Risk Group:**

**Outcome Variables**
The number of patients in the high-risk group at each health care facility is likely to be small. Using survival as an endpoint may not be ideal because the follow-up is short, and it may require a long time to enroll all patients and there would hardly be any right censoring. In other words, very few patients will survive pass the outcome evaluation time, thereby making survival as an endpoint to ineffective. Therefore, in our design we focused on reducing 30 days mortality rate.

We define the 30 days mortality as the primary outcome. Let \( Y = 1 \) indicates death of patient within 30 days (failure), and \( Y = 0 \) represents a person still alive on the 30th day (success). Accordingly, we calculated results based on the reduction of mortality rate from 80% or 70% in the standard arm to various rates (70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%) in the treatment arm.

**Stratification**

For the high-risk group, stratification is recommended as sample size is satisfied when using 30 days mortality as primary outcome. Enrolling sufficient patients within a given timeframe should not be an issue assuming a trial to be a multi-center trial.

For stratification, similar factors as discussed for the intermediate-risk group design are recommended, with some modification. We consider three factors for the high-risk group: one cardiovascular disease risk factor among diabetes, hypertension, and obesity (Yes/No), age (< 65 and \( \geq 65 \) years), and gender (male/female). The reason of selecting 65 years age as a cutoff point is that in a recent study of COVID-19, mortality rate for those who received mechanical ventilation in the age of 18 to 65 years was 76.4%, and for those over 65 years of age, the mortality rate was 97.2% (14). In the high-risk group, we did not further classify patients based on their stages (Stage 6 and 7) as we have done in the intermediate-risk group. The reason is that the number of patients in stage 7 is likely to be small, and it is not possible to stratify based on these two stages. However, technically it is ideal to stratify patients evenly in every arm based on their stages, but that is not achievable in this case. Since we have chosen other factors for stratification, if extreme bias occurs, then stage 7 patients should be dropped, and researchers should only perform analysis on stage 6 patients with 80% power.

**Interim Analysis**

For the high-risk group, two interim analyses are recommended. The reason is that the mortality rate in these patients is high and they need some innovative treatments. For example, convalescent plasma therapy has been widely attempted among the high-risk group. However, the levels of neutralizing antibodies in specific plasma preparation are likely to vary, leading to variable outcomes. Therefore, if during two interim analyses patients respond better to plasma with specific antibody titers, then the rest of the patients could be moved to the higher quality plasma quickly, so they have a higher chance of survival. In this design, first interim analysis is to be performed when one-third of total patient population has been enrolled, completed 30 days, and evaluated (\( p = 0.002 \)), second being performed when two-third of the patients are enrolled, completed 30 days, and evaluated (\( p = 0.014 \)), and the final analysis when all patients are enrolled and completed 30 days (\( p = 0.046 \)). Rho equals two is used in the G-rho spending function based on the consideration of being more conservative in interim analyses to ensure that the treatment is efficacious and avoid the chances of falsely declaring the treatment to be efficacious, which could mean huge losses in terms of resources invested and loss of lives. Sample sizes for one interim analysis are also calculated and are given in Tables 8 and 9 in the appendix for reference.
### Table 8
Required Sample Size for High-Risk Group Patients with One Interim Analysis.

| Power = 80% | Power = 90% |
|-------------|-------------|
| P1          | P0 = 80%    | P0 = 70%    | P0 = 80%    | P0 = 70%    |
| 70%         | 467         | NA          | 645         | NA          |
| 65%         | 218         | 2206        | 301         | 3051        |
| 60%         | 127         | 568         | 175         | 785         |
| 55%         | 83          | 257         | 114         | 355         |
| 50%         | 58          | 145         | 80          | 201         |
| 45%         | 42          | 93          | 58          | 128         |
| 40%         | 32          | 64          | 44          | 88          |
| 35%         | NA          | 46          | NA          | 63          |

P0: 30 days mortality rate in the standard arm. P1: 30 days mortality rate in the treatment arm.

N1: sample size for the standard care arm. N2: sample size for the treatment arm.

Probability of rejection at each look: 1st look $p < 0.010$, final look $p < 0.046$.

1:1 randomization; $\rho = 2.3$. NA = Not Applicable

### Table 9
Required Sample Size for High-Risk Group Patients with One Interim Analysis.

| Power = 80% | Power = 90% |
|-------------|-------------|
| P0 = 80%    | P0 = 70%    | P0 = 80%    | P0 = 70%    |
| P1          | N1 | N2 | Total | N1 | N2 | Total | N1 | N2 | Total |
| 70%         | 168 | 336 | 504  | NA | NA | NA   | 231 | 462 | 693  |
| 65%         | 77  | 154 | 231  | 817 | 1634| 2451 | 107 | 214 | 321  |
| 60%         | 45  | 90  | 135  | 208 | 416 | 624  | 62  | 124 | 186  |
| 55%         | 29  | 58  | 87   | 94  | 188 | 282  | 40  | 80  | 120  |
| 50%         | 20  | 40  | 60   | 53  | 106 | 159  | 28  | 56  | 84   |
| 45%         | 15  | 30  | 45   | 34  | 68  | 102  | 21  | 42  | 63   |
| 40%         | 12  | 24  | 36   | 24  | 48  | 72   | 16  | 32  | 48   |
| 35%         | NA  | NA  | NA   | 17  | 34  | 51   | NA  | NA  | NA   |

P0: 30 days mortality rate in the standard arm. P1: 30 days mortality rate in the treatment arm.

N1: sample size for the standard care arm. N2: sample size for the treatment arm.

Probability of rejection at each look: 1st look $p < 0.010$, final look $p < 0.046$.

1:2 randomization; $\rho = 2.3$. NA = Not Applicable

**Group Ratio**
For patients in the high-risk group, 1:2 randomization is recommended, because such patients are in danger and possibly have failed other treatments. Hence, they should be treated with whatever intervention available to improve their chances of survival. In addition, sample size is large enough to handle the 1:2 treatment allocation ratio when expecting a reduction of mortality rate from 70–55%. Estimated sample sizes for 1:1 randomization are also calculated and provided in the appendix for reference.

**Toxicity Monitoring**

For the high-risk group, no toxicity monitoring is necessary since mortality rate (between 70–97% has been reported across many health care facilities. With such a high death rate, it is not necessary to look at the toxicity level. Any intervention that could increase the chances of saving a patient should be utilized, regardless of treatable toxicities. In addition, two interim analyses are built in our design to help stop the trial early if any harmful events are detected.

**Results**

Result for Intermediate-Risk Group:

**Sample size calculation**

In the intermediate-risk group, we assume the baseline success rate is 40% in the standard arm (Standard Care) and increased success rate in the treatment group. In this section, the design with two built in interim analyses is discussed. Tables with required sample size for no or one interim analysis are provided in the appendix. All values are calculated with one-sided tests and using an un-pooled variance estimate. *EAST* software was used for sample size calculation (12).

Table 2 shows the required sample size for design with two interim analyses. Ideally, improvement of 20% of response rate with 90% power is recommended. For 1:1 randomization, each arm requires 105 patients. In 1:2 randomization, the standard care arm requires 79 patients while the treatment arm requires 158 patients. Comparing this table with Tables 3 and 4 in the appendix, sample size did not increase much from those with no or one interim analysis. Therefore, two interim analyses are recommended since it does not require lots of additional patients. We suggest inflating the sample size from the table by approximately 5% to account for the loss of information (such as dropout).
### Table 2
Required Sample Size for Intermediate-Risk Group Patients with Two Interim Analyses.

| Effect size | $\alpha = 0.05$ with 1:1 Group Ratio | $\alpha = 0.05$ with 1:2 Group Ratio |
|-------------|-------------------------------------|-------------------------------------|
|             | Power 80% | 90% | 80% | 90% | Power 80% | 90% | 80% | 90% |
| 10%         | N1 309    | 428 | 231 | 319 | N2 309    | 428 | 462 | 638 |
|             | Total 618 | 856 | 693 | 957 |             |     |     |     |
| 15%         | N1 137    | 189 | 102 | 141 | N2 137    | 189 | 204 | 282 |
|             | Total 274 | 378 | 306 | 423 |             |     |     |     |
| 20%         | N1 76     | 105 | 57  | 79  | N2 76     | 105 | 114 | 158 |
|             | Total 152 | 210 | 171 | 237 |             |     |     |     |
| 25%         | N1 48     | 66  | 36  | 50  | N2 48     | 66  | 72  | 100 |
|             | Total 96  | 132 | 108 | 150 |             |     |     |     |
| 30%         | N1 32     | 44  | 25  | 34  | N2 32     | 44  | 50  | 68  |
|             | Total 64  | 88  | 75  | 102 |             |     |     |     |
| 35%         | N1 22     | 31  | 18  | 24  | N2 22     | 31  | 36  | 48  |
|             | Total 44  | 62  | 54  | 72  |             |     |     |     |
| 40%         | N1 16     | 22  | 13  | 18  | N2 16     | 22  | 26  | 36  |
|             | Total 32  | 44  | 39  | 54  |             |     |     |     |

N1: sample size for the standard care arm. N2: sample size for the treatment arm.

Response rate = 40%.

Probability of rejection at each look: 1st look $p < 0.002$, 2nd look $p < 0.014$, final look $p < 0.046$.

**Bold** indicates recommended sample size with suggested parameters.
| Effect Size | Power | α = 0.05 with 1:1 Group Ratio | α = 0.05 with 1:2 Group Ratio |
|-------------|-------|-------------------------------|-------------------------------|
|             | 80%   | 90%                           | 80%                           | 90%                           |
| 10%         |       |                               |                               |
| N1          | 303   | 420                           | 226                           | 313                           |
| N2          | 303   | 420                           | 452                           | 626                           |
| Total       | 606   | 840                           | 678                           | 939                           |
| 15%         |       |                               |                               |
| N1          | 134   | 186                           | 100                           | 139                           |
| N2          | 134   | 186                           | 200                           | 278                           |
| Total       | 268   | 372                           | 300                           | 417                           |
| 20%         |       |                               |                               |
| N1          | 75    | 103                           | 56                            | 78                            |
| N2          | 75    | 103                           | 112                           | 156                           |
| Total       | 150   | 206                           | 168                           | 234                           |
| 25%         |       |                               |                               |
| N1          | 47    | 65                            | 35                            | 49                            |
| N2          | 47    | 65                            | 70                            | 98                            |
| Total       | 94    | 130                           | 105                           | 147                           |
| 30%         |       |                               |                               |
| N1          | 31    | 43                            | 24                            | 33                            |
| N2          | 31    | 43                            | 48                            | 66                            |
| Total       | 62    | 86                            | 72                            | 99                            |
| 35%         |       |                               |                               |
| N1          | 22    | 30                            | 17                            | 24                            |
| N2          | 22    | 30                            | 34                            | 48                            |
| Total       | 44    | 60                            | 51                            | 72                            |
| 40%         |       |                               |                               |
| N1          | 16    | 22                            | 13                            | 18                            |
| N2          | 16    | 22                            | 26                            | 36                            |
| Total       | 32    | 44                            | 39                            | 54                            |

N1: sample size for the standard care arm. N2: sample size for the treatment arm.  
Response rate = 40%.
Table 4
Required Sample Size for Intermediate-Risk Group Patients with One Interim Analysis.

| Effect Size | α = 0.05 with 1:1 Group Ratio | α = 0.05 with 1:2 Group Ratio |
|-------------|-------------------------------|-------------------------------|
|             | Power 80% | 90% | 80% | 90% |
| 10%         |           |     |     |     |
| N1          | 309       | 428 | 231 | 319 |
| N2          | 309       | 428 | 462 | 638 |
| Total       | 618       | 856 | 693 | 957 |
| 15%         |           |     |     |     |
| N1          | 137       | 189 | 102 | 141 |
| N2          | 137       | 189 | 204 | 282 |
| Total       | 274       | 378 | 306 | 423 |
| 20%         |           |     |     |     |
| N1          | 76        | 105 | 57  | 79  |
| N2          | 76        | 105 | 114 | 158 |
| Total       | 152       | 210 | 171 | 237 |
| 25%         |           |     |     |     |
| N1          | 48        | 66  | 36  | 50  |
| N2          | 48        | 66  | 72  | 100 |
| Total       | 96        | 132 | 108 | 150 |
| 30%         |           |     |     |     |
| N1          | 32        | 44  | 25  | 34  |
| N2          | 32        | 44  | 50  | 68  |
| Total       | 64        | 88  | 75  | 102 |
| 35%         |           |     |     |     |
| N1          | 22        | 31  | 18  | 24  |
| N2          | 22        | 31  | 36  | 48  |
| Total       | 44        | 62  | 54  | 72  |
| 40%         |           |     |     |     |
| N1          | 16        | 22  | 13  | 18  |
| N2          | 16        | 22  | 26  | 36  |
| Total       | 32        | 44  | 39  | 54  |

N1: sample size for the standard care arm. N2: sample size for the treatment arm.

Response rate = 40%.

Probability of rejection at each look: 1st look p < 0.01, final look p < 0.046.

**Toxicity Monitoring**

Since 1:2 randomization is recommended, we use the sample size value from Table 2 to compute toxicity boundaries (n = 79 and n = 158) at 25% toxicity level. A summarized toxicity boundary is presented in Table 5. In Table 5, if suppose overall number of subjects is 5 and out of 5 if there are 2 cases of toxicities, then the trial should stop because the toxicity boundary of 25% is exceeded. R computer program was used for toxicity boundary calculation (15). The full toxicity boundaries table can be found in the Appendix (Table 6).
Table 5
Abbreviated Toxicity Boundaries at Probability of Toxicity = 0.25 and α = 0.01.

| Maximum Number of Subjects | Number of Subjects with Toxicities |
|----------------------------|-----------------------------------|
| 5                          | 2                                 |
| 6                          | 3                                 |
| 8                          | 4                                 |
| 10                         | 5                                 |
| 12                         | 6                                 |
| 14                         | 7                                 |
| 16                         | 8                                 |
| 18                         | 9                                 |
| 20                         | 10                                |
| 23                         | 11                                |
| 25                         | 12                                |
| 28                         | 13                                |
| ...                        | ...                               |
| 55                         | 23                                |
| ...                        | ...                               |
| 79                         | 31                                |
| ...                        | ...                               |
| 107                        | 40                                |
| ...                        | ...                               |
| 158                        | 56                                |
| 161                        | 57                                |
| 162                        | 58                                |
Table 6
Full Toxicity Boundaries Table at Probability of Toxicity = 0.25 and α = 0.01.

| Maximum Number of Subjects | Number of Subjects with Toxicities |
|----------------------------|-----------------------------------|
| 5                          | 2                                 |
| 6                          | 3                                 |
| 8                          | 4                                 |
| 10                         | 5                                 |
| 12                         | 6                                 |
| 14                         | 7                                 |
| 16                         | 8                                 |
| 18                         | 9                                 |
| 20                         | 10                                |
| 23                         | 11                                |
| 25                         | 12                                |
| 28                         | 13                                |
| 30                         | 14                                |
| 33                         | 15                                |
| 36                         | 16                                |
| 38                         | 17                                |
| 41                         | 18                                |
| 44                         | 19                                |
| 47                         | 20                                |
| 49                         | 21                                |
| 52                         | 22                                |
| 55                         | 23                                |
| 58                         | 24                                |
| 61                         | 25                                |
| 64                         | 26                                |
| 67                         | 27                                |
| 70                         | 28                                |
| 73                         | 29                                |
| 76                         | 30                                |
| 79                         | 31                                |
| 82                         | 32                                |
### Maximum Number of Subjects vs. Number of Subjects with Toxicities

| Maximum Number of Subjects | Number of Subjects with Toxicities |
|----------------------------|-----------------------------------|
| 85                         | 33                                |
| 88                         | 34                                |
| 91                         | 35                                |
| 94                         | 36                                |
| 97                         | 37                                |
| 100                        | 38                                |
| 103                        | 39                                |
| 107                        | 40                                |
| 110                        | 41                                |
| 113                        | 42                                |
| 116                        | 43                                |
| 119                        | 44                                |
| 122                        | 45                                |
| 126                        | 46                                |
| 129                        | 47                                |
| 132                        | 48                                |
| 135                        | 49                                |
| 138                        | 50                                |
| 142                        | 51                                |
| 145                        | 52                                |
| 148                        | 53                                |
| 151                        | 54                                |
| 155                        | 55                                |
| 158                        | 56                                |
| 161                        | 57                                |
| 162                        | 58                                |

### Result for High-Risk Group:

#### Sample size calculation

For patients in the high-risk group, sample size is calculated based on different levels of improvement rate at 90% and 80% power with two interim analyses (Table 7). $P_0$ is the 30 days mortality rate in the standard arm and $P_1$ is the 30 days mortality rate in the treatment arm. The sample size is the estimated number of subjects that are required to bring down the mortality rate from $P_0$ to $P_1$. Typically, we recommend a reduction in mortality from 70–55% with 90% power. In this
case, 130 patients should be enrolled for the standard arm, and 260 should be enrolled for the treatment arm. When conducting two interim analyses, to start with, one should enroll 44 patients for the standard arm and 87 patients for the treatment arm, then perform the first look. Similar procedure should be used for the second and third look. For results with one interim analysis and 1:1 randomization, see Table 8 in the appendix. For results with one interim analysis and 1:2 randomization, see Table 9 in the appendix. For results with two interim analyses and 1:1 randomization, see Table 10 in the appendix. Note that all options have nearly identical required sample size, and therefore, performing two interim analyses would be more cost-effective from risk-benefit perspective such as the increased cost of recruiting additional patients. We suggest inflating the sample size from the table by approximately 5% to account for the loss of information (such as dropout).

### Table 7
Required Sample Size for High-Risk Group Patients with Two Interim Analyses.

| Power = 80% | Power = 90% |
|-------------|-------------|
| P0 = 80%    | P0 = 70%    | P0 = 80%    | P0 = 70%    |
| P1          | N1 | N2 | Total | N1 | N2 | Total | N1 | N2 | Total | N1 | N2 | Total |
| 70%         | 167 | 334 | 501 | NA | NA | NA | 291 | 402 | 693 | NA | NA | NA |
| 65%         | 77  | 154 | 231 | 816 | 1632 | 2448 | 106 | 212 | 318 | 1128 | 2256 | 3384 |
| 60%         | 45  | 90  | 135 | 208 | 416 | 624 | 61  | 122 | 183 | 288  | 576  | 864  |
| 55%         | 29  | 58  | 87  | 94  | 188 | 282 | 40  | 80  | 120 | 130  | 260  | 390  |
| 50%         | 20  | 40  | 60  | 53  | 106 | 159 | 28  | 56  | 84  | 73   | 146  | 219  |
| 45%         | 15  | 30  | 45  | 34  | 67  | 101 | 21  | 41  | 62  | 47   | 94   | 141  |
| 40%         | 12  | 24  | 36  | 24  | 47  | 71  | 16  | 31  | 47  | 32   | 64   | 96   |
| 35%         | NA  | NA  | NA  | 17  | 34  | 51  | NA  | NA  | NA  | 24   | 48   | 72   |

P0: 30 days mortality rate in the standard arm. P1: 30 days mortality rate in the treatment arm.
N1: sample size for the standard care arm. N2: sample size for the treatment arm.

Probability of rejection at each look: 1st look $p < 0.002$, 2nd look $p < 0.014$, final look $p < 0.046$.
1:2 randomization; $\rho = 2.0$.

**Bold** indicates recommended sample size with suggested parameters.
Table 10
Required Sample Size for High-Risk Group Patients with Two Interim Analyses.

| Power = 80% | Power = 90% |
|-------------|-------------|
| P1          | P0 = 80%    | P0 = 80%    | P0 = 70%    |
| 70%         | 466         | NA          | 645         | NA          |
| 65%         | 217         | 2204        | 300         | 3048        |
| 60%         | 126         | 567         | 175         | 784         |
| 55%         | 83          | 257         | 114         | 355         |
| 50%         | 58          | 145         | 80          | 201         |
| 45%         | 42          | 93          | 58          | 128         |
| 40%         | 32          | 63          | 44          | 88          |
| 35%         | NA          | 45          | NA          | 63          |

P0: 30 days mortality rate in standard arm. P1: 30 days mortality rate in treatment arm.
Probability of rejection at each look: 1st look $p < 0.002$, 2nd look $p < 0.014$, final look $p < 0.046$.
1:1 randomization; $\rho = 3.0$. NA = Not Applicable

Discussion

Designing a trial for testing the efficacy of therapeutic interventions for COVID-19 is challenging. The pandemic is new and there is little specific information about the virus and its adverse health effects. We do not yet clearly know the typical course of the infection, the range of susceptibility factors and the effects of co-morbid conditions. How the use of different medications, supplements and pharmaceuticals affect the severity of the infection also remain unknown. Whatever little information we currently have, is constantly being revised as new data become available. Nonetheless, it is important to develop streamlined clinical trial designed, with harmonized measures, questionnaires, biomarkers and clinical endpoints, so that the results of different clinical trials could be compared. This is critically important in current circumstance, where a large number of clinical trials need to conducted, as rapidly as possible and with extraordinary care to ensure that maximal information could be extracted from each trial and the results obtained could be compared meaningfully with other trials in the field to administer effective therapies as soon as possible.

Many factors need careful consideration in designing clinical trials, and critical decisions have to be made regarding which parameters to include and which tests should be conducted. In developing model clinical trial designs here, we gathered information from recently published manuscript and from frontline physicians’ opinion, while fully recognizing that these may need revision. However, based on currently available evidence, we have developed robust design that may require only minimal modification and updating for rapid implementation.

To aid rapid and robust clinical evaluation, our trials have been designed for feasibility and for minimizing the number of participants required. Even though ideally, for a balanced design many known factors should be considered for stratification, we have selected only the most basic demographic parameters, as too many strata require much larger sample size. On the basis of currently available evidence, the stratification factors considered in our design seem most appropriate and generally-applicable to us; however, investigators should pick the factors that are most suitable for their patients and for the specific requirements of the trial. The stratification factors that we include in our design – age, sex and cardiovascular disease risk seem fundamental to the etiology of the infection, which seems primarily to affect older
male individuals with pre-existing cardiovascular disease or cardiovascular disease risk (16). Reasons for the high susceptibility of individuals with cardiovascular disease risk for COVID-19 remain unclear and are under intense investigation, but it has been speculated that conditions associated with chronic unresolved inflammation – such as diabetes, obesity, cardiovascular disease, which are characterized by intrinsic immune dysfunction leading to inflammation may enhance the risk of severe infection and more severe outcomes (17). Although there are significant racial and ethnic differences in susceptibility to cardiovascular disease (18, 19), there is little evidence to support racial differences per se and not race-specific differences in cardiovascular disease burden affect COVID-19 severity. However, should emerging data indicate that race is an important determinant of the severity of infection or its outcomes, independent of pre-existing cardiovascular disease risk, it could be used for additional stratification of the patient population. Additionally, if a trial is designed to assess pulmonary or renal outcomes, stratification based on lung or kidney function may be important.

In the design of our clinical trials, we focused on primary outcomes. In general, mortality as the primary outcome seems appropriate at least for advanced stage patients, while for intermediate risk patients, event-free survival appears more appropriate. However, different primary endpoints may be considered, which along with appropriately selected secondary endpoints could provide important mechanistic information. Current evidence suggests that even though COVID-19 significantly impairs tissue function, much of the tissue injury is mediated by the resultant IL-6-driven cytokine storm that exacerbates pulmonary injury and may further damage other peripheral organs as has been reported for SARS (20–22). Therefore, an intervention designed to decrease viral load, may be only marginally efficacious in preventing clinical symptoms, even though it might lead to a significant decrease in viral load. Similarly, interventions targeted at pro-inflammatory cytokines (e.g., with antibodies) may not affect the viral load but significant attenuate the subsequent response and clinical outcomes. Hence, to understand such non-linear relationships between infection and response, it may be important to judiciously select a panel of biomarkers informative of the immune response and its resolution at different stages of clinical disease progression.

In addition to monitoring biochemical, physiological and clinical responses, investigators should also be attentive to toxicity due to the therapeutic intervention per se. However, deciding upon an optimal toxicity monitoring rate is problematic, especially in a patient population with a high death rate, as is the case with advanced stage COVID-19 patients. Generally, phase I clinical trials use a toxicity rate of 33%, i.e., the trial is discontinued if the intervention elicits toxicity in 33% patients. In the trial designs we describe here, we suggest a toxicity rate of 25%, which may be too high for phase II/III trials, but given the high lethality of COVID-19 infection, such high levels of toxicity may be tolerable, if eventually the treatment is effective in saving a patient's life. Codes for calculating toxicity boundaries have been published before (23).

We have designed our clinical trials with the expectation that the treatment or intervention is likely to be more effective than standard care, hence all standard tests in the work were conducted one-sided. This could be readily ascertained during the interim analysis. However, we suggest that the trials should only be stopped for efficacy, but not for futility. This will ensure that a trial is not ended early when not enough patients are enrolled, and that marginal effects that relate only to a specific subgroup of patients may not be apparent in all cases. Nevertheless, in some scenarios where the intervention is clearly not working or causing unacceptable toxicity, it may be appropriate to discontinue the trial and to test a different intervention. But usually this is difficult to establish and therefore care should be taken to continue the trial to its entirety, while monitoring closely to higher toxicity rates particularly in intermediate stage patients. Sample size increases only marginally if we add stop-for-futility into the design (Table 11). Although, this is not our focus in the manuscript, it is an option for those who are willing to incorporate that information into their design.
Table 11
Sample Size for High-Risk Group with Stopping for Futility and Efficacy with One Interim Analysis.

|          | Power = 80% |          | Power = 90% |
|----------|-------------|----------|-------------|
| P1       | P0 = 80%    | P0 = 70% | P0 = 80%    |
| 70%      | 477         | NA       | 661         |
| 65%      | 222         | 2255     | 308         |
| 60%      | 129         | 580      | 179         |
| 55%      | 84          | 262      | 117         |
| 50%      | 59          | 149      | 82          |
| 45%      | 43          | 95       | 60          |
| 40%      | 33          | 65       | 45          |
| 35%      | NA          | 47       | NA          |
| P0: 30 days mortality rate in standard arm. P1: 30 days mortality rate in treatment arm. |
| Probability of rejection for efficacy at each look: 1st look p < 0.01, final look p < 0.046. |
| Probability of rejection for futility: p > 0.480. |

We suggest inflating sample size very marginally (approximately 5%) from our calculation as normally an inflation of 10 to 20 percent is performed in regular clinical studies. The reason is that COVID-19 patients are unlikely to be lost in the follow up since it is a lethal disease, and enrolled patients in both the intermediate-risk group and high-risk group are likely to be quarantined in the hospital for an extended period of time. Therefore, we suggest inflation of the sample size by only 5% to account for unexpected events, such as suicide.

Conclusions

For the high-risk patient group, we recommend a clinical trials design incorporating a composite endpoints design with two interim analyses and three factors stratification. Given the precarious condition of patients in this group, no toxicity monitoring is needed. We suggest that for this group, the use of 1:2 randomization is ideal, and that a 15% reduction in the 30 days mortality rate (from 70% in the standard arm to 55% in the treatment arm) may be an optimal measure of acceptable efficacy.

For the intermediate-risk patient group, we suggest using a composite endpoints design with two interim analyses and four factors stratification. The use of 1:2 randomization is recommended for broader patient benefit. Toxicity monitoring is acceptable at 25% level. For clinical trials with this patient population, we suggest that it is optimal to use 90% power and an improvement of 20% response rate (from 40% in the standard arm to 60% in the treatment arm).

Appendix

Partial List of Drugs/Therapies Currently Under Investigation for COVID-19 Treatment

Andrographolide (*In Silico*)
Antisense RNA (*In Vitro*)

Bacillus Calmette–Guérin (BCG) (Vaccine Clinical Trial to start in Germany)

Benzy l-quinazolin-4-yl-amine (Potential Repurpose drug using Bioinformatics)

Bevacizumab (Clinical trial underway)

Camptothe cin (Potential Repurpose drug using Bioinformatics)

Carfilz omib (Possible candidate for inhibitory activities against SARS-CoV-2 main protease)

Chloroquine (phosphate) (Clinical trial results published)

Convalescent plasma therapy (Clinical study results published, and many are underway)

Degarelix (Clinical trial underway)

Didanosine (Potential Repurpose drug using Bioinformatics)

Duvelisib (Pilot trial underway)

Elbasvir (Possible candidate for inhibitory activities against SARS-CoV-2 main protease)

Eravacycline (Possible candidate for inhibitory activities against SARS-CoV-2 main protease)

Favilavir (Patients treated in China and another trial to start in Italy)

Favipiravir (Clinical trial results published)

Human immunoglobulin/Convalescent plasma Therapy (Clinical trial results published and many are underway)

Hydroxychloroquine plus Azithromycin) (Clinical trial results published)

Interferon lambda (Potential Therapeutic Intervention)

Interferons, Arbidol (currently recruiting patients in China)

Lopinavir/Ritonavir (Clinical trial results published)

Macrolides (MAC) (Potential drug candidate)

Methylprednisolone (Clinical study results published)

Nitazoxanide (Potential drug candidate)

Ozone (Potential therapeutic agent)

R0-90-7501 (Potential Repurpose drug using Bioinformatics)

Remdesivir (GS-5734) (GS-441524) (Clinical trial and compassionate study results published)

Sarilumab (Potential drug candidate)

Tocilizumab (Clinical study results published, Tocilizumab Vs. corticosteroids to start in Malaysia)
Valrubicin (Potential drug candidate)

Vitamin C (Clinical trial underway in Wuhan, China)

**Abbreviations**

COVID-19  
Coronavirus Disease  
CVD  
Cardiovascular Disease  
ECMO  
Extracorporeal Membrane Oxygenation  
HbA1C  
Hemoglobin A1C  
MI  
Myocardial Infarction  
ICU  
Intensive Care Unit  
IMV  
Invasive Mechanical Ventilation  
FDA  
United States Food and Drug Administration  
CTCAE  
Common Terminology Criteria for Adverse Events  
DSMB  
Data and Safety Monitoring Board

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

Not applicable
Funding

Qian was supported by the University of Louisville Fellowship.

N. Rai was partly supported with Wendell Cherry Chair in Clinical Trial Research Fund.

Author’s contributions

Conception & Design: SNR, CQ, JP, DKS

Data analysis: JP, DKS

Original Draft: SNR, CQ, AS, AB

Critical Input: AS, AB

Review & Editing: All Authors

Approval of the Final Version: All Authors

References

1. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. N Engl J Med. 2020 [cited 2020 May 5]; Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa2001282. DOI: 10.1056/NEJMoa2001282.

2. Wang Y, Fan G, Horby P, Hayden F, Li Q, Wu Q, et al. Comparative outcomes of adults hospitalized with seasonal influenza A or B virus infection: application of the 7-category ordinal scale. Open Forum Infect Dis. 2019 Feb [cited 2020 May 5]; 6(3): ofz053. Available from: https://academic.oup.com/od/article/6/3/ofz053/5320627. DOI: 10.1093/od/ofz053.

3. Merton K, Davies MJ, Vijapurkar U, Inman D, Meininger G. Achieving the composite endpoint of HbA1c, body weight, and systolic blood pressure reduction with canagliflozin in patients with type 2 diabetes. Curr Med Res Opin. 2018 Feb; 34(2): 313-318.

4. Stolker JM, Spertus JA, Cohen DJ, Jones PG, Jain KK, Bamberger E, et al. Rethinking composite end points in clinical trials: insights from patients and trialist. Circulation. 2014 Oct; 130(15):1254-61.

5. DeMets DL, Califf RM. Response to: combining composite endpoints: counterintuitive or a mathematical impossibility? Circulation. 2003 Mar; 107(9): e70.

6. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. Ann Intern Med. 2008 Nov; 149(9):612-7.

7. Srivastava DK, Rai SN, Pan J. Robustness of an odds-ratio test in a stratified group sequential trial with a binary outcome measure. Biom J. 2007 Jun; 49(3): 351-64.

8. Goyal P, Choi J. J, Pinheiro L. C, Schenck E. J, Chen R, Jabri, A, et al. Clinical characteristics of Covid-19 in New York City. N Engl J Med. 2020 [cited 2020 May 5]; Available from: https://doi.org/10.1056/NEJMc2010419. DOI: 10.1056/NEJMc2010419.

9. Reynolds HR, Adhikari S, Pulgarin C, Troxel AB, Iturrate E, Johnson SB, et al. Renin-Angiotensin-Aldosterone system inhibitors and risk of Covid-19. N Engl J Med. 2020 [cited 2020 May 5]; Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa2008975. DOI: 10.1056/NEJMoa2008975.
10. Meng Y, Wu P, Lu W, Liu K, Ma K, Huang L. Sex-specific clinical characteristics and prognosis of coronavirus disease-19 infection in Wuhan, China: A retrospective study of 168 severe patients. PLoS Pathog. 2020 Apr [cited 2020 May 5]; 16(4): e1008520. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008520. DOI: 10.1371/journal.ppat.1008520.

11. Jennison C, Turnbull B.W. Group sequential methods with applications to clinical trials. Chapman & Hall/CRC; 1999.

12. East 6. Statistical software for the design, simulation and monitoring clinical trials. 2016. Cytel Inc., Cambridge MA.

13. United States Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. 2017. https://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf. Accessed 5 May 2020.

14. Richardson S, Hirsch J. S, Narasimhan M, Crawford J. M, McGinn, T, Davidson, K. W, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020 Apr [cited 2020 May 5]; e206775. Available from: https://doi.org/10.1001/jama.2020.6775. DOI: 10.1001/jama.2020.6775.

15. R Core Team. R: A language and environment for statistical computing. 2018. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org. Accessed 5 May 2020.

16. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol. 2020 May; 109(5): 531-538.

17. Schett G, Sticherling M, Neurath MF. COVID-19: risk for cytokine targeting in chronic inflammatory diseases? Nat Rev Immunol. 2020 May; 20(5): 271-272.

18. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG. Et al. Cardiovascular health in African Americans: a scientific statement from the American Heart Association. Circulation. 2017 Nov; 136(21): e393-e423.

19. Brothers RM, Fadel PJ, Keller DM. Racial disparities in cardiovascular disease risk: mechanisms of vascular dysfunction. Am J Physiol Heart Circ Physiol. 2019 Oct; 317(4): H777-H789.

20. McGonagle D, Sharif K, O'Regan A, Bridgewood C. The role of cytokines including Interleukin-6 in COVID-19 induced pneumonia and macrophage activation syndrome-like disease. Autoimmun Rev. 2020 April [cited 2020 May 5]; 102537. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7195002/. DOI: 10.1016/j.autrev.2020.102537.

21. Pedersen SF, Ho YC. SARS-CoV-2: a storm is raging. J Clin Invest. 2020 May; 130(5): 2202-2205.

22. Zhang Y, Li J, Zhan Y, Wu L, Yu X, Zhang W, et al. Analysis of serum cytokines in patients with severe acute respiratory syndrome. Infect Immun. 2004 Aug; 72(8): 4410-5.

23. Ray HE, Rai SN. An evaluation of a simon 2-stage phase II clinical trial design incorporating continuous toxicity monitoring. Contemp Clin Trials. 2011 Feb; 32(3): 428-436.

**Figures**
Figure 1

Flowchart of Two Parallel Clinical Trials with Key Design Components