Cohort Multiple randomized controlled trials open-label of immune-modulatory drugs and other treatments in COVID-19 patients

General Statistical Analysis Plan for CORIMUNO-19 Trials

[To be adapted for particular trials whenever necessary]

Version 3.0

December 13, 2020

Redacted by Raphaël Porcher and Gabriel Baron
Validated by Philippe Ravaud

Coordinating Investigator:	**Pr Olivier Hermine,** Service Hématologie Adulte Necker Hospital, Paris, France Tel: +33 144495282 +33603707920 E-mail: olivier.hermine@aphp.fr
Co-coordinating Investigator	**Pr Xavier Mariette,** Service Rhumatologie Hôpital du Kremlin Bicêtre Tel: +33145213758 +33623268104 E-mail: xavier.mariette@aphp.fr **Immune COVID 19 group**
Scientific Director:	**Dr Pierre Louis Tharaux** INSERM U970, PARCC HEGP Tel +33689502948 E-mail: pierre-louis.tharaux@inserm.fr

Methodologist and statisticians:	**Pr Philippe Ravaud** Centre d'Epidémiologie Clinique. Hôpital Hôtel Dieu 1 Place du Parvis Notre Dame 75004 Paris E-mail: philippe.ravaud@aphp.fr **Statistician:** **Pr Raphael Porcher** Centre d'Epidémiologie Clinique. Hôpital Hôtel Dieu 1 Place du Parvis Notre Dame 75004 Paris E-mail: raphael.porcher@aphp.fr **Clinical Trial Unit :** **Pr Matthieu Resche-Rigon** Hôpital Saint-Louis 1 av Claude vellefaux 75010 Paris E-mail: matthieu.resche-rigon@u-paris.fr
Sponsor : AP-HP and by delegation:	Délégation à la Recherche Clinique et à l'Innovation (DRCI) Saint-Louis Hospital 1, avenue Claude Vellefaux, Paris, France DRCD-Siège : Cécile Kedzia - E-mail: cecile.kedzia@aphp.fr Damien Vanhoye - E-mail: damien.vanhoye@aphp.fr
Monitoring and data Management:	URC Lariboisière Saint-Louis Site Saint - Louis Hôpital Saint-Louis, 1 av Claude Vellefaux, 75010 PARIS **Pr Matthieu Resche-Rigon** E-mail: matthieu.resche-rigon@u-paris.fr **Nabil Raked** E-mail: nabil.raked@univ-paris-diderot.fr

1 Summary

	CORIMUNO-19 – [Drug name]
Diagnosis and inclusion and Exclusion criteria for the trial	**Inclusion Criteria for the trial:** 1. Patients included in the CORIMUNO-19 cohort 2. Patients belonging to one of the 2 following groups: - *Group 1: patients **not requiring ICU** at admission with moderate and severe pneumopathy according to the WHO **Criteria of severity of COVID pneumopathy.*** ***Moderate cases*** Cases meeting all of the following criteria: ● Showing fever and respiratory symptoms with radiological findings of pneumonia. ● Requiring between 3L/min and 5L/min of oxygen to maintain SpO2 >97% ***Severe cases*** Cases meeting any of the following criteria: ● Respiratory distress (≧30 breaths/ min); ● Oxygen saturation≤93% at rest in ambient air; or Oxygen saturation ≤97 % with O2 > 5L/min. ● PaO2/FiO2≦300mmHg - *Group 2: patients **requiring ICU** based on **Criteria of severity of COVID pneumopathy.*** ● Respiratory failure and requiring mechanical ventilation ● No do-not-resuscitate order (DNR order) **Exclusion Criteria for the trial:** 1. Patients with exclusion criteria to the CORIMUNO-19 cohort. 2. Known hypersensitivity to Tocilizumab or to any of their excipients. 3. Pregnancy 4. Current documented bacterial infection 5. Patient with any of the following laboratory results out of the ranges detailed below at screening should be discussed depending on the medication: a. Absolute neutrophil count (ANC) ≤ 1.0 x 10^9/L b. Haemoglobin level: no limitation c. Platelets (PLT) < 50 G /L SGOT or SGPT > 5N

Randomisation and Treatment procedures	Group defined by **requiring** ICU vs. **not requiring** ICU. All consecutive patients meeting the inclusion criteria will be randomized 1:1 either in the experimental arm or control arm in a set of 120 patients in total (60 in each arm), **stratified on the group**. Trials within each group are analyzed separately but are conducted simultaneously (with stratification of the randomization) for logistical reasons. If other subtrials are available, the inclusions will stop to allow inclusions in these other subtrials of the protocol and interim analysis. If the interim analysis indicates to continue the subtrial, a new set of 120 patients will be included on the same basis. If no other subtrial is available the inclusions will not be stopped in waiting for the interim analysis, given Tocilizumab is already a drug with known safety. Inclusions of new sets will stop when statistical analyses conclude on futility or efficacy or by DSMB decision. ● Group 1: patients **not requiring ICU** ○ Patients will be randomized to be offered **the active drug**. ○ Patients from the cohorts and with the same baseline characteristics will be used as controls (best standard of care) ● Group 2: patients **requiring ICU** ○ Patients will be randomized to receive **the active drug**. d. Patients from the cohorts with the same baseline characteristics will be used as controls
Duration of follow-up	○ 90 days
Criteria for efficacy	**Measures** A core set of clinical measures will be recorded daily for the first 2 weeks and then every week. The core measures include measures of OMS progression scale, oxygenation, mechanical ventilation. For patients who are eligible for an intervention trial (in both the intervention and control arms), this days measurement will include trial-specific measures related to the trial outcomes of interest. **Primary and secondary endpoints:** The primary endpoint and secondary endpoints will depend on the group of patients and tested medication. **For group 1 of patients *not requiring ICU*:** ***Co Primary Endpoints*** 1. Survival without needs of ventilator utilization (including non-invasive ventilation) at day 14. Thus, events considered are needing ventilator utilization (including Non-Invasive Ventilation, NIV), or death. New DNR order will be considered as an event at the date of the DNR. 2. Early endpoint: OMS progression scale <=5 at day 4, defined as follow:

OMS Progression scale	Descriptor	Score
Uninfected	Uninfected; no viral RNA detected	0
Ambulatory	Asymptomatic; viral RNA detected	1
Ambulatory	Symptomatic; Independent	2
Ambulatory	Symptomatic; Assistance needed	3
Hospitalized: mild disease	Hospitalized; No oxygen therapy	4
Hospitalized: mild disease	Hospitalized; oxygen by mask or nasal prongs	5
Hospitalized: severe disease	Hospitalized; oxygen by NIV or High flow	6
Hospitalized: severe disease	Intubation and Mechanical ventilation, pO2/FIO2>=150 OR SpO2/FIO2>=200	7
Hospitalized : severe disease	Mechanical ventilation, (pO2/FIO2<150 OR SpO2/FIO2<200) OR vasopressor (norepinephrine >0.3 microg/kg/min)	8
Hospitalized : severe disease	Mechanical ventilation, pO2/FIO2<150 AND vasopressors (norepinephrine >0.3 microg/kg/min), OR Dialysis OR ECMO	9
Death	Dead	10

Secondary end-points will be the WHO progression scale at 4, 7, and 14 days, overall survival at 14, 28, and 90 days, time to discharge, time to oxygen supply independency, time to negative viral excretion.
Biological parameters improvement:
Estimated GFR, CRP, myoglobin, CPK, cardiac troponin, ferritin, lactate, cell blood count, liver enzymes, LDH, D-Dimer, albumin, fibrinogen, triglycerides, coagulation tests, urine electrolyte, creatinuria, proteinuria, uricemia, IL6, procalcitonin, immunophenotype (Annexe 2), and exploratory tests (Annexe 3).

For group 2 of patients *requiring ICU*:

Co Primary Endpoints

1. Cumulative incidence of successful tracheal extubation (defined as duration extubation > 48h) at day 14. Death or DNR order will be considered as a competing event.
2. Early endpoint: WHO progression scale <=7 at day 4

Secondary endpoints will be WHO progression scale at 4, 7, and 14 days, overall survival at 14, 28, and 90 days, the 28-day ventilator free-days, respiratory acidosis at day 4 (arterial blood pH of <7.25 with a partial pressure of arterial carbon dioxide

	[Paco2] of ≥60 mm Hg for >6 hours), the evolution of PaO2/FiO2 ratio, time to oxygen supply independency, duration of hospitalization, time to negative viral excretion, time to ICU and hospital discharge. Biological parameters improvement (estimated GFR, CRP, cardiac troponin, urine electrolyte and creatinine, proteinuria, uricemia, IL6, myoglobin, KIM-1, NGAL, CPK, ferritin, lactate, cell blood count, liver enzymes, LDH, D-Dimer, albumin, fibrinogen, triglycerides, coagulation tests (including activated partial thromboplastin time), procalcitonin, immunophenotype (Annexe 2), and exploratory tests (Frozen samples Annexe 3). Rate of renal replacement therapy, ventilation parameters. **For each comorbidities group secondary criteria will be specifically addressed:** For each tested medication, specific markers of efficacy and safety may be used and will be defined.
Criteria of safety	● Number of serious adverse events ● Cumulative incidence of serious adverse events (SAEs) ● Cumulative incidence of Grade 3 and 4 AEs. Investigational medication discontinuation (for any reason)

Statistical Method	To maximize information from limited data generated, while allowing rapid decision, Bayesian monitoring of the trial based on the co-primary outcomes will be used. The overall strategy has been determined so as to control for a frequentist one-sided 5% type I error rate. The following methods pertain to the conduct and analysis of the subtrial in a given group of patients (group I or group II), which are analyzed separately with different primary outcomes, but conducted simultaneously (with stratified randomization) for logistical reasons. **The total sample size will be 120 (60 in each arm) at the interim analysis, and 240 (120 per arm) at the second analysis.** Since it is not possible to determine in advance how many patients will be recruited in each stratum, the sample sizes used for the following calculations are indicative, considering equally sized strata (groups). At the interim analysis, two posterior probabilities will be calculated: 1) the posterior probability of a lower event rate in the experimental than in the control arm (posterior probability of efficacy) and 2) the posterior probability of achieving at least a predefined effect corresponding to a hazard ratio of 0.85 (for time-to-event primary outcomes) or a risk difference of 5.5% (for binary co-primary outcomes) (posterior probability of sufficient efficacy). If the posterior probability of sufficient efficacy is less than 0.20, the trial can be stopped for futility. If the posterior probability of efficacy is higher than 0.99, the trial can be stopped for efficacy. Otherwise, the trial will continue with inclusion of additional patients, as predefined, and a final analysis is conducted with decision boundary at a posterior probability of efficacy > 0.95. Decision boundaries are non-binding, and the DSMB can recommend continuing recruitment, in the whole population or a subgroup. Final decision boundaries are then readapted to control for a one-sided type I error rate close to 5%. If the strata (groups I or II) are equally sized, the interim analysis should occur after 60 patients, and the second one with 120. This design (with only two stages) has then type I error rate 0.047 if event rates are 50% in each arm, and power 0.972 to detect a decrease from 0.50 to 0.20 and 0.739 to detect a decrease from 0.50 to 0.30. For interim monitoring, the analysis will be carried out according to the intention to treat (ITT) principle, i.e. each randomized participant will be analyzed in the group assigned to him/her by randomization, regardless of the actual treatment received or other protocol deviations. In particular, patients randomized while not meeting eligibility criteria will be kept in the analysis. In the cmRCT design, randomization occurs prior to offering an intervention, and some number of eligible patients who are randomly selected to be offered an intervention will not accept the offer. An intention to treat analysis could therefore dilute any treatment effects, and Relton et al. suggested using a complier average causal effect (CACE) analysis which provides unbiased estimates of the treatment effect for patients who comply with the protocol. ● At the final analysis stage, the ITT will be carried out, comparing all randomized patients in the intervention arm they were allocated to as described above, but a CACE analysis will be added, using an instrumental variable approach which assumes that a patient's decision not to accept the intervention will not affect the outcome (except through the intervention actually received).

2 Major amendments to the protocol

After an amendment following the first interim analysis of the CORIMUNO-SARI (sarilumab) trial, the version of the protocol version 5.0 at the date of April 6, 2020, redefined the groups of patients at inclusion (that should be analyzed separately) as:

- *Group 1: Cases meeting all of the following criteria*
- *Requiring more than 3L/min of oxygen*
- *OMS/WHO progression scale = 5*
- *No NIV or High flow*

- *Group 2: Cases meeting all of the following criteria*

- *Respiratory failure AND (requiring mechanical ventilation OR NIV OR High flow)*
- *OMS/WHO progression scale >=6*
- *No do-not-resuscitate order (DNR order)*

And primary outcomes were redefined as:

Co Primary Endpoints

Group 1:

Co-primary endpoints:

1. Survival without needs of ventilator utilization (including) at day 14. Thus, events considered are needing ventilator utilization (including no-invasive ventilation, NIV or high-flow), or death. New DNR order (if given after the inclusion of the patient) will be considered as an event at the date of the DNR.
2. Early endpoint: proportion of patients alive without non-invasive ventilation of high low at day 4 (WHO progression scale ≤ 5). A patient with new DNR order on day 4 will be considered as with a score > 5.

OMS/WHO Progression scale	Descriptor	Score
Uninfected	Uninfected; non-viral RNA detected	0
Ambulatory	Asymptomatic; viral RNA detected	1
Ambulatory	Symptomatic; Independent	2
Ambulatory	Symptomatic; Assistance needed	3
Hospitalized: mild disease	Hospitalized; No oxygen therapy	4
Hospitalized: mild disease	Hospitalized; oxygen by mask or nasal prongs	5
Hospitalized:	Hospitalized; oxygen by NIV or	6

severe disease	High flow	
Hospitalized: severe disease	Intubation and Mechanical ventilation, pO2/FIO2>=150 OR SpO2/FIO2>=200	7
Hospitalized: severe disease	Mechanical ventilation, (pO2/FIO2<150 OR SpO2/FIO2<200) OR vasopressors (norepinephrine >0.3 microg/kg/min)	8
Hospitalized: severe disease	Mechanical ventilation, pO2/FIO2<150 AND vasopressors (norepinephrine >0.3 microg/kg/min), OR Dialysis OR ECMO	9
Death	Dead	10

For group 2:

Co Primary Endpoints

1. Cumulative incidence of successful tracheal extubation (defined as duration extubation > 48h) at day 14 if patients have been intubated before day 14; or removal of NIV or high flow (for > 48h) if they were included under oxygen by NIV or High flow (score 6) and remained without intubation. Death or new DNR order (if given after the inclusion of the patient) will be considered as a competing event.
2. Early endpoint: proportion of patients with a decrease of WHO score of at least 1 point at day 4.

These modifications imply considering patients with non-invasive ventilation or high flow (WHO score 6) in group 2 rather than group 1, owing 1) to the possible severity of these patients that could be included a few hours before mechanical ventilation and 2) the possibility to include a patient under non-invasive ventilation (NIV) in the group 1, which would imply the realization of the longer-term primary outcome as soon as inclusion.

3 Analysis population

3.1 Flow diagram

At the final analysis of the trial, a flow chart will be constructed according to the CONSORT 2010 reporting guidelines. It will describe:

- The number of eligible patients, randomized patients and the number of patients who have actually followed the study;
- The intervention arm allocated per randomization;
- Early cessation of the intervention and their causes and drop-outs;

- The number of patients excluded from the analysis.

The number of randomized but ineligible patients, if any, will also be reported, as well as the reason for ineligibility.

3.2 Definition of the analysis population

For interim monitoring, the analysis will be carried out according to the intention to treat (ITT) principle, i.e. each randomized participant will be analyzed in the group assigned to him/her by randomization, regardless of the actual treatment received or other protocol deviations. In particular, patients randomized while not meeting eligibility criteria will be kept in the analysis. In the cmRCT design, randomization occurs prior to offering an intervention, and some number of eligible patients who are randomly selected to be offered an intervention will not accept the offer. An intention to treat analysis could therefore dilute any treatment effects, and Relton et al. suggested using a complier average causal effect (CACE) analysis which provides unbiased estimates of the treatment effect for patients who comply with the protocol.
At the final analysis stage, the ITT will be carried out, comparing all randomized patients in the intervention arm they were allocated to as described above, but a CACE analysis will be added, using an instrumental variable approach which assumes that a patient's decision not to accept the intervention will not affect the outcome (except through the intervention actually received). However, if no or only a few patients were followed in the cohort but declined the offer for randomization, then the CACE analysis will not be carried out.
No data will be analyzed for patients who have withdrawn their consent during the study and have expressed opposition to the analysis of their data. If necessary, the data concerning these patients that have been collected will be destroyed. The existence of these patients will nevertheless be documented in the study flow chart.

3.3 Sample size

The total sample size has been fixed for the whole trial at 120 (60 per arm) for the first formal interim analysis, and 240 (120 per arm) for the final analysis, but with an option to accrue 120 patients more (60 per am) depending on the recommendations of the DSMB (see below). Each group of patients (group 1, patients not requiring ICU, and group 2, patients requiring ICU) will however be analyzed separately.

4 Analysis principles

4.1 General principles for the analysis of outcomes

Although conducted as a single trial with randomization stratified on the group of patients (group 1, patients not requiring ICU, and group 2, patients requiring ICU) for operational and logistical reasons, the trials in each group are considered as separate trials, with different primary outcomes, and **are analyzed separately**.
Data analysis will be blinded to treatment allocation. Accordingly, when analyses are not symmetrical (e.g. probability of a lower event rate with experimental than control), two analyses will be performed, successively considering each arm as the experimental one.
The final results will be reported according to the recommendations of CONSORT 2010.
All outcomes will be analyzed in superiority analyses, and the final analyses will be adjusted for center as a random effect (randomization stratification). At the final analysis stage, secondary analyses will be carried out adjusting for the center in random effects models.
One crucial feature of the CORIMUNO-19 trials is to remain as flexible as possible, in an urgency context, when information may change quickly. The study, therefore, attempts to maximize information from limited data generated, while allowing rapid decision. This will be achieved by the use of Bayesian monitoring of the trial. While using a Bayesian approach,

where standard definition of type I and II error rate do not apply, the trial is also planned to control for frequentist (i.e. non-Bayesian) error rates. In particular, the overall strategy will be to control for a frequentist one-sided type I error rate close to 5% over one specific trial.
The primary efficacy analyses will therefore rely on computing the posterior distribution of the hazard ratio between the experimental and control arms for time-to-event co-primary outcomes and the posterior distributions of event rates in each arm for binary co-primary outcomes. From the latter, the posterior distribution of the difference in event rate will be derived. These posterior distributions will be graphically displayed, and summarized by their medians and two-sided 90% credibility intervals (the Bayesian counterparts of confidence intervals).
For secondary efficacy and safety outcomes, frequentist (i.e. non-Bayesian) analyses will be used. No correction for multiplicity and no hierarchical testing procedures are planned in analyzing secondary outcomes. These analyses will therefore be considered exploratory in nature.

4.2 Participants' characteristics at inclusion

The characteristics of patients collected at inclusion will be described globally and by randomization group, using means, standard deviations, medians, interquartile intervals, minimum and maximum for quantitative variables, and by their numbers and percentages by modality for qualitative variables.
The number of missing data for each variable will also be reported. No statistical tests for comparison between groups will be carried out.

4.3 Handling of missing or incoherent data

Given their nature and the trial settings, it is not be expected that primary outcome data would be missing. However, in the case some outcomes would be missing, binary missing outcomes will be treated as treatment failures in interim and primary final analyses, with an imputation by last value carried forward as a sensitivity analysis. For time-to-event outcomes, they will be naturally handled using methods for censored data. No imputation will be used for secondary efficacy and safety outcomes.

4.4 Statistical software

The analyses will be carried out using the R software version 3.6.1 or later (The R Foundation for Statistical Computing, Vienna, Austria), SAS version 9.4 or later (SAS Institute Cary, NC), and JAGS version 4.3.0 or later.

5 Co-primary outcome analysis

5.1 Original definitions

Two co-primary outcomes are used for each group of patients, one short-term outcome evaluated at 4 days, primarily used for trial monitoring, and one longer-term outcome evaluated at 14 days. For numbering the days, the day of inclusion is considered as day 1.

5.1.1 Group 1: patients not requiring ICU

1) Longer-term outcome: Survival without needs of ventilator utilization (including non-invasive ventilation, NIV) at day 14. Thus, events considered are needing ventilator utilization (including NIV), or death. New Do-Not-Resuscitate (DNR) order will be considered as an event at the date of the DNR;
2) Early outcome: OMS progression scale ≤ 5 at day 4, defined as follow:

OMS Progression scale	Descriptor	Score
Uninfected	Uninfected; non viral RNA detected	0
Ambulatory	Asymptomatic; viral RNA detected	1
Ambulatory	Symptomatic; Independent	2
Ambulatory	Symptomatic; Assistance needed	3
Hospitalized: mild disease	Hospitalized; No oxygen therapy	4
Hospitalized: mild disease	Hospitalized; oxygen by mask or nasal prongs	5
Hospitalized: severe disease	Hospitalized; oxygen by NIV or High flow	6
Hospitalized: severe disease	Intubation and Mechanical ventilation, pO2/FIO2>=150 OR SpO2/FIO2>=200	7
Hospitalized : severe disease	Mechanical ventilation, (pO2/FIO2<150 OR SpO2/FIO2<200) OR vasopressor (norepinephrine >0.3 microg/kg/min)	8
Hospitalized : severe disease	Mechanical ventilation, pO2/FIO2<150 AND vasopressors (norepinephrine >0.3 microg/kg/min), OR Dialysis OR ECMO	9
Death	Dead	10

5.1.2 Group 2: patients requiring ICU

1) Longer-term outcome: Cumulative incidence of successful tracheal extubation (defined as duration extubation > 48h) at day 14. Death or DNR order will be considered as a competing event;
2) Early outcome: OMS progression scale ≤ 7 at day 4.

5.2 Amended definitions

5.2.1 Group 1: patients with WHO score 5 at inclusion

1) Survival without needs of ventilator utilization (including non-invasive ventilation and high flow) at day 14. Thus, events considered are needing ventilator utilization (including Non-Invasive Ventilation, NIV, or high flow), or death. New DNR order (if given after the inclusion of the patient) will be considered as an event at the date of the DNR.
2) Early endpoint: proportion of patients alive without non-invasive ventilation of high low at day 4 (WHO progression scale ≤ 5). A patient with new DNR order on day 4 will be considered as with a score > 5.

OMS/WHO Progression scale	Descriptor	Score
Uninfected	Uninfected; non-viral RNA detected	0
Ambulatory	Asymptomatic; viral RNA detected	1
Ambulatory	Symptomatic; Independent	2
Ambulatory	Symptomatic; Assistance needed	3
Hospitalized: mild disease	Hospitalized; No oxygen therapy	4
Hospitalized: mild disease	Hospitalized; oxygen by mask or nasal prongs	5
Hospitalized severe disease	Hospitalized; oxygen by NIV or High flow	6
Hospitalized severe disease	Intubation and Mechanical ventilation, pO2/FIO2>=150 OR SpO2/FIO2>=200	7
Hospitalized : severe disease	Mechanical ventilation, (pO2/FIO2<150 OR SpO2/FIO2<200) OR vasopressors (norepinephrine >0.3 microg/kg/min)	8
Hospitalized : severe disease	Mechanical ventilation, pO2/FIO2<150 AND vasopressors (norepinephrine >0.3 microg/kg/min), OR Dialysis OR ECMO	9
Death	Dead	10

5.2.2 Group 2: patients with WHO score ≥ 6 at inclusion

1) Cumulative incidence of successful tracheal extubation (defined as duration extubation > 48h) at day 14 if patients have been intubated before day 14; or removal of NIV or high flow (for > 48h) if they were included under oxygen by NIV or High flow (score 6) and remained without intubation. Death or new DNR order (if given after the inclusion of the patient) will be considered as a competing event.
2) Early endpoint: proportion of patients with a decrease of WHO score of at least 1 point at day 4.

5.3 Trial monitoring

This section describes the Bayesian monitoring of the trial in one of the groups. Calculations have been made for a fixed sample size at the interim and final analysis (30 per arm and 60 per arm, respectively), but in practice, since the trial is conducted simultaneously in both groups, the numbers may differ. For simplicity, we did not plan to modify the decision boundaries according to the observed numbers of patients actually included in each group. Rather, the properties of the design (current table 1) will be re-evaluated taking the actual numbers into account.

We defined two co-primary outcomes, one time-to-event outcome evaluated up to day 14, and an early success outcome evaluated on day 4. Methods for trial monitoring have been developed for the early outcome because (1) short-term outcomes are obtained more quickly so are easier for early interim decision and (2) calculations of all possible outcomes are more tractable for binary outcomes. For analyses based on the hazard ratio, which allow accounting for all information gathered in the trial (even for patients who do not have the entire follow-up necessary to evaluate a binary outcome), the same decision boundaries will be used. It is not expected that the properties of the boundaries would be significantly different when using the posterior distribution of the hazard ratio. More comprehensive simulation studies will be performed to describe the properties of the design in an appendix to the protocol. Also, in all that follows, we assume the "event" corresponding to the outcome being detrimental to patients, so that effective treatment would lower the event rate, or achieve a hazard ratio $\theta < 1$. When the clinical definition of the outcome is opposite, then the analysis will be performed on the inverse (e.g. failure instead of success, or inverse of the hazard ratio $1/\theta$).

5.3.1 Interim analyses

Let us denote p_E and p_C the event rates in the experimental and control arms, respectively. At each analysis, the posterior probability of a lower event rate in the experimental than in the control arm is calculated, i.e. $P(p_E < p_C \mid \text{data})$, which we term the *posterior probability of efficacy*. The posterior probability $P(p_E < p_C - \delta \mid \text{data})$ is also computed, corresponding to the probability to achieve at least a δ treatment effect, termed the *posterior probability of sufficient efficacy*. To compute the probability of sufficient efficacy, we assumed that the hazard ratio for time-to-event outcomes should be at least 0.85, which translates to an event rate of 45.5% in the experimental arm when it is 50% in the control arm. Accordingly, δ was set to 0.055 for calculations with binary outcomes. The specification of the prior distribution is crucial. For this first trial conducted in the cmRCT, we want the conclusions to depend primarily on data from the trial, not on prior opinion. An uninformative prior for the hazard ratio will therefore be used. More precisely, the prior distribution of p_E and p_C will be set as a beta prior distribution with parameters 1 and 1, equivalent to a uniform distribution on the interval (0,1). This corresponds to a hypothetical situation where we would have data on two individuals treated with each arm strategy and observing that exactly 1 of the 2 experiencing the outcome.

For time-to-event outcomes, a Bayesian Cox model will be estimated using Markov chain Monte Carlo (MCMC) methods, using a Gaussian prior distribution with mean 0 and variance 10^6. The posterior probability of the hazard ratio θ will be used to define posterior probability of efficacy as $P(\theta < 1)$ and the posterior probability of sufficient efficacy $P(\theta < \eta)$, with η fixed at 0.85. The prior distributions used ensure very little influence of our prior opinion on conclusions.

5.3.2 Stopping rules

At each interim analysis, if the posterior probability of sufficient efficacy is less than 0.20, the trial could be stopped for futility upon decision of the DSMB (indicative and not binding futility

boundary). If the posterior probability of efficacy is higher than 0.99, then the trial may be stopped for efficacy (again this boundary is not binding and the DSMB may propose to continue the accrual based on other information, such as secondary outcomes or safety). The choice of interim monitoring for futility based on the posterior probability of sufficient efficacy and not the posterior probability of efficacy is justified by the need to increase the chance of early stopping for futility when information increases, if the experimental treatment is no better than the control. Conversely, keeping a constant futility boundary on the posterior probability of efficacy would decrease the chances of early stopping if additional analyses are performed, because under the null, as information increases, the posterior distribution of efficacy would converge to 0.5. This boundary is stricter than using a boundary on the posterior probability of efficacy (grey line on figure 1, left panel), but this choice is justified by the need to quickly identify treatments with a large effect.

At the interim analyses, the predictive probability of achieving success after inclusion of a total of 60 patients per arm (posterior probability of efficacy > 0.95) will also be computed for the short-term outcome, and the trial can be stopped for futility if it is less than 10%.

When no stopping for futility or efficacy is decided, additional patients are recruited in each arm. The final analysis will occur after final recruitment, and a posterior probability of efficacy higher than 0.95 will be considered as indicating efficacy.

Another option would be to continue accrual in a subgroup only (adaptive enrichment) according to the posterior probabilities in the different subgroups. If such a modification is implemented, then the SAP will be revised to accommodate such modifications.

The protocol also mentions additional interim analyses by the DSMB, without formal stopping rules. For these analyses, safety data will be presented, as well as posterior probabilities for both short-term and mid-term outcomes.

5.3.3 Frequentist properties of the design

Table 1 presents the properties of the design under different scenarios. Figure 1 displays the decision boundaries for the early outcome in the case 30 patients per arm have been recruited.

Table 1. Operational characteristics of the design under different scenarios.

	Failure rate p in each group			
Scenario	No effect	Very large effect	Large effect	Mild effect
Parameterizations	p_C=0.5, p_E=0.5	p_C=0.5, p_E=0.2	p_C=0.5, p_E=0.3	p_C=0.5, p_E=0.35
Corresponding hazard ratio	1	0.32	0.51	0.62
Probability of early stopping for futility	0.349	0.0017	0.023	0.057
Probability of early stopping for efficacy	0.0087	0.558	0.228	0.121
Probability of efficacy at 2nd stage	0.038	0.413	0.510	0.393
Overall probability of rejection	0.047	0.972	0.739	0.514

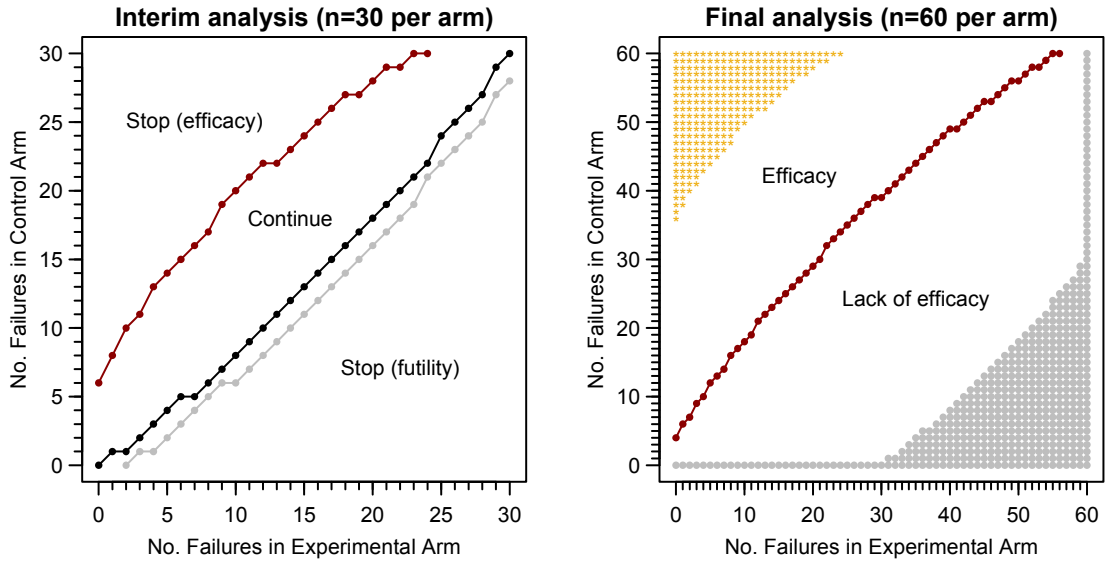

Figure 1. Decision boundaries for the interim and final analysis. Red lines indicate efficacy boundaries and black lines futility boundaries. On the left plot, the interim analysis is performed after the inclusion of 30 patients per arm, and the gray line indicates what the boundary would be if the posterior probability of efficacy was used to define futility instead of the posterior probability of sufficient efficacy. On the right plot, the final analysis after accrual of 30 more patients per arm is presented. Gloden stars indicate regions that should not occur if the decision boundaries are respected because the trial would have been stopped for efficacy at the interim analysis. Gray points indicate regions that should not occur if the decision boundaries are respected because the trial would have been stopped for futility at the interim analysis.

Table 2. Operational characteristics of the design under different scenarios for analysis of the time-to-event outcome. Results were obtained from 10,000 numerical simulation runs. We used exponential simulations, assuming a median survival with control of 14 days and accrual of 120 patients over 10 days, interim analysis at 10 days, and final analysis after 24 days (when the last patient would have attained 14 days follow-up).

	Failure rate p in each group		
Scenario	No effect	Very large effect	Large effect
Parameterizations	p_C=0.5, p_E=0.5	p_C=0.5, p_E=0.2	p_C=0.5, p_E=0.3
Corresponding hazard ratio	1	0.32	0.51
Probability of early stopping for efficacy	0.011	0.478	0.204
Probability of efficacy at 2nd stage	0.043	0.507	0.623
Overall probability of rejection	0.054	0.985	0.827

In the case the DSMB would deem results promising but not yet conclusive after inclusion of the final sample size (that we consider for illustration as a posterior probability of sufficient efficacy of 0.40 or more but a posterior probability of efficacy is of 0.97 or less), the protocol envisaged that 30 additional patients per arm could be recruited. The final decision boundary could be adapted to a posterior probability of efficacy > 0.963 to control the type I error rate. Table 2 summarizes the properties of such extension under the four previous scenarios and illustrates that this could have an important effect on the power in scenarios where the efficacy is less than anticipated.

Table 2. Operational characteristics of the design with extension to a third stage, under different scenarios. In this example, it is assumed that the DSMB would consider results to be promising if the posterior probability of sufficient efficacy of 0.40 or more but a posterior probability of efficacy is of 0.97 or less, and the final decision boundary is set to a posterior probability of efficacy > 0.963 to control the type I error rate.

	Failure rate p in each group			
Scenario	No effect	Very large effect	Large effect	Mild effect
Parameterizations	p_C=0.5, p_E=0.5	p_C=0.5, p_E=0.2	p_C=0.5, p_E=0.3	p_C=0.5, p_E=0.35
Probability of occurrence	0.307	0.046	0.313	0.460
Probability of efficacy at 3rd stage	0.018	0.043	0.209	0.221
Overall probability of rejection	0.050	0.994	0.848	0.631

5.3.4 Presentation of results

For unadjusted analyses, and for purpose of trial monitoring, the posterior distributions of the event rates in each group and of their difference will be graphically displayed, and summarized by their median and two-sided 90% credibility intervals. Similarly, for longer-term outcomes, the posterior distribution of the hazard ratio will be displayed and summarized by its median and two-sided 90% credibility intervals. Kaplan-Meier plots or cumulative incidence of the longer-term events will also be estimated in each arm, in a frequentist approach. Posterior probabilities of efficacy and sufficient efficacy will also be presented for both short-term event rates and longer-term outcomes.

5.4 Final analyses

For the short-term outcome, the posterior distributions of the difference in outcome rate and the odds ratio will be computed and summarized by their median and two-sided 90% and 95% credible intervals. The 90% level matches the 95% threshold for the posterior probability of efficacy, and the 95% level the more usual level. The posterior distribution of odds ratio adjusted for age and center (as a random effect) will be also estimated using MCMC and summarized in the same way.

For the long-term outcome, the posterior distribution of the hazard ratio both unadjusted and adjusted for age and center (as a random effect) will be calculated using MCMC and summarized by their median, and two-sided 90% and 95% credible intervals. For group 2, where the primary outcome is the cumulative incidence of extubation, the hazard ratio will be estimated by a Fine-Gray model (subdistribution hazard ratio).

Frequentist analysis will be also presented for both outcomes, only for the adjusted analyses, using a logistic model, a Cox model, and a Fine-Gray model, respectively.

5.4.1 Settings for Monte Carlo Markov Chain Bayesian analyses

The initial protocol specified using Gaussian prior distributions with mean 0 and variance 10^6 for the log hazard ratio. For adjusted analyses, the prior for the log hazard ratio for age is also a Gaussian prior, with mean 0 and variance 10^6. Four different chains with different starting values will be run, with a burn-in of 10,000 iterations, and 100,000 additional iterations, and a thinning interval of 10, leading to keeping 10,000 values per chain, 40,000 in total. The convergence of the models will be assessed using the Gelman-Rubin statistic and by visual inspection of the trace of coefficients

As a sensitivity analysis, we will investigate different prior distributions, with a flat prior with smaller variance (10^2) which makes less likely unrealistic treatment effects, two sceptic priors centred on 0 with variance set so that a P(HR < 0.2) = P(HR > 5) = 0.05 (SD 0.975) or P(HR < 0.2) = P(HR > 5) = 0.025 (SD 0.82), and two enthusiastic informative priors centred treatment effects observed on other studies:

- For CORIMUNO-TOCI1 (tocilizumab, group 1 trial): centered on half the log HR and the log HR reported for death in an observational study (Somers et al. doi:

https://doi.org/10.1101/2020.05.29.20117358), and the same variance as for the sceptic prior with SD 0.975;

- For CORIMUNO-ANA (anakinra, group 1 trial): centered on half the log HR and the log HR reported for death or mechanical ventilation in an observational study (Huet et al. Lancet Rheumatol 2020; 2: e393–400 https://doi.org/10.1016/S2665-9913(20)30164-8, which reported a hazard ratio of 0.22 in a similar population, and the same variance as for the sceptic prior with SD 0.975;
- For CORIMUNO-SARI1 (sarilumab, group 1 trial): centred on the log HR in the CORIMUNO-TOCI1 trial, and a prior SD of 0.975 or 0.82;
- For CORIMUNO-SARI2, CORIMUNO-TOCI2, and CORIMUNO-ANA2 (sarilumab tocilizumab and anakinra, group 2 trials): centered on a moderate (log HR 0.16, HR 1/0.85) or large (log HR 0.48, HR 1/0.62) effect, with SD 0.975.

5.5 Calculation of the outcome

The short-term primary outcome will simply use the values of WHO scores reported on day 4 (and day 1 in group 2). Missing data will be considered as a failure but an analysis of observed data and imputation by the last observation carried forward (LOCF) will be added.
For longer-term outcomes, discrepancies between the reported WHO scores and reported data for oxygen or ventilation status, for instance, which includes missing data, will be handled by considering the most severe scenario (for instance patients with WHO score 5 but noted as under mechanical ventilation will be considered as ventilated, and a patient noted as under nasal cannula but with a WHO score of 7 or more as under mechanical ventilation). Monitoring of such discrepancies will be carried out to limit at best their occurrence.
Moreover, since non-invasive ventilation or high flow may be more prone to center-specific practice or device ability, a sensitivity analysis only considering mechanical ventilation (i.e. survival without the need for mechanical ventilation) will be considered in group 1.
For the day 14 primary outcome, patients discharged alive before day 14 without information on respiratory status at day 14 will be considered as being alive without the need for ventilation at day 14 (or maximum theoretical follow-up if shorter than 14 days). A close data monitoring will be carried out to limit this situation as much as possible.
The definition of the outcomes in the protocol states that "New Do-Not-Resuscitate (DNR) orders" in group 1 and "DNR orders" in group 2 will be considered as events. The precise definition of "new DNR order" is set as DNR orders posterior to the date of randomization and that has been noted as having been effectively used to limit care.

5.6 Subgroup analyses

The protocol specified that, at the end of the study, subgroup analyses would be performed according to antiviral therapies at baseline. Moreover, interactions between experimental treatments and antiviral therapies will be explored and tested. These analyses will be performed using frequentist methods.
Additional subgroup analyses can be added post-hoc, in particular in the light of other published trials. In particular, post-hoc subgroup analyses will be carried according to the receipt of corticosteroids (in general), or specifically dexamethasone at baseline, and CRP levels (≤ 150 or > 150 mg/L). For group 2 trials, analyses according to the WHO score at baseline (6, or 7 or more), and delay from ICU admission (≤ 1 or > 1 day) will be added, per Scientific Committee recommendation.
When the number of events in one subgroup is less than five, no treatment effect will be computed for that subgroup.

6 Secondary efficacy outcomes analysis

6.1 Definitions

6.1.1 Group 1: patients not requiring ICU

- WHO progression scale at 4, 7, and 14 days
- Overall survival at 14, 28, and 90 days
- Time to discharge
- Time to oxygen supply independency
- ~~Time to negative viral excretion~~

Biological parameters improvement: Estimated GFR, CRP, myoglobin, CPK, cardiac hs troponin, ferritin, lactate, cell blood count, liver enzymes, LDH, D-Dimer, albumin, fibrinogen, triglycerides, coagulation tests, urine electrolyte, creatinuria, proteinuria, uricemia, IL-6, procalcitonin, immunophenotype, and exploratory tests.

6.1.2 Group 2: patients requiring ICU

- WHO progression scale at 4, 7, and 14 days
- Overall survival at 14, 28, and 90 days
- 28-day ventilator free-days
- ~~Respiratory acidosis at day 4 (arterial blood pH of <7.25 with a partial pressure of arterial carbon dioxide [$PaCO_2$] of ≥60 mm Hg for >6 hours)~~
- Evolution of PaO_2/FiO_2 ratio
- Time to oxygen supply independency
- ~~Duration of hospitalization~~
- ~~Time to negative viral excretion~~
- Time to ICU and hospital discharge.

6.2 Methods for analysis

6.2.1 Time-to-event outcomes

Time-to-event outcomes will be analyzed using Cox or Fine-Gray regression models adjusted for the same variables as the day 14 primary outcome; results will be expressed as hazard ratios with 95% confidence interval. Competing risks analyses (Fine-Gray model) will be used for time to discharge, time and time to oxygen supply independency, for which death will be considered as a competing event. When several timepoints are mentioned, separate models will be estimated at 14, 28, and 90 days. When no timepoints were mentioned in the protocol (e.g., time to oxygen supply independency, time to discharge), the outcome will be analysed at day 28 and 90. Point estimates of survival in each arm will be presented together with Kaplan-Meier survival curves.

6.2.2 WHO ordinal scale

For the WHO ordinal scale, Bayesian proportional odds models will be used to compare the distribution of ordinal scores at day 4, 7, and 14, adjusted for age and center, and a longitudinal version of the model with a time effect and a random subject effect will be used to analyze all scores up to day 14. The distribution of scores will be described at 4 (primary outcome), 7, and 14 days. For 14 days scores, a tolerance of plus/minus two days will be used, the value closest to 4 days being used, values before days 14 having precedence over values after day 14.

6.2.3 Ventilator-free days

For participants under invasive mechanical ventilation at randomization, 28-days ventilator-free days (VFDs) will be defined as (adapted from Schoenfeld, Bernard, ARDS Network.

Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit Care Med* 2002;30:1772–1777):

- VFDs = 0 if the subject dies within 28 days of randomization;
- VFDs = 28 – *x* if successfully liberated from ventilation *x* days after randomization;
- VFDs = 0 if the subject is still mechanically ventilated at day 28.

For patients not under invasive mechanical ventilation at randomization, VFDs will be counted from the date of initiation of mechanical ventilation, up to day 28 of randomization. If a subject does not undergo mechanical ventilation before (and including) day 28, then VFDs = 28.
The mean difference between randomization arms will be computed, together with bootstrap 95% confidence intervals, to accommodate an expected skewed distribution, with a peak at zero.
A separate analysis will be undertaken, restricted to patients ventilated at randomization (WHO score 7 or more).

6.2.4 Biological and physiological outcomes

For biological outcomes, only descriptive analyses will be performed.

6.2.5 Changes in outcomes

Time to negative viral excretion and respiratory acidosis at day 4 (in group 2) had been mentioned in the original protocol but are not recorded in the CRF, so they cannot be analyzed. Duration of hospitalization was mentioned for group 2, but it is redundant with time to hospital discharge.

6.3 Pooled analysis of IL-6 inhibitor Group 2 trials

Following the registration of several systematic reviews of the effect of IL-6 inhibitors on PROSPERO, a pooled analysis of the SARI2 and TOCI2 CORIMUNO-19 trials (sarilumab and tocilizumab, group 2 trials) will be undertaken.
For the day14 co-primary outcome and day 90 survival, a two-stage random-effects meta-analytic model with inverse-variance weighting will be first used to estimate the between-trial variance (tau square) by Paule-Mandel estimator. Estimates of the treatment effect in each trial obtained from Fine-Gray or Cox models adjusted for age and center (as a random effect) will be pooled.
In the absence of evidence of statistical heterogeneity, a pooled IL-6 inhibitor effect will be used with a one-stage approach. The model will adjust for trial as a fixed effect, center as a random-effect, and age at randomization as a fixed effect. All analyses will be carried out in a frequentist framework.
Subgroup analyses according to CRP levels (≤ 150 or > 150 mg/L) and delay from ICU admission (≤ 1 or > 1 day) will be added. The choice of one day from ICU admission follows from the REMAP-CAP eligibility criteria after a press release of tocilizumab efficacy (2020-11-19).

7 Safety analysis

7.1 Definitions

Adverse events are spontaneously declared on the CRF. For each adverse event, the following information is collected:

- Classification of the adverse event (AE) as a serious adverse event (SAE);
- Seriousness criteria for SAEs;
- Intensity (severity): mild, moderate, or severe;

- Start/end dates;
- Investigator judgment on relationship with the study treatment, concomitant treatment, pre-existing disease, and COVID-19;
- Modification of study treatment;
- Symptomatic treatment;
- Outcome.

Moreover, major safety endpoints are monitored: blood cells and platelets counts and liver transaminases, are monitored frequently, every three days systematically:

- Neutrophil count;
- Platelet count;
- Liver enzymes: ALT and AST;
- Occurrence of skin rashes;
- Systolic and diastolic blood pressure;
- Ventilator asynchronization.

7.2 Analysis

Adverse events and their characteristics will be described using numbers and percentages per treatment arm. The proportion of participants with each of the reported events, as well as the proportions of participants with at least one SAE, will be compared using Fisher's exact tests. The total number of AE/SAEs and SAEs will also be described for each arm, and compared using Poisson models (with a robust error variance if necessary).

8 Summary of changes since previous versions

8.1 Version 1.4 compared to the previous working version and version 1.0

- A new paragraph 5.4 has been introduced to better separate the final analysis and data presentation of the primary outcomes from the analyses carried interim analyses aiming at trial monitoring (paragraph 5.3). Subsequent paragraph numbering has been adapted accordingly.
- The original version of the protocol and SAP mentioned reporting 95% credibility intervals. Since the decision rules are one-sided (posterior probability of efficacy > 0.99 at the interim and > 0.95 at the final analysis), credibility intervals coherent with the decision rules would be one-sided 95% credibility intervals (though this was not formally specified when quickly drafting the protocol in a crisis situation). We considered it would be preferable to report two-sided credibility intervals, and therefore specified reporting two-sided 90% credibility intervals, which have the same upper bound and therefore allow the same conclusion. We however added two-sided 95% credibility intervals for the final analysis.
- Adjustment of analyses on age has been made explicit instead of "for major prognostic factors". The choice of age as the only adjustment factor (in addition to center, the randomization stratification variable) has been determined by the DSMB.
- Practical settings for the Bayesian analyses have been detailed.
- The use of a Fine-Gray model to estimate the hazard ratio of the longer-term outcome in group 2 has been made explicit. This choice is natural given the primary outcome is expressed as a cumulative incidence in a competing risks framework.
- For secondary outcomes, the analysis of the WHO score over time has been changed from the planned ranked ANCOVA approach to a longitudinal proportional odds model. This choice was determined because the latter has been advocated for analyzing the

WHO ordinal scale in the context of COVID-19 trials and because the large number of ties on this scale may limit the advantage of ranked ANCOVA. Of note, the proportional odds model is close to Wilcoxon rank-sum tests, but (1) provides an interpretable measure of treatment effect, and (2) allows for adjustment.
- Analysis of biological outcomes over time has been specified.

8.2 Version 2.0 compared to 1.4

- The handling of patients discharged alive before day 14 for the day 14 primary outcome (as alive without the need for ventilation) has been clarified. This was decided early for allowing interim analyses when day 14 outcome was not recorded for a majority of patients, while (1) ensuring the assumption of uninformative censoring would hold and (2) avoiding later event being unduly influential if those observations were censored.
- The mention of the JAGS software (and version) for Bayesian analyses has been added.
- A tolerance of plus/minus two days for defining day 14 WHO scores has been added.

8.3 Version 2.1 compared to 2.0

- Given no patient was followed-up in the CORIMUNO-19 cohort but declined inclusion to the CORIMUNO-19-TOCI trial (patients may have subsequently withdrawn consent but this applied to the cohort also), it has been noted that the compliers average causal effect analysis will not be carried out.

8.4 Version 3.0 compared to 2.1

- Added a table with frequentist operational characteristics of the design for the time-to-event outcome.
- The parameterization of enthusiastic informative priors in the sensitivity analysis to the priors used in the Bayesian analysis of the survival co-primary outcome, for the different CORIMUNO-19 trials subject to this SAP has been defined.
- Added details on the calculation of secondary outcomes (§6.2): clarified the date of analysis of outcomes with no timeframe in the protocol, and computation of ventilator-free days.
- Added a pooled analysis of CORIMUNO-SARI2 and CORIMUNO-TOCI2 (§6.3).