Statistical Data Analysis Plan

An International Multicenter, Adaptive, Randomized Double-Blind, Placebo-Controlled Trial of the Safety, Tolerability and Efficacy of Anti-Coronavirus Hyperimmune Intravenous Immunoglobulin for the Treatment of Adult Hospitalized Patients at Onset of Clinical Progression of COVID-19

Short Title: Inpatient Treatment with Anti-Coronavirus Immunoglobulin (ITAC)

INSIGHT Protocol Number: 013

1 Introduction

1.1 Objective of the Statistical Analysis Plan

The objective of this statistical analysis plan (SAP) is to provide a description of the analytic strategy and the statistical methods that will be used to analyze the data for the Inpatient Treatment with Anti-Coronavirus Immunoglobulin (ITAC) Phase III randomized, double-blind, placebo-controlled trial. The primary objective of the trial is to determine whether hyperimmune intravenous immunoglobulin (hIVIG) to SARS-CoV-2, derived from the plasma of individuals who recover and develop neutralizing antibodies, is a potentially safe and effective therapeutic approach to COVID-19 compared to placebo (each treatment is given with standard of care [SOC]). The primary endpoint of this trial is an ordinal outcome based on the patient's clinical status on Day 7.

In Version 1.0, this SAP provided:

- A short description of the study design (Section 1.4)
- Goals of the interim reviews by the independent DSMB and the planned format of the review meetings (Section 2)
- A description of the planned data analyses presented in the reports to the DSMB (Sections 3-8). Guidelines for stopping the study because of early proof of efficacy, futility, or harm are described in Sections 7.1, 7.2, and 7.3, respectively.
- A description of data summaries to be provided to study leadership to aid in monitoring trial conduct and data quality; these data summaries, which will be regularly updated and posted to the INSIGHT website, will be restricted to enrollment (Section 3), baseline data (Section 4), and summaries of follow-up data completeness (Section 8).

Version 2.0 of the SAP was prepared prior to unblinding of the data by statisticians and other members of the core ITAC team who have been blinded to interim treatment comparisons for the duration of the trial.

Below we briefly summarize the status of the trial and some key blinded data that informed the preparation of this updated SAP (Version 2.0).

1.2 Trial Status and Information That Informed the Updated SAP

1.2.1 Trial Status

Version 1.0 of the protocol was used for the duration of the trial. On November 24, 2020 a letter of amendment was issued that extended the exclusion criteria in the trial to "Prior receipt of any SARS-CoV-2 monoclonal antibody treatments at any time." The DSMB carried out two full reviews of the protocol, on November 24, 2020 and on January 5, 2021. Following each review, the DSMB recommended the study continue as planned. Between these full reviews and until enrollment was completed, the DSMB also reviewed safety data on a weekly basis that was provided to them by the unblinded statisticians.

Following the DSMB meeting on January 5, 2021, the DSMB approved the provision of the pooled (both treatment groups combined) category proportions to blinded statisticians in order to re-estimate sample size. Using the observed pooled proportions, power was estimated to be 0.83 for the planned sample size of 500. Based on this, no change in sample size was recommended.

1.2.2 Enrollment Summary

The first participant was enrolled on October 8, 2020; the last participant was enrolled on February 10, 2021. On February 3, 2021 we notified the central IRB for ITAC, Advarra, that the final accrual would be more than planned and that it likely would be between 550 and 600 participants. A total of 593 participants were enrolled by 63 sites in 11 countries (93 more participants than were planned). These 63 sites were provided study drug infusion bags following randomization by 47 study site pharmacies. The number enrolled by site ranged from 1 to 41 participants. Twenty three sites enrolled 1-4 participants; 17 sites enrolled 5-9 participants; 11 enrolled 10-14; and 12 sites enrolled 15 or more participants.

Two participants did not meet strict eligibility criteria. One participant was enrolled 13 days after symptom onset, and the second participant had a condition that did not allow venipuncture; this participant did not receive an infusion.

1.2.3 Summary of Baseline Characteristics

Selected baseline characteristics that were considered (as of March 23, 2021) in revising the SAP are given below.

1. As noted above, participants are from 11 countries. Numbers enrolled by country are: Denmark (77), Germany (10), Spain (65), Greece (70), UK (19), Indonesia (33), Argentina (4), Israel (6), Japan (15), Nigeria (41), and U.S. (253).
2. Median (25^{th}, 75^{th} percentile) of age is 59 (50, 70) years.
3. Median (25^{th}, 75^{th} percentile) of symptom onset is 8 (6, 10) days.
4. Oxygen status (above pre-COVID-19 requirement): 28% were not receiving oxygen; 34% were receiving < 4 L/min; 28% were receiving ≥ 4 L/min; and 10% were receiving high-flow oxygen.
5. 541 (93%) received remdesivir prior to randomization or on the same day as randomization as part of SOC.
6. 329 (56%) were receiving corticosteroids at entry; 48 (84%) among those on high-flow oxygen and 229 (62%) among those on supplemental oxygen.
7. 355 (60%) were receiving heparin (prophylactic, intermediate, or therapeutic dose) at entry.
8. Median (10^{th}, 25^{th}, 75^{th}, 90^{th} percentile) BMI is 29.8 (23.0, 25.8, 34.7, 40.2) kg/m^2.
9. Selected diagnoses (% reported) collected as part of a medical history and an elevated BMI are summarized below:
    - Asthma (10%)
    - Cerebrovascular event (1%)
    - COPD (6%)
    - Diabetes (28%)

- Heart failure (4.6%)
- Hepatic impairment (2%)
- HIV (2%)
- Hypertension requiring medication (42%)
- Immunosuppressive disorder other than HIV (1%)
- Malignancy (4%)
- MI (3%)
- Renal impairment (7%)
- BMI ≥ 30 kg/m^2 (49%)

1.2.4 hIVIG/Placebo Assignment and Completeness of Infusions

As indicated in the protocol, each study site pharmacy was assigned an hIVIG product/matching placebo to use. The product assignment is summarized below:

Product Manufacturer	No. of Study Site Pharmacies	No. of Lots of hIVIG Used	No. of Participants
CSL Behring	11	5	155
Emergent	11	3	153
Grifols	11	22	146
Takeda	14	14	139
Total	47	44	593

Lot potency was measured by Texcell for each of the 44 lots used. The potency measures were reported in AY/mL units; these levels were multiplied by 1.73674589 to obtain units in IU/mL. The median (25^{th}, 75^{th} percentile) potency level of the 44 lots of hIVIG used was 1220 (893, 1442) IU/mL.

Twelve participants did not receive an infusion. Ten of these participants withdrew consent prior to being infused; for one participant, venous access could not be achieved; and one participant refused the infusion but continued in follow-up.

1.2.5 Summary of Pooled (Both Treatment Groups Combined) Follow-up Results

The last 28 day follow-up was to be completed by March 14, 2021.

Selected follow-up data for both treatment groups combined were considered.

The median (25^{th}, 75^{th} percentile) of time to discharge from randomization is 6 (4, 9) days.

During regular investigator meeting, it became clear that some sites were retaining participants in the hospital for public health reasons and/or for the collection of the Day 7 primary outcome data. There were 7 sites (all that enrolled 5 or more participants) in 6 countries that enrolled a total of 93 participants where no one was discharged before day 7.

The Day 3 NEW score is missing for 82 participants (13.8% of randomized participants). The score can only be determined for hospitalized participants and most of the missing data is for participants discharged before Day 3.

A total of 581 participants received a full or partial infusion; as noted above, 12 (2%) participants were not infused.

Among participants meeting the eligibility criteria and who received an infusion, there are 7 participants missing the 7-category primary ordinal endpoint that is assessed on Day 7 (see table below). Six of these 7 participants were discharged prior to day 7 with no further contact. One participant was transferred to another hospital before day 7. Details of the participants missing the Day 7 endpoint due to reasons other than missing forms are below.

Pt	Study Day of Discharge	Discharged to	Oxygen use on day of discharge	Study Day last known alive	Additional information
1	Day 2	Home	None	Day 2	
2	Day 4	Home	2.0 l/min	Day 4	
3	Day 3	Home	None	Day 28	Paramedics visited pt at home to collect Day 7 and Day 28 specimens. No other study data was collected.
4	Day 1	Home	1.5 l/min	Day 5	
5	Day 3	Home	4.0 l/min	Day 7	
6	Day 2	Transferred to another hospital	None	Day 27	Transferred to another hospital due to pt preference. Known alive day 27, but no other contact.
7	Day 5	Home	None	Day 25	Pt did not want to be contacted anymore, but allowed access of medical records

Among all randomized participants, 18 (3.0%) are missing the Day 7 primary endpoint, and 36 (6.1%) have unknown survival status on Day 28. Of these, 11 withdrew consent, all before the infusion.

1.3 Summary of Changes to SAP

Based on the above information and additional data expected we made the following changes to the SAP:

1. The secondary outcomes of hospitalization status at days 7, 14, and 28, time to discharge, and days alive outside of the hospital at Day 28 will be supplemented with the following additional outcomes: i) time to discharge or being able to independently undertake usual activities with minimal or no

symptoms; ii) time to being able to independently undertake usual activities with minimal or no symptoms (discharge status will be ignored); iii) the binary outcome of hospitalization will also be defined as alive and either discharged from the hospital or being able to independently undertake usual activities with minimal or no symptoms; iv) days alive and able to independently undertake usual activities with minimal or no symptoms at Day 28; and v) days alive and out of the hospital or able to independently undertake usual activities with minimal or no symptoms at Day 28 (whichever lead to greatest time).

Additionally, as a sensitivity analysis for the analysis of time-to-discharge and days outside the hospital we will exclude participants from the 7 sites where no participants were discharged prior to Day 7.

In all of the outcomes related to “discharge”, “discharge” will refer to home, to a rehabilitation center or to a post-acute care facility.

2. The change in NEW score from baseline to Day 3 will not be considered as a secondary outcome because it is only collected for hospitalized participants, and the Day 3 New score is missing for 80 participants (nearly all the participants with missing data had been discharged before Day 3).

3. A key subgroup analysis defined in the protocol is according to duration of symptoms. The primary ordinal outcome will be summarized for the following approximately equal subgroups (<6, 6-7, 8-9, and 10-12 days). The presence of a treatment by subgroup interaction will be estimated in 2 ways, with a 1 degree of freedom (df) test with duration of symptoms included in the ordinal regression model as a continuous variable and with a 3 df test with indicators for categories of duration of symptoms in the regression model.

4. Another important subgroup is by age. The median (25^{th}, 75^{th} percentile) of age is 59 (50, 70) years. Therefore, we will present age in approximate quartiles (<50, 50-59, 60-69, and ≥ 70) years.

5. For the subgroup by geographic region the following subgroups will be defined: U.S.; Europe, UK, or Israel; and Argentina, Indonesia, Japan or Nigeria.

6. Subgroup analyses by chronic conditions will be carried out for individual conditions which have prevalence at least 5% at baseline (asthma, COPD, diabetes, hypertension requiring medication, renal impairment, and BMI ≥ 40 kg/m^2).

7. The following other subgroups that combine chronic conditions and concomitant treatments will be carried out:
    a. Participants with and without compromised immune function; participants with HIV, an immunosuppressive condition other than HIV, or taking antirejection medication, immune modulators, or biologic treatment for

autoimmune disease or cancer will be considered to have compromised immune function.
b. BMI < 40 and ≥ 40 kg/m^2 according to history of diabetes (4 groups)
c. Hypertension with and without history of other metabolic and vascular co-morbidity (4 groups): i) no hypertension or other metabolic/vascular co-morbidity; ii) hypertension without metabolic/vascular co-morbidity; iii) metabolic/vascular condition without hypertension; and iv) hypertension and a metabolic/vascular co-morbidity. Metabolic/vascular co-morbidities include a history of diabetes, a cerebrovascular event, heart failure, or an MI or acute coronary syndrome.
d. Number of vascular co-morbidities (0, 1, 2, 3+)
e. Quartile of Charlson Comorbidity Index (for conditions assessed)
f. Risk calculator for vaccine prioritization (JHU)
g. Quartile of disease progression risk score for Day 7 outcome that considers baseline antigen and antibody levels, age, gender, duration of symptoms, oxygen saturation, ordinal category, NEW score, and history of chronic health conditions.

8. Subgroup analyses by concomitant medications will be carried out for corticosteroids, overall and in combination with oxygen requirements (baseline ordinal scale).

9. Planned subgroup analyses by baseline antigen and antibody levels and antibody comparisons during follow-up have been added to the SAP as a new section.

10. The analysis population for the efficacy outcomes has been changed from the intention-to-treat (ITT) population to a modified intention-to-treat (mITT).

1.4 Description of the Study Design

This section is adapted from the ITAC protocol version 1.0. *Italicized sections have been added as part of this update.*

Design

This study is an adaptive, randomized, double-blind, placebo-controlled trial to evaluate the safety, tolerability and efficacy of hIVIG in adult participants who are hospitalized with COVID-19 with symptoms for no more than 12 days, and who do not have life-threatening organ dysfunction or organ failure. Remdesivir will be provided to participants in both the hIVIG and placebo groups as SOC unless contraindicated for an individual participant.

Primary Objective and Primary Outcome

The primary objective is to compare the clinical status of participants in the hIVIG + SOC and placebo + SOC groups on Day 7 using an ordinal outcome with seven mutually exclusive categories. On Day 7, the worst of the seven categories the

participant was in that day will constitute the primary outcome. The seven categories are:

7. Death
6. End-organ failure
5. Life-threatening end-organ dysfunction
4. Serious end-organ dysfunction
3. Moderate end-organ dysfunction
2. Limiting symptoms due to COVID-19
1. No limiting symptoms due to COVID-19

Appendix F of the protocol provides clinical definitions of each category.

The primary ordinal outcome captures the range of severity experienced by hospitalized patients with COVID-19, recognizing that end-organ manifestations in addition to pneumonia and acute respiratory distress syndrome are increasingly emerging as significant contributors to morbidity. The ordinal outcome includes both pulmonary manifestations as assessed in prior COVID-19 trials and additional components representing key non-pulmonary outcomes; the latter are highlighted as *"extra-pulmonary"* in the guidance table (Appendix F). The primary endpoint will include both pulmonary and extra-pulmonary components, while the pulmonary manifestation scale only will be reported as a secondary endpoint.

Day 7 was chosen for the timing of the primary endpoint for several reasons based on the following assumptions. The impact of hIVIG on disease progression may not be immediate; a few days may be needed to see the effects on clinical outcomes as measured by the ordinal outcome. Also, transient treatment effects that are no longer present at Day 7 may be clinically less relevant. Assessment of the ordinal outcome at a later time point may result in a diminished treatment difference because spontaneous recovery from COVID-19 may have begun in many participants. Also, antibody differences between the treatment groups, an important biologic mechanism for observing a clinical benefit, are assumed to be greatest during the first week after infusion.

Lastly, use of Day 7 to characterize the clinical severity of participants in seven categories as studied here, results in a distribution of participants in the placebo group for the ordinal outcome that is sufficiently granular and not overly skewed to the most severe or least severe categories and, therefore, provides good power for comparing the two treatment groups with a feasible sample size given the difficulty in producing large quantities of hIVIG.

The primary and secondary objectives of ITAC are addressed by pooling the four hIVIG products and making comparisons with the corresponding placebo groups. To justify this pooling, each hIVIG lot prepared has a neutralizing potency that provides an

appreciable dose margin over convalescent plasma. A standard dose of 400 mg/kg is used for each hIVIG product (see section 8.1.2 of the protocol).

Key Secondary Outcomes

A number of secondary endpoints to assess safety and efficacy have been specified. Four endpoints are defined as key secondary outcomes: 1) a composite of death, end-organ failure, or life threatening end-organ dysfunction (categories 5-7 of the primary ordinal outcome at Day 7); 2) time to the two most favorable categories of the primary ordinal outcome; 3) percentage in two most favorable categories of the ordinal outcome at Day 7; and hospitalization status at Day 14.

Mortality, adverse events (AEs), including infusion reactions, and biological correlates of therapeutic activity are also assessed. Because there is no established endpoint for evaluating the clinical efficacy of treatments for COVID-19, other clinically relevant outcomes, including outcomes used in other COVID-19 treatment trials, will be recorded. Thus, the randomized groups (hIVIG + SOC versus placebo + SOC) can be compared for multiple outcomes, and results can be compared or combined with other trials. A list of secondary outcomes is given in Section 7.4.

Study Treatments

Anti-Coronavirus hIVIG is a human hyperimmune product of the purified gamma globulin (IgG) fraction of human plasma containing polyvalent neutralizing antibodies to SARS-CoV-2. hIVIG is prepared from pooled plasma collected from healthy, adult donors who have recovered from SARS-CoV-2 infection. Four hIVIG products (Emergent BioSolutions, Grifols Therapeutics, Inc., Takeda Pharmaceuticals, and CSL Bering) will be used in this trial.

Duration

All participants will be followed for 28 days. A subsample of participants is followed for 90 days. If the trial goes to completion, the primary analysis will be completed after all randomized participants are followed for 28 days.

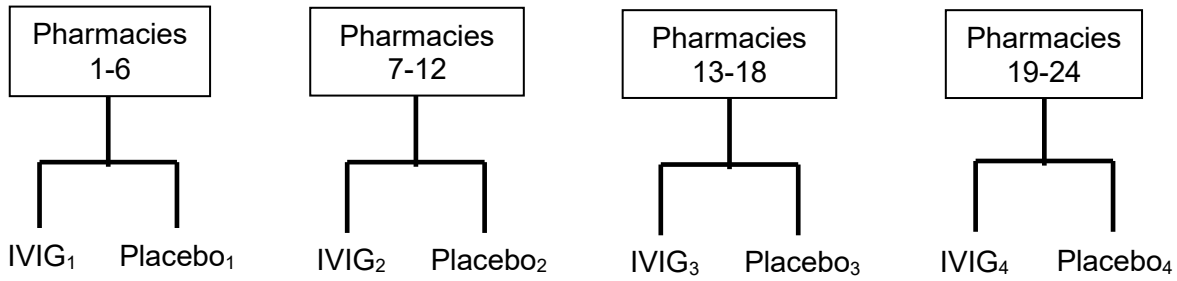
Randomization

Randomization will be stratified by site pharmacy (clinical sites may share a pharmacy). Participants will be randomized (1:1) to a single infusion of hIVIG + SOC or placebo + SOC on the day of randomization (Day 0).

Hyperimmune IVIG will be manufactured by four companies; to simplify logistics related to the supply of hIVIG to clinical sites and to take advantage of the randomization, which is stratified by site pharmacy, the same hIVIG product will be provided to a given site pharmacy for the duration of the trial to the extent possible. This is illustrated in Figure 1 with an example that assumes that there are 24 site pharmacies and the supply of each of hIVIG products will be the same.

Within each stratum permuted block randomization will be used to generate treatment assignments.

FIGURE 1. EXAMPLE OF ALLOCATION OF HIVIG TO 24 SITE PHARMACIES



Sample Size

The planned sample size is 500 participants (250 per group).

The following assumptions were made in estimating the required sample size.

a. *The primary analysis will be modified intention-to-treat (mITT).*
b. A proportional odds model with indicators for the six cut-offs corresponding to using any of categories 1 to 6 as cut-offs for determining clinical improvement, treatment group (hIVIG versus placebo), baseline severity of illness as defined by the ordinal outcome, two-way interactions between baseline severity of illness and the six cut-offs, hIVIG product/matching placebo used, and two-way interactions between hIVIG product/matching placebo used and the six cut-offs will be used to estimate the odds ratio (OR).
c. Type 1 error = 0.05 (2-sided) and power = 0.80.
d. The clinical status of participants in the placebo group at Day 7 is assumed as shown in the third column in Table 1. Since both randomized treatment groups will receive remdesivir as SOC (unless contraindicated), these percentages were estimated using Day 7 data from the ACTT-1 trial for a subgroup of patients similar to ours (the subgroup of participants who entered ACTT-1 in categories 4 and 5 of their eight-category ordinal outcome for disease severity and were randomized to the remdesivir group).
e. We assumed an OR (hIVIG/placebo) of 1.61 for a more favorable outcome. This corresponds to the percentage of participants in the hIVIG group at Day 7 shown in each level of the ordinal scale given in the second column in Table 1 below. For example, the percentage of participants in the two most favorable categories would be increased to 65.4% in the hIVIG group from 54.0% in the placebo group (an 11.4 percentage point increase from the placebo group). Conversely, the percentage of participants in the four most severe categories would decrease to 19.4% from 28.1% in the placebo group. The same proportional improvement was assumed across the ordinal scale.
f. Sample size depends on a number of assumptions, including the hypothesized odds ratio, the number of categories in the ordinal outcome, and the distribution of responses for the placebo group.[i] Hypothesized odds ratios closer to 1.0 correspond to a smaller treatment effect and require a larger sample size to maintain 80% power. The final sample size was chosen

after consideration of a range of odds ratios and of category percentages for the placebo group.

g. *Based on the category percentages in Table 1, the estimated sample size is 494. This was increased to 500 to allow for a small number of participants who may be randomized but not receive the study infusion or meet strict eligibility criteria. These participants will be excluded from the mITT analysis.*

A planned blinded sample size re-estimation that utilized the observed pooled (both treatment groups combined) category percentages confirmed that the sample size of 500 participants would provide the planned power for detecting an odds ratio of 1.61.

TABLE 1. HYPOTHESIZED PERCENTAGE OF PARTICIPANTS IN EACH CATEGORY ON DAY 7 IN THE HIVIG AND PLACEBO GROUPS BASED ON AFOREMENTIONED ASSUMPTIONS

Category	hIVIG + SOC Group	Placebo + SOC Group
7. Death	0.6	1.0
6. End-organ failure	4.0	6.3
5. Life-threatening end-organ dysfunction	4.2	6.3
4. Serious end-organ dysfunction	10.6	14.5
3. Moderate end-organ dysfunction.	15.1	17.9
2. Limiting symptoms due to COVID-19	57.6	49.0
1. No limiting symptoms due to COVID-19	7.8	5.0
Total	100.0	100.0

For the key subgroup defined according to duration of symptoms at entry, in addition to analysis by quartile, the OR for the primary endpoint will be estimated for the participants in the lower 3 quartiles. Assuming the category percentages in Table 1, with an estimated 444 participants (75% of 593) (222 per treatment group), an OR of a more favorable outcome on hIVIG compared to placebo of 1.61 can be detected with 77% power.

The study is not powered to detect treatment differences in mortality because the mortality is expected to be low given the eligibility criteria and duration of follow-up.

The following outcomes are defined as key secondary outcomes:

Composite of death, end-organ failure, or life-threatening end-organ dysfunction (categories 5-7 of the ordinal outcome) at Day 7: This composite outcome comprises the most severe three categories of the ordinal outcome. Decreasing the probability that a participant enters one of these disease states and remains there through Day 7 has high clinical significance. *Comparing the hIVIG+SOC versus the placebo+SOC groups for the proportion of participants in the three worst categories on Day 7, a total sample*

size of 579 participants (the estimated number of participants in the mITT analysis) is sufficient to detect a decrease to 6.6% in the hIVIG group compared with 13.6% in the placebo group (difference 7.0%) with 80% power. A decrease from 13.6% to 8.8% as in Table 1 (OR = 1.61) can be detected with power of 45%.

Time to the two most favorable categories of the primary ordinal outcome (first occurrence): We expect that by Day 28, almost all participants will be discharged from the hospital. Similarly, we expect most participants will be in in one of the two most favorable categories of the primary ordinal outcome by Day 28. In the ACTT-1 trial, in the subset of participants who entered the trial with disease severity similar to our eligibility criteria (ACTT-1 ordinal outcome categories 4 and 5), 94.7% had been discharged from the hospital by Day 28. This percentage was similar for the ACTT-1 definition of "recovery" that includes a small percentage of participants who were hospitalized but no longer requiring medical care. Comparing the hIVIG versus placebo groups for time to the 2 most favorable categories, our study is powered to detect a relative rate ratio (RRR) of 1.3 with 80% power and a significance level of 0.05. *The power calculations assume that the RRR is approximately constant to Day 28, the overall cumulative percentage in the two most favorable categories (pooled across treatment groups) by Day 28 is 81% and that between 2.5 and 3% withdraw consent or are lost to follow-up by Day 28.*

Two most favorable categories at Day 7: *Comparing the hIVIG versus the placebo+SOC groups for the percentage in the two most favorable categories (a binary outcome) on Day 7, the total sample size of 579 participants is sufficient to detect an increase in the this percentage from 54% in the placebo group to 65.4% in the placebo group (as in Table 1) with 80% power.*

Hospitalization status at day 14: *The study has greater than 80% power to detect an increase to 87% in the hIVIG group compared with 77% in the placebo group for the percentage dischargedl at Day 14. Estimates from the ACTT-1 trial used for the placebo group were 51% discharged on Day 7, and 77% on Day 14, for participants that were similar to ours and who were randomized to the remdesivir arm (confidential data; personal communication). Power calculations assume that the treatment groups are compared by mITT.*

Data and Safety Monitoring

An independent Data and Safety Monitoring Board (DSMB) will review interim data and use pre-specified guidelines for early termination of the trial or protocol modification.

2 Interim DSMB Reviews: Goals and Format

Goals of the interim reviews:

- Protect the safety of study participants.
- Advise on stopping or modifying the trial for efficacy, for patient safety in case of emerging data on harm, or for futility.
- Review the conduct of the trial

The DSMB will conduct frequent safety reviews. The DSMB will review safety data for the first 20 to 30 participants randomized after they have been followed for 7 days. Thereafter, the DSMB will be asked to review safety data at 30 day intervals. The blinded sample size re-estimation will occur after 150 participants have been followed for 7 days. Futility reviews will be presented to the DSMB for the primary endpoint after 50% of information time (based on the percentage of participants who have completed 7 days of follow-up) is available.

The DSMB may request interim reports on safety and efficacy at any time.

Review meetings will typically consist of an Executive session (optional; closed), open session, closed session, and a second open session to give feedback to study leadership (optional).

Masking of treatment group labels in interim reports: In the open reports, any data reports will be pooled across the two treatment groups. In the closed reports, treatment group labels will be masked; for example as “Group A” versus “Group B”. The treatment group labels will be consistent across all analyses and over subsequent reports. With each closed report, the DSMB will receive a separate, encrypted file that unmasks the treatment group labels. This procedure ensures that the DSMB has the full information to weigh benefit versus harm.

Open report to the DSMB

The open reports will contain:

- A synopsis of the trial design and current status of the trial
- Responses of the study team to DSMB recommendations
- A summary prepared by the study leadership including any relevant emerging data from other studies
- Data summaries for enrollment (including enrollment by hIVIG product/matching placebo group) and eligibility violations (Section 3), baseline characteristics (Section 5) and protocol deviations
- Summary reports for data completeness and study conduct (including number infused), pooled across treatment groups (see Sections 6 and 8)
- Unanticipated problems

All data summaries in the open report will be pooled across treatment groups. The open report will be prepared by the blinded statisticians in cooperation with the unblinded statisticians. In addition to the DSMB, open reports will be provided to the study team, and posted on the study website following the DSMB meeting for access by study investigators.

While the study is ongoing, summaries by treatment group, and comparisons of the hIVIG versus placebo are restricted to the confidential closed report to the DSMB.

Additionally, all summaries of follow-up data other than the data completeness and study conduct reports (pooled across the two treatment groups) will be restricted to the confidential closed report. For the planned sample size re-estimation, the pooled proportion of participants in each level of the seven-level ordinal outcome at Day 7 will be provided to the blinded study statisticians and study leadership.

Closed report to the DSMB

All data summaries in the closed report will be by (masked) treatment group. Comparisons between treatment groups will be by intention-to-treat among those randomized (efficacy outcomes) and among those receiving any infusion (safety outcomes). Specific details are given in the indicated sections. The closed reports for a full review will contain:

- Specific data summaries requested by the DSMB or study leadership
- Data summaries in the open report, by treatment group (enrollment, baseline characteristics, eligibility violations)
- Data summaries to assess safety of the investigational treatment including infusion reactions, AEs, SAEs, deaths, composite primary safety outcome are described in Section 7.2. Data summaries for the primary “efficacy outcomes”, and selected secondary outcomes will also be included in each report because these data contain information about the risk/benefit profile of hIVIG. Analyses are described in Sections 7.1, 7.4, and 7.5.
- Summaries on data completeness and study conduct, described in Section 8.
- Interim monitoring boundaries for the primary safety outcomes (Section 7.2).
- Interim monitoring boundaries for efficacy when sufficient data have accrued (Section 7.1).
- Futility analyses when sufficient data have accrued (Section 7.3).
- Listings of incident (new or increase in severity from baseline) grade 3 and 4 adverse events, serious adverse events (SAE), unanticipated problems (UP), suspected unexpected serious adverse reactions (SUSAR), and deaths (Section 7.2).

3 Enrollment

For the open report, the following enrollment and eligibility summaries will be provided:

- Enrollment over calendar time: plot by day or week, cumulative and increments
- Enrollment by country: number (%)

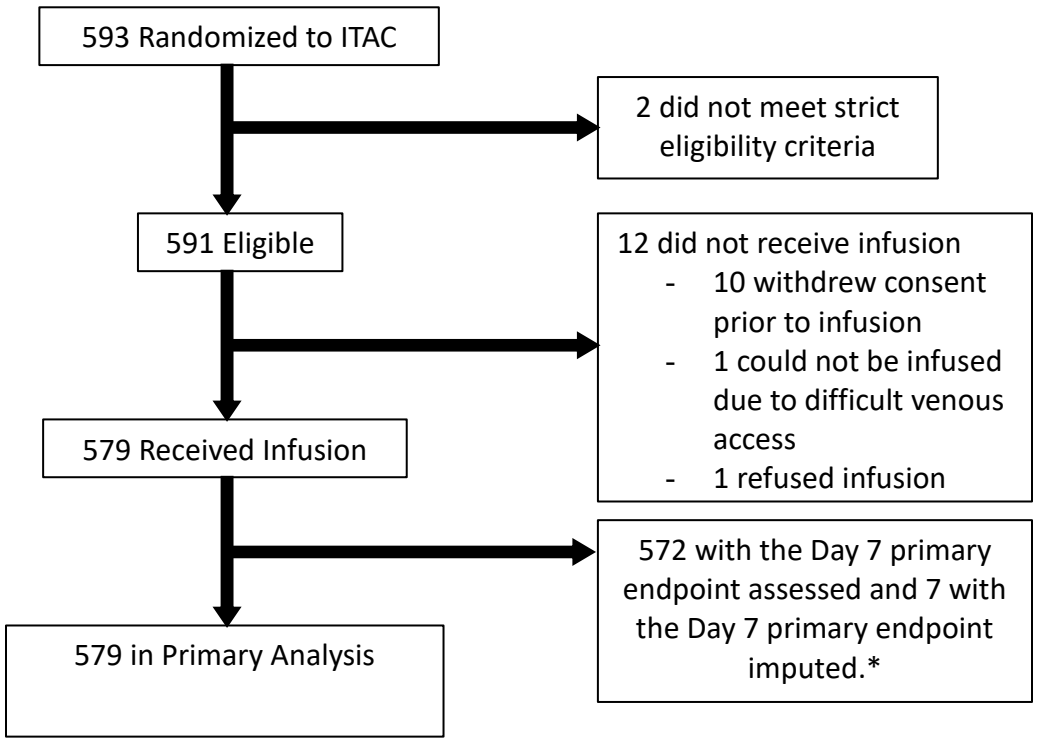
These summaries will be provided overall and by study product/matching placebo randomization stratum.

For the closed report, enrollment will be summarized by treatment group.

Eligibility violations will be reported as protocol deviations (see Section 8).

4 Analysis Populations

All analyses for both efficacy and safety outcomes, will be carried out for a modified intention-to-treat (mITT) population. This mITT population will include all participants who met the eligibility criteria (2 participants did not) AND who received an infusion (12 participants were not infused, 10 who withdrew consent before the infusion, one for whom venous access could not easily be achieved, and one who refused the infusion but continued in follow-up). The mITT population includes 579 participants, 97.6% of the 593 participants randomized.



*6 were discharged with no further information, 1 was transferred to another hospital

5 Baseline Characteristics

Baseline characteristics will be based on information collected on baseline and screening forms.

Baseline characteristics will be summarized by randomized treatment group and overall. Unless noted otherwise, categorical variables will be summarized with frequency (percentage) in each category, and continuous variables will be summarized with

median (25^{th}, 75^{th} percentile) and/or mean (standard deviation). The following characteristics will be reported. Whether the variable will be summarized as a continuous or categorical covariate (and the categorization used) is given in brackets as needed.

- Demographics
  - Age [<50, 50-59, 60-69, ≥70 years; and summary as continuous variable]
  - Sex at birth [male, female]
  - *Race/Ethnic group: [Asian, Black, Latino/Hispanic, white, other]*
  - Country of enrollment
  - *Geographic region*
- COVID-19 related characteristics
  - *Duration of symptoms prior to enrollment (<6, 6-7, 8-9, 10-12 days)*
  - Use of remdesivir prior to enrollment
  - Ordinal outcome category
  - National Early Warning Score (NEWS) [summary as continuous variable]
  - Oxygen saturation
  - Respiratory function scale (modified Borg dyspnea scale; continuous outcome)
  - Receipt of SARS-CoV-2 vaccination (active or control)
  - Upper respiratory SARS-CoV-2 viral RNA
- Other clinical characteristics
  - *Concomitant treatments including corticosteroids and antiplatelet/anticoagulant medications*
  - *Corticosteroid use according to oxygen requirements*
  - History of chronic conditions (heart failure, diabetes, asthma, chronic obstructive pulmonary disease, hypertension requiring medication, renal impairment, hepatic impairment, malignancy, MI, stroke)
  - Requirement of continuous chronic supplemental oxygen
  - *Body mass index (BMI) [<25, 25.0-29.9, 30-39, 40+ kg/m^2]*
  - Pregnancy
  - *Participants with and without compromised immune function; participants with HIV, an immunosuppressive condition other than HIV, or taking antirejection*

medication, immune modulators, or biologic treatment for autoimmune disease or cancer will be considered to have compromised immune function.

- *BMI < 40 and ≥ 40 kg/m^2 according to history of diabetes (4 groups)*
- *Hypertension with and without history of other metabolic and vascular co-morbidity (4 groups): i) no hypertension or other metabolic/vascular co-morbidity; ii) hypertension without a metabolic/vascular co-morbidity; iii) metabolic/vascular co-morbidity without hypertension; and iv) hypertension and a metabolic/vascular co-morbidity. Metabolic/vascular co-morbidities include a history of diabetes, a cerebrovascular event, heart failure, or an MI or acute coronary syndrome.*
- *Number of vascular co-morbidities (0, 1, 2, 3+)*
- *Quartile of Charlson Comorbidity Index (for conditions assessed)*
- *Risk calculator for vaccine prioritization (JHU)*
- *Quartile of risk score for Day 7 outcome that considers baseline antigen and antibody levels, age, gender, duration of symptoms, oxygen saturation, ordinal categorty, NEW score, and history of chronic health conditions.*

- Laboratory values [as continuous outcomes and grade 3 or 4 abnormalities according to the *DAIDS AE Grading Table*]

6 Administration of Study Treatment

These data are an important part of the safety reviews by the DSMB, with particular emphasis on infusion-related reactions and symptoms occurring during or within up to 2 hours after the infusion. These reactions and symptoms will be graded according to the DAIDS AE Grading Table.

The administration of study treatment is also an essential element of study conduct. Several summaries, pooled across treatment groups, will be included in the open report or provided to study leadership. Any summaries of adverse events or infusion-related reactions are restricted to the closed report.

Following the completion of the trial, the summaries below will be used to describe the infusions given to each treatment group (hIVIG and control). Some of the summaries will be carried out for all participants who meet the strict eligibility criteria; safety summaries will be summarized for participants in the mITT population.

These summaries will be stratified by study product/matching placebo group.

- Number and percentage of participants receiving complete infusion, partial infusion, or not infused among all randomized participants.

- Number and percentage of participants for whom infusion was interrupted.
- Number and percentage of participants with infusion-related reactions and symptoms (reported during the infusion or within 2 hours after the infusion), by grade.
- Number and percentage of participants with an incident grade 3 or 4 AE, SAE, UP or SUSAR on the day of infusion. Types of AEs will be summarized by system organ class and by grade.
- Number and percentage of participants who received:
    - Prior to infusion, medication to prevent infusion reactions, and type of medication among all randomized participants
    - During or within 2 hours after infusion, medication to treat infusion reaction, and type of medication
- The day the infusion began (same day as randomization, next day, > 1 day after randomization), and time between randomization and beginning of infusion (median hours, 25^{th}, 75^{th} percentiles).
- Among participants receiving full infusion, duration of infusion (median minutes, 25^{th}, 75^{th} percentiles).
- Time from pooling of infusion bag (beginning of preparation of study product/matching placebo by the pharmacist) to the end of the infusion.
- *Actual dose received, infusion rate, infusion volume, and percentage who received the 400 mL dose (dose was capped at 400 mL corresponding to those who weighed 100 kg or greater).*
- Remdesivir:
    - Number and percentage of participants who received (any) remdesivir, and number of days remdesivir was administered: median, 25^{th}, 75^{th} percentiles, distribution (< 5 days, 5 days, > 5 days).
    - Number and percentage of participants who received remdesivir prior to the day of randomization, and number of doses (median, 25^{th}, 75^{th} percentiles).
    - On the day of randomization: Number and percent of participants who received remdesivir prior to the hIVIG/placebo; after the hIVIG/placebo; no remdesivir.

For these outcomes related to administration of study product, treatment groups will be compared using stratified Cochran Mantel Haenszel test stratified by study product/matching placebo group for binary outcomes and Wilcoxon rank-sum test for continuous outcomes.

Section 7.2 outlines halting rules for pausing the trial due to infusion related adverse events.

7 Statistical Analyses

All analyses will utilize 2-sided tests with a 0.05 significance level. Analyses will compare hIVIG treatment to placebo pooled across study products using the mITT population unless stated otherwise.

7.1 Primary Efficacy Analysis

The primary efficacy endpoint is the seven-level ordinal outcome described in Section 1.4.

For the primary endpoint, the percent of participants in the seven categories of the ordinal outcome will be compared. A proportional odds model will be used to estimate a summary OR (the ratio of the cumulative odds of being in a better category of the ordinal outcome for hIVIG versus placebo). The model uses the cumulative probabilities of being in any of categories 1 up to a threshold (or cutoff) to define six cumulative odds corresponding to cutoffs at categories 1,2,..,6. The model will include a single indicator for treatment, indicators for the participant's clinical state at entry (categories of ordinal outcome), and its two-way interactions with the six cutoffs for each of the six cumulative odds of improvement. The model will also include indicators for which of the four study product/matching placebo group was used and their two-way interactions with the six cutoffs. The primary test statistic will be a Wald test of the coefficient for the treatment indicator.

For the primary endpoint analysis only, multiple imputation based on baseline and follow-up data will be used to estimate participant status at Day 7 for participants with missing data on the ordinal outcome. Imputation will not be performed for secondary endpoints or for subgroup analyses of the primary endpoint.

Of the 7 participants with a missing Day 7 primary ordinal outcome, 6 were known to have been discharged. For these participants it will be assumed that they are in one of the three most favorable categories (3 of the 6 discharged participants were on oxygen the day of discharge). For this imputation, for those discharged by Day 7, in addition to treatment group, the following baseline covariates will be considered: age, clinical status based on the ordinal outcome at enrollment, duration of symptoms, presence of any comorbidity, and NEW score. In addition to the baseline covariates, the day of discharge after randomization, and oxygen status on the day of discharge will be used in the imputation.

For the participant who was transferred to another hospital on Day 2 and who was known to be alive on Day 27, the ordinal category on Day 2 will be imputed for the Day 7 primary outcome.

We will impute ten data sets; parameter estimates (e.g., the summary odds ratio) from the 10 multiply imputed datasets will be combined using Rubin's combining rules.

Interim monitoring boundaries for superiority

The DSMB is to consider a recommendation for stopping the trial early for efficacy only if there is clear and convincing evidence of superiority of the hIVIG versus the control group with respect to the primary outcome.

For monitoring superiority, the Lan-DeMets spending function analogue of the O'Brien-Fleming boundaries will be used, with a 1-sided 0.025 level of significance over multiple looks. The boundary for harm is asymmetric, requiring less evidence to stop for harm than for superiority, described in Section 7.2. For computing the Lan-DeMets boundary, the information fraction at each interim analysis will be the observed total number of participants with Day 7 ordinal outcome data divided by the planned number of participants.

It is important that each hIVIG product is well-represented in the number of participants enrolled at the time of a recommendation by the DSMB to stop early due to convincing evidence of efficacy. Thus, we recommend that early termination for efficacy not be considered until at least 250 participants have been enrolled and at the time such a decision is made the DSMB also consider the number assigned each product. As a guideline we recommend that at least 20% of the information (number with a Day 7 outcome), be from each hIVIG product.

At each interim analysis the following will be provided:

- The value of the primary test statistic (Wald test statistic defined as the standardized estimate of the summary log OR) from the cumulative logistic regression model, plotted over information time, at the current DSMB review, and the corresponding values of the test statistic presented at the previous reviews. The graph will also show the O'Brien-Fleming boundary with Lan-DeMets alpha-spending function. Boundaries will be shown for a one-sided test with alpha=0.025 for superiority of hIVIG, and an asymmetric, Haybittle-Peto boundary for harm (2.5 standard deviations for the first 100 participants; 2 standard deviations thereafter).
- The primary safety outcome is a composite of grade 3 and 4 AEs, SAEs, or death through Day 7, as described in Section 7.2 below (primary safety outcome). Along with the overall primary outcome, this measure will be used to assess whether benefits of the treatment outweigh the risk.
- Estimated proportion of participants in each level of the Day 7 ordinal scale by treatment group.
- The summary ORs from fitting a similar cumulative logistic regression model for the ordinal outcome at Days 3, 5, 14, and 28.
- History of the estimated ORs from the cumulative logistic regression model with 95% CIs and p-values at previous DSMB reviews, as presented.

Assessment of model assumptions

A test for the proportional odds assumption will be made from a model that allows different effect estimates for the hIVIG versus placebo according to the cut-off of the ordinal scale (a partial proportional odds model). That is, a cumulative logistic regression model with the same terms as above but including two-way interactions between the treatment indicator and the six cut-offs. A composite likelihood ratio test will be used to determine if any of the additional terms are significantly different from zero (i.e., a test of the proportional odds assumption). Even if the proportional odds assumption is violated, the overall summary OR will be the basis for inference for the primary analysis.

Sample size re-estimation

The sample size re-estimation was conducted by blinded statisticians and did not use any information on the treatment effect. The purpose was to re-estimate the sample size needed to obtain 80% power with an assumed odds ratio of 1.61 based on the distribution of the ordinal outcome at Day 7 observed in the trial pooled across study arms (as opposed to the hypothesized distribution informed by data from the ACTT-1 trial).

Sensitivity Analyses

- In order to further characterize the summary OR and any deviations from proportional odds, separate ORs will be estimated for each of the six dichotomized definitions of improvement that can be formulated from the components of the ordinal outcome. That is, we will fit six separate logistic regression models for each of the six dichotomized definitions of improvement. These models will include an indicator for treatment arm, study product/matching placebo group, and baseline ordinal outcome group.
- Categories 3 and 4 of the primary ordinal outcome differ in part by the amount of supplemental oxygen required, and a single cut point (4 liters/minute) defines the difference. Since, together, these two categories of the ordinal outcome are expected to include approximately 30% of participants, an analysis that combines these two categories (resulting in a six-category ordinal outcome) will be carried out to supplement the primary analysis using the same methods described above.
- *The adjusted cumulative logistic regression model for the final analysis will be refitted but exclude the 7 participants for whom the Day 7 outcome is missing (i.e., no multiple imputation will be performed).*
- *The adjusted cumulative logistic regression model for the final analysis will be refitted and include all participants randomized with a Day 7 primary outcome.*
- *The adjusted cumulative logistic regression model for the final analysis will be refitted and include a "worst case imputation". This was recommended by the FDA. Among the 593 participants randomized, 18 (3.0%) are missing the Day 7 outcome. Of these 18 participants, partial follow-up data are available for 7*

participants. For 11 participants, all who were not infused, there are no follow-up data. Given the double-blind nature of the trial, a worse case analysis (e.g., imputation of category 1 for those in the placebo group and category 7, death, in the hIVIG group) for these participants will not be done. These 11 participants will be excluded from this sensitivity analysis. Among the other 7 participants, 6 were discharged before Day 7. For these 6 participants, we assumed those in the placebo group were in category 1 and those in the hIVIG group were in category 4. For the other participant who was transferred to another hospital on Day 2 and was known to be alive on Day 27, category 1 will be imputed on Day 7 if assigned to the placebo group, and category 2 will be imputed on Day 7 if assigned to the hIVIG group. This is the same category of ordinal outcome the participant was in on Day 2 (i.e., no change is assumed).

- The adjusted cumulative logistic regression model will be refitted but will include geographic region as an additional covariate.

7.2 Safety Analyses

The following safety and tolerability outcomes will be analyzed:

- The **primary safety endpoint** is a composite of grade 3 and 4 adverse events, SAEs, or death through Day 7. The percent of participants experiencing the composite safety outcome will be compared. A logistic regression model will be used to estimate an OR for hIVIG versus placebo. The logistic regression model will include indicators for treatment group, study product/matching placebo group, and baseline ordinal outcome category. Over the first 28 days, the cumulative proportion of participants with a SAE or death will be estimated using Kaplan-Meier curves, by treatment group. The hazard ratio (HR) for hIVIG versus placebo will be estimated with a 95% CI using a Cox proportional hazards model with an indicator for treatment group, stratified by study product/matching placebo group and baseline ordinal outcome category.

  Stopping boundaries for harm: A Haybittle-Peto boundary of 2.5 standard deviations (SD) for the first 100 participants enrolled and 2.0 SD afterwards will be used as a guideline for harm. The SD refers to the standard deviation of the test statistic (standardized estimate of the summary log OR). With this boundary, less evidence is needed for stopping a trial early due to harm compared with stopping for efficacy.

  Similar to the efficacy analysis, the observed value of the test statistic for the primary safety outcome will be plotted over information time, for the current data, along with the boundaries and the values presented at previous DSMB reviews.

- Safety analyses will also include infusion reactions collected during or within 2 hours after the infusion of hIVIG/placebo. Percentages of participants who experience infusion reactions or prematurely terminated infusions will be

summarized by treatment groups and compared as described in the preceding section.

- Other safety analyses will be conducted including the following outcomes:
  - Composite of grade 3 and 4 adverse events, SAEs, or death through Day 7 excluding exempt events
  - Composite of grade 3 and 4 adverse events, SAEs, or death through Day 7 by MedDRA system organ class (SOC)
  - Composite of SAE or death through Day 28 by MedDRA SOC
  - Each component of the primary safety outcome analysed separately (deaths, SAEs nonexmpt, SAEs exempt, and grade 3 and 4 adverse events)
  - A composite of SAEs and death through Day 28 (including and excluding exempt events and each component separately analyzed)
  - Prevalence of clinical AEs of any grade on Days 0, 1, 3, 7, and 28; AEs will be summarized by grade and day, and by MedDRA® system organ class and grade. (AEs present on those days).
  - Summaries of UPs and SUSARS, and listings of SAEs, UPs, SUSARs and deaths.
  - Change in laboratory test values from baseline to Day 7, and incidence of grade 3 and 4 laboratory abnormalities at Day 7.

AEs will be coded with MedDRA, version 23.1. In addition to the tables, listing of participants with grade 3 or 4 events, SAEs or deaths will be provided with the MedDRA preferred term (PT) and study day of AE.

Further safety assessments may be considered including by study product/matching placebo group (see Section 7.5).

Because the infusion volume in this protocol is significant (250 mL for remdesivir and up to 400 mL for hIVIG/placebo), as a guideline, the DSMB will be asked to consider halting enrollment if more than 5% of participants experience a grade 3 or 4 infusion AE or if more than 10% do not complete the infusion due to an AE(s). This will be informed by the lower bound of the 95% confidence interval which will not be adjusted for multiplicity. When this occurs, differences will be compared by randomized group. If the study is temporarily halted or stopped for safety reasons, institutional review boards/ethics committees will be informed.

7.3 Monitoring for Futility

To assess futility, conditional power calculations based on an unadjusted model (as was done for the original power calculations) for the Day 7 ordinal outcome will be presented under a range of scenarios. In the primary futility analysis, it will be assumed that the treatment effect for the future, as yet unobserved follow-up, will be as hypothesized in

the study design (adjusted OR = 1.61). As secondary analysis, the treatment effect for future follow-up will be assumed to be similar to the observed effect. Additional scenarios may be provided. Typical futility guidelines recommend stopping a trial when conditional power (assuming the originally hypothesized treatment effect for the future, as yet unobserved follow-up) is below 10%-20%.

As a guideline, futility will first be assessed after 50% of the planned number of participants have Day 7 ordinal outcome data, and a value of 20% will be suggested as a threshold for the conditional power. Conditional power will be computed using the test statistic for the treatment indicator in a cumulative logistic regression model.

Decisions to terminate the study for futility will include a broad assessment of the risk/benefit trade-off in addition to these guidelines.

7.4 Secondary Outcomes

The protocol defines a number of secondary endpoints in addition to the two key endpoints described in the previous section. These analyses will be carried out for the final report. No adjustment for multiplicity for all the treatment comparisons for the secondary outcomes will be made; they are supportive to the primary endpoint analysis.

Selected secondary endpoints may also be analyzed for the interim monitoring report to help evaluate the safety and efficacy of hIVIG.

Multiple imputation will only be used for data missing on the primary endpoint; all secondary endpoints will use complete case.

Below, the secondary outcomes from the protocol are cited, with a short description of the analysis methods. The secondary outcomes are grouped by analysis methods and are not listed in order of importance.

1. The primary ordinal outcome on Days 3, 5, 14 and 28.

2. Pulmonary only components of the primary outcome measure at Days 3, 5, 7, 14 and 28

3. Thrombotic components of the primary outcome measure (stroke, myocardial infarction, venous and arterial thrombosis or embolism, plus disseminated intravascular coagulation) at Days 3, 5, 7, 14, and 28.

4. Clinical organ dysfunction

5. The ordinal outcome similar to the one used in the ACCT trial of remdesivir will be used to compare treatment groups at Day 7. The ACCT ordinal outcome included 8 categories. We can only approximate the categories because data were not collected concerning hospitalization for infection-control reasons. It is similar to the ITAC primary ordinal outcome but

considers hospitalization status. The following 8 categories will be defined for this ordinal outcome:

- Not hospitalized, no limiting symptoms due to COVID-19
- Not hospitalized, limiting symptoms due to COVID-19, home oxygen requirement, or both
- Hospitalized, no limiting symptoms due to COVID-19, not requiring supplemental oxygen
- Hospitalized, limiting symptoms due to COVID-19, not requiring supplemental oxygen
- Hospitalized, requiring any supplemental oxygen
- Hospitalized, requiring non-invasive ventilation or use of high-flow oxygen devices.
- Hospitalized, requiring invasive mechanical ventilation or ECMO
- Death

6. Percentage in 2 most favorable categories at Day 7.

7. Hospitalization status at Days 14 and 28.

8. Hospitalization status (a binary outcome, alive and discharged from the hospital to home or rehabilitation OR able to independently undertake usual activities with minimal or no symptoms versus dead or hospitalized AND unable to independently undertake usual activities) at Days 7, 14 and 28.

   For outcomes 1-8, the proportion of participants in each category will be summarized by treatment group. For the ordinal outcomes, a summary OR will be estimated using a cumulative logistic regression model which includes a single indicator for treatment, indicators for the participant's clinical state at entry (categories of ordinal outcome), its two-way interactions with the six cutoffs for each of the six cumulative odds of improvement, indicators for which of the four study product/matching placebo group was used and their two-way interactions with the six cutoffs. For the binary outcomes, a summary OR will be estimated using a logistic regression model which includes a single indicator for treatment, indicators for the participant's clinical state at entry (categories of ordinal outcome), and indicators for which of the four study product/matching placebo group was used. For interim analyses we will consider clinical organ dysfunction as a binary composite endpoint at Day 7 and Day 28. At the end of the trial we will take into account the severity of each organ dysfunction by developing a weighting scheme by examining the pooled association of each item with subsequent death.

9. All-cause mortality through Day 28.

10. Time to the three least favorable categories of the primary outcome measure.

For outcomes 9 and 10, the cumulative incidence of death (outcome 9) or the three least favorable categories (outcome 10) by treatment group will be estimated using Kaplan-Meier methods. The treatment groups will be compared using a log-rank test. A summary HR comparing the treatment groups will be estimated using a Cox proportional hazards model. These models will be stratified by study product/matching placebo group and baseline ordinal outcome group.

11. Time to the two most favorable categories of the primary outcome measure.

12. Time to discharge (this is similar to the recovery outcome used in the ACTT-1 trial)

13. Time to discharge or being able to independently undertake usual activities with minimal or no symptoms

14. Time to being able to independently undertake usual activities with minimal or no symptoms (discharge status will be ignored)

    For outcomes 11-14, time-to-event methods that take into account the competing risk of death will be used. Specifically, Gray's test with rho=0, the Fine-Gray model, and the Aalen-Johansen estimator for the cumulative incidence curve are the competing risk equivalents to the log-rank test, Cox proportional hazards model, and the Kaplan-Meier estimator for the cumulative proportion of participants with the event, respectively. Gray's test and the Fine-Gray model will be stratified by study product/matching placebo group and baseline ordinal category.

    Additionally, as a sensitivity analysis for the analysis of time-to-discharge and days outside the hospital we will exclude participants from the 5 sites where no participants were discharged prior to Day 7.

15. Days alive outside of a hospital through Day 28

16. Days alive and able to independently undertake usual activities with minimal or no symptoms through Day 28

17. Days alive and out of the hospital or able to independently undertake usual activities with minimal or no symptoms through Day 28 (whichever lead to greatest time)

    These outcomes (15-17) have been used in other trials of therapeutics for COVID-19. We will sum the number of days that each individual spends outside a short-term acute care hospital up to 28 days. A person who dies within 28 days will be assigned a value 0, consistent with the approach taken

in trials of intensive care-based interventions. We will present the mean and median days by group. We will test the hypothesis of no mean difference between arms using methods for continuous outcomes (ANCOVA models), with baseline baseline ordinal outcome group and study product/matching placebo group as covariates. Because the residual distribution is unlikely to be normally distributed, we will use robust or sandwich standard errors. This analysis will be undertaken only when complete follow-up data are available. The outcome does have limitations due to its handling of death and withdrawal. We expect that there will be minimal missing data but may use multiple imputation for the final analysis.

18. Change in immunoglobulin levels (IgG, IgG subclasses, IgM, IgA) and neutralizing antibody titres from baseline to Days 1, 2, 3, 7, 28 and 90.

    Longitudinal random effects models will be used to summarize log-transformed antibody level differences between the hIVIG and placebo groups at Days 1, 3, 7 and 28 of follow-up. Baseline antibody levels will be included as a covariate in these models. For the subset of participants for whom blood is collected at Day 90, antibody levels will be compared.

    Exploratory analyses will be carried out if hIVIG is efficacious to determine if antibody differences post-infusion on Days 1 and 3 predict the primary outcome. First, post-infusion antibody level (perhaps also considering the pre-infusion level or change from baseline) will be considered as a predictor of the Day 7 ordinal outcome using a proportional odds model. If there is evidence that the post-infusion level or change predicts the Day 7 outcome, a model which includes the post-infusion level and treatment will be fit to determine the impact on the treatment estimate without the antibody level in the model. Second, we will examine the association between antibody treatment differences post-infusion by treatment group and the Day 7 summary OR from the cumulative logistic model. This could be done for each product separately (four data points) and/or according to the potency of each lot, grouped into more than four groups.

7.5 Subgroup Analyses

Subgroup analyses for the primary seven-category ordinal outcome (primary efficacy outcome), as well as for the primary safety outcome (Grade 3 and 4 adverse events, SAE or death through Day 7) will be performed to determine whether and how the treatment effect (hIVIG versus placebo) differs qualitatively across various subgroups defined at baseline, and whether there are safety concerns in specific subgroups.

The protocol denotes the subgroup analysis by the duration of symptoms at study entry as the "key subgroup analysis." In addition to the analysis by 4 categories of duration of symptoms (<6, 6-7, 8-9, 10-12 days), an analysis based on the upper quartile of symptom duration (e.g., less than or equal to the 75th percentile versus greater than the

75th percentile) (2 categories). The upper quartile is 10-12 days (75% of participants will have symptom duration < 10 days). In ACCT-1, a more severely ill target population than studied here, there was no limit to the duration of symptoms and the median was 9 days (25th and 75th percentile, 6 – 12 days). The quartile definitions for duration of symptoms will be determined following the completion of enrollment. For interim analyses, the quartiles will be determined based on interim data. For those with shorter duration of time since symptom onset, the treatment effect is hypothesized to be greater than among participants who have had symptoms for a longer period of time. A global test for heterogeneity of the treatment effect across the symptom duration subgroups will be carried out in 2 ways: 1) by adding the interaction between symptom duration as a continuous variable (1 df test) and the treatment group to the model; and 2) by adding the interaction between categories of symptom duration and the treatment group to the model (3 df test).

Other important subgroups include subgroups by disease severity, by age, and by pre-existing conditions. *A priori* we have no reason to believe the clinical efficacy or safety of hIVIG compared to placebo will be substantially different in relative terms in any of the following subgroups considered. Subgroup analyses for the primary efficacy endpoint and safety endpoint will use the adjusted (cumulative) logistic models described earlier. ORs with 95% CIs comparing the treatment group versus control will be estimated for each subgroup. Global tests for heterogeneity of the treatment effect across subgroups will be carried out by adding the interaction between the subgroup indicator and the treatment group indicator to the model. In case the subgroup was formed by categorizing a continuous variable, the interaction term will be formed between the subgroup indicator and the continuous variable.

Subgroup analyses will not be adjusted for multiple comparisons; they are supportive to the primary endpoint analysis. Subgroup analyses will be interpreted with caution due to limited power and uncontrolled type I error.

Subgroup analyses will be performed for a number of baseline factors. Unless otherwise stated, continuous outcomes will be summarized in quartiles. The following subgroups will be considered:

- *Age (18-49, 50-59, 60-69, 70+ years)*
- Sex at birth (male, female)
- Race/ethnicity (Asian, Black, Latino/Hispanic, White, other)
- *BMI [<25, 25.0-29.9, 30-39, 40+ kg/m^2]*
- *Presence of chronic medical conditions which had greater than 5% prevalence at baseline (diabetes, hypertension, COPD, asthma, renal impairment)*
- *Geographic location (U.S.; Europe, UK, or Israel; and Argentina, Indonesia, Japan or Nigeria)*
- Upper respiratory SARS-CoV-2 viral load

- Oxygen saturation level
- Dyspnea severity (Modified Borg dyspnea scale)
- Organ/respiratory dysfunction category based on ordinal primary outcome
- NEWS
- *Participants with and without compromised immune function; participants with HIV, an immunosuppressive condition other than HIV, or taking antirejection medication, immune modulators, or biologic treatment for autoimmune disease or cancer will be considered to have compromised immune function.*
- *BMI < 40 and ≥ 40 kg/m^2 according to history of diabetes (4 groups)*
- *Hypertension with and without history of other metabolic and vascular co-morbidity (4 groups): i) no hypertension or other metabolic/vascular co-morbidity; ii) hypertension without metabolic/vascular co-morbidity; iii) metabolic/vascular condition without hypertension; and iv) hypertension and a metabolic/vascular co-morbidity. Metabolic/vascular co-morbidities include a history of diabetes, a cerebrovascular event, heart failure, or an MI or acute coronary syndrome.*
- *Number of vascular co-morbidities (0, 1, 2, 3+)*
- *Quartile of Charlson Comorbidity Index (for conditions assessed)*
- *Risk calculator for vaccine prioritization (JHU)*
- *Corticosteroids, overall and in combination with oxygen requirements (ordinal category at baseline)*
- *Use of antiplatelet/anticoagulant therapy (prophylactic heparin, intermediate or therapeutic heparin or other anticoagulant therapy, none)*
- *Quartiles of disease progression risk score (defined using pooled treatment groups with the following baseline predictors of the Day 7 ordinal outcome: age, gender, antigen and antibody level, duration of symptoms, oxygen saturation level, ordinal outcome category at entry, NEWS, and chronic health conditions). .*

We will analyse the association between lot potency and the primary outcome for participants assigned active hIVIG using a cumulative logistic regression model adjusted for hIVIG product and baseline ordinal category. The aim of this analysis will be to determine if the primary outcome varies by range in potency among the various lots of hIVIG used.

Additionally, we will explore if the association between lot potency and the primary outcome differs by baseline antibody titre. The potency of each lot will be measured by

Texcell. In addition to the subgroup analyses above, a subgroup analysis by lot potency (by tertiles) will be carried out. Participants in the placebo group will be classified according to the lot potency they would have received had they been randomly assigned to the hIVIG group.

If there is a beneficial effect of hIVIG compared to placebo, in order to support regulatory claims for each hIVIG product used, sensitivity analyses comparing each hIVIG product to its matching placebo will be carried out for key efficacy and safety endpoints. First, we will assess the evidence for any difference in effect among the products. A test for heterogeneity of the treatment effect across study products will be carried out by adding the interaction between study product/matching placebo group and the treatment group indicator to the model. In addition to the models for efficacy and safety endpoints described above, we will also test for heterogeneity of the treatment effect across study product/matching placebo group after adjusting for lot potency. In one of the subgroup analyses by hIVIG product, the subgroups will be further divided by the potency level of the hIVIG lots used for that product (e.g., above and below the median level of lot potency for the product).

Second, we will obtain estimates of the effect for each product using two general approaches. a) We will estimate the effect for each hIVIG product using only data from participants receiving that particular product/matching placebo (e.g., separate analyses for each study product). b) We will use multisource exchangeability models[ii] (MEMs) to dynamically borrow information from other study products to improve estimation of the efficacy and safety of a single study product. MEMs work by enumerating all possible exchangeability patterns between the data sources (here data from different study products) and obtain a single posterior distribution for the parameter of interest using Bayesian model averaging. The key benefit of this approach is that it can borrow differentially from different study products. Estimates of the effect for each product will be done both adjusting for and not adjusting for lot potency.

These analyses will consider that each hIVIG product will be used by a different group of clinical sites (i.e., each comparison will represent a small multi-center trial), and that power will likely be very low for all of the outcomes. Because the hIVIG product that each site receives is not randomized, any comparisons of the efficacy among hIVIG products should be interpreted cautiously. These analyses are referred to as sensitivity analyses because overall therapeutic efficacy and safety will be based on the pooled analysis of the four hIVIG products with placebo.

7.6 Analyses of Stored Specimens

NIH plans to measure antibody and antigen levels on plasma samples from ITAC. Antibody levels will be determined using kits made by Bio-Rad, which measures total (IgA, IgG, and IgM) anti-nucleoprotein (NP), and by GenScript (anti-spike neutralizing antibody surrogate), which measures a subset of antibodies capable of inhibiting binding by spike proteins. Antigen levels will be determined in plasma using an assay

made by Quanterix.

Results of the Bio-Rad antibody measurement are reported in terms of “specimen ratios”. Specimen ratios are defined as the specimen optical density (OD) divided by the optical density of the cut-off control R4 (OD_MR4). According to the manufacturer, specimen ratios less than 0.8 are considered negative, those with a specimen ratio between 0.8 and 1.0 are considered equivocal, and those greater than 1.0 are considered positive for the presence of anti-SARS-CoV-2 antibodies.

Results of the GenScript antibody assay are reported as binding inhibition percentages. For this assay, levels less than 30% are considered negative and those ≥ 30% are considered positive.

Results of the quantitative Quanterix assay are reported in pg/mL. The lower level of detection is 3 pg/mL.

Subgroup analyses will be carried out using the antigen and antibody data at baseline and the Day 7 pulmonary and pulmonary+ ordinal outcomes at Day 7.

Our hypothesis is as follows: Patients with negative or lower positive antibody levels will benefit more from hIVIG compared to placebo than patients with higher antibody levels. Furthermore, those with lower antibody levels AND with higher antigen levels, will benefit more from hIVIG compared to placebo than other subgroups categorized by both antibody and antigen levels.

8 Data Completeness and Study Conduct

The primary outcome (seven-level ordinal outcome) will be assessed daily through Day 28. In-person visits are scheduled on Days 1, 2, 3, 7, and 28, when blood is collected (plasma and serum); other visits on Days 5 and 14 may be conducted by phone. Participants at selected sites will return for a visit 90 days after randomization to obtain a blood draw; this subset will comprise all participants at selected sites where return for a later visit is practical for participants.

Data completeness and study conduct reports will be provided by treatment group (for the closed report) and pooled across treatment groups (for the open report). Data summaries for the infusion of hIVIG/placebo on Day 0 are described in Section 6; several of those reports are also relevant for monitoring study conduct and will be included in the open report or provided to study leadership, pooled across treatment groups.

9 The following data summaries will be provided:

- Number, percent and type of protocol deviations. Specific protocol deviations are reported in the protocol.

- Expected and observed number (% of expected) of participants who completed visits on Days 1, 2, 3, 5, 7, 14, 28, and 90.
- Ascertainment of the primary outcome: Expected and observed number (% of expected) of participants with outcome status for the ordinal outcome (Days 0-7, 14, and 28).
- Expected and observed number (% of expected) of participants with known vital status at Days 0-7, 14, and 28.
- Number and percent of participants who withdrew consent, or with missing primary outcome data for other reasons will be summarized.
- Listing of participants who withdrew consent, including dates of randomization, study product/matching placebo group, receipt of study treatment, date of withdrawal, and reason of withdrawal.
- Length of follow-up: Median, 25^{th}, 75^{th} percentiles, range and distribution
- Collection of specimens: Expected and observed number (% of expected) of participants with specimens collected as specified by the protocol, by visit.

A visit counts as “expected” if the visit window has closed or the data have been received.

The summaries for the final report will be provided for the mITT population unless otherwise stated.

10 Addendum to Statistical Analysis Plan for European Medicines Agency

During the review by EMA of INSIGHT 013 (ITAC), An International Multicenter, Adaptive, Randomized Double-Blind, Placebo-Controlled Trial of the Safety, Tolerability and Efficacy of Anti-Coronavirus Hyperimmune Intravenous Immunoglobulin for the Treatment of Adult Hospitalized Patients at Onset of Clinical Progression of COVID-19, it was requested that the data analysis plan be revised to make the 7-category pulmonary ordinal outcome at day 7 (a secondary endpoint) the primary endpoint for submissions to the EMA instead of the 7-category ordinal outcome specified in the protocol (see section 4 and Appendix F of the protocol) which also includes a range of organ dysfunction in addition to respiratory dysfunction. The protocol-defined primary endpoint would be a secondary endpoint for EMA submissions.

An addendum to Version 1.0 of the Statistical Analysis Plan was prepared on January 11, 2021 to address this request from the EMA. It is reproduced here for ease of reference.

The statistical data analysis plan based on the protocol, dated 25 September 2020, will not be changed as it reflects the protocol. For the EMA submission, this document will be submitted with the protocol and the statistical data analysis plan. Should the data analysis plan be modified before the end of the study, this document will be submitted with the updated statistical data analysis plan.

For the EMA submission, the ordinal outcome shown in the table at the end of this section will be used as the primary endpoint. Sample size assumptions, as stated in section 5.5 of the protocol, are not expected to differ for this outcome compared to those stated in the protocol for the primary endpoint. This assumption is supported by results from another trial, INSIGHT 014 (TICO), for which similar endpoints were used. A paper describing the findings from TICO at day 5 for both ordinal outcomes reported that only 2 of 311 participants were in different categories of the two ordinal outcomes (ACTIV-3/TICO LY-CoV555 Study Group, N Engl J Med 2020, doi: 10.1056/NEJMoa20331.30).

The planned data analysis for the pulmonary ordinal outcome described in section 7 of the Statistical Data Analysis Plan for ITAC will be identical to those stated for the protocol-defined primary ordinal outcome. Likewise, the assessment of model assumptions and sensitivity analyses specified in the Statistical Data Analysis Plan will be carried out for pulmonary ordinal outcome as well as the protocol-defined primary endpoint.

Interim monitoring guidelines specified in the protocol for the primary endpoint will not change. Stopping boundaries for substantial evidence of benefit, for harm, and for futility will be based on the protocol-defined primary endpoint.

Sample size re-estimation was recently carried out using pooled (both treatment groups combined) category percentages for the protocol-defined primary endpoint. This re-estimation confirmed that 500 participants will be sufficient to detect an OR of 1.61 with 80% power.

This document was prepared by blinded statisticians and reviewed by the blinded protocol team. This document will be used with the protocol and the statistical data analysis plan to prepare the final study report for the EMA.

TABLE. CLINICAL CATEGORICAL DEFINITIONS FOR PULMONARY ORDINAL OUTCOME

Each participant is categorized in the highest applicable category.

Ordinal Category	Categorical Description	Categorical Definition*
7	Death	Death
6	End-organ failure	Currently requiring invasive assisted ventilation, extracorporeal membrane oxygenation, mechanical circulatory support, vasopressor therapy or renal replacement therapy
5	Life-threatening end-organ dysfunction	Currently requiring non-invasive assisted ventilation or high-flow oxygen or
4	Serious end-organ dysfunction	Currently requiring supplemental oxygen (≥ 4 liters/min, or ≥ 4 liters/min above premorbid requirements**) but not high-flow oxygen
3	Moderate end-organ dysfunction	Requiring supplemental oxygen < 4 liters/min, or < 4 liters/min above premorbid requirements**
2	Limiting symptoms due to COVID-19	Symptomatic and currently unable to independently undertake usual activities
1	No limiting symptoms due to COVID-19	Can independently undertake usual activities with minimal or no symptoms

**Continued hospitalization or presence in a particular category of inpatient facility (e.g. intensive care or high dependency) is not used to divide these categories, as indication for continued hospitalization among recovering COVID patients is intrinsically subjective, in part determined by social and financial factors, and varies markedly across the globe.*

*** Premorbid requirement refers to requirements prior to the development of COVID-19, for example in patients with chronic obstructive pulmonary disease, other chronic pulmonary diseases, or oxygen requirements related to altitude.*

References

[i] Whitehead J. Sample size calculations for ordered categorical data. Stat Med 1993; 12: 2257-2271.

[ii] Kaizer AM, Koopmeiners JS, Hobbs BP. Bayesian hierarchical modeling based on multisource exchangeability. Biostatistics. 2018;19(2):169-184. doi:10.1093/biostatistics/kxx031

Kaizer AM, Koopmeiners JS, Kane MJ, Roychoudhury S, Hong DS, Hobbs BP. Basket designs: Statistical considerations for oncology trials. JCO Precision Oncology. 2019; 3:1-9.

Hobbs BP, Landin R. Bayesian basket trial design with exchangeability monitoring. Statistics in medicine. 2018; 37(25):3557-72.