Statistical analysis plan

Principal Investigator:	John Stone, MD, MPH
Protocol Number and Title:	2020P001159 Tocilizumab to Prevent the Progression of Hypoxemic Respiratory Failure in Hospitalized Non-Critically Ill Patients with COVID-19
Protocol Version and Date:	Version 4.0, 18 August 2020
Author(s):	Nora Horick, Principal Biostatistician Brian Healy, Biostatistician Andrea Foulkes, Senior Biostatistician
SAP Version:	Version 1.0
SAP Version Date:	18 August 2020

Revision History

Version #	**Date** (dd-mmm-yyyy)	**Document Owner**	**Revision Summary**
Version 1.0	18-Aug-2020	Nora Horick	Initial Release Version

Signature

I confirm that I have reviewed this document and agree with the content.

APPROVALS	
DocuSigned by: Nora Horick Signer Name: Nora Horick Signing Reason: I am the author of this document Signing Time: 8/18/2020 | 09:54:32 PDT 63D8A3FE7842495DABACF095A00AE49E	8/18/2020 | 09:54:50 PDT
Nora Horick, MS **Principal Biostatistician**	**Date**
DocuSigned by: Andrea Foulkes Signer Name: Andrea Foulkes Signing Reason: I am the author of this document Signing Time: 8/18/2020 | 13:33:49 PDT 60EA309BB4F64D8598CD0961379BFD9B	8/18/2020 | 13:33:59 PDT
Andrea Foulkes, ScD **Senior Biostatistician**	**Date**
DocuSigned by: John H. Stone, M.D. Signer Name: John H. Stone, M.D. Signing Reason: I approve this document Signing Time: 8/18/2020 | 22:03:59 BST CFF18BA818A54E55A0B14CDEE90A26F7	8/18/2020 | 22:04:12 BST
John H. Stone, M.D., MPH **Principal Investigator**	**Date**

Table of Contents

1 Study design and conduct

1.1 Overview

As of April 2020, COVID-19 had been confirmed in more than 1 million people worldwide, with an estimated symptomatic case fatality ratio of around 1.4%. [1,2] There remains an urgent need for effective treatment to curtail the rate of respiratory failure, the leading cause of mortality in COVID-19 disease. With increasing numbers of patients requiring intensive unit level care and mechanical ventilation, some nations are having to triage patients for ventilatory support due to limited resources and healthcare systems around the world being stretched to the point of collapse, identifying interventions that could prevent the development of respiratory failure for these patients is critical.

Clinical data suggest an immunologic link between COVID-19 and immune dysregulation resulting in macrophage activation syndrome (MAS). Clinical trials are already underway studying the role of immunomodulatory therapy including modulation of IL-1 and IL-6 and downstream pathways in the setting of CAR-T induced MAS (NCT04150913, NCT04071366) and agents such anakinra and tocilizumab have been used in this context with promising results and good safety profiles. Based on the MGH experience thus far with COVID-19, the need for mechanical ventilation has been approximately 30%. We propose a trial of IL-6 receptor blockade with tocilizumab given early in disease course to try to prevent progression of COVID-19.

This is a prospective placebo-controlled, blinded, randomized controlled trial at seven Boston area hospitals: the MGH, the Brigham & Women's Hospital, North Shore Medical Center, Newton-Wellesley Hospital, Boston Medical Center, the Lahey Hospital and Medical Center, and St. Elizabeth's Hospital.

1.2 Study Objectives

The objective of the study is to determine whether the use of early tocilizumab can decrease progression of COVID-19 associated respiratory failure and death.

1.3 Study Population

Study eligibility and exclusion criteria are provided in **Table 1**.

Table 1: Study participant inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
1. Age > 18 and < 86 years old 2. Male or female gender 3. Confirmed SARS-CoV-2 infection by nasopharyngeal swab PCR or serum assay for IgM antibody 4. Requiring hospital but not mechanical ventilation, with oxygen supplementation not greater than 10L delivered by any device 5. Evidence of severe COVID-19 (at least 2 of the following): • Fever > 38C within 72 hours • Pulmonary infiltrate on chest X ray • Need for supplemental O2 to maintain saturation > 92%	1. Unable to provide verbal informed consent or have verbal agreement to participate through attestation and signature of a Witness required, as outlined in the Partners IRB's Table for Consenting in COVID Research that is More than Minimal Risk. 2. Patients between the ages of 79 and 86 will be excluded if they have NYHA Class III/IV heart failure, insulin-dependent diabetes mellitus, angina, or treatment of a malignancy (excluding non-melanoma skin cancer) within six months 3. Uncontrolled bacterial, fungal, or non-COVID viral infection 4. Active tuberculosis

AND at least 1 of the following: • Ferritin > 500 ng/ml • CRP > 50 mg/L • LDH >250 U/L • D-dimer > 1000 ng/mL	5. Any prior investigational immunosuppressive therapy within 28-days or 3 half-lives of the agent (for instance with biologic or JAK inhibitor) 6. Any concurrent immunosuppressive medication that the PI believes would put the patient at higher risk 7. Receipt of intravenous tocilizumab for the treatment of a non-COVID condition within three weeks of the first COVID symptom 8. History of hypersensitivity to tocilizumab 9. Any concurrent immunosuppressive medication that the PI believes would put the patient at higher risk 10. Treatment with other biologic or small-molecule immunosuppressive therapy such as IL1R-antagonism, JAK inhibition, or other agents. 11. Treatment with convalescent plasma 12. History of diverticulitis or bowel perforation 13. ANC <500, Platelets <50,000 14. AST/ALT > 5X ULN 15. Women who are pregnant or planning to get pregnant in the next 90 days

1.4 Treatment Allocation and Concealment

Individuals determined eligible based on the study inclusion and exclusion criteria are randomized to receive tocilizumab 8 mg/kg x 1 (not to exceed 800 mg) vs placebo in a 2:1 randomization ratio using randomly permuted blocks of size 3 and 6. Randomizations are stratified by study site.

The randomization schedule is generated using Sealed Envelope, a web-based randomization service, by the unblinded study Biostatistician. The randomization schedule is provided to the research pharmacist designated for the study in each trial center. The research pharmacist prepares the study drug for infusion on the day of dosing based on the randomization schedule and labels the infusion bag with the subject number but no treatment assignment information. The investigators, study coordinators, study subjects are blinded to the treatment assignment and do not have access to the randomization schedule.

The unblinded treatment information can be provided to the DSMB to facilitate the evaluation of any clinically important increase in the rate of a serious suspected adverse reaction when the treatment information is required to determine if an expedited safety report must be submitted to regulatory agencies. The unblinded biostatistician would provide the treatment assignment of that subject.

If knowledge of the study drug ingredients is needed to manage the subject's condition, the investigator may contact the research pharmacist at the site to obtain the treatment assignment. In the event of an unblinding request, Dr. Stone will also be notified before unblinding is performed. If unblinding occurs for any reason, the time and reason for breaking the blind will be recorded. The number of patients unblinded by arm will be reported.

The unblinded statisticians may review clinical data and identify missing data fields and forward the issues to the study team to facilitate DSMB review. The unblinded statistician may also provide findings on data that are inconsistent across various Case Report Forms. These findings are to be forwarded to the clinical data monitor for query issuance and resolution. The unblinded biostatisticians do not discuss any results or other information that may inadvertently unblind the study team until the database is locked and treatment assignments are unblinded.

1.5 Responsibilities

Dr. John Stone and colleagues have designed the study protocol. Dr. Stone is responsible for the conduct of the trial. The Data Coordinating Center (DCC) is managed by Ana Fernandes and Dr. Naomi Serling-Boyd in collaboration with the MGH Biostatistics Center. Nora Horick, MS and Brian Healy, PhD, are the unblinded Biostatisticians. Andrea Foulkes, PhD is the blinded Biostatistician for the trial. The DCC and the MGH Biostatistics Center are responsible for the oversight of data collection and analysis. The responsibilities of the DCC include:

1. Development and implementation of the data flow, schedules for transferring data from sites, and data tracking;
2. Development of procedures for data entry, error identification, and error correction;
3. Adverse event monitoring and reporting to the DSMB;
4. Site monitoring via Electronic Data Capture (EDC) to ensure adherence to the protocol and procedures;
5. Quality control procedures;
6. Creating reports - enrollment, adverse events, participant status (e.g., withdrawals) by site; and
7. Trial data analysis

1.6 Timing of Statistical Analyses

The final analysis of safety and efficacy endpoints will be conducted when 1) randomization is complete; 2) all randomized subjects have completed treatment, withdrawn from the study, or died; and 3) and all surviving randomized subjects have completed the 28-day study follow-up.

2 Study Endpoints

2.1 Efficacy endpoints

The primary endpoint is the time from administration of the investigational agent (or placebo) to requiring mechanical ventilation and intubation, or death for subjects who die prior to intubation.

The secondary efficacy endpoints are:

1. Time from administration of the investigational medication (or placebo) to at least one point worsening on the clinical improvement scale (Table 2) for subjects requiring supplemental oxygen (score >= 3) at baseline, or at least two point worsening otherwise (score = 2 at baseline).
2. Time from administration of the investigational agent (or placebo) to absence of the need for supplemental oxygen among those who require at least supplemental oxygen at baseline.

Table 2: Ordinal Clinical Improvement Scale

Clinical Improvement Scale	
1	Discharged (or "ready for discharge" as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or <= 2L supplemental oxygen)
2	Non-ICU hospital ward (or "ready for hospital ward") not requiring supplemental oxygen
3	Non-ICU hospital ward (or "ready for hospital ward") requiring supplemental oxygen
4	ICU or non-ICU hospital ward, requiring non-invasive ventilation or high-flow oxygen
5	ICU, requiring intubation and mechanical ventilation
6	ICU, requiring ECMO or mechanical ventilation and additional organ support (e.g. vasopressors, renal replacement therapy)
7	Death

2.2 Tertiary and exploratory endpoints

Additional tertiary endpoints are:

- Time to first improvement from baseline of at least 2 points (or the maximum amount) on the ordinal scale given in Table 2.
- Ordinal Clinical Improvement Scale (Table 2) score at day: 4, 7, 14, 21, and 28.
- Time from initiation of supplemental oxygen to end of supplemental oxygen use during 28-day study follow-up period.
- Time from administration of the investigational agent (or placebo) to death.
- Mortality at 28 days after administration of investigational agent (or placebo).
- Time from administration of the investigational medication (or placebo) to intubation.
- Duration of mechanical ventilation during 28-day study follow-up period.
- ICU admission or death among those not in the ICU at the time of administration of investigational agent (or placebo).
- Time from administration of the investigational medication (or placebo) to hospital discharge.

The change over time for the following exploratory endpoints will be evaluated:

- Biomarkers: ferritin, LDH, CRP, D-dimer, ESR, troponin, NT-proBNP, IL-6, and procalcitonin.
- Cytokine profiling, including rapid IL-6 assessment.
- Clinically relevant inflammatory biomarkers including TREM-1, procalcitonin, and Pro-ADM.
- Measures of cardiac injury (troponin- difference in peak troponin level) and cardiac dysfunction (NTproBNP-difference in peak NTproBNP level).
- Viral titers.

- Cell subsets for functional and transcriptional immunophenotyping.

2.3 Safety Endpoints

The proportion of adverse events graded by CTCAE v5.0 will be evaluated.

3 Sample Size Determination

The primary endpoint is the rate of requirement for invasive mechanical ventilation. The control group is assumed to have a 30% chance of requiring invasive mechanical ventilation by 28 days, which corresponds to a 70% chance of not requiring mechanical ventilation. Our assumption is that the investigational treatment tocilizumab will increase the likelihood that a patient will not require mechanical ventilation to 85%. With a total of 278 subjects (185 randomized to tocilizumab, 93 randomized to standard care), we will have 85% power to demonstrate such a difference, assuming two-sided tests and an alpha of 0.05. With a total of 243 subjects (163 randomized to tocilizumab, 80 randomized to standard care), we will have 80% power to demonstrate such a difference, assuming two-sided tests and an alpha of 0.05. At the outset, the target enrollment was 278 patients to achieve 85% power. However, the enrollment rate significantly slowed as the pandemic surge waned in the Boston area, and in early June the decision was made to reduce the target enrollment to 243 (80% power) and the protocol was amended to reflect this change.

An interim analysis was to be performed when approximately 50% of the subjects had enrolled or approximately 40% of subjects had completed Day 28 or withdrawn prior to Day 28. Both efficacy and futility of the study were to be assessed at the time of the interim analysis. However, due to the rapid study initiation and enrollment, two thirds of the target N had already been enrolled and more than 50% of the subjects had already completed the 28-day follow up at the time of the anticipated interim analysis. Because the anticipated completion of enrollment was only two weeks away at the time of the DSMB meeting, the interim analysis for efficacy was not conducted. Rather, the DSMB reviewed the accumulated safety events in the first 180 patients enrolled. This change of conduct is described in the protocol amendment.

4 General statistical considerations

4.1 Statistical Software

All statistical analyses will be performed using SAS (SAS Institute, NC, USA) and R (R Foundation for Statistical Computing, Vienna, Austria).

4.2 Analysis Population

The following analysis samples are defined for safety and efficacy analysis:

- Intent-to-Treat (ITT) Sample: Subjects who are randomized regardless of treatment adherence or availability of follow-up data will be included in the intention-to-treat analysis set (ITT). All analyses of the ITT will be based on each subject's randomized treatment assignment.
- Modified Intent-to-Treat (mITT) Sample: Randomized subjects who receive any amount of the study drug before intubation or death.
- Per-Protocol (PP) Sample: Randomized subjects who receive the full dose of the study drug before intubation or death. Patients who do not receive the full dose of the study drug will be excluded from the PP population. Subjects who had a major protocol deviation that may impact the validity of the efficacy analysis are excluded from the PP population

- Safety Sample: Randomized subjects who receive any amount of the study drug. Safety analyses will be based on the medication that was actually dispensed to each subject.

The primary efficacy analysis and summary level tables on patient characterisitics will be based on the mITT sample.

4.3 Subject Disposition

Subject disposition data will be listed. A disposition table will present, by treatment arm and overall, the number and/or percentage of subjects who signed the informed consent and entered the study (i.e., were screened, screen failed and randomized), completed study drug administration, withdrew from the study, completed the study, and discontinued treatment after randomization. The reasons for early withdrawal after randomization will be summarized.

Assignment to the analysis sets (ITT, mITT, PP and Safety) will be summarized.

4.4 Demographic and Baseline Characteristics

Baseline characteristics including pre-existing conditions, medications and demographic information will be summarized by treatment group using descriptive statistics and visual displays. Descriptive statistics for continuous variables will include the number of subjects, mean, standard deviation, median, first and third quartiles, minimum and maximum values for the observed value, and change from baseline. Analyses of categorical variables will include calculations of frequencies and percentages.

4.5 Transformations

Continuous data that are strongly rightward skewed (skewness greater than 3) will be log transformed prior to analysis.

4.6 Multiplicity Adjustments

A single primary analysis is planned and the criterion for statistical significance will control the type-1 error rate at a level of 0.05. Testing of secondary efficacy endpoints will be performed using a Bonferroni-Holm correction to ensure an overall two-sided type 1 error rate of less than 0.05. Nominal p-values will also be reported. No correction for multiple comparisons will be used for the tertiary and exploratory analyses.

4.7 Missing Data

If a participant is lost to follow-up prior to 28 days, all time-to-event outcomes will be censored at the time of last contact. Date and time will be used to define event times whenever it is possible to do so without imputing the time value; otherwise only dates will be used.

4.8 Stratification

All analyses will be stratified by site and a combined treatment effect will be estimated. If a study site has low enrollment (<12 individuals) then data on this site will be combined with the smallest site with at least 12 individuals prior to stratified analysis.

4.9 Re-admission

Data collected during re-admissions occurring within the 28-day study period will be used to define outcome variables.

5 Efficacy Analysis

5.1 Primary and secondary endpoints

Primary endpoint: time from administration of the investigational agent (or placebo) to requiring mechanical ventilation and intubation, or death for subjects who die prior to intubation. The treatment groups will be compared using a log-rank test stratified by study site, and the p-value from the log-rank test will be the primary p-value. Subjects who do not have either event by the end of the follow-up period will be censored at 28 days. Subjects missing 28 day follow-up will be censored at last contact. The difference between the treatment groups will be estimated using the hazard ratio from a stratified Cox proportional hazards model. The type I error rate for this study will be 0.05, and the primary outcome will be tested at this level. The 95% confidence interval for the hazard ratio will be reported in addition to the p-value.

Secondary endpoint: Time from administration of the investigational medication (or placebo) to at least one point worsening on the clinical improvement scale (Table 2) for subjects requiring supplemental oxygen (score >= 3) at baseline, or at least two point worsening otherwise (score = 2). We will compare the groups using a stratified log-rank test and estimate the hazard ratio comparing the groups using a stratified Cox proportional hazards model. This composite endpoint based on the clinical ordinal scale given in Table 2 is defined as: a) time to progressing to a score of 4-7 for individuals who start with a score of 2 or 3; or b) progressing to a score of 5-7 for individuals starting with a score of 4.

Secondary endpoint: Time from administration of the investigational agent (or placebo) to absence of the need for supplemental oxygen among those who require at least supplemental oxygen at baseline. Time to absence of the need for supplemental oxygen will be measured from time of investigational treatment administration. Only subjects on supplemental oxygen at the time of randomization will contribute to this analysis. The groups will be compared using a stratified log-rank test and stratified Cox proportional hazards model and subjects who die prior to reaching this endpoint, including subjects whom supplemental oxygen was discontinued as a part of comfort measures preceding death, will be censored at 29 days.

5.2 Tertiary and exploratory endpoints

Time to first improvement from administration of the investigational agent (or placebo) of at least 2 points (or the maximum amount) on the ordinal scale given in Table 2. Time to improvement will be assessed by changes in subject status, ranked on the ordinal scale shown in Table 2. Time to improvement will be measured from administration of the investigational agent (or placebo) to improvement of 2 points or more. Subjects who die prior to reaching this endpoint will be censored at 29 days. The groups will be compared using a log-rank test stratified on study site. The hazard ratio comparing the groups will be estimated using a stratified Cox proportional hazards model.

Ordinal Clinical Improvement Scale (Table 2) score at day: 4, 7, 14, 21, and 28. The raw ordinal scale scores at days 4, 7, 14, 21, and 28 in subjects treated with the investigational agent vs. placebo will be compared at each time point using a random intercept proportional odds logistic regression model. This model will be used to

model all measurements together to estimate the differences between the treatment groups at each time point. This model will include a fixed effect for time, treatment group, time by treatment group interaction, and study site. The parameter of interest will be the time by treatment interaction term which will equal the difference in the change with time comparing the treatment groups. Ordinal scale score at discharge will be carried forward following discharge to reflect that the subject remained discharged and recalculated if needed to reflect post-discharge re-hospitalization or death.

Time from initiation of supplemental oxygen to end of supplemental oxygen use during 28-day study follow-up period. The duration of supplemental oxygen will be compared between the groups. For this analysis, we will include all subjects in the analysis by assigning all subjects who did not receive supplemental oxygen a value of 0. Subjects who died prior to discontinuation of supplemental oxygen or for whom supplemental oxygen was discontinued as a part of comfort measures preceding death will be given a value of the number of days from when supplemental oxygen began until the end of the follow-up period. The groups will be compared using a stratified Wilcoxon rank sum test.

Time from administration of the investigational agent (or placebo) to death. Time to death will be measured from the time the investigational medication is administered until the time of the subject's death. The groups will be compared using a stratified log-rank test and a stratified Cox proportional hazards model will be used to estimate the hazard ratio comparing the groups.

Mortality at 28 days after administration of investigational agent (or placebo). Mortality at 28 days will be compared using a Mantel-Haenszel test to allow stratification on study site. The relative risk will be estimated using the Mantel-Haenszel method. If we have missing mortality data on any subjects, we will estimate the proportion of subjects who died in each treatment group using the estimate from the Kaplan-Meier curve in each group. Then, we will compare the two groups using the approaches described in Klein et al. (citation: Klein JP, Logan B, Harhoff M, Anderson PK. Analyzing survival curves at a fixed point in time. Statistics in Medicine. 2007;26:4505–4519.)

Time from administration of the investigational medication (or placebo) to intubation. Time to intubation will be measured from time of investigational treatment administration to the time of intubation. For this analysis, death will be treated as a competing risk. The analysis will compare the cause-specific hazard in the treatment groups using a Cox proportional hazards model. We will also compare the cumulative incidence functions between groups using the approach of Fine and Gray.

Time from initiation of mechanical ventilation to end of mechanical ventilation during 28-day study follow-up period. The duration of mechanical ventilation will be compared between the groups using two approaches. First, we will include all subjects in the analysis by assigning all subjects who were not intubated a value of 0. Subjects who died following intubation will be given a value of the number of days from when mechanical ventilation began until the end of the follow-up period. The groups will be compared using a stratified Wilcoxon rank sum test. Second, we will analyze only subjects who were intubated and compare the time on mechanical ventilation using a stratified log-rank test. Subjects who die without being taken off the ventilator will be censored at a duration of mechanical ventilation longer than the longest time.

ICU admission or death among those not in the ICU at the time of administration of investigational agent (or placebo). The proportion of subjects who require ICU admission or die between baseline and 28 days will be measured as the number of subjects who require ICU admission during their hospitalization or die over the number of evaluable subjects (i.e., the number of subjects not in the ICU at the time of investigational treatment

administration). The groups will be compared using a Mantel-Haenszel test to allow stratification on study site. The relative risk will be estimated using the Mantel-Haenszel method.

Time from administration of the investigational medication (or placebo) to initial hospital discharge. The time to initial discharge from the hospital in subjects, measured from the time of investigational treatment administration to time of discharge, will be compared using a stratified log-rank test. Subjects who die prior to reaching this endpoint will be censored at 29 days.

Exploratory endpoints:

We will estimate the change over time in nine biomarkers (ferritin, LDH, CRP, D-dimer, ESR, troponin, NT-proBNP, IL-6, and procalcitonin) using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups.

We will estimate the change over time in cytokine profiling, including rapid IL-6 assessment, using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups.

We will estimate the change over time in clinically relevant inflammatory biomarkers including TREM-1, procalcitonin, and Pro-ADM using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups.

We will estimate the change over time in measures of cardiac injury (troponin- difference in peak troponin level) and cardiac dysfunction (NTproBNP-difference in peak NTproBNP level) using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups. We will also compare the time to serious cardiac arrhythmias between the treatment groups using a stratified log-rank test.

We will estimate the change over time in viral titers using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups.

We will estimate the change over time in cell subsets for functional and transcriptional immunophenotyping using a linear mixed effects model with an unstructured covariance matrix, a fixed effect for time, treatment group, time by treatment group interaction, study site and the study site by time interaction. The parameters of interest will be the time by treatment interaction terms which will equal the difference in the change with time comparing the treatment groups.

5.3 Sensitivity Analysis

For the primary endpoint, sensitivity to stratified analysis will be evaluated by applying an analysis that is not stratified by site.

For all endpoints corresponding to improvement (e.g. time to first improvement from baseline of at least 2 points (or the maximum amount) on the ordinal scale given in Table 2) sensitivity to inclusion of patients who improve and subsequently worsen will be performed by removing these individuals from analysis.

Primary and secondary endpoint analysis estimates will be reported for PP samples.

5.4 Subgroup Analysis

Differences in treatment efficacy for primary the endpoint will be evaluated in multiple pre-defined subgroups defined by sex, race/ethnicity, age category (<65, >=65) , obesity (BMI>=30), diabetes and baseline labs, incuding IL-6, CRP, Ferritin and D-dimer. These differences in treatment efficacy will be assessed using the stratified Cox proportional hazards model including a fixed effect for treatment, a fixed effect for each of the pre-defined subgroups and an interaction between treatment and the subgroup. The interaction will represent the difference in the treatment effect. Both subgroup specific hazard ratios and the p-value for the treatment by subgroup interaction will be reported.

6 Safety Analysis

Safety and tolerability will be estimated in the Safety Sample. Safety and tolerability, defined as adverse events (AEs) graded by CTCAE v5.0, will be compared using a Mantel-Haenszel test to allow stratification on study site. The relative risk will be estimated using the Mantel-Haenszel method.

Only treatment-emergent AEs, defined as AEs with date of onset on or after the time of treatment administration, will be reported. AEs reported on the Day 28 follow-up form will be included.

AEs will be summarized by treatment group, severity (serious vs non-serious), grade, and relationship to study medication as indicated by the investigator. The following rules will be applied:

- The number and proportion of patients with an AE reported on one or more study days will be summarized for each AE catgory.
- AE grade will be defined as the highest grade reported for that AE category
- AE relatedness to tocilizumab will be defined as the highest degree of relatedness reported for each AE category.
- AEs in the following categories will only be reported if grade 3 or higher: neutropenia, thrombocytopenia, infection, bleeding, AST/ALT elevation
- We will grade AST/ALT elevations. The grading only applies to post-day 1 values. The grading algorithm involves first determining whether the day 1 value is elevated (y/n) based on an upper limit for normal (ULN) that is both site- and gender-specific—see Table 3. Grade for post-day 1 elevations is determined by the ratio of the current value to (1) the ULN if day 1 value is not elevated or (2) the day 1 value if the day 1 value is elevated as described in Table 4.

Table 3: Upper limits of normal (ULN) values for AST and ALT by study site and gender

Site	Gender	AST_ULN	ALT_ULN
BILH	Male	40	40
BILH	Female	40	35
BMC	Male	39	67
BMC	Female	39	67
BWH	Male	50	50

BWH	Female	50	50
MGH	Male	40	55
MGH	Female	32	33
NSMC	Male	41	50
NSMC	Female	41	35
NWH	Male	40	49
NWH	Female	40	49
SEMC	Male	41	63
SEMC	Female	41	54

Table 4: Grading algoritm for post-baseline AST/ALT elevations

Grade	**Grade if normal baseline**		**Grade if abnormal baseline**
1	**ULN - 3x ULN**		**1.5x - 3x baseline**
2	**>3x ULN - 5x ULN**		**>3x - 5x baseline**
3	**>5x ULN - 20x ULN**		**>5x - 20x baseline**
4	**>20x ULN**		**>20x baseline**

AEs of special interest include the following:

- Death
- Infections
- Myocardial infarction
- Gastrointestinal perforation
- Hypersensitivity reaction to infusion
- Deep venous thrombosis
- Pulmonary embolism
- Stroke
- Seizure
- Abnormal liver function test
- Neutropenia
- Thrombocytopenia
- Bleeding events
- Malignancy
- Demyelinating disorder