NECTAR

Novel Experimental COVID Therapies Affecting Host Response

Statistical Analysis Plan

CONNECTS Master Protocol for Clinical Trials targeting macro-, micro-immuno-thrombosis, vascular hyperinflammation, and hypercoagulability and renin-angiotensin-aldosterone system (RAAS) in hospitalized patients with COVID-19 (ACTIV 4 Host Tissue)

Version 1.4
Nov 14, 2023

Study Chair:

Sean P. Collins, MD, MSc.
Vanderbilt University Medical Center
1313 21st Ave S #801
Nashville, TN 37232
Phone: 615-936-0087
Email: sean.collins@vumc.org

Principal Investigator, Clinical Coordinating Center (CCC):

Wesley H. Self, MD, MPH
Vanderbilt University Medical Center
1313 21st Ave S
Nashville, TN 37232
Phone: 615-936-0253
Email: wesley.self@vumc.org

Principal Investigator, Data Coordinating Center (DCC)

Matthew S. Shotwell, PhD
Vanderbilt University School of Medicine
2525 West End Avenue
Suite 11000
Nashville, TN 37203
Phone: 615-875-3397
Email: matt.shotwell@vumc.org

Revision History:

Ver.	Date	Authors	Summary of Revisions:
1.0	2/14/22	M. Shotwell	Initial version
1.1	3/8/22	M. Shotwell	• Simplified interim stopping rule • Updated simulation results and decision thresholds at interim and final analyses • Clarified purpose of AngioNECTAR • Additional details about sensitivity analyses • Summarized sample size reassessment process • Additional details of final analysis procedure
1.2	7/28/22	M. Shotwell	• Clarifications to address FDA comments • Additional tipping point analysis to evaluate effect of partially observed outcomes on efficacy conclusion • Additional supplemental appendices
1.3	3/2/22	M. Shotwell	• Clarifications to address FDA comments
1.4	11/14/23	M. Shotwell	• Clarifications to address FDA comments • Additional sensitivity analysis to evaluate treatment efficacy if the proportional odds assumption is violated.

Signatures:

Name:  **Date:** 2024-09-03

ABBREVIATIONS AND ACRONYMS

ACE	Angiotensin-Converting Enzyme
AE	Adverse Event
ANG	Angiotensin
API	Application Programming Interface
ARDS	Acute Respiratory Distress Syndrome
BD	Becton Dickinson
BID	Biospecimen Identity
BCL	Biorepository and Central Laboratory
BMP	Basic Metabolic Panel
CBC	Complete Blood Count
CCC	Clinical Coordinating Center
CDE	Common Data Element
CDISC	Clinical Data Interchange Standards Consortium
CDASH	Clinical Data Acquisition Standards Harmonization
CFR	Code of Federal Regulations
COVID-19	Coronavirus Disease 2019
CRF	Case Report Form
CRP	C-Reactive Protein
CTOM	Clinical Trial Operation Manager
CTCAE	Common Terminology Criteria for Adverse Events
DAG	Data Access Group
DAIDS	Division of Acquired Immunodeficiency Syndrome
DCC	Data Coordinating Center
DOB	Date of Birth
DOR	Delegation of Responsibilities
DR	Disaster Recovery
DSMB	Data Safety Monitoring Board
DSMP	Data Safety Monitoring Plan
eConsent	Electronic Consent
eCRF	Electronic Case Report Forms
EDC	Electronic Data Capture
EMR	Electronic Medical Record
FAQ	Frequently Asked Question
FDA	Food and Drug Administration
GDPR	General Data Protection Regulation
GUID	Globally Unique Identifier
HT	Host Tissue
hsTn	High-sensitivity Troponins
ICF	Informed Consent Form
ICH GCP	International Conference of Harmonization Good Clinical Practice
ID	Identity
IMV	Invasive Mechanical Ventilation
INR	International Normalized Ratio
IP	Investigational Product

IRB	Institutional Review Board
IT	Information Technology
ITT	Intent-To-Treat
KSP	Key Study Personnel
LAR	Legally Authorized Representative
LFT	Liver Function Test
MOP	Manual of Procedures
NAT	Nucleic Acid Test
NHLBI	National Heart, Lung, and Blood Institute
NINDS	National Institute of Neurological Disorders and Stroke
NTproBNP	N-terminal prohormone B-type natriuretic peptide
N3C	National COVID Cohort Collaborative
PDF	Portable Data Format
PI	Principal Investigator
PSESE	Protocol-Specified Exempt Serious Events
PT	Prothrombin Time
PTT	Partial Thromboplastin Time
REDCap	Research Electronic Data Capture
RTI	Research Triangle Institute
SAP	Statistical Analysis Plan
SAE	Serious Adverse Event
sIRB	Single Institutional Review Board
SOC	Standard of Care
SOFA	Sequential Organ Failure Assessment
UAT	User Acceptance Testing
VCC	Vanderbilt Coordinating Center
VMP	Validation Master Plan
VUMC	Vanderbilt University Medical Center
VUMC-IT	Vanderbilt University Medical Center – Information Technology
VICTR-ORI	Vanderbilt Institute for Clinical and Translational Research – Office of Research Informatics
WHO	World Health Organization

1 INTRODUCTION

The Statistical Analysis Plan (SAP) was developed by DCC statisticians in collaboration with study team leadership and NHLBI representatives. The SAP describes treatment arms, analysis datasets, all outcomes and planned analyses, randomization procedure and algorithm, decision thresholds and interim stopping rules, design and results of simulations to determine power and sample size and demonstrate study operating characteristics, procedures for handling missing data, and any other information that is essential to carry out all statistical analyses.

1.1 AngioNECTAR SAP

AngioNECTAR is a mechanistic sub-study that will utilize biospecimens collected as part of ACTIV 4 Host Tissue and complement the clinical information obtained in our primary analysis. This sub-study will examine the effects of study therapies on biomarkers of the Renin-Angiotensin-Aldosterone-System. Statistical analyses associated with the AngioNECTAR sub-study will be designed and implemented by AngioNECTAR PI D. Clark Files, MD, and Co-Investigators Mark Chappell, PhD and Chris Schaich, PhD. A separate SAP for the AngioNECTAR sub-study will be finalized by the AngioNECTAR investigators prior to unblinding of the active/placebo status for sub-study participants.

1.2 SAP Approval and Revision

The SAP will be reviewed and approved by the ACTIV 4 Host Tissue stakeholders listed below prior to the first interim analysis for any arm:

- ACTIV 4 Host Tissue Study Chair: Sean Collins, MD
- ACTIV 4 Host Tissue DCC PI: Matthew S. Shotwell, PhD
- NHLBI Statistician: James Troendle, PhD

Amendments to the SAP must also be approved by the stakeholders listed above. Amendments must be version controlled and numbered. All revisions will be summarized briefly, including the changes made, new version number, and the author of the changes.

2 STUDY DESIGN

2.1 Summary

The ACTIV 4 Host Tissue master protocol describes a common approach to studies of blinded, placebo-controlled therapeutic approaches of host-tissue targeted therapies in hospitalized COVID-19 patients. The Master Protocol is designed to be flexible in the number of study arms, to have a common placebo group, and to allow for stopping and adding of new therapies, while using a common approach to design, analysis, and implementation.

2.2 Study Arms and Pooled Placebo

The ACTIV 4 Host Tissue platform consists of multiple study arms that represent distinct drug therapies. During the randomization process, each participant is assigned a study arm and either the active drug or a matching placebo. The statistical analyses described herein will be implemented separately for each study arm. However, placebo participants will be pooled across arms. For each study arm, the placebo comparator group will consist of all placebo participants that were *eligible* for that study arm at the time of randomization. A participant is considered eligible for a study arm if assignment to that arm was a possible outcome of randomization. Participants that decline to participate in any one or more study arms prior to

randomization will be treated as ineligible for those arms. The randomization process is designed to ensure balance in each active drug group versus the corresponding placebo comparator group.

2.3 Randomization

Participants are randomized individually at enrollment using a central electronic system. The permuted block method, with stratification by study site and study arm eligibility is used to generate treatment assignments. An eligibility stratum is the collection of study arms for which a participant is eligible. Stratification by site ensures balance across the active and pooled placebo comparator groups at regular enrollment intervals at each site, thus mitigating the impact of site heterogeneity on assessments of treatment effect. Each block contains a multiple of $m(m+1)$ assignments, where m is the number of study arms in the corresponding eligibility stratum. Within each block there are an equal number of allocations across study arms and, for each study arm, there are m active and 1 placebo assignments. For example, in the TXA127 and TRV027 eligibility stratum, each block consists of the following allocations, or multiples thereof:

Study Arm	Placebo/Active
TXA127	Active
TXA127	Active
TXA127	Placebo
TRV027	Active
TRV027	Active
TRV027	Placebo

Thus, within each block, assignments are balanced across study arms, and the active assignments are balanced with the pooled placebo assignments. The block size multiple is either 1 or 2, selected uniformly at random for each block.

2.4 Blinding

For organizational purposes, the randomized assignment comprises two distinct pieces of information: 1) study arm, and 2) active vs. placebo assignment. The study arm is not blinded, whereas the active/placebo assignment is blinded from participants and investigators (other than unblinded personnel as required for study operations, data quality/analysis, and safety). Blinding will remain in place until all participants have completed the study, all data quality monitoring is complete, and the database is locked.

3 OUTCOMES

3.1 Primary Outcome

The primary outcome for the ACTIV 4 Host Tissue platform is oxygen free days (OFD) at day 28. OFD will be calculated as the number of calendar days during the first 28 days after randomization during which the patient was alive and not receiving supplemental oxygen therapy. Participants who chronically used supplemental oxygen prior to their COVID-19 illness will be considered oxygen free when their use of supplemental oxygen does not exceed the level of oxygen support (measured in daily L/min·h by nasal canula) used prior to COVID-19 illness. Supplemental oxygen therapy includes the following: supplemental oxygen by nasal cannula, supplemental oxygen by face mask, high flow nasal cannula (HFNC), non-invasive ventilation (NIV), invasive mechanical ventilation (IMV), or extracorporeal membrane

oxygenation (ECMO). The day of randomization is defined as day 0. Starting with study day 1 (the day after randomization) and continuing for 28 days, study personnel will document whether the participant received supplemental oxygen therapy on each day for any duration of time. Use of supplemental oxygen at home after discharge will be assessed via telephone follow-up calls to the participant or surrogates. OFD will be calculated as 28 minus the number of days between and including the first and last days of supplemental oxygen use during the first 28 days after randomization. OFD will be coded as -1 for patients who died on or before study day 28. Hence, OFD may take any integer value between -1 and 28. OFD is an ordered categorical (i.e., ordinal) outcome that may be interpreted as a count of days. Additional details about calculating OFDs may be found in the SAP appendix (see Appendix: Algorithm to Compute Primary Outcome).

3.2 Secondary Outcomes

Listed below are the ACTIV 4 Host Tissue platform secondary outcomes. The "Test Order" field indicates the order in which key secondary outcomes will be tested, using the fixed-sequence method, to control the familywise type-I error probability across the primary and key secondary outcomes.

Description	Type	Test Order	Analysis Method
Alive and oxygen free at day 14	Binary		LogR
Alive and oxygen free at day 28	Binary		LogR
Alive and respiratory failure-free at day 14	Binary		LogR
Alive and respiratory failure-free at day 28	Binary	1	LogR
Alive and free of new IMV at day 14	Binary		LogR
Alive and free of new IMV at day 28	Binary		LogR
Mortality in-hospital	Binary		LogR
Mortality at day 28	Binary	3	LogR
Mortality at day 60	Binary		LogR
Mortality at day 90	Binary		LogR
WHO 8-point ordinal scale at day 14	Ordinal		POLR
WHO 8-point ordinal scale at day 28	Ordinal	2	POLR
WHO 8-point ordinal scale at day 60	Ordinal		POLR
Hospital-free days at day 28	Ordinal		POLR
Respiratory failure-free days at day 28	Ordinal		POLR
Ventilator-free days at day 28	Ordinal		POLR

LogR – Logistic Regression; POLR – Proportional Odds Logistic Regression

The WHO 8-point ordinal scale is defined as most severe clinical status among the following on the day of assessment:

1. Ambulatory – Not hospitalized, no limitation of activities
2. Ambulatory – Not hospitalized with limitation of activities or home oxygen therapy
3. Hospitalized Mild Disease – Hospitalized, no oxygen therapy
4. Hospitalized Mild Disease – Oxygen by mask or nasal prongs
5. Hospitalized Severe Disease – Non-invasive ventilation of high-flow oxygen
6. Hospitalized Severe Disease – IMV
7. Hospitalized Severe Disease – IMV + organ support with-vasopressors, RRT, or ECMO
8. Dead

Alive and respiratory failure-free at day 28, the WHO 8-point ordinal scale at day 28, and Mortality at day 28 are key secondary outcomes that will be treated as a family for testing

purposes, even though the studies will not be adequately powered to detect anything but a very strong treatment effect on these outcomes. A supplementary analysis to assess the evidence that treatment lowers the risk of death in a way that is consistent with its effect on nonfatal outcomes will be performed. A respiratory failure-free day is defined as a day alive without the use of HFNC, NIV, IMV, or ECMO.

3.3 Safety Outcomes

Safety outcomes include the following events, assessed daily during hospitalization or intermittently following hospital discharge. For each event, we will analyze two composite binary outcomes: 1) the occurrence of one or more such events by the end of study day 7 and 2) the occurrence of one or more such events by the end of study day 28.

Description	Type	Analysis Method
Hypotension	Binary	LogR
Allergic reaction, rash, or angioedema	Binary	LogR
Incident renal replacement therapy	Binary	LogR
Other PSESE	Binary	LogR

LogR – Logistic Regression

Hypotension is defined by low arterial blood pressure leading to either [1] initiation or increase in vasopressor therapy, [2] administration of a fluid bolus of 500 ml or more, or [3] modification of the dose or discontinuation of the study drug.

3.4 Exploratory Outcomes

Exploratory outcomes will include (at least) the following:

Description	Type	Analysis Method
Change in troponin during hospitalization	Quantitative	LinR
Change in NT-proBNP	Quantitative	LinR
Change in RAAS mechanistic biomarkers: 1. AngII 2. Ang(1-7) 3. Plasma renin activity 4. Aldosterone 5. ACE 6. ACE2	Quantitative	LinR
Change in serum creatinine	Quantitative	LinR
Change in eGFR	Quantitative	LinR
Acute kidney injury (KDIGO criteria)	Ordinal	POLR

LinR – Linear Regression; POLR – Proportional Odds Logistic Regression

Exploratory outcomes may be collected at just a subset of sites.

4 ANALYSIS DATASETS

For each study arm, the following analysis datasets will be produced using records for participants that were assigned to the active drug group and placebo participants that were *eligible* for the active drug group at the time of randomization:

Modified intention-to-treat dataset: The mITT analysis dataset will include all randomized participants grouped by study arm and active/placebo assignment at randomization, regardless of subsequent compliance or protocol violations, with the following exceptions: 1. Participants who have not received the study drug assigned at randomization will be excluded. 2. Participants who were randomized and later found to be ineligible based on assessments initiated prior to randomization will be excluded. All statistical analyses will be implemented using mITT dataset unless otherwise explicitly specified in this statistical analysis plan.

Intention-to-treat dataset: The intention-to-treat (ITT) analysis dataset will consist of all randomized participants grouped by study arm and active/placebo assignment at randomization regardless of subsequent compliance or protocol violations.

Safety dataset: The safety analysis dataset will consist of all participants grouped by the drug(s) received.

5 EFFICACY TESTING & FAMILYWISE TYPE-I ERROR CONTROL

Efficacy regarding the primary outcome and each key secondary outcome will be tested using a one-sided method that ensures no more than a 2.5% chance of a type-I error. The fixed-sequence method will be used to control the familywise type-I error probability at 2.5% for the family of primary and key secondary outcomes.[1] Specifically, a conclusion of efficacy regarding the primary outcome will be required prior to testing the first designated key secondary outcome. Each subsequent key secondary outcome, in the designated order, will take place only if the preceding key secondary outcome demonstrates efficacy. This approach provides strong control of the familywise type-I error probability at 2.5% for the family of primary and key secondary outcomes. No other statistical hypothesis tests will be made regarding other secondary, safety, or exploratory outcomes. P-values associated with certain null hypothesis tests may be provided for descriptive purposes, or to fulfill special requests, e.g., for DSMB safety assessments.

6 ANALYSIS OF THE PRIMARY OUTCOME

The effect of the active drug versus placebo will be quantified using an odds ratio – the primary estimand – which quantifies the treatment effect on the odds of greater oxygen-free days at day 28. Based on the behavior of similar outcomes in prior trials,[2-6] we anticipate the distribution of the primary outcome to be irregular, with peaks around -1 to 0 and between 22 and 28 days. Thus, we will use a flexible semi-parametric approach for the primary outcome analysis. Estimation and inferences about the odds ratio will be made using Bayesian proportional odds (PO) logistic regression methods, adjusting for the active drug vs placebo indicator variable, age group (18-30, 31-65, >65 years), sex at birth, and WHO COVID ordinal outcome score at baseline (4, 5, and 6-7).[7] Evidence for efficacy will be quantified using the posterior probability that the active drug versus placebo odds ratio is greater than one (i.e., treatment is associated with greater oxygen free days at day 28). This is denoted the “efficacy probability” or $P(\mathrm{OR} > 1|\mathrm{Data})$, where OR represents the odds ratio, and Data represents the mITT analysis dataset. The “inferiority/harm probability” is defined as $P(\mathrm{OR} \leq 1|\mathrm{Data})$. The primary analysis will be implemented separately for each study arm, where the placebo comparator group will consist of placebo participants that were eligible for the corresponding study arm at randomization, regardless of the study arm assigned. The primary and supplementary estimates will be presented with 95% credible intervals.

6.1 Statistical Model

The PO model can be written in terms of the covariates X and an outcome variable Y, where probabilities of outcome value y or greater $\Pr(Y \geq y|X) = \mathrm{expit}(\alpha_y + X\beta)$ where α_y is the intercept for outcome value y and expit is the logistic (inverse logit) transformation and the columns of matrix X contain coded baseline covariates and the active/placebo treatment indicator. β represents the log odds ratio (OR) associated with the effects of covariates and group assignment. Specifically, the group assignment odds ratio represents the relative effect of treatment versus placebo on the odds $\Pr(Y \geq y|X)/(1 - \Pr(Y \geq y|X))$, for any value y.

A flat prior distribution will be used for all PO model parameters. This ensures that the estimate of the primary estimand will be free of influence from an informative prior, and the Bayesian maximum *a posteriori* estimate will be identical to the maximum likelihood estimate (see Appendix: Cumulative Logit Model). The posterior distribution for the log odds ratio will be approximated using the Laplace method.[8] Use of a flat prior ensures the Laplace-approximated posterior distribution is identical to the asymptotic sampling distribution of the maximum likelihood estimate; in both cases a normal distribution centered at the estimate with variance-covariance equal to the negative inverse Hessian of the log likelihood function (inverse observed Fisher information; see Appendix: Laplace Approximation). All statistical inferences about the odds ratio will be made using this method. Statistical uncertainty about supplementary estimands (e.g., treatment difference in the median of the primary outcome) will be quantified using the delta method.[9] Given the investigational nature of the agents tested by this platform, there is insufficient information upon which to justify a more informative prior. The flat prior approach ensures that Bayesian inferences regarding the efficacy of study agents are based exclusively on the data collected in the ACTIV 4 Host Tissue platform.

6.2 Loss to Follow-up, Censoring, and Intercurrent Events

Participants who withdraw consent prior to data collection, or for whom there is no partial information about the primary outcome, will not be excluded from analysis. We will strive to avoid loss to follow-up by making repeated attempts to contact participants or otherwise retrieve participant records. If loss-to-follow-up cannot be avoided, and the information needed to compute the primary endpoint is partially known (i.e., censored), we will use a likelihood-based method to account for this censoring. For example, if a study participant received supplemental oxygen every day during the 10-day period after randomization, but is then lost to follow-up, the primary outcome is only partially known (i.e., OFDs ≤ 18 in this example). The PO model provides a convenient mechanism to account for this and other types of censoring using a likelihood-based approach.[10] For observations that are fully observed, the log likelihood contribution is $l(\alpha, \beta;\ y, x) = \log \Pr(Y = y|X = x)$. For observations that are left censored at y (e.g., ≤ 18 OFDs), the log likelihood contribution is $l(\alpha, \beta;\ y, x) = \log \Pr(Y \leq y|X = x)$. The latter is conveniently computed by substituting $1 - \mathrm{expit}(\alpha_y + x\beta)$. More complex partially observed outcomes (e.g., right or interval censored) are modeled in a similar manner.

All primary analyses will be implemented using the mITT analysis dataset. The intercurrent event of death will be coded as a special value in the primary outcome (i.e., composite strategy). No other intercurrent events will affect the primary outcome assessment (i.e., treatment policy strategy).[11]

Participant age, sex, and WHO COVID scale at baseline are subject to source verification monitoring. Thus, we do not anticipate missing covariate data.

6.3 Planned Interim and Final Analyses, Early Stopping, and Type-I Error Control

Two planned interim analyses will occur separately for each study arm when the number of participants with complete 28-day follow-up (or were deceased, withdrawn, or lost-to-follow-up by day 28) reaches 33% and 67% of maximum enrollment for that arm. Interim analyses will be executed by unblinded personnel only. Participant records that inform the primary outcome must undergo monitoring prior to interim (and final) analysis. At each interim analysis, a study arm may be stopped early if there is evidence for inferiority/harm. Enrollment in the trial will be stopped early if the posterior probability for inferiority/harm exceeds 0.95.

Final analysis will occur once enrollment, follow-up, and the required monitoring are completed. Should additional data be collected after enrollment is halted at an interim analysis, the final analysis will incorporate this additional data. If the trial was stopped early at an interim analysis due to evidence of inferiority/harm, a conclusion of inferiority/harm will be indicated if the posterior probability for inferiority/harm remains greater than 0.95 at the final analysis. If the trial was not stopped early at an interim analysis due to evidence of inferiority/harm, efficacy will be indicated if the posterior probability for efficacy regarding the primary outcome exceeds a threshold as follows: For studies under this master protocol, the efficacy threshold was selected using statistical simulation to ensure a type-I error probability of 2.5% for each study arm. In all other scenarios, the trial is inconclusive.

6.4 Supplementary Efficacy Estimands

The PO model is attractive for the analysis of ordinal and quantitative response variables, such as the primary outcome, because they directly model the cumulative distribution function from which the mean, median, other percentiles, and cumulative probabilities of the primary outcome, stratified by treatment group, are easily derived.[12] In addition to the odds ratio, the effects of treatment versus placebo will be quantified using the difference in mean, difference in median, and differences in clinically relevant proportions associated with the primary outcome: mortality at day 28: $\Pr(Y = -1|X)$, and oxygen requirement every day until day 28: $\Pr(Y = 0|X)$, adjusted to the modal value for each covariate. These important and clinically meaningful supplementary estimands will be used to describe and communicate the treatment effect. The posterior distribution for each of the supplementary estimands is readily computed using standard Bayesian methods.

6.5 Sensitivity and Supplementary Analyses

Sensitivity and supplemental analyses will be implemented at the final analysis.

Most regression methods, including proportional odds logistic regression, Cox proportional hazards regression, and linear (mean) regression, assume that the effects of the independent variables are consistent across the outcome distribution. Violation of this assumption is a complex situation that implies a heterogeneous treatment effect. For example, the active drug group might experience more frequent extreme outcomes (death and 28 oxygen-free days) versus the placebo group. In this type of situation, much like differential treatment effects observed across participant strata such as sex, no single summary fully describes the treatment effect, and the differential treatment effects must be carefully examined and interpreted in their totality. The *proportional odds assumption* of the PO model specifies that the effect of treatment on the odds that $\mathrm{Y} \geq 3$ (measured as an odds ratio versus placebo) is the same relative effect as for $\mathrm{Y} \geq 4$. However, even when the PO assumption is strongly violated, the estimated OR remains a simple function of the Wilcoxon-Mann-Whitney U-statistic, namely the probability that a randomly chosen patient on treatment B has a higher response than a randomly chosen

patient on treatment A,[13] the *probability index* or *concordance probability*. In addition, under the null hypothesis, the PO assumption is always satisfied. Thus, statistical testing based on the odds ratio, as estimated using the PO model, has the specified type-I error rate and provides a reasonable global assessment of treatment effectiveness, regardless of violations of the proportional odds assumption. However, derived quantities such as the difference in means may be more sensitive to violations of the PO assumption. Deviations from proportional odds will be examined by separately estimating the odds ratio for each possible dichotomization (that preserves ordering) of the primary outcome (e.g., alive versus dead at day 28, alive and oxygen free for at least 10 days at day 28 versus alive and oxygen free for fewer than 10 days or dead at day 28, etc.), in a planned sensitivity analysis. These analyses will be implemented using the logistic regression method described below (see *Logistic Regression (LogR)*). No hypothesis testing will be implemented regarding the PO assumption. This sensitivity analysis serves the following two purposes: 1) to assess for evidence of violation of the proportional odds assumption and 2) to estimate the differential treatment effects (with 95% credible interval) when the proportional odds assumption is relaxed. The latter provides the information needed to interpret the treatment effects should there be evidence of violation of the proportional odds assumption. In addition, as a key secondary outcome, we will quantify the effect of treatment on 28-day mortality, which directly addresses the possibility that treatment affects 28-day mortality differently than cumulative oxygen use within the first 28 days. This analysis enables us to detect non-proportional effects of the treatment on the two major components of the primary outcome, mortality, and oxygen-free days.

As a supplementary analysis to inform decision-making in the event there is evidence of violation of the PO assumption, we will implement an alternative primary outcome using a partial proportional odds (PPO) model[14], where the effect of the study intervention and each covariate will be allowed to vary across all possible order-preserving dichotomizations of the primary outcome. These covariate effects will be “unconstrained.” However, to ensure estimability of this model, and to shrink toward the PO model, a weakly informative prior will be assigned to shrink each dichotomization effect toward the common effect. The log odds ratio for each dichotomization will have a normal prior distribution centered at the common log odds ratio with standard deviation 0.354, which assigns 95% prior probability that each dichotomization odds ratio falls within a factor of 0.5 and 2.0 of the common odds ratio. The estimated common odds ratio from the PPO model will be evaluated in a manner similar to the common odds ratio in the primary analysis using the PO model. Specifically, we will report the estimated common odds ratio with 95% credible interval and the posterior probability for efficacy. Due to limitations of existing software to implement the PPO model, partially observed values of the primary outcome will be treated as missing.

Analysis of partially observed or missing outcome data requires assumptions regarding the mechanism by which censoring and missing values arise. The likelihood method described above, and other similar methods such as multiple imputation assume that missing values occur at random (i.e., missing at random or MAR). However, because censored and missing values cannot be observed, assumptions about the missingness mechanism are not verifiable. In order to assess the sensitivity of study findings to violations of this assumption, we will conduct additional sensitivity analyses by reproducing the primary analysis under alternative assumptions regarding the mechanism for missing values. Specifically, we will perform sensitivity analyses that vary assumptions about the missing outcomes on the two treatment arms separately. These analyses will consider the following two scenarios: 1 “missing favors inefficacy”) each partially observed primary outcome in the placebo group will be assumed to have taken the highest/best possible value, whereas each partially observed primary outcome in the intervention group will be assumed to have taken the lowest/worst possible value, and 2

"missing favors efficacy") each partially observed primary outcome in the placebo group will be assumed to have taken the lowest/worst possible value, whereas each partially observed primary outcome in the intervention group will be assumed to have taken the highest/best possible value. These analyses will be implemented using the primary analysis methodology, including an assessment of hypothesis testing outcomes. For any trial under this platform, if there is a conclusion of efficacy at the final analysis, and the conclusion would have been different under the "missing favors inefficacy" scenario, then an additional tipping-point analysis will be implemented to estimate the association between the degree to which missing values must favor inefficacy versus the probability the trial would have failed to conclude efficacy. In these analyses, the partially observed outcomes will be randomly imputed under the assumption that partially observed outcomes favor the inefficacy conclusion by a specified amount. The degree to which the partially observed outcomes favor inefficacy will be encoded using an odds ratio that adjusts the outcome probabilities conditional on the participant covariates, using the maximum a posteriori (MAP) estimate at the final analysis. These probabilities will then be used to randomly sample the outcome for imputation purposes. For partially observed outcomes that exclude some levels of the outcome, the sampling probabilities for the excluded levels will be set to zero and the remaining probabilities normalized to sum to one. After sampling the outcome for all partially observed outcomes, the primary analysis will then be implemented using the imputed outcome data and the study conclusion recorded. This process will be repeated 1000 times and the probability of a trial conclusion other than efficacy will be calculated using a Monte-Carlo estimate. This process will again be repeated for a range of odds ratios encoding the degree to which the partially observed outcomes favor inefficacy. Specifically, this will be guided by two parameters: $\alpha 0$ and $\alpha 1$, where $\alpha 0$ is the log odds ratio of more oxygen free days comparing partially missing versus non-missing placebo arm participants and $\alpha 1$ is the log odds ratio of a higher score comparing partially missing versus non-missing active arm participants. We will vary both parameters across the tipping point analysis within in the following range (-0.5, -0.25, 0.25, 0.5) (i.e., 16 total scenarios). We chose this range of parameters because together they induce a treatment effect in the partially observed participants that ranges from reasonably pessimistic to reasonably optimistic. If β is the observed log odds ratio (active versus control) in the analysis, then the treatment effect in the partially observed participants would range from β-1 to β+1. This range (2 on the log odds ratio scale) is four times the anticipated treatment effect for which the study is powered (i.e., to detect a log odds ratio equal to 0.5, or an odds ratio equal to 1.65). The results of this sensitivity analysis will be summarized graphically.

Co-enrollment in other studies testing COVID-19 therapeutics may occur. Co-enrollment may affect the treatment effect estimates if there is effect modification associated with co-enrollment. We expect co-enrollment to occur in fewer than 5% of patients enrolled in the trial. However, because the decision to co-enroll is not affected by the treatment assignment in ACTIV 4 Host Tissue, co-enrollment will not favor any particular treatment. In addition, due to its rarity, we expect co-enrollment to have little impact on the estimated treatment effects, even when there is effect modification.

Differential treatment effect, also referred to as heterogeneity of treatment effect, refers to differences in efficacy as a function of pre-existing patient characteristics such as baseline variables. This is often assessed by forming subgroups or using an interaction analysis. Supplemental interaction analyses will be implemented to examine the potential for differential treatment effect. Differential treatment effect will be examined in strata defined by (but not limited to) respiratory support category at enrollment, status of co-enrollment in an open label clinical trial of antiplatelet agents (ACTIV 4a), age category, SARS-CoV-2 vaccination status, passive immunity status, co-enrollment in other studies, and concomitant use of study drug and

other medications during the study drug administration period. These analyses will be implemented using a modified version of the primary analysis method, where the treatment effect will be estimated separately for each level of the stratification variable. Stratum-specific treatment effect estimates will be presented with 95% Bayesian confidence interval. No formal hypothesis testing will be implemented for these analyses. Studies under this master protocol will be sized only for assessing efficacy using the primary analysis. Thus, there may be inadequate power to examine differential treatment.

6.6 Sample Size and Decision Thresholds

The maximum number of participants to be enrolled in sub studies under the Master Protocol is 600 participants per trial, resulting in approximately 300 patients per active treatment arm, and 300 patients in the matching placebo arm. The placebo arm will be shared across all active treatment arms. We expect placebo participants to continue to accrue for as long as there are additional treatments to test and cases to enroll.

Type-I error and power regarding the analysis of the primary outcome was assessed based on the pooled (across all active and placebo arms) distribution of the primary outcome among the first 100 participants to complete follow-up and monitoring. The efficacy threshold was identified using statistical simulation under the null hypothesis to ensure the study operating characteristics achieve design specifications. Pooled and blinded summaries of oxygen-free days at day 28 were used to approximate the distribution of the oxygen free days in the placebo group. Based on these data, the anticipated frequency distribution, mean, and median of oxygen-free days (OFDs) for the placebo group, and for the treatment group under hypothetical effect sizes computed using the PO model are displayed in the table below.

		Inferiority		**Superiority**						
OFDs / Odds Ratio	**Placebo**	**0.67**	**0.80**	**1.40**	**1.45**	**1.50**	**1.55**	**1.60**	**1.65**	**1.70**
Mean	8.8	6.6	7.5	10.8	11.1	11.3	11.5	11.7	11.9	12.0
Median	0	0	0	6.5	7.5	9.0	10.0	10.5	12.5	14.5
P(OFDs >= 22)	0.19	0.14	0.16	0.25	0.26	0.26	0.27	0.28	0.29	0.29
Proportion:										
-1 (death)	0.235	0.316	0.279	0.181	0.176	0.171	0.166	0.162	0.158	0.154
0	0.296	0.314	0.309	0.268	0.264	0.261	0.257	0.254	0.251	0.247
1	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
---	---	---	---	---	---	---	---	---	---	---
27	0.050	0.034	0.041	0.069	0.072	0.074	0.076	0.078	0.081	0.083
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Based on these data and effect size scenarios, a series of statistical simulations were implemented to examine the operating characteristics of the statistical study design described above, including the plan for randomization, statistical analysis method, interim analysis, and final assessments of efficacy using the odds ratio. In each simulation, participant age group, sex, and baseline WHO COVID severity score were randomly sampled with replacement from the values observed, and their effects on the primary outcome were simulated to match the estimated effects of age group, sex, and WHO score on the primary outcome among the first 100 participants. In order to assess the potential impact of attrition and loss-to-follow-up, partially observed oxygen free days were simulated to match the observed frequency of partially observed outcomes, which occurred in 12% of the first 100 participants. To encode attrition, a subset of the simulated study participants was selected at random, each with probability 0.12.

The primary outcome for each selected participant was encoded as partially observed by assuming that oxygen free days may have taken any value between -1 and a randomly sampled value ranging from the simulated oxygen free days to 28. For example, if the simulated oxygen free days is 10, then a value between 10 and 28 is sampled uniformly at random and this value is treated at the upper limit for the partially observed oxygen free days. This pattern of partially observed oxygen free days closely resembles the patterns observed among the first 100 participants. All simulation analyses, including those associated with interim and final assessment of efficacy and inferiority were implemented using the methods described above for the analysis of the primary outcome.

Simulation under the null hypothesis was used to select the efficacy threshold for the final analysis. The efficacy threshold was selected to ensure no more than 2.5% type-I error. In this simulation, 10000 replicates were used to ensure ~0.31% simulation margin of error in estimating the type-I error rate. The efficacy threshold was identified as 0.976. A final analysis will occur once enrollment, follow-up, and the required monitoring are completed for all participants. Should additional data be collected after enrollment is halted at an interim analysis, the final analysis will incorporate this additional data. If enrollment was halted at an interim analysis due to evidence of inferiority/harm, a conclusion of inferiority/harm will be indicated if the posterior probability for inferiority/harm remains greater than 0.95 at the final analysis. If the trial was not stopped early at an interim analysis due to evidence of inferiority/harm, efficacy will be indicated if the posterior probability for efficacy regarding the primary outcome exceeds 0.976 at the final analysis. If neither condition is met for a conclusion of efficacy or inferiority/harm at the final analysis, the trial is inconclusive. The efficacy and inferiority/harm thresholds will be applied as described in the table below.

Analysis	**Condition**	**Action**
Interim analysis	Inferiority/harm probability > 0.950	Halt enrollment
Final analysis	Inferiority/harm probability ≤ 0.950 at all interim analyses and efficacy probability > 0.976	Conclude efficacy
Final analysis	Inferiority/harm probability > 0.950	Conclude inferiority/harm

Using the selected efficacy and inferiority/harm thresholds, the results of 10000 simulations under the null hypothesis, and 1000 simulations per inferiority/efficacy scenario are summarized in the table below. In these simulations, the type-I error probability was 2.47%. The frequency of stopping early for inferiority under the null was 8.6% (5.3% at the first interim analysis, and 3.2% at the second interim analysis). *A maximum sample size of 600 participants per trial provides greater than 85% power to detect an odds ratio of 1.65, corresponding to a 3.1-day difference in mean OFDs, and a 7.8 percentage point reduction in 28-day mortality.* Differences larger than 2 ventilator-free days on average have been considered clinically important in prior trials.[2-4] *Thus, the minimum detectable effect with 85% power (MDE85) is an odds ratio of 1.65.* The frequency of stopping early for inferiority when there was an effect larger than OR=1.40 was <1%. When the simulated treatment was inferior/harmful relative to placebo, at OR=0.67, a conclusion of inferiority/harm occurred in 83.3% of simulated trials (39.1%at the first interim, 27.9% at the second interim, and 16.3% at the final analysis), and the average half-sample size was 193.9 per arm.

	Null	**Inferiority**		**Superiority**						
OFDs / Odds Ratio	**1.00**	**0.67**	**0.80**	**1.40**	**1.45**	**1.50**	**1.55**	**1.60**	**1.65**	**1.70**
Pr(Efficacy)	**0.025**	0.000	0.001	**0.552**	**0.631**	**0.705**	**0.782**	**0.826**	**0.856**	**0.893**

	Null	Inferiority		Superiority						
OFDs / Odds Ratio	**1.00**	**0.67**	**0.80**	**1.40**	**1.45**	**1.50**	**1.55**	**1.60**	**1.65**	**1.70**
Pr(Inferiority)	0.108	0.833	0.508	0.003	0.002	0.001	0.000	0.000	0.001	0.000
Pr(Inconclusive)	0.867	0.167	0.491	0.445	0.366	0.294	0.218	0.173	0.143	0.107
Average(N/2)	286.1	193.9	242.0	299.4	299.8	299.8	300.0	300.0	299.8	300.0

In order to characterize the effect of uncertainty in the distribution of the OFD outcome on the type-I error probability, simulations under the null hypothesis were twice repeated assuming a "mild" and "severe" distribution for the OFD outcome. The mild and severe distributions were selected such that the unadjusted mortality rate ranged ± 3% relative to the initial simulation. The results of 1000 simulations in each of the mild placebo and severe placebo scenarios are summarized in the table below. In these simulations, the type-I error probability was 2.5% and 2.3%.

	Severe	Mild
	OR = 1.00	**OR = 1.00**
Mortality rate	0.266	0.206
Pr(Efficacy)	**0.023**	**0.025**
Pr(Inferiority)	0.0.119	0.117
Pr(Inconclusive)	0.858	0.858
Average(N)	284.0	286.4

Prior to the start of enrollment, initial sample size assessments were based on pooled and blinded summaries of OFDs from the PassItOn (convalescent plasma) trial of patients hospitalized for COVID-19. The inclusion and exclusion criteria for PassItOn are similar to that for ACTIV 4 Host Tissue. In these initial assessments, the estimated MDE85 was OR=1.55. Statistical power was subsequently reassessed using OFDs summaries in the first 100 participants enrolled in ACTIV 4 Host Tissue, which demonstrated a more severe distribution relative to PassItOn participants (23.6% vs 17.6% mortality). The estimated MDE85 was OR=1.65 at the time of sample size reassessment. However, additional information from blinded summaries of the first 200 enrolled participants are consistent with the distribution of OFDs observed in PassItOn (18.6% vs 17.6% mortality). After discussion of these findings among the blinded study investigators and study sponsor, it was determined that statistical power was sufficient and no sample size adjustment was warranted.

7 ANALYSIS OF SECONDARY, EXPLORATORY, AND SAFETY OUTCOMES

Final analysis of the secondary, exploratory, and safety outcomes will be implemented separately for each study arm by comparing each active drug group with the corresponding pooled placebo comparator group. The effect of active agent versus placebo on the odds of binary and ordinal outcomes will be quantified using logistic and proportional odds logistic regression. Quantitative outcomes will be analyzed using a linear regression method. In order to preserve consistency across statistical analyses, we will uniformly apply a Bayesian approach using flat priors. Odds ratio, hazard ratio, and differences in mean estimates will be presented with a 95% credible interval.

7.1 Statistical Methods for Secondary, Exploratory, and Safety Analyses

The methods described below will be applied uniformly to the examine the effect of each active drug versus the placebo comparator on the secondary, exploratory, and safety outcomes, as appropriate.

7.1.1 Proportional Odds Logistic Regression (POLR)

Ordinal secondary, exploratory, and safety outcomes will be analyzed using a method similar to that described above for the analysis of the primary outcome, using proportional odds logistic regression (POLR), and adjusting for participant age group, sex, and WHO COVID ordinal severity at baseline. The effect of the active drug versus placebo will be presented using an odds ratio which quantifies the treatment effect on the odds of greater values of the ordinal outcome. The odds ratio will be presented with 95% credible interval. A flat prior distribution will be used for all model parameters. The posterior distribution for the log odds ratio will be approximated using the Laplace method. All statistical inferences about the odds ratio will be made using this method. The proportional odds assumption means that the odds-ratio has the same interpretation for all dichotomizations (that preserve ordering) of the ordinal outcome. The repeated dichotomization method, as described for the analysis of the primary outcome, will be used to assess for severe violations of the proportional odds assumptions. Missing or partially observed outcomes will be handled using the likelihood method as described for the primary analysis (see *Loss to Follow-up, Censoring, and Intercurrent Events*).

7.1.2 Logistic Regression (LogR)

Binary secondary, exploratory, and safety outcomes will be analyzed using logistic regression (LogR), and adjusting for participant age group, sex, and WHO COVID ordinal severity at baseline. The effect of the active drug versus placebo will be presented using an odds ratio which quantifies the treatment effect on the odds of outcome occurrence. The odds ratio will be presented with 95% credible interval. In addition, to facilitate clinical interpretability and meaningfulness, the difference in proportions corresponding to the most common (modal) values of the adjustment variables will be presented with 95% credible interval. A flat prior distribution will be used for all model parameters. The posterior distribution for the log odds ratio will be approximated using the Laplace method. All statistical inferences about the odds ratio and other posterior quantities will be made using this method. Missing outcomes will be handled using the likelihood method as described for the primary analysis (see *Loss to Follow-up, Censoring, and Intercurrent Events*).

7.1.3 Linear Regression (LinR)

Quantitative exploratory will be analyzed using linear regression (LinR), and adjusting for participant age group, sex, and WHO COVID ordinal severity at baseline. The effect of the active drug versus placebo will be presented using a difference in means. The difference in means will be presented with 95% credible interval. A flat prior distribution will be used for all model parameters. The posterior distribution for the difference in means will be approximated using the Laplace method. All statistical inferences about the difference in means will be made using this method. Graphical regression diagnostics, including normal Q-Q plots, will be used to assess for severe violations of the linear regression assumptions. Missing exploratory outcomes will be omitted from linear regression analyses.

7.1.4 Key Secondary Outcome Testing Procedure

A fixed-sequence testing approach will be used to preserve the type-I error rate across tests of the primary and key secondary outcomes. The key secondary outcomes will be tested in the specified order (see *Secondary Outcomes*). This approach provides strong control of the

familywise type-I error rate for the family of primary and key secondary outcomes. No other formal hypothesis tests will be made regarding the secondary, exploratory, or safety outcomes.

All key secondary outcomes use Bayesian logistic regression with a flat prior. Thus, the log odds ratio estimate is also a maximum likelihood estimate (MLE). At the final analysis (only) for each arm and key secondary outcome, efficacy will be indicated using a one-sided likelihood-based Wald test, to ensure a type-I error probability of 2.5% for each test. Specifically, a one-sided test of the null hypothesis (log OR = 0) will be computed by approximating the asymptotic distribution of the MLE under the null hypothesis: a Gaussian distribution with mean zero and variance equal to the inverse observed Fisher information. For descriptive purposes, evidence for efficacy will also be quantified using the posterior probability that the efficacy odds ratio is greater than one (i.e., treatment is associated with greater odds of a favorable outcome). This is denoted the "posterior probability for efficacy" or $P(\mathrm{OR} > 1|\mathrm{Data})$, where OR represents the odds ratio, and Data represents the mITT analysis dataset.

7.2 Analysis of Safety, Adherence, and Retention Outcomes for DSMB Review

Monitoring and reporting of safety events will be conducted continuously as described in the Data and Safety Monitoring Plan. Records will undergo monitoring for a two-week period (at minimum) prior to interim analysis for inferiority or futility. However, all records, regardless of monitoring status, will be used in enrollment, demographic, and safety summaries for DSMB safety reporting. Agent-specific safety and toxicity endpoints (if any) are detailed in that therapy's appendix. The frequencies of PSESEs, adverse events, mortality, and other safety endpoints will be reported. Screening, enrollment, withdrawal, loss-to-follow-up, mortality, study completion, hospitalization status and discharge location will be summarized in a similar manner. All safety-related protocol violations will be listed in the DSMB report. Receipt of planned therapy and adverse events will be recorded on case report forms and monitored continuously. Study drug stoppages and adverse events will be summarized and reported to the DSMB.

8 DATA FLOW, SHARING, AND ARCHIVING

8.1 Requests for secondary use of the data

Requests for secondary use of study data must adhere to review, approval, and provision processes developed by ACTIV 4 Host Tissue leadership and must comply with all applicable rules and regulations. All study data will be de-identified prior to sharing for secondary use.

8.2 Data flow for final and interim analyses

All data necessary for interim analyses, final analyses, and DSMB reporting will be exported from the EDC using the REDCap API. A custom R script will be used to both export the data and perform the interim analyses.

8.3 Archival data model

Data will remain in the production database. At the time of data locking, all users will be moved to read only access or removed, or as specified in the Data Management Plan.

8.4 Final analysis procedure

Once a study arm has completed enrollment, follow-up, and monitoring for all participants, all records that contribute to final analyses will be locked. Final analysis will be executed promptly after data lock, regardless of the status of other study arms. Blinded personnel will remain

blinded to the active/placebo status for individual participants until all arms that share blinded information with the completed arm have also been completed and their records locked. Final analyses will be executed by unblinded personnel only. Reporting of final analyses should avoid revealing the blinded treatment assignment for individual participants.

9 REFERENCES

1. Weins BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Statistics, 2003; 2:211-215.
2. Schoenfeld DA, Bernard GR, ARDS Network. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. Crit Care Med. 2002 Aug;30(8):1772–1777. PMID: 12163791
3. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, Connors AF, Hite RD, Harabin AL. Comparison of two fluid-management strategies in acute lung injury. N Engl J Med. 2006 Jun 15;354(24):2564–2575. PMID: 16714767
4. Laterre P-F, Berry SM, Blemings A, Carlsen JE, François B, Graves T, Jacobsen K, Lewis RJ, Opal SM, Perner A, Pickkers P, Russell JA, Windeløv NA, Yealy DM, Asfar P, Bestle MH, Muller G, Bruel C, Brulé N, Decruyenaere J, Dive A-M, Dugernier T, Krell K, Lefrant J-Y, Megarbane B, Mercier E, Mira J-P, Quenot J-P, Rasmussen BS, Thorsen-Meyer H-C, Vander Laenen M, Vang ML, Vignon P, Vinatier I, Wichmann S, Wittebole X, Kjølbye AL, Angus DC, SEPSIS-ACT Investigators. Effect of Selepressin vs Placebo on Ventilator- and Vasopressor-Free Days in Patients With Septic Shock: The SEPSIS-ACT Randomized Clinical Trial. JAMA. 2019 Oct 2; PMCID: PMC6802260
5. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, Slovis CM, Lindsell CJ, Ehrenfeld JM, Siew ED, Shaw AD, Bernard GR, Rice TW, SALT-ED Investigators. Balanced Crystalloids versus Saline in Noncritically Ill Adults. N Engl J Med. 2018 01;378(9):819–828. PMCID: PMC5846618
6. Self WH, Semler MW, Leither LM, Casey JD, Angus DC, Brower RG, Chang SY, Collins SP, Eppensteiner JC, Filbin MR, Files DC, Gibbs KW, Ginde AA, Gong MN, Harrell FE, Hayden DL, Hough CL, Johnson NJ, Khan A, Lindsell CJ, Matthay MA, Moss M, Park PK, Rice TW, Robinson BRH, Schoenfeld DA, Shapiro NI, Steingrub JS, Ulysse CA, Weissman A, Yealy DM, Thompson BT, Brown SM, National Heart, Lung, and Blood Institute PETAL Clinical Trials Network, Steingrub J, Smithline H, Tiru B, Tidswell M, Kozikowski L, Thornton-Thompson S, De Souza L, Hou P, Baron R, Massaro A, Aisiku I, Fredenburgh L, Seethala R, Johnsky L, Riker R, Seder D, May T, Baumann M, Eldridge A, Lord C, Shapiro N, Talmor D, O'Mara T, Kirk C, Harrison K, Kurt L, Schermerhorn M, Banner-Goodspeed V, Boyle K, Dubosh N, Filbin M, Hibbert K, Parry B, Lavin-Parsons K, Pulido N, Lilley B, Lodenstein C, Margolin J, Brait K, Jones A, Galbraith J, Peacock R, Nandi U, Wachs T, Matthay M, Liu K, Kangelaris K, Wang R, Calfee C, Yee K, Hendey G, Chang S, Lim G, Qadir N, Tam A, Beutler R, Levitt J, Wilson J, Rogers A, Vojnik R, Roque J, Albertson T, Chenoweth J, Adams J, Pearson S, Juarez M, Almasri E, Fayed M, Hughes A, Hillard S, Huebinger R, Wang H, Vidales E, Patel B, Ginde A, Moss M, Baduashvili A, McKeehan J, Finck L, Higgins C, Howell M, Douglas I, Haukoos J, Hiller T, Lyle C, Cupelo A, Caruso E, Camacho C, Gravitz S, Finigan J, Griesmer C, Park P, Hyzy R, Nelson K, McDonough K, Olbrich N, Williams M, Kapoor R, Nash J, Willig M, Ford H, Gardner-Gray J, Ramesh M, Moses M, Ng Gong M, Aboodi M, Asghar A, Amosu O, Torres M, Kaur S, Chen J-T, Hope A, Lopez B, Rosales K, Young You J, Mosier J, Hypes C, Natt B, Borg B, Salvagio Campbell E, Hite RD,

Hudock K, Cresie A, Alhasan F, Gomez-Arroyo J, Duggal A, Mehkri O, Hastings A, Sahoo D, Abi Fadel F, Gole S, Shaner V, Wimer A, Meli Y, King A, Terndrup T, Exline M, Pannu S, Robart E, Karow S, Hough C, Robinson B, Johnson N, Henning D, Campo M, Gundel S, Seghal S, Katsandres S, Dean S, Khan A, Krol O, Jouzestani M, Huynh P, Weissman A, Yealy D, Scholl D, Adams P, McVerry B, Huang D, Angus D, Schooler J, Moore S, Files C, Miller C, Gibbs K, LaRose M, Flores L, Koehler L, Morse C, Sanders J, Langford C, Nanney K, MdalaGausi M, Yeboah P, Morris P, Sturgill J, Seif S, Cassity E, Dhar S, de Wit M, Mason J, Goodwin A, Hall G, Grady A, Chamberlain A, Brown S, Bledsoe J, Leither L, Peltan I, Starr N, Fergus M, Aston V, Montgomery Q, Smith R, Merrill M, Brown K, Armbruster B, Harris E, Middleton E, Paine R, Johnson S, Barrios M, Eppensteiner J, Limkakeng A, McGowan L, Porter T, Bouffler A, Leahy JC, deBoisblanc B, Lammi M, Happel K, Lauto P, Self W, Casey J, Semler M, Collins S, Harrell F, Lindsell C, Rice T, Stubblefield W, Gray C, Johnson J, Roth M, Hays M, Torr D, Zakaria A, Schoenfeld D, Thompson T, Hayden D, Ringwood N, Oldmixon C, Ulysse C, Morse R, Muzikansky A, Fitzgerald L, Whitaker S, Lagakos A, Brower R, Reineck L, Aggarwal N, Bienstock K, Freemer M, Maclawiw M, Weinmann G, Morrison L, Gillespie M, Kryscio R, Brodie D, Zareba W, Rompalo A, Boeckh M, Parsons P, Christie J, Hall J, Horton N, Zoloth L, Dickert N, Diercks D. Effect of Hydroxychloroquine on Clinical Status at 14 Days in Hospitalized Patients With COVID-19: A Randomized Clinical Trial. JAMA. 2020 Dec 1;324(21):2165–2176. PMCID: PMC7653542

7. Harrell, Frank. General COVID-19 Therapeutics Trial Design [Internet]. 2020. Available from: https://hbiostat.org/proj/covid19/statdesign.html
8. Zhenming S, McCullagh P. Laplace approximation of high dimensional integrals. Journal of the Royal Statistical Society. 1995;57(4):749–760.
9. Doob J. The limiting distributions of certain statistics. The Annals of Mathematical Statistics. 1935 Sep 1;6(3):160–169.
10. Simpson DG, Carroll RJ, Zhou H, Guth DJ. Interval Censoring and Marginal Analysis in Ordinal Regression. Journal of Agricultural, Biological, and Environmental Statistics. 1996 Sep;1(3):354.
11. Food and Drug Administration. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials [Internet]. 2017. Available from: https://www.fda.gov/media/108698/download
12. Liu Q, Shepherd BE, Li C, Harrell FE. Modeling continuous response variables using ordinal regression. Stat Med. 2017 Nov 30;36(27):4316–4335. PMCID: PMC5675816
13. Harrell, Frank. Violation of Proportional Odds is Not Fatal | Statistical Thinking [Internet]. 2020. Available from: https://www.fharrell.com/post/po/
14. P Peterson, B. and Harrell Jr, F.E., 1990. Partial proportional odds models for ordinal response variables. Journal of the Royal Statistical Society: Series C (Applied Statistics), 39(2), pp.205-217.

10 APPENDIX: ALGORITHM TO COMPUTE PRIMARY OUTCOME

The primary outcome is oxygen-free days at study day 28. It can take values -1, 0, 2, …, 27, 28. When computing oxygen free days, the “outcome” for each participant should be a length 30 vector of zeros and ones that indicate which of the 30 possible values (-1, 0, 2, …, 27, 28) that OFDs could take for that participant. This representation allows for arbitrary censoring of the outcome. For example [0,1,1,0,0,...,0] indicates that OFDs could be either 0 or 1. If there is loss-to-follow-up, withdrawal, or missing follow-up information, there can be interval censoring. The algorithm below is designed to compute OFDs in this representation.

- If participant was deceased by study day 28, OFDs is [1,0,0,0,0,...,0]
- For study day 1 through 28, compute whether or not supplemental oxygen was used (code with “yes” or “no”), or if supplemental oxygen use was uncertain (code with “?”).
    - For our purposes supplemental oxygen means oxygen use that exceeds any pre-enrollment home oxygen use. Home oxygen use is recorded in the “Medical History” form in variables mhco2, mhio2, and the amount (L/m) in field home_ox. If a participant had not used pre-enrollment home oxygen, then it should be assumed that all hospital and post-discharge use of oxygen counts against oxygen-free days. If a participant had used pre-enrollment home oxygen, then only the supplemental oxygen use that exceeds the amount used at home should count against oxygen-free days. If the participant is in the inpatient phase of the study and using standard supplemental oxygen (o2type = “O2 by mask or nasal prongs”), then the L/m recorded on the vitals signs form (o2_lpm_cannula_sofa) must exceed the amount used at home (home_ox). If hospital oxygen use takes any other value except “No O2 therapy” and “O2 by mask or nasal prongs”, then that study day should count against oxygen free-days.
    - If the participant is in the outpatient phase of the study (i.e., after discharge from the enrollment admission or after 28 days, whichever comes first), but is not hospitalized, then only the post-discharge home oxygen use that exceeds the amount used at home prior to enrollment (if any) will count against oxygen-free days. The phone script and outpatient form are designed to record only the home oxygen use that exceeds any pre-hospitalization oxygen use.
    - If the participant is in the outpatient phase of the study, but is hospitalized, the branching logic on the outpatient form determines whether the participant had used oxygen. Any hospital oxygen use during the outpatient phase counts against oxygen-free days.
    - If the preceding calculations cannot be made for any particular study day, then the supplemental oxygen status is “?” for that study day.
- The preceding step results in “yes”, “no”, or “?” for each study day 1 through 28.
    - If there are no “?” values, then OFDs is 28 minus the number of days between and including the days of the first “yes” and the last “yes”.
    - If there are “?” before the first “yes” or after the last yes, then OFDs is partially observed and multiple values are possible. To compute the possible values, consider each possible pair of first ‘yes’ and last ‘yes’ days, and compute the associated OFDs.

  OFDs should be represented as a vector of length 30, one element for each value that OFDs can take: -1, 0, 1, …, 27, 28. There should be a 1 for each element that OFD that is possible for this participant, and a zero otherwise. The -1 (first) element should take a value 0 if the participant was known to be alive at day 28 and 1 otherwise.

11 APPENDIX: CUMULATIVE LOGIT MODEL

11.1 Model Formulation

The cumulative logit model can be written in terms of the covariates X and an ordinal outcome variable G, where probabilities of outcome value g or smaller are modeled as follows

$$\Pr(G \leq g|X) = \text{expit}(\alpha_g - X\beta). \quad (1)$$

Without loss of generality, an outcome with p levels may be coded using the first p integers, such that g may take on the values $1, \ldots, p$. In the expression above, α_g is a scalar intercept, expit is the logistic (inverse logit) transformation, and the vector X contains coded baseline covariates and the active/placebo treatment indicator. The model has intercepts for each of the first $p-1$ outcome levels, and the intercepts must be ordered: $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{p-1}$. The ordering of intercepts ensures that the probabilities $\Pr(G \leq g|X)$ are monotonically increasing in g. The parameter vector β represents the log odds ratios (OR) associated with the effects of covariates and group assignment. Specifically, the group assignment odds ratio represents the relative effect of treatment versus placebo on the odds $\Pr(G > g|X)/(1 - \Pr(G > g|X))$, for each of the first $p-1$ values that G may take.

The $p-1$ linear predictors $\alpha_g - X\beta$ represent the logit transformed cumulative probabilities associated with the first $p-1$ levels of the ordinal outcome, adjusted for the effects of covariates X. The probabilities that the outcome takes a specific value g, adjusted for covariates X, is derived as follows:

$$\Pr(G = g|X) = \mathrm{expit}\big(\alpha_g - X\beta\big) - \mathrm{expit}\big(\alpha_{g-1} - X\beta\big), \tag{2}$$

where $\mathrm{expit}(\alpha_0 - X\beta)$ and $\mathrm{expit}\big(\alpha_p - X\beta\big)$ are defined to be 0 and 1, respectively.
When there are partially observed ordinal outcomes, it is convenient to recode the outcome as a vector $Y = [Y_1, \ldots, Y_p]$, such that $Y_g = 1$ if $G = g$ or when g is one of the values that G might have taken if the outcome were fully observed, and $Y_g = 0$ otherwise. Thus, the cumulative logit model may be written as follows

$$\Pr\big(Y_1 = 1 \cup Y_2 = 1 \cup \cdots \cup Y_g = 1\big|X\big) = \Pr(G \leq g|X) = \mathrm{expit}\big(\alpha_g - X\beta\big). \tag{3}$$

Denote a sample of covariate vectors $x_1, \ldots, x_N$ and outcomes $g_1, \ldots, g_N$, and corresponding outcome vectors $y_1, \ldots, y_N$, where $y_i = [y_{i1}, \ldots, y_{ip}]$. Using this representation, partially observed outcomes are encoded by assigning a value 1 to each element of y_i that the outcome g_i might have taken if fully observed. For example, if g_i might have taken values 1 or 2, but other values were not possible, then y_i would be coded $y_i = [1, 1, 0, \ldots, 0]$. Further denote the collection of model parameters $\theta = \big[\alpha_1, \ldots, \alpha_{p-1}, \beta\big]$. Using this notation, the observed data likelihood is as follows:

$$L(\theta|y_1 \ldots y_N, x_1 \ldots x_N) = \prod_{i=1}^{N} L_i(\theta|y_i, x_i) = \prod_{i=1}^{N}\sum_{j=1}^{p} I\big(y_{ij} = 1\big)\Pr(Y_j = 1\,|X = x_i), \tag{4}$$

where $I(\cdot)$ is the indicator function that takes a value 1 when its argument is true, and 0 otherwise.

In a Bayesian analysis, the posterior density function is proportional to the likelihood function multiplied by the prior density function as follows:

$$P(\theta|y_1 \ldots y_N, x_1 \ldots x_N) \propto L(\theta|y_1 \ldots y_N, x_1 \ldots x_N)P(\theta) \tag{5}$$

A flat prior distribution, where $P(\theta) \propto 1$, is used for all model parameters. Thus, the posterior density is proportional to the likelihood function.

Each supplemental estimand is a treatment difference in one of the following summaries of the adjusted outcome distribution: mean, median, proportion experiencing mortality by day 28, and proportion receiving oxygen every day until day 28. Supplemental estimands are computed using the probabilities expressed in equations (1) and (2) above, adjusted to the modal value for each covariate: age group 31-64, WHO COVID-19 Score value 4 (hospitalized receiving oxygen via nasal cannula), and male sex. The adjusted means are computed as the linear combination of all possible outcome values and their associated adjusted probabilities. The adjusted medians are computed as the smallest value of g such that the $\Pr(G \geq g|X) \geq 0.5$. The adjusted proportions of participants experiencing mortality by day 28 or oxygen requirement every day until day 28 are computed as $\Pr(G = -1|X)$ and $\Pr(G = 0|X)$, respectively.

11.2 Model-based Statistical Inferences

The posterior distribution for the log odds ratio and any other required parameter is approximated using the Laplace method. A flat prior ensures the Laplace-approximated posterior distribution is identical to the approximate sampling distribution of the maximum likelihood estimate for θ; in both cases a normal distribution centered at the estimate (i.e., the maximum likelihood estimate or equivalently the maximum *a posteriori* estimate) with variance-covariance equal to the negative inverse Hessian of the log likelihood function (inverse observed Fisher information) evaluated at the estimate (see "Appendix: Laplace Approximation"). All statistical inferences about the odds ratio and derivative quantities (including all supplementary estimands) will be made using this method.

11.3 Model Fitting and Computation

The cumulative logit model is implemented in the R code file "clm_model.R". Readers should examine the `clm_fit` function first, which is the entry point for model fitting, and then examine other functions as they are called by `clm_fit`. The function `clm_fit` takes as arguments the matrix of coded covariates `x`, and a matrix of coded outcomes `y`. Each matrix has one row per record (i.e., study participant). The covariate matrix has one column per coded covariate (e.g., age group has three levels and thus requires two columns to distinguish the levels), and the outcome matrix has one column per value that the outcome might take. The cells of the outcome matrix `y` contain the values y_{ij} as defined above (see "Model Formulation").

In practice, when one or more levels of an ordinal outcome are not observed in the analysis data set, some of the model intercepts are not estimable (i.e., there is no unique set of model intercepts that maximizes the likelihood/posterior density function). To overcome this, each outcome level is characterized as "estimable" if there is at least one record in the analysis data set where that level is observed and no other level was possible (i.e., ignoring partially observed outcomes), and "not estimable" otherwise. Levels of the outcome that are not estimable are collapsed with the nearest adjacent estimable level to form a new level, e.g., levels 3, 4, and 5 may be collapsed to form level "3|5". When levels are collapsed, if any collapsed level was possible as part of a partially observed outcome, then the collapsed level is considered possible as well. This functionality is implemented by the function `clm_collapse`, which is called by `clm_fit` prior to any model fitting.

The estimate of θ is found by maximizing the log of the posterior density function (i.e., a maximum *a posteriori* estimate, or MAP for short) defined in expression (5). Note that the normalizing constant in expression (5) is not needed to identify the MAP estimate, nor is it necessary to form a Laplace approximation to the posterior density. The estimate of θ is found using an iterative optimization algorithm, and the associated observed Fisher information is estimated using a finite difference method. These calculations are implemented using the R function `optim`, which uses the quasi-Newton "BFGS" method (Byrd, Lu, Nocedal, and Zhu,

1995, A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 1190–1208. doi: 10.1137/0916069), and is built-in as part of the "stats" package for R (R Core Team, 2022, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). The initial values for β are set to zero. Initial values for the model intercepts are generated by first calculating the fraction of each observed outcome level (i.e., an initial estimate of $\Pr(G = g|X)$ where $\beta = 0$), and then applying the inverse of expression (2) as follows:

$$\alpha_g^{\text{init}} = \text{logit}\left(\sum_{k=1}^{g} \frac{\sum_i y_{ik}}{\sum_i \sum_j y_{ij}}\right). \quad (6)$$

The initial values calculations for the model intercepts are implemented by the function `clm_alpha_init`. Starting at the initial values, the `optim` function iteratively maximizes the `clm_optim` function, which computes the log of the posterior density function given by expression (5). The `clm_optim` function calls the `clm_loglik` and `clm_logpri` functions, which evaluate the log of the likelihood function given by expression (4) and log of the prior density function (defined to be zero for a flat prior), respectively. The `clm_loglik` function calls `clm_predict` which computes, for each record, the linear predictors, $\alpha_g - X\beta = \text{logit}\Pr(G \leq g|X)$, and the associated covariate adjusted probabilities for each ordinal outcome level $\Pr(G = g|X)$. The `clm_predict` function calls `alphs_to_probs` to convert the logit cumulative probabilities to level specific probabilities according to expression (2). The `probs_to_alphs` function computes the inverse of `alphs_to_probs`.

The `clm_fit` function returns a model fit object that contains a model convergence assessment, the MAP estimate for θ, and the estimated Hessian of the log posterior density function evaluated at the estimate. The MAP estimate and Hessian are sufficient to define the Laplace (Normal) approximation to the posterior density, and are used to compute posterior cumulative probabilities as follows

$$\Pr(\theta_k \leq q|y_1 \ldots y_N, x_1 \ldots x_N) = \Phi\left(\frac{q-\hat{\theta}_k}{\sqrt{[-H^{-1}]_{kk}}}\right),$$

where H is the estimated Hessian, $\hat{\theta}$ is the MAP estimate, and Φ is the standard normal cumulative density function. This is implemented by the `clm_ppost` function for specified scalar elements θ_k. Notably, this function is used to compute the posterior probabilities used for decision-making at the interim and final analyses.

For supplementary estimands, $g(\theta)$, that are smooth scalar functions of θ (i.e., treatment difference in the mean of the primary outcome, and treatment difference in the probabilities associated with outcome categories -1 and 0), the posterior distribution will be approximated using the delta method, for example, to compute posterior cumulative probabilities as follows:

$$\Pr(g(\theta) \leq q|y_1 \ldots y_N, x_1 \ldots x_N) = \Phi\left(\frac{q-g(\hat{\theta})}{\sqrt{g'(\hat{\theta})^T[-H^{-1}]g'(\hat{\theta})}}\right),$$

Where $g'(\hat{\theta})$ is the gradient of $g(\cdot)$ evaluated at $\hat{\theta}$, which is approximated numerically using a finite difference method. For non-smooth scalar functions of θ (i.e., treatment difference in the median of the primary outcome), the posterior distribution will be identified using a Monte Carlo

method; by generating 10000 realizations from the posterior distribution for θ, and evaluating the supplementary estimand using those realizations. For either approach, an equal-tailed, level $(1-\alpha)$ credible interval will then be identified by selecting the $\alpha/2$ and $1-\alpha/2$ quantiles of the approximate posterior distribution. The functions `clm_crint_delta` and `clm_crint_montecarlo` compute credible intervals for supplementary estimands using the two methods described above, respectively. The adjusted outcome mean and median calculations are implemented by functions `mean_xp` and `quantile_xp`, respectively. The four supplementary estimands are implemented by functions defined in the R code file "supplemental_estimands.R"

The four supplementary estimands include the treatment difference in mean and median of the primary outcome, and the treatment difference in probabilities associated with outcome levels -1 and 0. Each of these estimands will be adjusted to the most common (modal) value for each covariate. The mean and median estimates are defined as the mean and median of the distribution defined by the cumulative probabilities associated with each outcome level, adjusted for covariates.

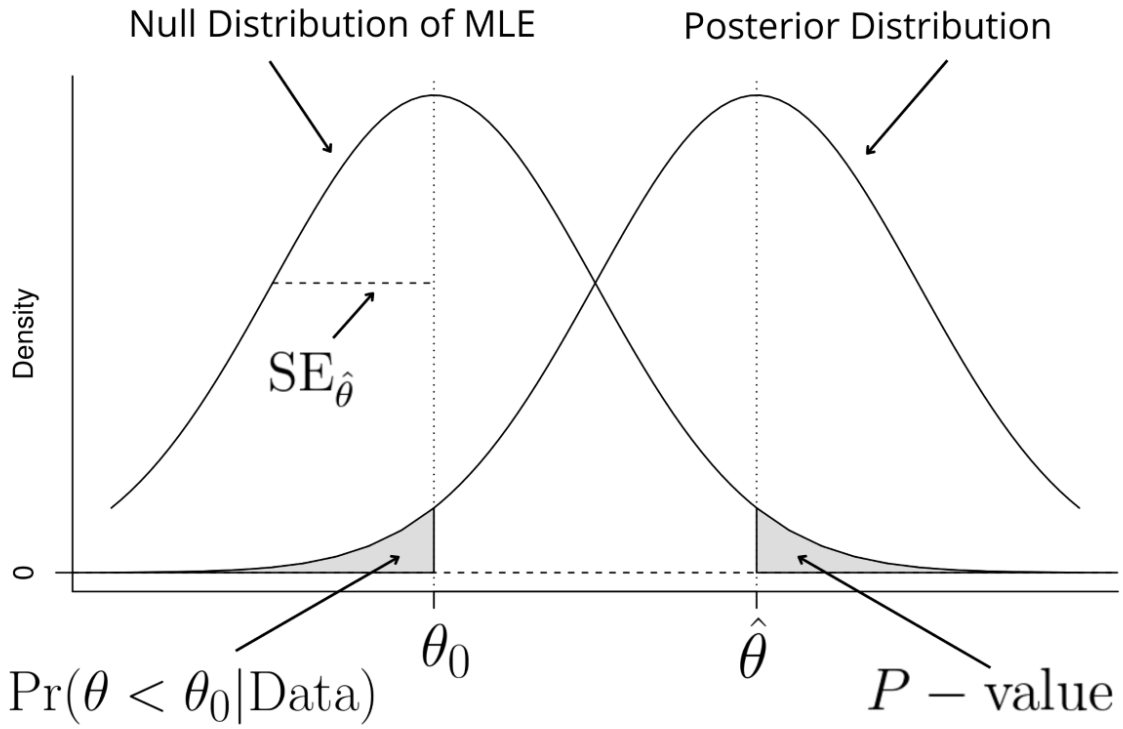
12 APPENDIX: KEY SECONDARY OUTCOME TESTING PROCEDURE

Each trial in the ACTIV 4 Host Tissue platform will separately use a fixed sequence method to control the familywise type-I error probability, i.e., the probability of erroneously concluding efficacy of the trial intervention with respect to any one or more of the primary and key secondary outcomes. Specifically, a conclusion of efficacy regarding the primary outcome will be required prior to testing the first designated key secondary outcome. Each subsequent key secondary outcome, in the designated order, will take place only if the preceding key secondary outcome demonstrates efficacy. This approach provides strong control of the familywise type-I error probability for the family of primary and key secondary outcomes. For weak familywise type-I error control (i.e., under the assumption that the intervention effect is null for all tests in the family), the fixed sequence method requires only that the test of the primary outcome (i.e., the outcome tested first) have the specified type-I error rate. For strong type-I error control, the fixed sequence procedure requires that each individual test in the sequence have the desired type-I error probability, 2.5% for trials under the ACTIV 4 Host Tissue platform. Because the test of efficacy associated with the primary outcome has adaptive elements, including interim analyses, a statistical simulation (as described in the "Statistical Analysis Plan") was implemented to identify the test characteristics that ensure a 2.5% type-I error probability for that test. Each key secondary outcome is tested for efficacy only at the final analysis. Thus, type-I error control for the key secondary outcomes relies on established theoretical arguments and methods.

All key secondary outcomes use Bayesian logistic regression or proportional odds logistic regression. If key secondary outcome testing is required under the fixed sequence procedure, efficacy will be concluded if the posterior probability for efficacy (P(OR > 1|Data) for Alive and respiratory failure-free at day 28, and P(OR < 1 | Data) for WHO 8-point ordinal scale at day 28 and Mortality at day 28) exceeds 0.975.

Because a flat prior is used, and the posterior is computed using a Laplace approximation, the maximum *a posteriori* estimate of the log odds ratio is identical to the maximum likelihood estimate (MLE), and the Laplace approximated posterior distribution is identical to the approximate sampling distribution of the MLE: a normal distribution with mean equal to the estimate and variance-covariance equal to the inverse observed Fisher information (see Appendix: Laplace Approximation). In conventional frequentist testing, efficacy is indicated

when the estimate exceeds a critical value selected such that the frequency of this occurring under the null hypothesis is 0.025. Because of the equivalence between the approximate posterior and MLE sampling distributions, setting the posterior probability for efficacy threshold to 0.975 ensures that any estimate meeting this threshold must also exceed the critical value that ensures less than 2.5% type-I error frequency. The figure below illustrates this concept:



13 APPENDIX: LAPLACE APPROXIMATION

Let random variables $Y_1 \ldots Y_N$ represent an independent and identically distributed sample from a probability distribution with density function $f(Y|\theta)$, and define $y_1 \ldots y_N$ as realizations of this sample. If $f(Y|\theta)$ is derived from a regression model, then the density function may also condition on covariates (elsewhere denoted X and x). However, covariate information is not pertinent to the derivations below, and are omitted for clarity. The likelihood function is defined as follows:

$$L(\theta|y_1 \ldots y_N) = \prod_{i=1}^{N} f(y_i|\theta) \tag{1}$$

The natural log of the likelihood function is defined as follows:

$$\ell(\theta|y_1 \ldots y_N) = \sum_{i=1}^{N} \log f(y_i|\theta) \tag{2}$$

In Bayesian analysis, the posterior density function is proportional to the likelihood function multiplied by the prior density function as follows:

$$P(\theta|y_1 \dots y_N) \propto L(\theta|y_1 \dots y_N)P(\theta) \tag{3}$$

13.1 Equivalence of MAP and MLE with Flat Prior

A "flat prior" density function is defined to be proportional to 1 for all values of θ. Thus, when a flat prior is specified, the posterior density function is proportional to the likelihood function. In addition, the maximum *a posteriori* (MAP) estimator of θ is also a maximum likelihood estimator (MLE):

$$\hat{\theta} = \arg\max_{\theta} P(\theta|y_1 \dots y_N) = \arg\max_{\theta} L(\theta|y_1 \dots y_N) = \arg\max_{\theta} \ell(\theta|y_1 \dots y_N) \tag{4}$$

13.2 Asymptotic Normality of MLE

Under regularity conditions, the MLE converges in distribution to a normal distribution:

$$\hat{\theta} \xrightarrow{d} N(\theta_0, I^{-1}) \tag{5}$$

where θ_0 is the true but unknown value of θ, and I is the Fisher information:

$$I = E_{\theta_0}\left[-\frac{\partial^2}{\partial\theta^2}\ell(\theta_0|Y_1 \dots Y_N)\right] \tag{6}$$

In practice, because θ_0 is unknown, inferences about θ_0 are made by substituting $\hat{\theta}$ in place of θ_0 and the observed information is substituted in place of the Fisher information:

$$\hat{\theta} \sim N\left(\hat{\theta}, \hat{I}^{-1}\right) \tag{7}$$

The observed information is the negative Hessian of the log likelihood function evaluated at $\hat{\theta}$:

$$\hat{I} = \left[-\frac{\partial^2}{\partial\theta^2}\ell(\theta|y_1 \dots y_N)\right]_{\theta=\hat{\theta}} \tag{8}$$

13.3 Laplace Approximation to Posterior

The Laplace approximation to a posterior density function (or any density function) is based on a two-term Taylor expansion of the natural log of the density function about $\hat{\theta}$:

$$q(\theta) \approx q\left(\hat{\theta}\right) + \left(\theta - \hat{\theta}\right)q'\left(\hat{\theta}\right) + \frac{1}{2}\left(\theta - \hat{\theta}\right)^T q''\left(\hat{\theta}\right)\left(\theta - \hat{\theta}\right) \tag{9}$$

where $q(\theta)$ is the log posterior density function and $q'\left(\hat{\theta}\right)$ and $q''\left(\hat{\theta}\right)$ are the gradient and Hessian of $q(\theta)$, respectively, evaluated at $\hat{\theta}$. When a flat prior is used, $q(\theta)$ is equal to the log likelihood function plus a constant c:

$$q(\theta) = \log P(\theta|y_1 \dots y_N) = \ell(\theta|y_1 \dots y_N) + c \tag{10}$$

Because $\hat{\theta}$ is defined to be the MAP estimate, $q'(\hat{\theta}) = 0$. Thus, expression (9) simplifies:

$$q(\theta) \approx -\frac{1}{2}(\theta - \hat{\theta})^T[-q''(\hat{\theta})](\theta - \hat{\theta}) + c \quad (11)$$

where the negative Hessian is identical to the observed information when a flat prior is used:

$$-q''(\hat{\theta}) = \left[-\frac{\partial^2}{\partial\theta^2}\log P(\theta|y_1 \dots y_N)\right]_{\theta=\hat{\theta}} = \left[-\frac{\partial^2}{\partial\theta^2}\ell(\theta|y_1 \dots y_N)\right]_{\theta=\hat{\theta}} = \hat{I} \quad (12)$$

Exponentiating expression (11) demonstrates that the Laplace approximation to the posterior density must be a normal density with mean $\hat{\theta}$ and variance-covariance $\hat{I}^{-1}$. This is identical to the asymptotic sampling distribution of the MLE given in expression (7):

$$(\theta|y_1 \dots y_N) \sim N(\hat{\theta}, \hat{I}^{-1}) \quad (13)$$

Under regularity conditions, the Bernstein-von Mises theorem provides asymptotic guarantees regarding the quality of the Laplace approximation.