ACTIV-6
COVID-19 Outpatient Randomized Trial to Evaluate Efficacy of Repurposed Medications

STATISTICAL ANALYSIS PLAN (SAP)

Funding Agency	National Center for Advancing Translational Sciences
Date	2022-04-01
Version	5.0
Principal Investigator	Adrian Hernandez, MD, MHS Duke Clinical Research Institute 200 Morris St. Durham, NC 27701 Phone: 919-668-7515 Email: adrian.hernandez@duke.edu
Co-Principal Investigator/IND Sponsor:	Susanna Naggie, MD, MHS Duke Clinical Research Institute 200 Morris St. Durham, NC 27701 Phone: 919-684-2584 Email: susanna.naggie@duke.edu
Data Coordinating Center	Chris Lindsell, PhD Vanderbilt University School of Medicine 2525 West End Avenue Suite 1100, Rm 11129 Nashville, TN 37203 Phone: 615-343-9867 Email: chris.lindsell@vumc.org
Statisticians	Thomas G. Stewart, PhD (DCC) Frank Harrell, PhD (DCC)

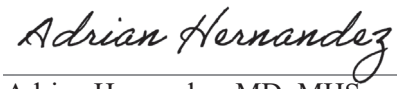
Approval Signatures

By signing below, I indicate that I have reviewed this document in its entirety and approve its contents.

Statistician:

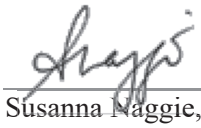
Thomas G. Stewart, PhD — 2022-04-01 Approval Date

Principle Investigator:



Adrian Hernandez, MD, MHS — 4/8/2022 Approval Date

Co-Principal Investigator/IND Sponsor:



Susanna Naggie, MD, MHS — April 15 2022 Approval Date

Contents

1 Summary

The key features of the trial are summarized in detail in the study protocol. The following table restates some of the key study features from the protocol.

Feature	**Description**
Study Design	Multicenter, blinded, placebo-controlled randomized clinical trial
Sample Size	A maximum of 1200
Primary Aim	To evaluate the effectiveness of repurposed medications [study drug(s)] in reducing symptoms of non-hospitalized participants with mild to moderate COVID-19
Primary Outcome	Appendix specific. Either (a) time to recovery or the composite endpoint of (b) hospitalization or death
Primary Estimand	Appendix specific.
Secondary Outcomes	• Mean time unwell • Hospitalization or death (Day 14 and Day 28), if not primary • Time to recovery (within 28 days), if not primary • Mortality (Day 28) • Hospitalization, urgent care, emergency room visit, or death (Day 28) • COVID Clinical Progression Scale (Day 7, Day 14, and Day 28) • Modified Patient-Reported Outcomes Measurement Information System (PROMIS)-29 (Day 7, Day 14, Day 28, and Day 90)
Inference approach	Bayesian posterior probabilities, point estimates, and credible intervals
Planned screening analyses	N = 300 and possibly 600
Screening endpoint guidance quantity	$P(screening\ endpoint\ at\ N = 1200 > 0)$
Planned interim analyses	If screening endpoint met, primary endpoint interim analysis at possibly N=300, 600, 900
Planned interim rules	Futility: posterior predictive $P(\text{primary endpoint efficacy at N} = 1200) < 0.05$ Efficacy: posterior P(primary endpoint efficacy) > 0.95
Type I error control	The combination of the screening endpoint, primary endpoint efficacy and futility rules, schedule of interim analyses, and choice of priors result in a simulated type I error rate of 0.05
Power	The minimal detectable effect at 80% power is a time to recovery hazard ratio of 1.2.

Study populations	• Modified Intention to Treat (mITT) defined as all participants who receive study drug/placebo. • Safety population defined as all participants in the mITT population who report taking at least one dose of study drug or matching placebo

2 Schedule of planned analyses

Fixed enrollment triggers will be used for interim analyses. A screening analysis will occur after enrollment and completion of 14 day follow up of approximately 300 participants in a study arm (150 in study drug arm and 150 in placebo arm). If deemed appropriate based on the screening endpoint, a primary endpoint interim may occur immediately at N=300. Subsequent interim analyses may occur at N=600 or 900. If the study (appendix) has not terminated early for futility or efficacy at previous interim analyses, a final analysis will occur at N=1200. Individual study drugs may require different sample sizes, and the sample sizes may be adjusted based on the results of interim analyses.

3 Analysis plan for the primary endpoint

The primary endpoint is appendix specific. Possible endpoints include (a) time to recovery (within 28 days) or (b) hospitalization or death (within 28 days).

3.1 Time to recovery (within 28 days)

3.1.1 Definition

Time to recovery is defined in accordance with the sustained clinical recovery endpoint described in FDA guidance document for COVID-19 clinical trials (https://www.fda.gov/media/137926/download). Specifically, day of recovery is the first day on which the participant reports three consecutive days without COVID-19 symptoms.

3.1.2 Estimand

The Fine-Gray approach will be used to compare each intervention to placebo with respect to the cumulative incidence of recovery, accounting for the semi-competing risk of mortality (Fine and Gray 1999). The treatment effect parameter is estimated from a piece-wise proportional hazards regression model of the recovery endpoint in which subjects who die are retained in the risk set.

3.1.3 Model

The proportional hazards regression model will be constructed using a smooth, flexible parametric baseline hazard function. The model will be estimated within the Bayesian framework. The following covariates will be included in the regression component of the model:

1. Randomization assignment
2. Age as a restricted cubic spline with 3 knots
3. Gender
4. Duration of symptoms prior to treatment
5. Calendar time as restricted cubic spline with 4 knots
6. Vaccination status
7. Geographic region (Northeast, Midwest, South, West)
8. Call center indicator
9. Additional appendix-specific covariates relevant to baseline disease severity and patient risk.

3.1.4 Priors

The regression parameter for randomization assignment will be assigned a normal prior distribution with mean zero and standard deviation 0.1, selected so that type I error is bounded within 0.05. All other regression parameters will have a non-informative prior.

3.2 All cause hospitalization or death (within 28 days)

3.2.1 Definition

If a participant is hospitalized for any reason or dies from any cause within 28 days of receipt of study drug, then the participant will have experienced the all cause hospitalization or death endpoint.

3.2.2 Estimand

The difference in hospitalization or death rates will be summarized as an odds ratio estimated from a Bayesian logistic regression model.

3.2.3 Model

The Bayesian logistic regression model will include the following covariates:

1. Randomization assignment
2. Age as a restricted cubic spline with 3 knots
3. Gender
4. Duration of symptoms prior to treatment
5. Calendar time as restricted cubic spline with 4 knots
6. Vaccination status
7. Geographic region (Northeast, Midwest, South, West)
8. Call center indicator
9. Additional appendix-specific covariates relevant to baseline disease severity and patient risk.

3.3 Missing data

Despite efforts to accurately collect all baseline and outcome data, some missingness may occur. For time to recovery, participants who are lost to follow-up will by censored at the time of last follow-up.

Missing data in the covariates will be handled with imputation and posterior stacking. If the percentage of observations with missing covariates exceeds 5%, then multiple imputation (predictive mean matching) will be used. Otherwise, single imputation (conditional mean of the complete cases) will be used.

3.4 Heterogeneity of treatment effect

Heterogeneity of treatment effect will be evaluated for the following variables:

1. Vaccination status
2. Calendar time (pandemic epoch) as a restricted cubic spline with 3 knots
3. Duration of symptoms (time from symptom onset) as a restricted cubic spline with 3 knots
4. Age as a restricted cubic spline with 3 knots
5. BMI as a restricted cubic spline with 3 knots
6. Patient reported symptom severity (none, mild, moderate, severe)

Heterogeneity of treatment effect will be assessed by estimating model posterior probabilities for a full model including interaction term(s) with randomization assignment and a reduced model without the interaction term(s). The reduced model, M_r, is the primary endpoint regression model with the addition

of a variable in question (if not already in the model). For example, for vaccination status, two models are:

$$M_r: \quad \beta_1 \text{trt} + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma_1 \text{vaccinated}$$

$$M_f: \quad \beta_1 \text{trt} + \cdots + \beta_k x_k + \gamma_1 \text{vaccinated} + \gamma_2 \text{trt} \times \text{vaccinated}$$

Model posterior probabilities, $P(M_f|y)$ and $P(M_r|y)$, will be estimated using non-skeptical, non-informative priors. Because the analyses for heterogeneity of treatment effect are exploratory, there are no decision thresholds for reportable findings.

4 Analysis of screening endpoint

4.1 Concordance-, discordance-, equivalence- probabilities

The OR and Wilcoxon test are translations of the treatment effect quantities known as concordance- and discordance- probabilities, which is why the Wilcoxon test can be performed by estimating a common OR fit with an ordinal regression model without covariates. In the setting of a treatment assignment variable, the concordance-, discordance-, and equivalence- probabilities are defined as a comparison of outcomes from a randomly selected subject in the intervention arm (denoted by A) and a similarly selected subject in the non-intervention arm (denoted by B):

$$\begin{aligned} \text{concordance} &= P(\text{intervention outcome better}) = P(Y_A < Y_B) \\ \text{discordance} &= P(\text{non-intervention outcome better}) = P(Y_A > Y_B) \\ \text{equivalence} &= P(\text{outcomes identical}) = P(Y_A = Y_B). \end{aligned}$$

4.2 Days Benefit

Days benefit is a summary measure of the treatment effect built on the concordance and discordance probabilities. As such, it is related to the common OR (and the Wilcoxon test), but it does not require the proportional odds assumption. It is a summary of treatment effect measured in days, which is an absolute and clinically relevant scale. The measure captures any degree of reduction in symptom burden or decreases in clinical event rates.

Days benefit is the cumulative difference of concordance and discordance probabilities over the course of follow-up. It is interpreted as a difference in days. Specifically, consider the outcome trajectory of study subjects assigned intervention to subjects assigned placebo, calculating (a) the expected number of days that a subject on treatment has a better outcome than placebo and (b) the expected number of days that the subject on placebo has a better outcome than intervention. The difference between the two counts is the days benefit.

To illustrate the definition, consider an hypothetical 14-day profile of outcome probabilities displayed in panels A and B in the figure below. For the sake of illustration, the outcome probabilities are different between the intervention and non-intervention arms. The colors in the profile plot denote distinct symptom and outcome scale values. At the bottom of the plot (in blue) is “no symptoms”. At the top (in red) is “death”. The intermediate scale values are ordered in between the extremes. The height of each box represents the probability of the corresponding outcome score. For example, on day 1, there is an approximate 0.18 probability that a patient in the intervention arm will report no symptoms.

The probabilities in panels A and B can be translated into concordance-, discordance-, and equivalence-probabilities, which are shown in panel C. Consider the comparison of a randomly selected patient in the intervention arm and a randomly selected patient in the non-intervention arm. Concordance occurs when

the outcome for the intervention arm patient is better than the outcome of the patient in the non-intervention. Discordance is the opposite event, and equivalence occurs when the outcome scores are equivalent between the two.

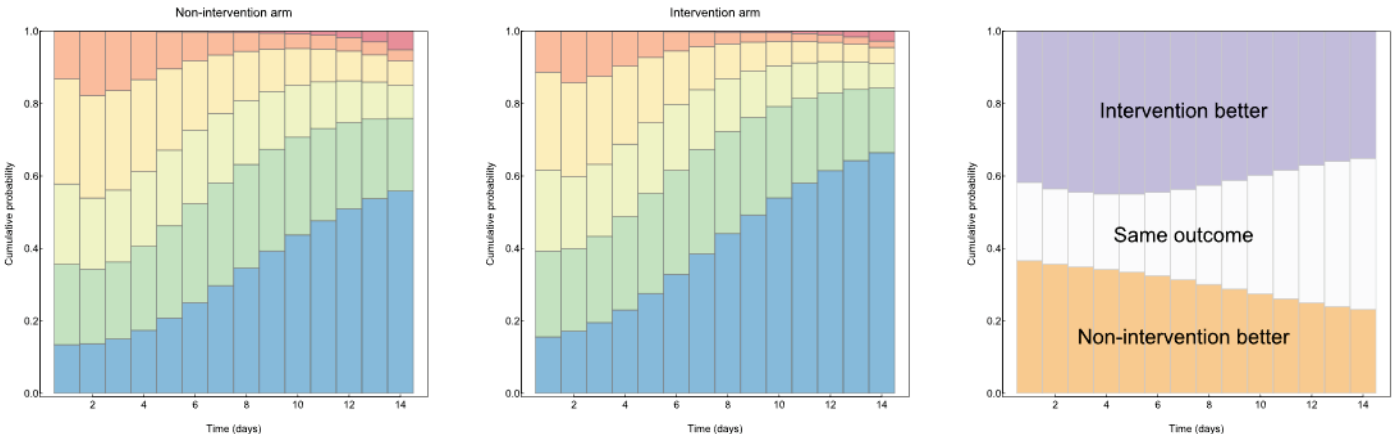


Figure 1. Illustration of Days Benefit

Days benefit, restated in terms of concordance and discordance probabilities, is the difference between (a) the probability that the intervention is better and (b) the probability that the non-intervention is better, summed over all the days of follow-up. In panel C, it is the area of the purple region minus the area of the orange region. Mathematically, it is expressed as

$$\text{days benefit} = \sum_{t=1}^{14} P(\text{intervention better}|t) - P(\text{non-intervention better}|t),$$

or equivalently,

$$\text{days benefit} = \text{follow-up time} \times [P(\text{intervention better}) - P(\text{non-intervention better})].$$

This quantity has the benefit of being measured in days, a scale that is readily understood and directly related to the practical impact of the treatment. An estimate of 3 days of benefit corresponds to a surplus of 3 days with better outcomes than the alternative treatment as captured by the primary endpoint.

The measure can also be modified to estimate how the treatment effect changes over time or over baseline symptom status. Estimates of expected benefit conditional on baseline covariate values can also be calculated, and the change in expected days benefit as a function of covariate values can be displayed in a partial effect plot.

4.3 Model

Consider a K category ordinal outcome assessed at T time points. Let Y denote the vector of random outcomes $Y_1, \ldots, Y_T$. The lower-case notation y and $y_1, \ldots, y_T$ denotes the observed values. Let trt indicate treatment assignment of intervention or placebo and let y_0 be the baseline symptom and outcome scale. The probability model for the outcome can be decomposed into conditional components:

$$\begin{aligned} P(Y = y_1, y_2, \ldots, y_T|y_0, \text{trt}) = {} & P(Y_T = y_T|y_{T-1}, y_{T-2}, \ldots, y_1, y_0, \text{trt}) \\ & \times P(Y_{T-1} = y_{T-1}|y_{T-2}, \ldots, y_1, y_0, \text{trt}) \cdots P(Y_3 = y_3|y_2, y_1, y_0, \text{trt}) \\ & \times P(Y_2 = y_2|y_1, y_0, \text{trt}) P(Y_1 = y_1|y_0, \text{trt}) \end{aligned}$$

The model for analysis builds on three assumptions/modeling choices:

1. **Markov Property**. The Markov Property assumes that the conditional probability of today's outcome only depends on yesterday's outcome and not the entire trajectory of outcomes. Mathematically, this is expressed as $P(Y_t = y_t|y_{t-1}, y_{t-2}, \dots, y_1, y_0, \text{trt}) = P(Y_t = y_t|y_{t-1}, \text{trt})$
2. **Smoothness in time and state**. The effect of treatment over time and the effect of the previous outcome (including the interaction of time and previous outcome) can be described adequately well with continuous functions like restricted cubic splines or polynomials in the cumulative log odds scale.
3. **Non-proportionality**. The model does not assume proportionality; however, it does contain non-proportionality parameters with skeptical priors.

We discuss each of these modeling choices in turn. First, the Markov assumption simplifies the expression of the probability model. The conditional probability components are only a function of the previous outcome and treatment assignment.

$$P(Y = y_1, y_2, \dots, y_T|y_0, \text{trt}) = P(Y_T = y_T|y_{T-1}, \text{trt})P(Y_{T-1} = y_{T-1}|y_{T-2}, \text{trt}) \cdots P(Y_1 = y_1|y_0, \text{trt})$$

This is a common modeling choice when analyzing transition state outcomes, as it reduces the complexity of the model by reducing the number of model parameters.

Like the Markov assumption, the choice to impose smoothness on the effects of time and previous outcome state reduces the number of parameters to estimate. It allows the effect estimates to borrow information from adjacent time points and previous outcomes. This choice also has implications for missing outcomes, which is likely to be important in this trial as a substantial proportion of patients are unlikely to report an outcome for every day of follow-up.

The conditional probability components are modeled on the cumulative log odds scale, specifically

$$\log \frac{P(Y_t \geq k|y_{t-1}, \text{trt})}{1 - P(Y_t \geq k|y_{t-1}, \text{trt})} = \text{regression function}$$

Ignoring the longitudinal nature of the data for a moment, the simplest form of the model is

$$\text{for k} = 2, \dots, \text{K:} \quad \log \frac{P(Y_t \geq k|\text{trt})}{1 - P(Y_t \geq k|\text{trt})} = \begin{cases} \alpha_k \text{ if trt} = 0 \\ \beta_k \text{ if trt} = 1 \end{cases}$$

$$\text{where } \alpha_2 \geq \alpha_3 \geq \cdots \geq \alpha_K$$
$$\text{and } \beta_2 \geq \beta_3 \geq \cdots \geq \beta_K$$

The α parameters relate to the estimated probabilities of the outcome states among subjects in the placebo arm (trt=0), and the β parameters capture the same for subjects in the intervention arm (trt =1). It is common to assume that $\alpha_2 - \beta_2 = \alpha_3 - \beta_3 = \cdots \alpha_K - \beta_K$ as a way to estimate a single common treatment effect. This assumption is known as the proportional odds assumption.

4.3.1 Effect of previous state and time

The above model, as currently written, does not include a term for either the previous state or time on treatment. The most flexible approach would be to specify a set of model parameters for each previous state and time point.

$$\text{for k } = 2, \ldots, \text{ K: } \quad \log\frac{P(Y_t \geq k|\text{trt})}{1 - P(Y_t \geq k|\text{trt})} = \begin{cases} \alpha_{t,y_{t-1},k} \text{ if trt } = 0 \\ \beta_{t,y_{t-1},k} \text{ if trt } = 1 \end{cases}$$

$$\text{where } \alpha_{t,y_{t-1},2} \geq \alpha_{t,y_{t-1},3} \geq \cdots \geq \alpha_{t,y_{t-1},K}$$

$$\beta_{t,y_{t-1},2} \geq \beta_{t,y_{t-1},3} \geq \cdots \geq \beta_{t,y_{t-1},K}$$

If the number of possible outcome states is small, such a model may be appropriate. However, each additional outcome state increases the number of model parameters. When the number of states is moderate to large, modeling the log odds as a flexible function of the previous state, y_{t-1}, will cap the number of parameters and constrain the log odds to vary smoothly over the range of previous state values.

To capture the effect of the previous state in the model, smooth functions like polynomials or restricted cubic splines replace the constant terms. For the sake of notational ease, we used the notation $s(t, y_{t-1}, d)$ to denote a smooth function of d parameters capturing the effect of time and previous state. Note that because outcome state K is death, the transition probabilities are known when $y_{t-1} = K$.

$$\text{for } k = 2, \ldots, K \text{ and } \quad y_{t-1} < K\text{:}$$

$$\log\frac{P(Y_t \geq k|y_{t-1},\text{trt})}{1 - P(Y_t \geq k|y_{t-1},\text{trt})} = \begin{cases} s_{\alpha_k}(t, y_{t-1}, d) \; if \; trt \; = 0 \\ s_{\beta_k}(t, y_{t-1}, d) \; if \; trt \; = 1 \end{cases}$$

$$\text{where } s_{\alpha_2}(t, y_{t-1}, d) \geq s_{\alpha_3}(t, y_{t-1}, d) \geq \cdots \geq s_{\alpha_K}(t, y_{t-1}, d)$$
$$\text{and } \quad s_{\beta_2}(t, y_{t-1}, d) \geq s_{\beta_3}(t, y_{t-1}, d) \geq \cdots \geq s_{\beta_K}(t, y_{t-1}, d)$$

Were it not for the constraints imposed by the ordinal model, regression-type surfaces capturing the effects of time and previous state could be used directly for s_{α_k} or s_{β_k}. In was follows, we describe functions of time and previous state which include regression-like components which also conform to the constraints. Specifically, the regression components are constructed from the cross product of smooth basis functions. The quantity d represents the total number of parameters that define the surface. Note that $d = d_y \times d_t$, where d_y is the number or basis functions along the previous state dimension and d_t is the number along the time dimension. The quantities d_y and d_t need not be the same, as one dimension may warrant additional basis functions to capture more complex, nonlinear effects. If the individual basis functions are of the form

$$s(x) = \theta_1 + \theta_2 s_2(x) + \theta_3 s_3(x) + \cdots + \theta_d s_d(x),$$

then the cross product of the basis functions, say s^t for time and s^p for previous state, is

$$\sum_{i=1}^{d_y} \sum_{j=1}^{d_t} \phi_{i,j} \, s_i^p(y_{t-1}) s_j^t(t).$$

Each surface is defined sequentially, starting with s_{α_2} (or s_{β_2}).

$$s_{\alpha_2}(t, y_{t-1}, d) = \sum\sum \phi_{i,j}^{\alpha_2} s_i^p(y_{t-1}) s_j^t(t)$$

$$s_{\alpha_3}(t, y_{t-1}, d) = s_{\alpha_2}(t, y_{t-1}, d) - exp \sum\sum \phi_{i,j}^{\alpha_3} s_i^p(y_{t-1}) s_j^t(t)$$

$$s_{\alpha_4}(t, y_{t-1}, d) = s_{\alpha_3}(t, y_{t-1}, d) - \exp \sum\sum \phi_{i,j}^{\alpha_4} s_i^p(y_{t-1}) s_j^t(t)$$

and, in general for s_{α_K}, k>2,

$$s_{\alpha_k}(t, y_{t-1}, d) = s_{\alpha_{k-1}}(t, y_{t-1}, d) - \exp \sum\sum \phi_{i,j}^{\alpha_k} s_i^p(y_{t-1}) s_j^t(t).$$

Defining the surfaces in this way ensure that $s_{\alpha_{k-1}}(t, y_{t-1}, d) \geq s_{\alpha_k}(t, y_{t-1}, d)$. The same construction holds for s_{β_k}.

4.3.2 Constraints and number of parameters

Without the smooth functional relationship over time and previous state, there would be $2T(K-1)^2$ parameters.

$$\underbrace{2}_{\text{trt groups}} \times \underbrace{T}_{\text{time points}} \times \underbrace{K-1}_{\text{previous states}} \times \underbrace{K-1}_{\text{intercepts}} = 2T(K-1)^2$$

With the smooth functional relationships described above, there are $2d(K-1)$ parameters.

$$\underbrace{2}_{\text{trt groups}} \times \underbrace{d}_{\substack{\text{time and} \\ \text{previous state surface}}} \times \underbrace{K-1}_{\text{intercepts}} = 2d(K-1)$$

The smallest possible model is with $d = 3$, though to allow for more flexible effect surfaces, d will be larger. Note that the parameters are highly constrained, and the degrees of freedom in the parameter space will be noticeably less.

4.4 Method of inference

Model parameters and subsequent summaries of treatment effect will be estimated with the Bayesian posterior distribution. The posterior distribution is the normalized product of the likelihood model implied by the model described above and a set of prior distributions (which will be specified in the next section). Point estimates will be generated as the mean of the marginal posterior distribution and interval estimates will be calculated from 95% highest posterior density limits.

4.5 Likelihood

The vector of model parameters $\phi = (\phi_\alpha, \phi_\beta)$ is a combination of parameters that define the smooth surfaces of the cumulative log odds intercepts for the non-intervention and intervention arms. Denoting $\gamma_{t,y_{t-1},y_t,\text{trt}} = \alpha_{t,y_{t-1},y_t}$ or β_{t,y_{t-1},y_t} as the generic intercept term, then the likelihood can be expressed as

$$P(Y = y_1, y_2, \ldots, y_T | y_0, \text{trt}) = \prod_{t=1}^{T} P(Y_t = y_t | y_{t-1}, t, \text{trt})$$

$$= \prod_{t=1}^{T} \begin{cases} \dfrac{1}{1+e^{-\gamma_{t,y_{t-1},K,trt}}} & if\ y_t = K \\ 1 - \dfrac{1}{1+e^{-\gamma_{t,y_{t-1},2,trt}}} & if\ y_t = 1 \\ \dfrac{1}{1+e^{-\gamma_{t,y_{t-1},y_t,trt}}} - \dfrac{1}{1+e^{-\gamma_{t+1,y_{t-1},y_t+1,trt}}} & otherwise \end{cases}$$

The overall log likelihood is the log of the product, resulting in a summation of the logged likelihood contribution of all subjects.

4.6 Priors

The quantities of inferential interest are not the raw parameters that characterize the smooth surfaces that capture the effects of time and previous state on the cumulative log odds intercepts. Rather, the quantities of interest are parameters derived from the raw parameters, such as the transition probabilities and the degree of non-proportionality in the treatment effect for any specific combination of time and previous state. As such, informative priors will be primarily placed on derived quantities to reflect prior beliefs about non-proportionality.

Specifically, the informative model priors are:

$$\text{non proportionality} \sim Exp(2) \text{ for each time and previous state}$$

The non-proportionality quantities are defined for each combination of time and previous state.

$$\text{Let } \Delta_{t,y_{t-1},k} = \alpha_{t,y_{t-1},k} - \beta_{t,y_{t-1},k},$$

$$\text{then non proportionality, } np_{t,y_{t-1}} = var\left(\Delta_{t,y_{t-1},2}, \Delta_{t,y_{t-1},3}, \cdots, \Delta_{t,y_{t-1},K}\right).$$

The non-informative priors:

$$\phi_{i,j}^{\gamma_k} \sim N(0,10), \text{ the regression parameters}$$

$$\text{days benefit} \sim N(0,4)$$

4.7 Decision thresholds

Decisions regarding screening are based on the posterior predictive distribution of expected days benefit. There is no formal decision threshold for inefficacy or harm because those findings are superseded by the futility rule.

4.7.1 Guidance to proceed to evaluation of primary endpoint

The $posterior\ predictive\ P(\text{days benefit at N=1200} > 0|\text{interim data})$ will be calculated as a guidance quantity. The trial has been calibrated so that the trial maintains the 0.05 bound on type I error regardless of how decision-makers choose to interpret the screening endpoint analysis.

4.7.2 Futility

A futility probability will be calculated at each interim analysis from the posterior predictive distribution. If the $posterior\ predictive\ \text{P(concluding efficacy at N=1200|interim data)}$ is sufficiently small, then decision-makers may terminate the trial early because it is unlikely that the study will demonstrate a positive benefit of treatment. Even if the futility information is disregarded, the type I error is still bounded above by 0.05.

4.8 Missing outcome data

All available symptom scale values within the 14 day follow-up period will be used in the analysis. Each subject will contribute to the likelihood calculation for each observed outcome. For example, if outcome values are observed for days 1, 3, and 12, then the likelihood contribution will be

$$P(Y_1 = y_1, Y_3 = y_3, Y_{12} = y_{12} | y_0, \text{trt}) = P(Y_{12} = y_{12} | y_3, \text{trt}) P(Y_3 = y_3 | y_1, \text{trt}) P(Y_1 = y_1 | y_0, \text{trt})$$

where the probabilities conditional on outcomes from more than a day prior are calculated using the Chapman-Kolmogorov equation for n-step transition probabilities:

$$P(Y_t = y_t | y_{t-r}) = \sum_{i_1=1}^{K} \sum_{i_2=1}^{K} \cdots \sum_{i_{r-1}=1}^{K} P(Y_t = y_t | y_{t-1} = i_1) P(Y_{t-1} = i_1 | y_{t-2} = i_2) \cdots P\left(Y_{t-(r-1)} = i_{r-1} \middle| y_{t-r} = y_{t-r}\right)$$

4.9 Covariate adjustment

All the decision-making quantities are marginalized estimates from the covariate adjusted model. Specifically, for general covariates x_1 to x_p, the model is

$$\text{for k} = 2, \ldots, \text{K:} \quad \log \frac{P(Y_t \geq k | \text{trt})}{1 - P(Y_t \geq k | \text{trt})} = \theta_1 x_1 + \cdots + \theta_2 x_p + \begin{cases} \alpha_{t, y_{t-1}, k} & \text{if trt} = 0 \\ \beta_{t, y_{t-1}, k} & \text{if trt} = 1 \end{cases}$$

$$\text{where } \alpha_{t, y_{t-1}, 2} \geq \alpha_{t, y_{t-1}, 3} \geq \cdots \geq \alpha_{t, y_{t-1}, K}$$

$$\beta_{t, y_{t-1}, 2} \geq \beta_{t, y_{t-1}, 3} \geq \cdots \geq \beta_{t, y_{t-1}, K}$$

For example, the covariates will include:

1. Age
2. Gender
3. Duration of symptoms prior to treatment
4. Additional appendix specific covariates relevant to baseline disease severity and patient risk.

Note that the covariates are included in the model relying on a a proportional odds assumption.

Summary measures, particularly of the screening endpoint treatment effect, will be calculated by marginalizing over the observed distribution of covariate values. For example, expected days benefit would be calculated by averaging the conditional quantity over observed covariate combinations.

$$\text{Days benefit} = \frac{1}{N} \sum_{i=1}^{N} \text{Days benefit} \mid x_{i,1}, \cdots, x_{i,p}$$

5 Model diagnostics and sensitivity analyses

The standard suite of model diagnostics for Bayesian models will be implemented, including graphical and analytical checks of the adequacy of the posterior samples, the model specification, and its predictions.

Imputation, as an alternative to the complete data likelihood approach to missing data described in an earlier section, will be performed. Further, a tipping point analysis will be provided which will estimate the degree to which differential missingness could change the conclusions of the analysis.

6 Analysis plan for secondary endpoints

The following table provides a summary of the planned secondary endpoints, the associated estimands, and an analysis approach.

Endpoint	Estimand	Analysis method
Mean time unwell	Difference in means	Screening endpoint model
Hospitalization or death (Day 14 and Day 28)	Model based estimates of the treatment effect odds ratio If the number of events is less than 30, a descriptive analysis will be performed.	Logistic regression
Mortality (Day 28)	Model based estimates of the treatment effect odds ratio If the number of events is less than 30, a descriptive analysis will be performed.	Logistic regression
Mortality (time to event)	Model based estimates of the treatment effect hazard ratio If the number of events is less than 30, a descriptive analysis will be performed.	Cox regression
Hospitalization, urgent care, emergency room visit, or death (time to event within 28 days)	Model based estimates of the treatment effect hazard ratio If the number of events is less than 30, a descriptive analysis will be performed.	Cox regression
COVID Clinical Progression Scale (Day 7, Day 14, and Day 28)	Treatment effect odds ratio	Cumulative probability ordinal regression with logit link
Modified Patient-Reported Outcomes Measurement Information System (PROMIS)-29 (Day 7, Day 14, Day 28, and Day 90)	Treatment effect odds ratio	Cumulative probability ordinal regression with logit link

6.1 Schedule of planned analyses

Secondary endpoints will be analyzed as part of the final analysis.

6.2 Definitions and analysis details

6.2.1 Mean time unwell

Mean time unwell is the expected number of days during 14 day follow-up that a patient is unwell, as defined by a dichotomization of the primary endpoint. The summary measure will be reported for all

possible dichotomizations. Mean time unwell for each individual treatment arm is estimated from the primary endpoint model, as is the difference in mean time unwell. The resulting point estimate and 95% credible interval of the difference in means will be generated from the marginalized (over covariates) posterior distribution. The posterior probability that the difference in means exceeds zero will also be reported.

6.2.2 Hospitalization or death (Day 14 and Day 28)

All cause hospitalization or death is hospitalization or death for any reason within the specified time window.

A Bayesian logistic regression model with flat priors will model the time to hospitalization or death. The regression model will include treatment assignment, age as a restricted cubic spline with 3 knots, and other covariates included in the primary endpoint model. Model-based estimates of the odds of the outcome on or before day 14 and day 28 will be generated. The resulting point estimate of the odds ratio comparing non-intervention and intervention arms, 95% credible interval, and the probability that the odds ratio is less than 1 will be generated from the posterior distribution marginalized over observed covariates.

6.2.3 Mortality (Day 28)

Like the hospitalization or death endpoint, time to death or loss-to-follow-up will be recorded for each subject. A Cox proportional hazards regression with flat priors will model the time to death endpoint from which model-based estimates of the odds of death within 28 days will be calculated. The baseline hazard will be degree 5 M-spline function. The resulting point estimate of the odds ratio comparing cases to controls, 95% credible interval, and the posterior probability that the odds ratio is less than 1 will be reported.

6.2.4 Hospitalization, urgent care, emergency room visit, or death (Day 28)

Time to hospitalization, urgent care, or emergency room visit is the date of the first qualifying event. Subjects that do not experience the event are censored on the day of last follow-up. This endpoint is analyzed like the mortality endpoint.

6.2.5 Symptom Count (Day 14)

Day 14 symptom count is the number of symptoms reported on day 14. If symptom resolution occurred prior to day 14 and a symptom count is not available for day 14, then the value of this endpoint is 0. If symptom count on day 14 is not available, then the nearest reported symptom count within 3 days may be used.

6.2.6 COVID Clinical Progression Scale (Day 7, Day 14, and Day 28)

The COVID clinical progression scale is the most severe of the following outcomes during the respective study day.

1. No limitation of activities
2. Limitation of activities
3. Hospitalized, no oxygen therapy
4. Hospitalized, on oxygen by mask or nasal prongs
5. Hospitalized, on non-invasive ventilation or high-flow oxygen
6. Hospitalized, on intubation and mechanical ventilation
7. Hospitalized, on ventilation + additional organ support – pressors, RRT, ECMO
8. Death

This endpoint will be analyzed with a cumulative probability ordinal regression with logit link (also called an ordered logistic model). Flat priors will be used. The regression model will include treatment assignment, age as a restricted cubic spline with 3 knots, and other relevant covariates. The resulting point estimate of the odds ratio comparing cases to controls, 95% credible interval, and the probability that the odds ratio is less than 1 will be generated from the posterior distribution marginalized over observed covariates.

Unless positively identified as dead or hospitalized, subjects are assumed to be 2 or lower on the scale. Subjects that are missing the endpoint are included in the analysis as a partially observed endpoint. (The contribution to the likelihood is the sum of both probabilities.)

6.2.7 Modified Patient-Reported Outcomes Measurement Information System (PROMIS)-29 (Day 7, Day 14, Day 28, and Day 90)

The PROMIS-29 is set of domain-specific subscales. The overall score and individual subscale scores will be analyzed with a cumulative probability ordinal regression with logit link (also called an ordered logistic model). Flat priors will be used. The regression model will include treatment assignment, age as a restricted cubic spline with 3 knots, and other relevant covariates. The resulting point estimate of the odds ratio comparing cases to controls, 95% credible interval, and the probability that the odds ratio is less than 1 will be generated from the posterior distribution marginalized over observed covariates. Subjects with missing responses will be excluded from the analysis.

7 Trial characteristics

7.1 Simulation Study

Rather than mimic the analysis of the screening endpoint followed by the primary endpoint, the following simulation study considers the time to recovery endpoint as if it were evaluated at each possible interim analysis, N=300, 600, 900, and 1200. This scenario represents the analysis sequence with largest type I error. By bounding this scenario to a type I error rate of 0.05, the global type I error rate of the study will also be bounded by 0.05.

In addition to the conservativeness introduced by bounding the worst-case scenario, this approach is also conservative because it does not account for the screening phase or futility rules, both of which put downward pressure on the type I error rate.

7.1.1 Constructing a data generation model for time to event

A model of time to event was constructed using an approach similar to method of moments with the Weibull distribution. Specifically, the parameters of the Weibull distribution were selected so that the resulting mean recovery time was 15 days from symptom onset and the 75th percentile of recovery was 22 days. The resulting distribution matches commonly reported metrics for recovery and the recovery data already collected within ACTIV-6. The Weibull distribution with shape 1.8 and scale 18.4 serves as the data generation model (the baseline model) for the simulations that follow. The recovery outcomes for patients in the intervention arm were generated from the baseline model modified with an additional treatment effect parameter (log relative hazard) which altered the outcome profile relative to the non-intervention arm. By increasing or decreasing the treatment effect parameter (β), the outcome profile of subjects in the intervention arm was improved or degraded. The outcome profiles of the intervention and non-intervention arms were identical when the treatment effect parameter was set to zero. For the sake of

simulation, the treatment effect assumes a proportional hazards treatment effect, and the treatment effect hazard ratio is $HR = e^{\beta}$.

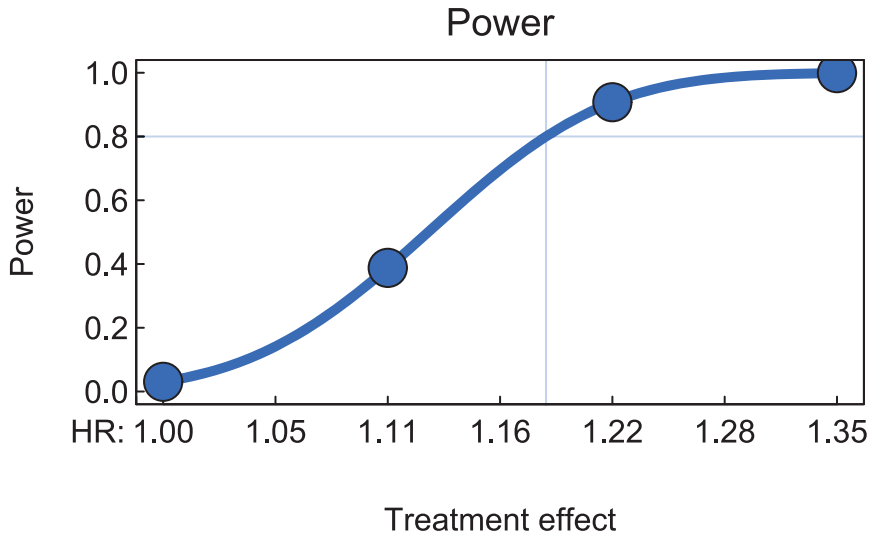
7.1.2 Summary of design choices and decision thresholds

There are a number of design choices and decision thresholds which can be adjusted to reduce the risk of type I error. The following table summarizes the choices made for this trial.

Design Choice	
Schedule of analyses	At N=300, 600, 900, 1200
Treatment effect summary	Hazard ratio for the treatment assignment variable
Efficacy definition	Hazard ratio exceeds zero
Efficacy threshold	Posterior probability of efficacy exceeds 0.95
Futility threshold	(not incorporated into simulation) Posterior predictive probability of efficacy at N=1200 is not more than 0.05.
Treatment effect prior distribution	Normal distribution with mean 0 and standard deviation 0.1.

7.2 Type I error control and power

Suppose the treatment effect HR is 1.5. In order to calculate the power in this setting, 5000 datasets were generated from the data generation model described above. Each was analyzed with the Bayesian model described in section 3.1.3, starting with the first 300 subjects, then 600, and so on until 1200 subjects. The first triggering event (if there was one) plus the corresponding sample size was recorded for each dataset. If no triggering event occurred by the final analysis, the dataset was said to have triggered the max sample size constraint. Thus, each dataset generated one of two possible outcomes: efficacy or max sample size. The value of the standard deviation of the prior for the treatment effect was selected so that type I error was capped at 0.05. Power was calculated as the proportion of datasets that generated an efficacy outcome. The following figure shows the power curve over a range of treatment effect sizes.



7.3 Supplementary details of simulation

7.3.1 The primary estimand

To summarize the differences between groups, we use the proportional hazards regression model. The transformed regression parameter $\exp(\beta_1)$ is the hazard ratio and is the primary quantity for inference.

$$\log \text{relative hazard} = \beta_1 \text{trt}$$

7.3.2 The estimation procedure

As has been described here, here, and here, the Cox model can be estimated from a Poisson generalized linear model. The dataset can be restructured so that the Poisson model generates identical estimates of the treatment effect as those achieved with direct estimation using the Cox model. That is, the restructured data combined with the following regression (where α_t denotes an estimate for each inter-event time interval) will generate identical estimates (and standard errors) of β_1.

$$\log(\lambda) = \alpha_t + \beta_1 x$$

Because β_1 is the quantity of primary interest, the estimation of several intercepts (α_t) is a hassle, especially when our end goal is a Bayesian posterior distribution. However, additional manipulation of the dataset will admit a much simpler Poisson GLM which results in approximate but highly accurate estimates of β_1. The intercept estimates are replaced with a single intercept (β_0) and an offset term, ($\log(z)$).

$$\log(\lambda) = \beta_0 + \beta_1 x + log(z)$$

For details about the connection between the Poisson GLM and the Cox model, see the documents linked above. For demonstration purposes, we show the estimates of β_1 from the Cox, Poisson GLM, and Poisson GLM with offset.

7.3.3 Posterior distribution of the treatment effect

The purpose of this document is to explain the derivation of the posterior distribution for β_1 in a Bayesian analysis. The jumping-off point is the Poisson GLM model with offset.

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i + log(z_i)$$

The prior for the treatment effect is potentially informative, with variance θ. The prior for the intercept is an improper flat prior.

$$\beta_0 \sim \text{flat prior}$$

$$\beta_1 \sim N(0, \theta)$$

The log posterior is of the form

$$\beta_0 \sum_i y_i + \beta_1 \sum_i y_i x_i - \sum_i e^{\beta_0 + \beta_1 x_i + \log(z_i)} - \frac{1}{2\theta}\beta_1^2 + \text{constant.}$$

To integrate out β_0, note that the posterior can be expressed as

$$e^{A\beta_0} e^{-Be^{\beta_0}} C$$

With

$$A = \sum_i y_i$$
$$B = \sum_i e^{\beta_1 x_i + \log(z_i)}$$
$$\log C = \beta_1 \sum_i y_i x_i - \frac{1}{2\theta}\beta_1^2 + \text{constant}$$

and

$$\int_{-\infty}^{\infty} e^{A\beta_0}\, e^{-Be^{\beta_0}} C\, d\beta_0 = B^{-A}\Gamma(A)C$$

The log posterior of β_1 (denoted $\log f\,(\beta_1|x,y)$) is

$$-\left[\sum_i y_i\right]\log\left[\sum_i e^{\beta_1 x_i + \log(z_i)}\right] + \beta_1 \sum_i y_i x_i - \frac{1}{2\theta}\beta_1^2 + \text{new constant}$$

7.3.4 The second approximation

To approximate the posterior with a normal distribution, we use a Taylor series. Specifically,

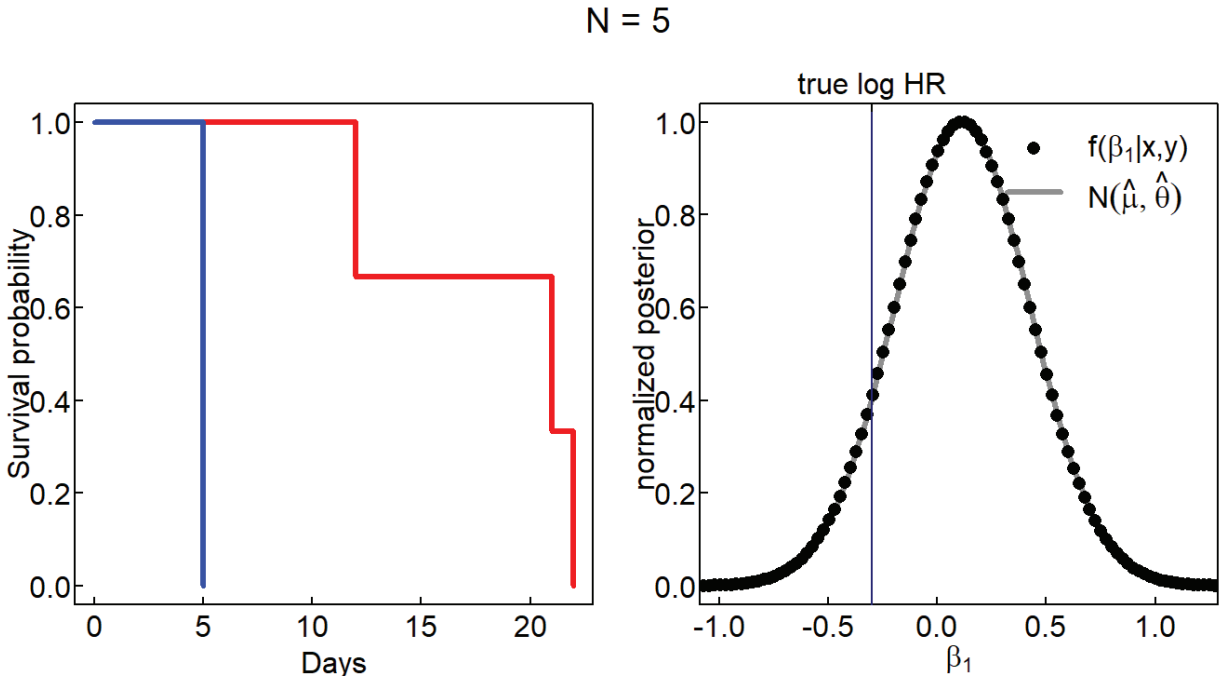
$$\hat{\mu} = \text{root}\ \frac{d}{d\beta_1}\log f\,(\beta_1|x,y)$$

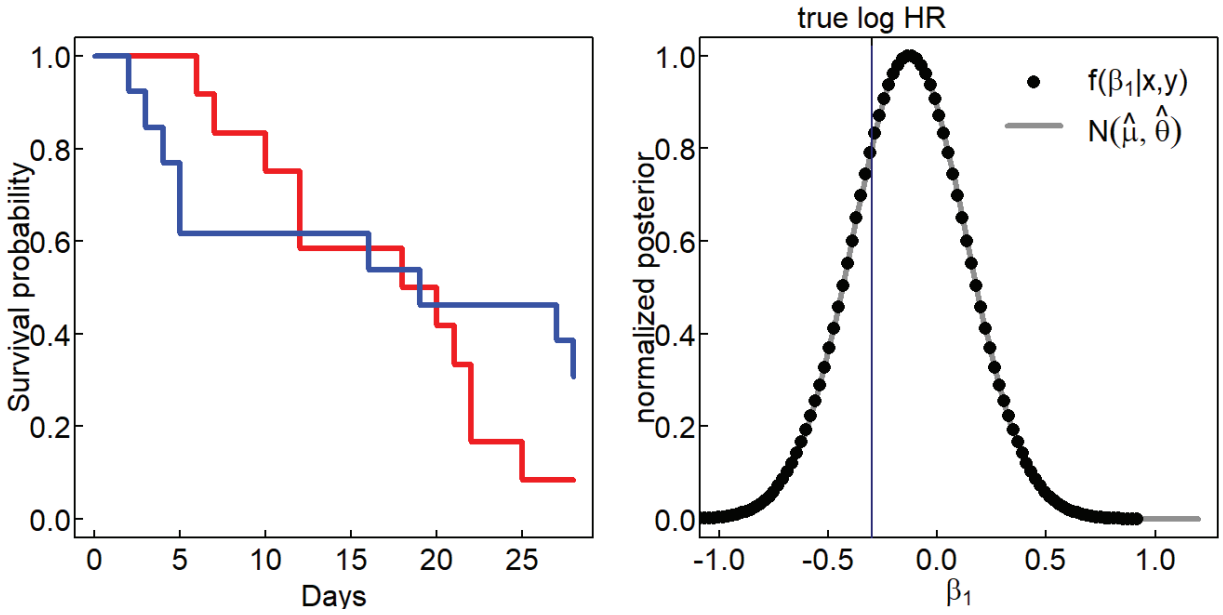
$$\hat{\theta} = -\left[\frac{d^2}{d\beta_1^2}\log f\,(\beta_1|x,y)\bigg|_{\hat{\mu}}\right]^{-1}.$$

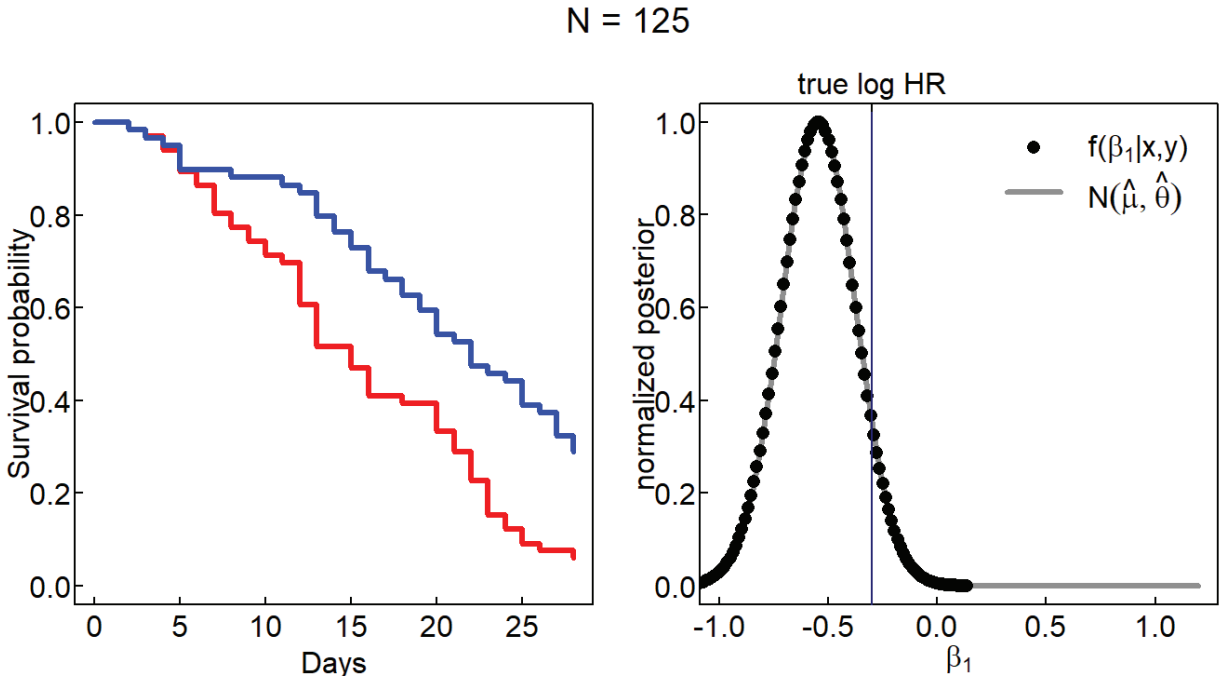
Because the first and second derivatives of the log posterior are readily available, and the root is extremely easy to calculate, the computation time for estimating the posterior distribution is dramatically reduced compared to sampling approaches.

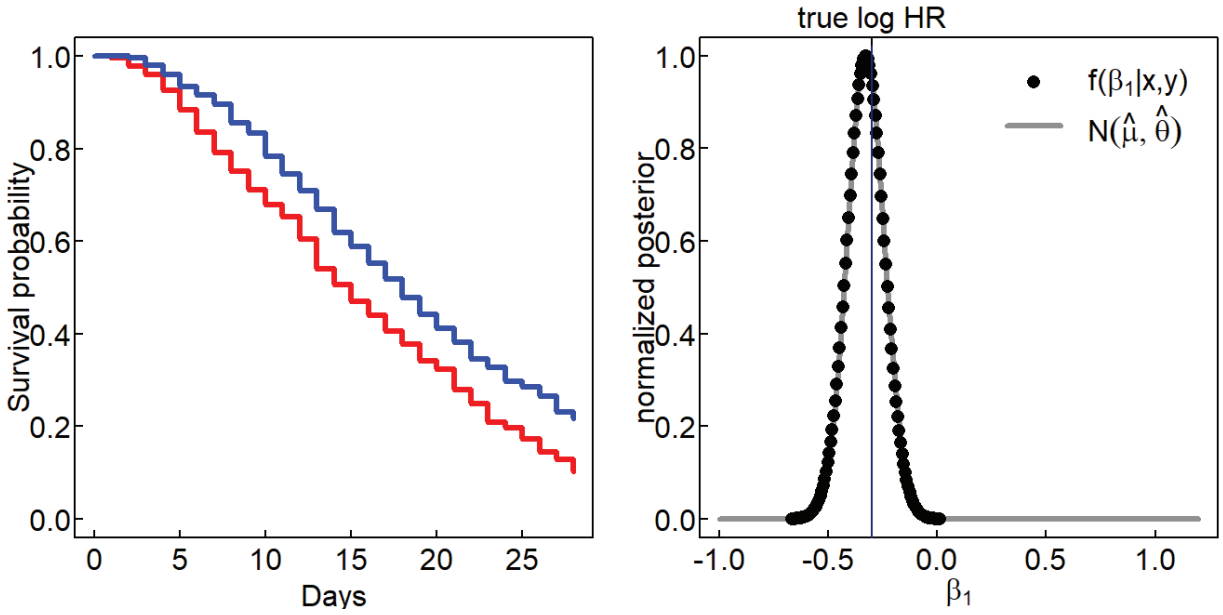
7.3.5 How good is the approximation?

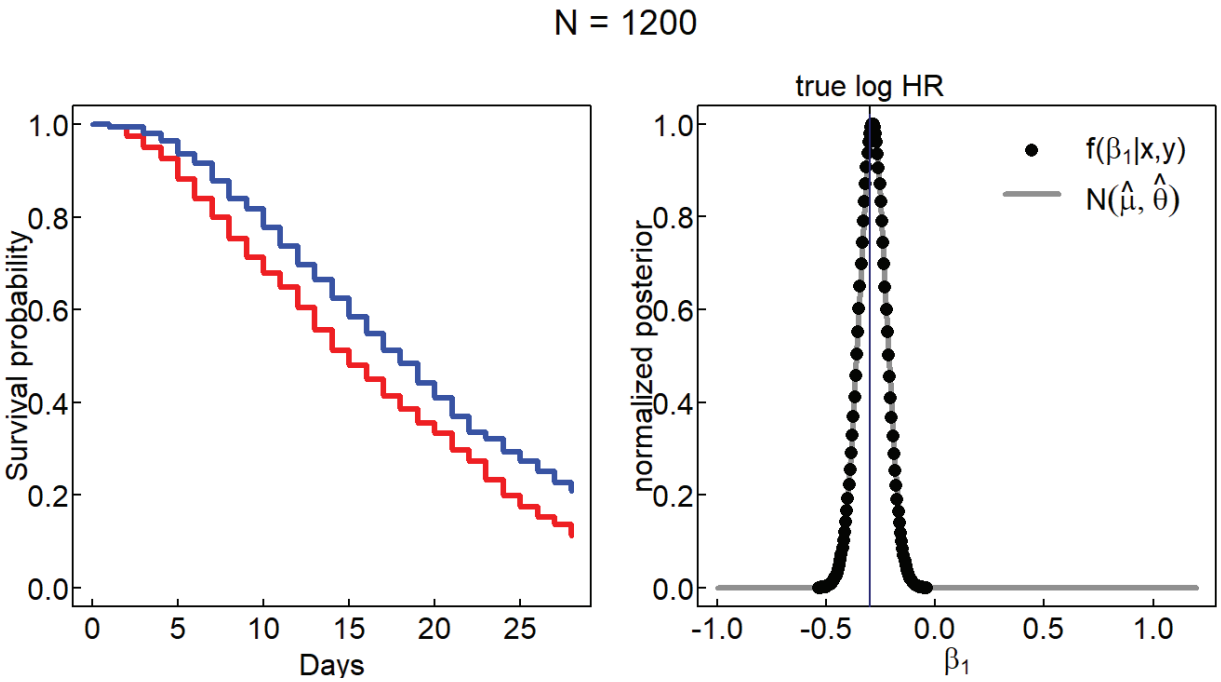
First, we analyze a single dataset of increasing size. On the left of each plot, we show the Kaplan-Meier curves for the survival data; on the right we show the posterior function overlaid by the normal approximation. Because our expression for the posterior is only proportional to a distribution (that is, it isn't normalized to have unit area), we *normalize* the expression to have unit mode. Likewise, the same type of normalization was implemented for the normal approximation. In this example, a skeptical prior with variance 0.1 is used.


N = 5
Survival probability
Days
true log HR
normalized posterior
f(β1|x,y)
N(μ̂, θ̂)
β1

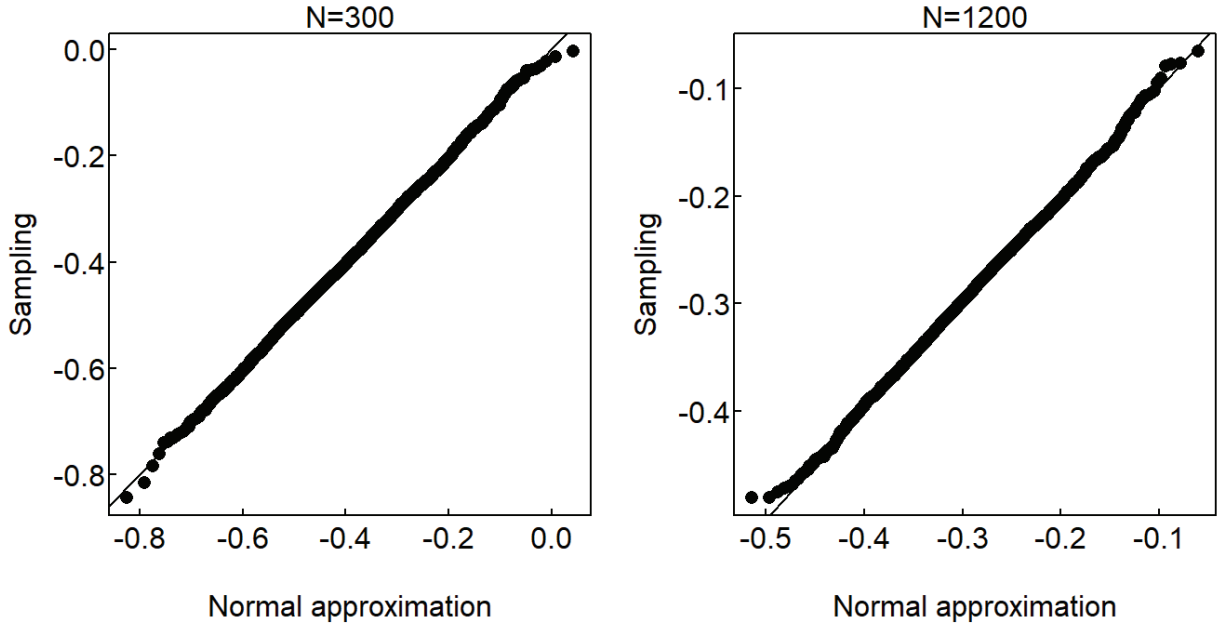

N = 25
Survival probability
Days
true log HR
normalized posterior
f(β1|x,y)
N(μ̂, θ̂)
β1


N = 125
Survival probability
Days
true log HR
normalized posterior
f(β1|x,y)
N(μ̂, θ̂)
β1


N = 625
Survival probability
Days
true log HR
normalized posterior
f(β1|x,y)
N(μ̂, θ̂)
β1

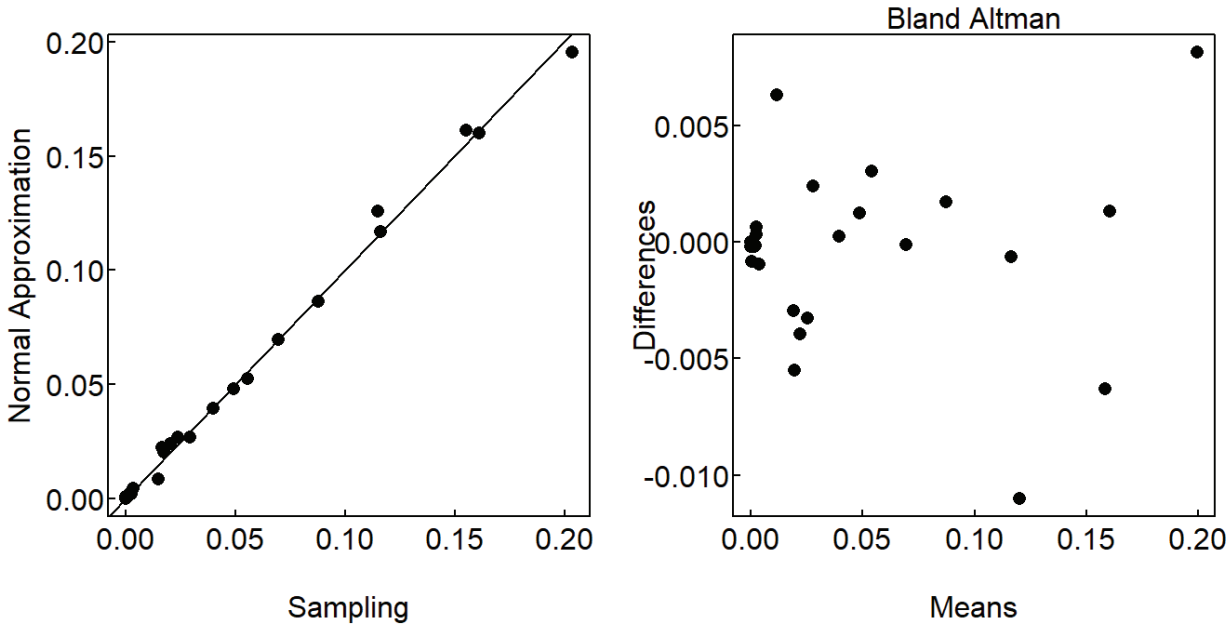


Second, we compare the approximate posterior solution to the `brms (R-package)` sampling solution. In the plot below, we have a QQ-type plot. The quantiles from the normal approximation are paired with the empirical quantiles of the 4000 posterior draws. The plot is generated for both N=300 and N=1200.



The approximate normal posterior distribution matches extremely well with the posterior distribution generated from sampling.

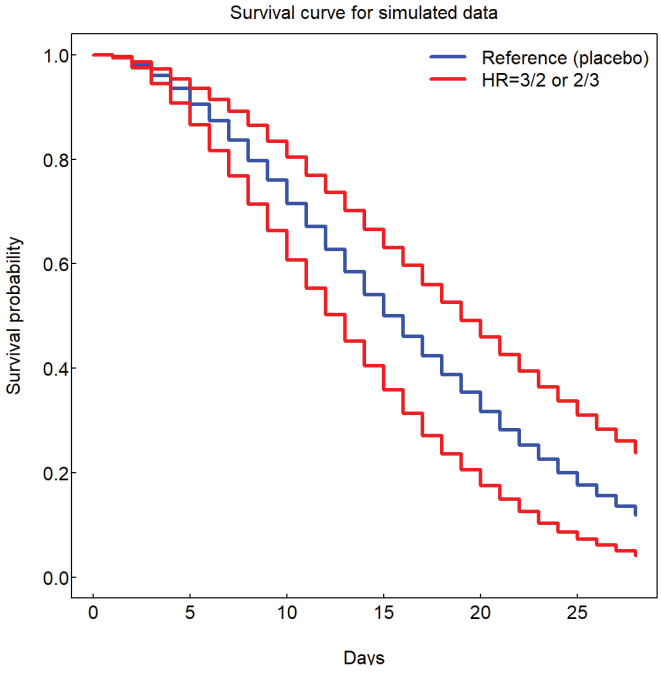
Third, we compare the posterior probabilities from the approximate solution to that of the sampling method in 25 datasets. Specifically, we calculate $P(\beta_1 > 0)$.



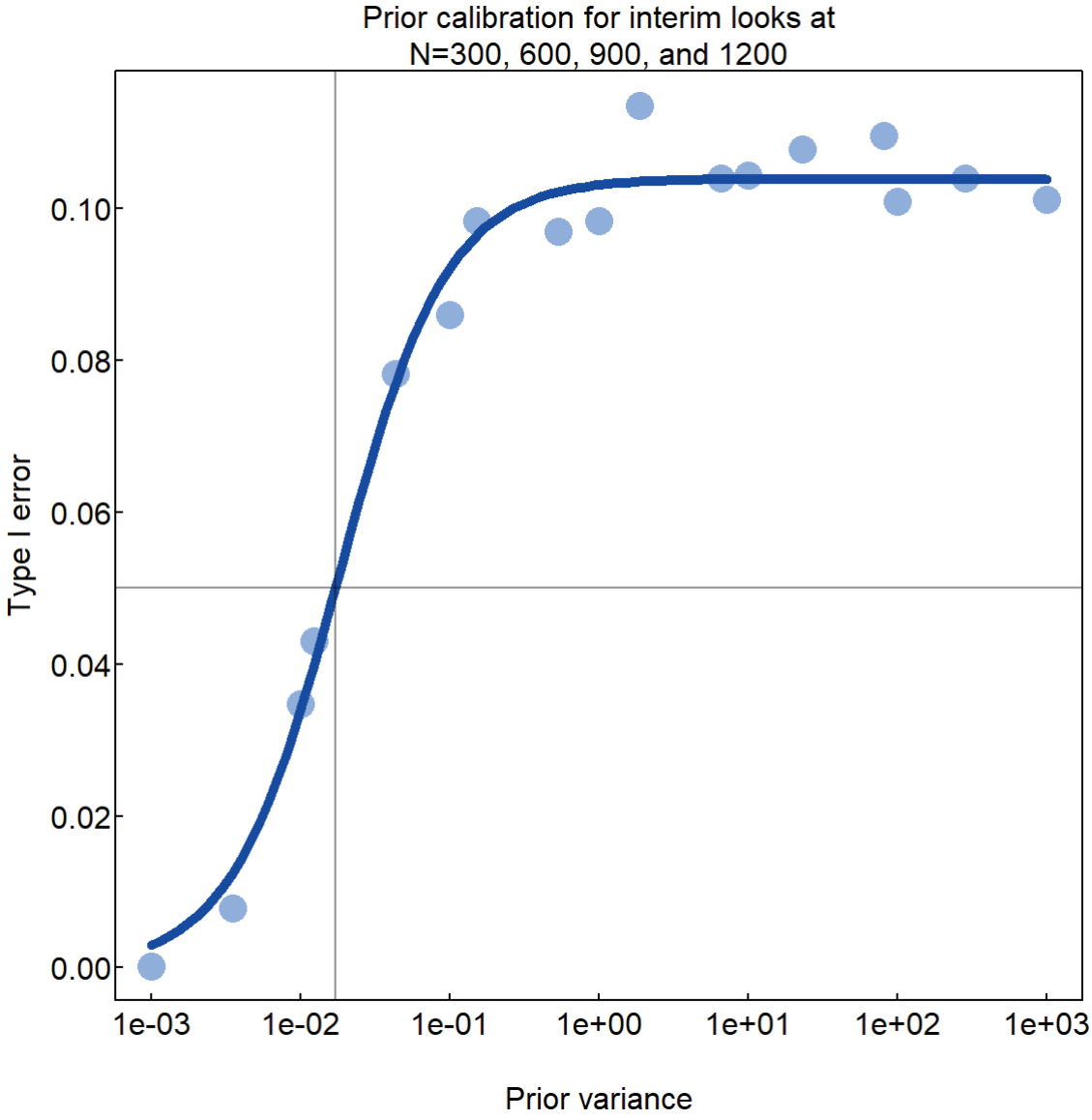
7.3.6 Calibration of type I error

In a trial with multiple looks, the variance of the treatment effect prior can be selected to achieve a desired type I error rate.

Simulated outcome data were generated from a Weibull distribution. Censoring occurred after 28 days. The resulting survival curve for the reference (placebo) arm is depicted in blue in the figure below. In red, we show the survival curves when the hazard ratio is 3/2 or 2/3. The reference survival curve was selected so that approximately 88% of subject experienced an event within 28 days.

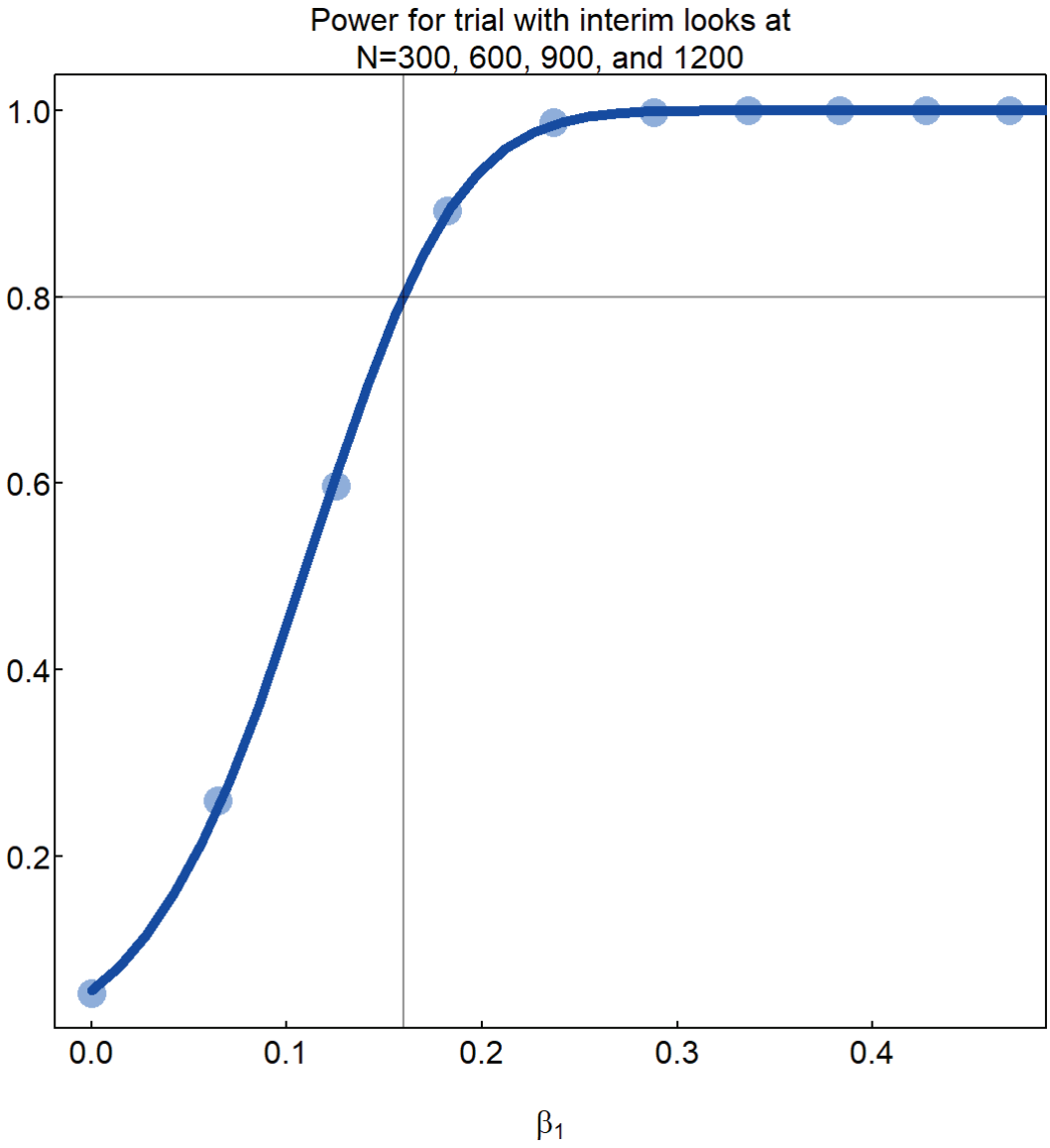


In the figure below, we show the results of a simulation study which indicate that a prior variance of 0.017 will achieve an error rate of 0.05. Note that this simulation study does not reflect a futility rule, which if implemented, may reduce the type I error rate and allow for a less skeptical prior.



7.3.7 Power

Using the prior standard deviation identified in the calibration plot above, we perform a simulation study to get at the power of a trial with looks at 300, 600, 900, and 1200 subjects. The minimal detectable difference at 80% power is HR=1.17 when the data follow the distribution used in this simulation.


Power for trial with interim looks at
N=300, 600, 900, and 1200
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
β1