Department	:	Biostatistics and programming
Information Type	:	Statistical Analysis Plan (SAP)

Title	:	A Randomized, Placebo-Controlled, Double-Blind, Efficacy and Safety Study of Allogeneic HB-adMSCs for the treatment of COVID-19
Product	:	HB-adMSCs Hope Biosciences Adipose Derived Mesenchymal Stem Cells
Effective Date	:	30-Jun-2020

Description:

- The purpose of this SAP is to describe the planned analyses and output to be included in the Clinical Study Report for Protocol HBCOV03.
- This SAP is intended to describe the planned safety, efficacy and tolerability analyses required for the study.
- This SAP is to convey the content of the complete statistical analysis deliverables.

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the analyses to be included in the Clinical Study Report for Protocol HBCOVID03:

2. SUMMARY OF KEY PROTOCOL INFORMATION

2.1. Study Objective(s) and Endpoint(s)

Objectives	Endpoints
Primary Objective	**Primary Endpoint**
• To investigate the safety and efficacy of HB-adMSCs in the treatment of patients with COVID-19 by decreasing inflammation, improving oxygenation, and decreasing time to return to room air (RTRA)	• Change from baseline in inflammatory markers (TNF-alpha, IL-10, IL-6, C-Reactive protein) • Change from baseline oxygenation • Time to return to room air (RTRA) and proportion of patients with room air status
Secondary Objectives	**Secondary Endpoints**

Objectives	Endpoints
• To investigate the safety and efficacy of HB-adMSCs in improving clinical status of patients suspected to have COVID-19).	• Change in clinical laboratories • Change from baseline Chest X-ray • Change from baseline CT scan • Change from baseline in exploratory markers (D-dimer, myoglobin, troponin, creatinine kinase, serum ferritin, CD3CD56, CD4/CD8, VEGF) • Time to negative PCR test results • Change from baseline in the 7-point ordinal scale • Cumulative incidence of serious adverse events (SAEs) or adverse events (AEs)

2.2. Study Design

Overview of Study Design and Key Features	
Randomized, double-blind, placebo-controlled, parallel group study	
Design Features	• Randomized, double-blind, Placebo controlled study. • 100 Subjects • 2 Treatment groups ▪ Group 1 → HB- adMSCs ▪ Group 2 → Placebo • Duration of treatment: up to 7 days, and overall study includes screening (up to 7 weeks prior to randomisation), and 4 treatment visits after randomisation scheduled at Infusion 1 (Day 0), Infusion 2 (Day 3), Infusion 3 (Day 7), and Infusion 4 (Day 10). Follow-up at day 28 (EOS).
Dosing	• HB- adMSCs (Hope Biosciences adipose derived mesenchymal stem cells) is to be administered as an IV infusion with a dose of 100 million cells • Placebo is to be administered as an IV infusion. Contains Saline solution 0.9%
Treatment Assignment	• Participants will be randomised 3:2 to receive HB-adMSCs active treatment or Placebo.

2.3. Synopsis

HBCOVID03 is a phase 2 randomized, placebo-controlled, double-blind clinical trial to assess efficacy of allogeneic adipose-derived mesenchymal stem cells (HB-HB-adMSCs) to treat symptoms of Corona Virus-19 (COVID-19) in patients suspected to have COVID-19. Developing effective interventions for COVID-19 requires incremental improvement of theoretically sound treatments based on systematically accruing data. Often such incremental development is hampered by statistical tools not appropriate to the task. Classical,

Frequentist statistics have advanced the field, but are less informative for the initial evaluation of a new treatment. The reliance of the Frequentist framework on dichotomous, null hypothesis-testing provides some control of the error rate in the context of multiple repeated trials; however, this is not what early-phase treatment testing requires. Developing nascent treatments requires investigators to bet on an alternative hypothesis. Investigators evaluating a theoretically sound intervention want to know the probability that the approach confers some level of benefit given the observed data: that is, they want to know the probability that the alternative hypothesis is true. While Frequentist inference does not directly address this issue, Bayesian statistical inference provides a principled approach to answer this question. Indeed, addressing the so-called "Pipeline Problem" in developing clinical applications, the FDA has indicated that Bayesian statistics offers one avenue for improved methodological efficiency.[1–7]

Decision-making, based on an initial treatment trial, is assisted by estimates of the probability of an effect of some specified magnitude. These statements, not part of the conventional, Frequentist statistical lexicon, are accessible via Bayesian approaches, particularly with small sample sizes.[8,9] Detailed descriptions of Bayesian statistical reasoning exist elsewhere[10,11]. Succinctly, Frequentist models estimate the probability of observing the data (or data more extreme) given that the null hypothesis is true; Bayesian analyses estimate the probability of the alternative hypothesis given the observed data[12]. Bayesian probability estimates incorporate prior information about plausible parameter values (i.e., the prior distribution) and the observed data (i.e., the likelihood). Combining these two distributions forms the posterior distribution which permits evaluation of the probability that the true value of the parameter falls in some range.

2.4. Statistical Hypotheses

The null hypothesis is treatment with HB-adMSCs does not cause changes in primary and secondary endpoints (see outcomes) for suspected Corona Virus Disease. The alternative hypothesis for this study is that HB-adMSC treatment does result in changes from baseline values of primary and secondary endpoints in patients with suspected Corona Virus Disease. Hypotheses will primarily use Bayesian inference. (Any frequentist analyses performed as supplement will consider the hypothesis two-sided and tested at the 5% significance level.)

3. PLANNED ANALYSES

3.1. Final Analyses

The final planned primary analyses will be performed after the completion of the following sequential steps:

1. All participants have completed (or withdrawn from) the study as defined in the protocol.
2. All required database cleaning activities have been completed and final database release and database lock has been declared by Data Management.

4. ANALYSIS POPULATIONS

Population	Definition / Criteria	Analyses Evaluated
Safety analysis set	• All randomised subjects who received at least one dose of HB-adMSCs infusion or placebo. • If participants receive a treatment different to their randomized treatment, they will be analysed according to the treatment actually received.	• Safety • Study Population
Efficacy analysis set	• All randomized participants who received all 4 infusions of HB-adMSCs or placebo. • Participants will be analysed according to their randomized treatment.	• Efficacy

5. CONSIDERATIONS FOR DATA ANALYSES AND DATA HANDLING CONVENTIONS

5.1. Treatment Group Comparability

Initial analyses examining group differences for baseline variables will use cross-tabulation, ANOVAs, and examination of correlations between baseline variables and specified outcomes. For the purposes of evaluating the comparibility of groups, a posterior probability of > 95% will constitute evidence for statistically reliable differences. Baseline or demographic variables on which group differences are detected (excluding stratification variables), and which are correlated with outcomes, meet the definition of confounders[13,14], and will result in two sets of analyses: one in which the relevant variable is included as a covariate, and one in which it is not. This will permit determination of the degree to which any group differences might confound conclusions regarding treatment.

5.2. Study Treatment & Sub-group Display Descriptors

Treatment Group Descriptions		
Protocol treatment arm	**Data Displays for Reporting**	
Treatment Arm	**Treatment Arm**	**Order in TLF**
Arm 1	HB-adMSCs 100MM	1
Arm 2	Placebo	2

Treatment comparisons will be displayed as follows using the descriptors as specified in the statistical analysis:

HB-adMSCs 100MM vs. Placebo

5.3. Baseline Definitions

For all endpoints, the baseline value will be the latest pre-treatment assessment visit with a non-missing value. i.e., If an assessment has been made both at screening visit (Visit 1), and Day 0, infusion 1 visit (Visit 2, Day 0), the value from the Day 0 visit is used as the baseline value. If the value measured at the Day 0 visit is missing and the assessment also has been made at screening, then the screening value is used as the baseline value.

Unless otherwise stated, if baseline data is missing, no derivation will be performed and baseline will be set to "missing".

5.3.1. Change from Baseline

Definition	Reporting Details
Change from Baseline	= Post-Dose Visit Value – Baseline
% Change from Baseline	= 100 x [(Post-Dose Visit Value – Baseline) / Baseline]

NOTES:

- The baseline will be displayed along with Visit name on all summary displays.

5.4. Examination of Covariates, Other Strata and Subgroups

5.4.1. Covariates and Other Strata

Baseline will be included as a covariate in the efficacy analyses, wherever applicable. Stratification variables used at randomization (site, illness severity, pre-existing condition, ethnicity, age group) will be included as covariates.

5.3.2 Examination of subgroups

Secondary analyses will evaluate heterogeneity of treatment effect as a function of gender, age, ethnicity and strata defined by pre-existing condition and illness severity. The approach to subgroup analyses will utilize skeptical, informative priors to increase the likelihood of future replication[15,16].

5.5. Multiple Comparisons and Multiplicity

In keeping with sound Bayesian analytic principles, salient error rates/operating characteristics for confirmatory analyses in each component of the trial (provided in the sample size justification section) result from Monte Carlo simulation. Any secondary analyses for which issues of multiplicity might be a consideration will use weakly, informative priors to regularize all estimates. Means for these regularizing priors will be centered on the null hypothesis, with variances determined by the scale of the data and credible effects previously reported in the literature in the most closely analogous studies. Of note is the principle that the more informative the priors are (based on credible estimates of these effects), (1) the greater the degree of regularization, (2) the more conservative the estimates, and (3) the more likely the results are replicate outside of the current sample. Indeed for any observed results, the Bayesian approach makes it possible to determine the sensitivity of the results to prior assumptions; the degree of prior skepticism an observer would require before dismissing the estimated treatment effect[17]. In the interest of transparency and reproducibility, the resulting reports will provide the specifications of Bayesian model priors used for all secondary analyses.

5.6. Interim analysis

Interim analyses of primary and secondary efficacy data were not performed.

5.7. General Considerations

All continuous measurements will be summarised descriptively at each visit by treatment using observed data. Summary of continuous variables will be presented using N, Range, Mean (minimum, maximum), Standard Error of mean (SE), Standard Deviation (SD) (for normally distributed date), or Median and interquartile range (IQR) for non-normally distributed data. The categorical variables will be presented using number and percentage based on N.

For measurements over time, mean values will be plotted to explore the trajectory over time. Observed data will be used as the basis for plotting data along with bars as standard error (SE) or Standard Deviation (SD), if not otherwise specified.

The primary analytical approach will involve Bayesian statistical methods to assess the probability that an effect of HB-adMSCs 100mm exists (relative to placebo). Our analytical plan is shaped by the limitations of conventional (Frequentist) methods for addressing this question and the advantages of a Bayesian approach for assessing the probability that a given strategy might successfully be expanded into a larger-scale program for the treatment of COVID-19 symptoms. This data, valuable in its own right, can justify the commitment of resources needed for such an expansion. Further, current uncertainty regarding the probability of worse status on the inflammatory markers, change form baseline oxygenation, and time to return to room air (RTRA) as a function of HB-adMSC 100mm is readily incorporated in a Bayesian approach permitting more robust trial planning and design.

The Bayesian approach addresses these questions: (1) “Among patients with suspected COVID-19, what is the probability that allogeneic HB-adMSC 100mm confers benefit relative to placebo on status on inflammatory markers (C-Reactive protein, TNF-alpha, IL-6, IL-10), change from baseline oxygenation, and time to return to room air (RTRA) at treatment Day 10?” (2) “What is the best estimate of these effects?” and (3) “What is their precision?” By estimating the probability that such effects exist, we are assessing the probability that the alternative hypothesis is true; a probability that is, by definition, not accessible to Frequentist methods. The FDA has discussed the use of Bayesian statistical methods to make decisions regarding the efficacy of new treatments as an alternative to Frequentist methods in developing clinical applications[1,2,4,5,11,18]. The current proposal will provide the optimal, unbiased estimates for the benefit conferred by allogeneic HB-adMSC, while also estimating the probability that such effects exist. Posterior distributions can then be used as informative priors for continued monitoring in expansions of treatments and treatment strategies exhibiting initial promise.

Presentation of results from the statistical models will include the estimated conditional/marginal mean treatment effects. For all endpoints analysed statistically, estimated mean treatment differences will be presented together with 95% credible intervals and posterior probabilities,

HB-adMSCs 100MM vs Placebo

All safety evaluations will be presented using Safety Analysis Set. The efficacy analysis will be presented using Efficacy analysis set.

Individual safety and efficacy parameters will be listed.

6. STUDY POPULATION ANALYSES

6.1. Overview of Planned Study Population Analyses

The study population analyses will be based on the Safety analysis set.

Study population analyses including analyses of subject disposition, demographic and baseline characteristics, medical history, prior and concomitant medications.

Coefficients and Bayesian posterior probabilities for treatment differences is displayed for baseline characteristics.

Details of the planned displays are presented in **Error! Reference source not found.**.

Categorical variables will be summarized by the number and percentage of subjects. Continuous parameters will be summarized by N, Range, Mean (minimum, maximum), Standard Error of mean (SE), Standard Deviation (SD) (for normally distributed data), or Median and interquartile range (IQR) for non-normally distributed data unless otherwise specified.

Disposition summary includes: subject screened, randomized and disposition at end of study – Day 28 along with reasons for withdrawals will be presented based on number of subjects and percentage. Subjects in different analysis populations will also be presented.

The screen failure reasons will be presented based on number of subjects and percentage. The percentage will be calculated based on total number of screened subjects.

7. EFFICACY ANALYSES

7.1. Primary Efficacy Analyses

7.1.1. Endpoint / Variables

The primary efficacy endpoints are changes from baseline in the following outcomes at each post-baseline visit:

- Inflammatory markers (TNF-alpha, IL-10, IL-6, C-Reactive protein)
- Oxygenation
- Time to return to room air (RTRA) and proportion of patients with room air status

7.1.2. Summary Measure

Mean values (descriptive, unadjusted for covariates) and conditional means from the Bayesian generalized linear (mixed) models (estimated, adjusted for covariates) will be provided for each endpoint.

7.1.3. Population of Interest

The primary efficacy analyses will be based on the efficacy analysis set population, unless otherwise specified.

7.1.4. Statistical Analyses / Methods

Details of the planned displays are provided in **Error! Reference source not found.**.

7.1.4.1. Statistical Methodology Specification

Initial analyses examining group differences for baseline variables will use cross-tabulation, ANOVAs, and examination of correlations between baseline variables and specified outcomes. For the purposes of evaluating the comparability of groups, a posterior probability of ≥ 75% will constitute evidence for statistically reliable differences. Baseline or demographic variables on which group differences are detected, and which are correlated with outcomes, meet the definition of confounders[13,14] and will result in two sets of analyses: one in which only the stratification variables are included as covariates, and another where the relevant variable is included as a covariate. This will permit determination of the degree to which any group differences might confound conclusions regarding treatment. A third set of analyses will model the effect of treatment group (controlling for stratification variables) on each outcome, focusing on the subgroup of patients who survived through the last infusion visit where statistically feasible.

Broadly, the data analytic strategy will use generalized linear, multilevel, and survival/cox proportional hazard models (R v 4.1, 2021; brms v 2.17; RStan v 2.21)[19–21] for discrete, continuous, and time-to-event outcomes. Multilevel generalized linear modeling to account for clustering of patients within site and repeated observations within patients will evaluate continuous, dichotomous, and count data. All analyses will address potential missingness through joint modeling of observed outcomes and the missing data, an approach robust to ignorable missingness[22] (i.e., MCAR and MAR). Sensitivity analyses will evaluate robustness of analytic conclusions to missing data. Non-ignorable missing data patterns will be addressed through pattern-mixture modeling methods[7].

Evaluation of posterior distributions will permit statements regarding the probability that effects of varying magnitudes exist, given the data. Specification of diffuse, neutral priors will reflect the initial uncertainty regarding effect sizes. For all generalized linear models, priors for regression coefficients will be specified as ~Normal (μ=0, σ^2=1 x 10), and priors for the levels one and two error variances will be specified as ~Half- Normal (μ=0, σ^2=1 x 10^3). The choice of prior distribution for level two variances will follow Gelman's recommendations from the literature[23]. Priors for the comparison of proportions will be specified as ~Beta (α=1.0, β=1.0).

7.1.4.2. Data Analytic Models

All primary endpoint efficacy measurements available at post-baseline at scheduled measurements (Visits 2 through 5 [infusions 1 through 4] for inflammatory markers, oxygenation, and room air status) will be analysed in a generalized linear mixed model. The model will predict each outcome as a function of the interaction between the fixed factors treatment group and time, controlling for lower order effects of time and treatment group. Models will control for stratification variables (see Covariates and Other Strata), and subject will be included as random effects.

7.2. Secondary Efficacy Analyses

7.2.1. Endpoints / Variables

Change from baseline through Visit 5 (Day 10):

- Change in clinical laboratories
- Change from baseline in exploratory markers (D-dimer, myoglobin, troponin, creatinine kinase, serum ferritin, CD3CD56, CD4/CD8, VEGF)

Change from baseline through Visit 6 (Day 28):

- Change from baseline Chest X-ray (available at Visit 6 only)
- Change from baseline CT scan (available at Visit 6 only)
- Time to negative PCR test results
- Change from baseline in the 7-point ordinal scale

Cumulative:

- Cumulative incidence of serious adverse events (SAEs) or adverse events (AEs)

7.2.2. Summary Measures

Mean values (descriptive, unadjusted for covariates) and conditional means from the Bayesian generalized linear (mixed) models (estimated, adjusted for covariates) will be provided for each endpoint.

7.2.3. Population of Interest

The secondary efficacy analyses will be based on the Efficacy analysis set population, unless otherwise specified.

7.2.4. Statistical Analyses / Methods

Details of the planned displays are provided in **Error! Reference source not found.**.

7.2.4.1. Statistical Methodology Specification

Principles will follow the same procedures as the Primary Efficacy analyses (see 7.1.4.1). Three sets of analyses will be performed: the first adjusting for stratification variables, a second set adjusting for variables found to be confounding (see 7.1.4.1), and a third set focusing on the subgroup of surviving patients as of the last infusion where statistically feasible. Bayesian generalized multilevel models will evaluate the differential trajectories of each secondary endpoint as a function of treatment, time, and the interaction of treatment and time.

Additional analyses will evaluate each outcome cross-sectionally at each time point as a function of treatment. Though theses analyses are exploratory in nature, and Type I Error can be less of an issue in the in the Bayesian context, we will evaluate the robustness of any identified effects by utilizing priors specified for the purpose of regularizing estimates. Regularization will render any identified effects more likely to demonstrate replication in future samples. Both double-exponential and the horseshoe distributed priors will provide regularization following hyper-prior distributions: Half-Student's t and Chi-square distributions respectively[24–26]. Specifying the degrees of freedom for each of these hyper-priors constrains the degree of shrinkage the regularization approach can impose on coefficients. A priori we set this value at degrees of freedom = 1 for each hyper-prior as suggested by Bürkner[21]. Priors essentially form a set of assumptions that require evaluation[27]. As such we will evaluate a family of hyper-priors on the shrinkage parameters identified by the specified degrees of freedom for both the double-exponential and horseshoe priors[28]. An initial random split of the data into a training set (70%) and test set (30%) will provide a basis for model building and model evaluation. Within the training set, analyses will examine both double-exponential and horseshoe priors for hyper-parameters ranging from one to at least ten degrees of freedom using k-fold cross-validation to determine the best model based as defined by the ELPD (or associated quantities including the KFOLD-IC, LOOCV-IC and WAIC)[29]v. The R package "brms" contains functions for specifying each type of hyper-prior, and in combination with the R package "loo" provide facilities for k-fold cross-validation[21,29]. Regarding all other model parameters (i.e. those not subject to regularization) intercepts will use ~Normal ($\mu = 0$, $\sigma^2 = 1 \times 10^3$) priors, level-one error variances will use ~Inverse Gamma (shape = 2.001, scale = 0.0001). Level-two variances will follow recommendations by Gelman[23] (i.e. ~Half-T (df = 3, mean = 0, standard deviation = 10).

7.2.4.2. Data Analytic Models

All secondary endpoint efficacy measurements available at post-baseline at scheduled measurements (Visits 2 through 5 [infusions 1 through 4] for exploratory markers and clinical labs; Visits 2 through 6 [infusion 1 through end of treatment] for chest X ray, CT scan, and ordinal scale score) will be analysed in a generalized linear mixed model. The model will predict each outcome as a function of the interaction between the fixed factors treatment group and time, controlling for lower order effects of time and treatment group. Models will control for stratification variables (see Covariates and Other Strata), and subject will be included as random effects. AEs/SAEs and AE severity will be analyzed as dichotomous outcomes (whether patient exhibited AE/SAE at any time during the study). AE/SAE attribution to study drug will be analyzed as count outcomes. All models will control for stratification variables.

8. SAFETY ANALYSES

The safety analyses will be based on the Safety analysis set, unless otherwise specified.

8.1. Adverse Events Analyses

AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA).

A treatment emergent adverse event (TEAE) is defined as an event that has onset date on or after the first day of exposure to infusion treatment and no later than 3 days after the last day of infusion treatment. Here the first day of exposure is defined as the first day of exposure to infusion treatment.

Treatment Adverse events (TAES) are summarised descriptively, whereas non-TEAEs are presented in listings. TAE data will be displayed in terms of the number of subjects with at least one event (N), percentage of subjects with at least one event (%) and the number of events (E).

Summaries of TAEs and of serious AEs will be presented as an overview including all AEs, serious AEs, AEs by severity, AEs by relation to treatment, action to AEs and treatment advised, and outcome of AEs.

Furthermore, summary tables based on system organ class and preferred terms are made for:

- All TAEs
- Serious AEs
- AEs leading to withdrawal of study

Individual adverse events will be listed.

The details of the planned displays are in **Error! Reference source not found.**.

8.2. Clinical Laboratory Analyses

Laboratory evaluations including the analyses of Biochemistry laboratory tests (includes Comprehensive Metabolic Profile), Hematology laboratory tests (Complete Blood Count (CBC) and Coagulation Panel) and Urinalysis. The details of the planned displays are in **Error! Reference source not found.**.

All laboratory parameters, including numerical urine analysis parameters will be summarised descriptively. Categorical urine analysis results will be summarized using count and percentage based on subjects.

Results of urine pregnancy test will be listed in individual subject data listings only.

Individual laboratory evaluations will be listed. In addition, a listing containing individual subject laboratory values outside the normal reference ranges will be provided.

Data recorded at unscheduled assessments will not be included in tables and figures but will be listed.

8.3. Other Safety Analyses

The analyses of non-laboratory safety test results including physical examination and vital signs. The details of the planned displays are presented in **Error! Reference source not found.**.

Physical Examination and Vital signs will be summarized using count and percentage based on subjects. The vital signs based on visit and change from baseline will be summarized using descriptive statistics.

Individual Vital signs, Physical Examination evaluations will be listed.

9. REFERENCES

1. FDA. *Innovation or Stagnation, Challenge and Opportunity on the Critical Path to New Medical Products*.; 2004. http://wayback.archive-it.org/7993/20180125035414/https:/www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ucm076689.htm.

2. FDA. *Innovation or Stagnation, Critical Path Opportunities List*.; 2006. http://wayback.archive-it.org/7993/20180125035449/https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM077258.pdf.

3. Irony T. Bayesian Medical Device Clinical Trials in the Regulatory Setting. In: *Bayesian Biostatistics Conference, University of Texas M.D. Anderson Cancer Center*. Houston; 2008.

4. O'Neill RT. FDA's Critical Path Initiative: A Perspective on Contributions of Biostatistics. *Biometrical J*. 2006;48(4):559-564. doi:10.1002/bimj.200510237

5. Temple R. How FDA currently makes decisions on clinical studies. *Clin Trials*. 2005;2(4):276-281. doi:10.1191/1740774505cn097oa

6. Woodcock J. FDA introductory comments: clinical studies design and evaluation issues. *Clin Trials*. 2005;2(4):273-275. doi:10.1191/1740774505cn096oa

7. Fitzmaurice GM. Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics*. 2000;1(2):141-156. doi:10.1093/biostatistics/1.2.141

8. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *BMJ*. 1995;311(7020):1621-1625. doi:10.1136/bmj.311.7020.1621

9. Spiegelhalter, Myles, Jones, Abrams. Bayesian methods in health technology assessment: a review. *Health Technol Assess (Rockv)*.

2000;4(38). doi:10.3310/hta4380

10. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov*. 2006;5(1):27-36. doi:10.1038/nrd1927

11. Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials*. 2005;2(4):282-290. doi:10.1191/1740774505cn098oa

12. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol*. 2009;62(1):13-21.e5. doi:10.1016/j.jclinepi.2008.07.006

13. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Stat Med*. 2002;21(19):2917-2930. doi:10.1002/sim.1296

14. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355(9209):1064-1069. doi:10.1016/S0140-6736(00)02039-0

15. Dixon DO, Simon R. Bayesian subset analysis. *Biometrics*. 1991;47(3):871. doi:10.2307/2532645

16. Dixon DO, Simon R. Bayesian subset analysis in a colorectal cancer clinical trial. *Stat Med*. 1992;11(1):13-22. doi:10.1002/sim.4780110104

17. Parmar MKB, Ungerleider RS, Simon R. Assessing whether to perform a confirmatory randomized clinical trial. *JNCI J Natl Cancer Inst*. 1996;88(22):1645-1651. doi:10.1093/jnci/88.22.1645

18. Berry DA. Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. *Clin Trials*. 2005;2(4):295-300. doi:10.1191/1740774505cn100oa

19. R Core Team. R: A Language and Environment for Statistical Computing. 2021.

20. Stan Development Team. RStan: the R interface to Stan. 2020.

21. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw*. 2017;80(1):1-28.

22. Best NG, Thomas A. Bayesian Graphical Models and Software for GLMs. In: Marcel Dekker; 2000.

23. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal*. 2006;1(3):515-534. doi:10.1214/06-BA117A

24. Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc*. 2008;103(482):681-686. doi:10.1198/016214508000000337

25. Carvalho CM, Polson NG, Scott JG. The horseshoe estimator for sparse signals. *Biometrika*. 2010;97(2):465-480. doi:10.1093/biomet/asq017

26. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron J Stat*. 2017;11(2):5018-5051. doi:10.1214/17-EJS1337SI

27. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press; 2006. doi:DOI: 10.1017/CBO9780511790942

28. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: John Wiley & Sons, Ltd; 2003.

29. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27(5):1413-1432. doi:10.1007/s11222-016-9696-4