# A Novel Bayesian Approach to Designing Dose Ranging Clinical Trials: A More Efficient Alternative to Traditional Approaches

Jianjun Gan[1], Amy Cutrell[2],

[1]GlaxoSmithKline, Five Moore Drive, Research Triangle Park, NC 27713
[2] GlaxoSmithKline, Five Moore Drive, Research Triangle Park, NC 27713

## Abstract

The goal here is to present a Bayesian study design that provides the information needed for dose ranging decision-making in a highly efficient manner. This will be accomplished through the examination of a simulated case study: A three-arm, dose ranging trial employing a Bayesian design to compare the primary endpoint (virologic response) of two dose arms of new antiviral Compound+ backbone to current standard of care. The weak informative prior is used for the distribution of the response rate for two tested dose arms. For standard of care, a strong informative prior for the response rate allows for a reduction in the sample size for the control arm while maintaining overall power and controlling type I error for the primary comparisons. Also the Bayesian predictive probability will be used to determine whether one test dose arm needs to be dropped after approximately 50% subjects are enrolled. Simulation results demonstrating the properties of the design will be shown as well.

**Key Words:** Bayesian, dose ranging, prediction probability, Clinical Trials

## 1. Introduction

Formulating new pharmaceutical treatments is an expensive endeavor. According to a Forbes article, one clinical trial can cost as much as $100 million. Clinical trial failure is one of the reasons why drug development is so costly, especially the high failure rate in Phase II and III studies. Pharmaceutical R&D is pushing on several fronts, including deliver more with less resource, innovation on top of innovation, streamline development and crisper decision making. The use of Bayesian approaches in clinical research has been increasing over the years because the approach is well suited to adapting to new information collected during a trail. For this reason the Bayesian approach may allow for quick decision and more efficient. Bayesian statistics can be helpful for all three phases of drug development. During phase I, scientist can use Bayesian methods to find the maximum tolerated dose. In Phase II, adaptive trail design can help reduce the number of subjects whom they expose to ineffective treatments, and can also take the right dose into Phase III. Bayesian techniques can also integrate phase II and III confirmatory trails into a more seamless execution.

Phase II studies usually play a pivotal role in drug development [1, 2]. The main purpose for Phase II clinical trials is to determine whether a new treatment demonstrated sufficient efficacy to warrant further investigation. In this paper, we will illustrate how to

use innovative Bayesian design for the three arms Phase IIb dose ranging trials with binary endpoints.

## 2. Bayesian Design for Dose Ranging Study

The example in this paper will be a phase II multicenter, Parallel-Group, Randomized, Dose Ranging Study, the primary endpoint will be antiviral activity response rate (RR, undetected viral RNA at certain weeks) of dose 1 and Dose 2 of investigate drug (Treatment A triple arm: Treatment A+Backbone), and compare with positive control triple group (Treatment C triple arm: Treatment C+Backbone).

### 1.1 **Traditional Three-arm trial setting:**

The endpoint for the primary efficacy comparison is the proportion of subjects with undetected viral load RAN after certain weeks. For example, the sample size of 50 subjects per arm is chosen to ensure a high probability (more like 95%) that a dosage regimen with truly poor response (i.e. at least 10% worse than the other dose) will not be selected for further study on the basis of the analysis of primary endpoint, while allowing for the formal consideration of other measures of antiviral activity in dose selection should response rates on the primary endpoint be similar across dosage regimens. The figure below shows the probability of identifying the better of two doses on the basis of the primary efficacy analysis as a function of a) the difference in response rates between the two doses and b) as a function of the sample size. If the response rate on the least effective dose in any pair is 20% lower than on the better dose, there is a high probability (92%) that the primary endpoint will identify the better dose. If the difference between the two doses is 10% then the primary endpoint will not pick up the difference reliably (62% chance of a correct decision and a 33% chance of no decision).

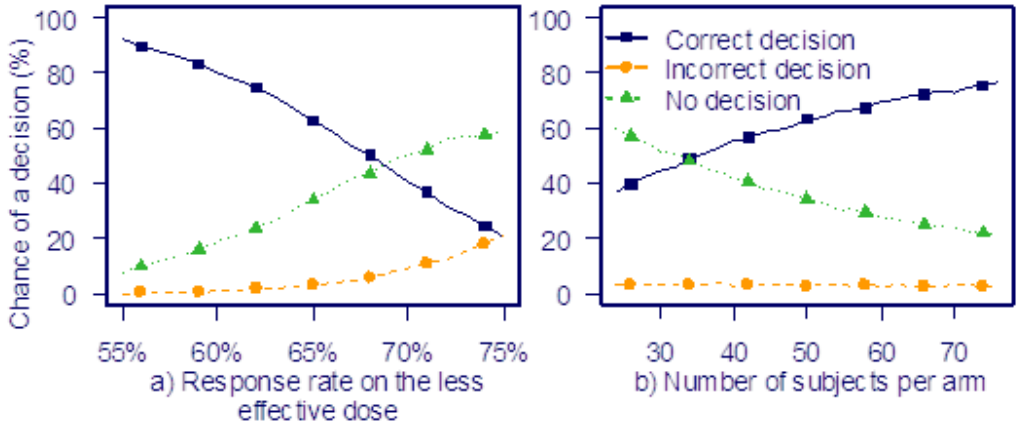


**Figure 1:** The probability of correctly identifying the better of two doses: a) when there are fifty subjects per arm and the response rate on the higher dose is 75% b) when the response rates on the two arms are 65% and 75% respectively.

This traditional three arms design only compares the different doses of Treatment A triple arm, and there is no head to head comparison with the control group, the decision also does not make use of any historical information. The Bayesian analysis expresses uncertainty about a parameter in terms of probability. A prior is formed that characterizes the level of knowledge about a parameter before data is collected. Combining with the prior information, the sample size for each arm can be reduced from 50 to 40, and can

only enroll 20 subjects for the control group. Also the proposed Bayesian design will do direct comparison of each treatment A triple arm dose 1 and 2 to the control arm. The built-in interim and final analysis will use probabilities to guide the directions and conclusion more efficiently.

## 2.1 Bayesian Alternative:

The sample size of 40 subjects per treatment A triple arm (dose 1 and dose 2) and 20 subjects per control triple (treatment C) arm was chosen to ensure a high probability that a treatment arm with truly poorer response will not be selected for further study, while allowing for the formal consideration of other factors in dose selection should efficacy be similar across treatment arms.

The study is designed to test the comparability of Treatment A triple arms to positive control triple Arm C. For this analysis, both treatment A and C subjects are concomitantly administered with some backbone to current standard of care. Thus, a demonstration of the comparability of Treatment A versus Treatment C demonstrates the comparability of Treatment A triple therapy with the SoC (standard of care). The primary comparison of interest will be performed using a Bayesian probability model. If the posterior probability that the difference greater than -8% is large (i.e., ≥95%), then sufficient statistical evidence has been provided for the positive outcome. A population RR of 52% in the Treatment A triple therapy group (compared with a control group RR of 60%) will result in a rejection of the null hypothesis with a probability of approximately 0.04 (type I error). The given sample size is unlikely to select a random sample that would falsely conclude that Treatment A triple therapy is comparable with Control triple therapy if the response rates are truly 52% versus 60%, respectively. If the Treatment A triple therapy yields an RR of greater than 80%, then there is a high probability of rejecting the null hypothesis and correctly concluding that Treatment A triple therapy is at least as good as Control arm .

To claim positive outcome, the following hypotheses will be tested:

$\mathbf{H_0}$: RR for Treatment A triple arm ≤ RR for Control triple arm C -8%

$\mathbf{H_1}$: RR for Treatment A triple arm > RR for Control triple arm C-8%

Historical performance of the backbone and Treatment C triple therapy yields RR of approximately 8% and 60%, respectively. Therefore, the null hypothesis utilizes a margin of 8% for establishing comparability. The observed RR on Treatment A was approximately 75%; we anticipate that the RR in this study will be greater than 70%.

**Table 1:** The probability of a positive outcome assuming the true Response Rate for Treatment A triple therapy

| *True RR for Treatment A* | *50%* | *52%* | *55%* | *60%* | *65%* | *70%* | *75%* | *80%* | *85%* |
|---|---|---|---|---|---|---|---|---|---|
| *Probability of Positive Outcome* | *2.5%* | 4% | *7.5%* | *18%* | *35%* | *56%* | *76%* | *91%* | *98%* |

Here type I error is less than 4% (When actual RR for Treatment A is less than 52%), which is well controlled less than 5%. The calculation above is assuming that the RR for control arm is 60%.

### *2.1.1 Model Assumptions and Sensitivity*

Incorporation of prior beliefs and information about population parameters is a required part of any Bayesian probability model. The assumptions made will help provide a more reliable estimate when the prior beliefs are combined with the observed data than the data alone as long as the beliefs are reasonable. Therefore, the trail will utilize a Beta(20,13) distribution. The median is 60%, with 95% probability that control triple therapy RR is between 44% and 76% based on historical data.

Furthermore, a less informative prior belief is assumed for the RR rate for Treatment A triple therapy but assumes rates roughly similar to what we observed from previous POC (proof of concept study). Therefore, the trial will conservatively utilize a Beta(11,6) distribution. The median is 65%, with 95% probability that Treatment A triple therapy RR is between 41% and 85%. The prior distributions are plotted below as Figure 2.

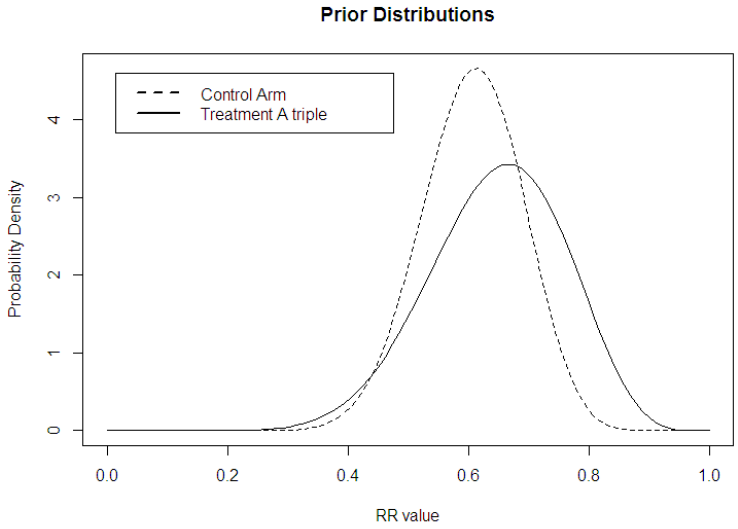


**Figure 2:** Plot of prior distributions for RR for Treatment A triple and control triple arms.

The posterior probability limit is used as a measure of evidence in support of a hypothesis. Selection of the critical value to base a decision of comparability can be made so that the type I error rate is controlled and the desired power profile is attained.
The total number of subjects attaining RR in each treatment arm (experimental and active control) is modelled as independent binomially distributed random variables. Conjugate beta priors are selected to reflect the beliefs about the RR previously described.

The operating characteristic (power) curve is used to evaluate any decision criteria (Bayesian and non-Bayesian alike).

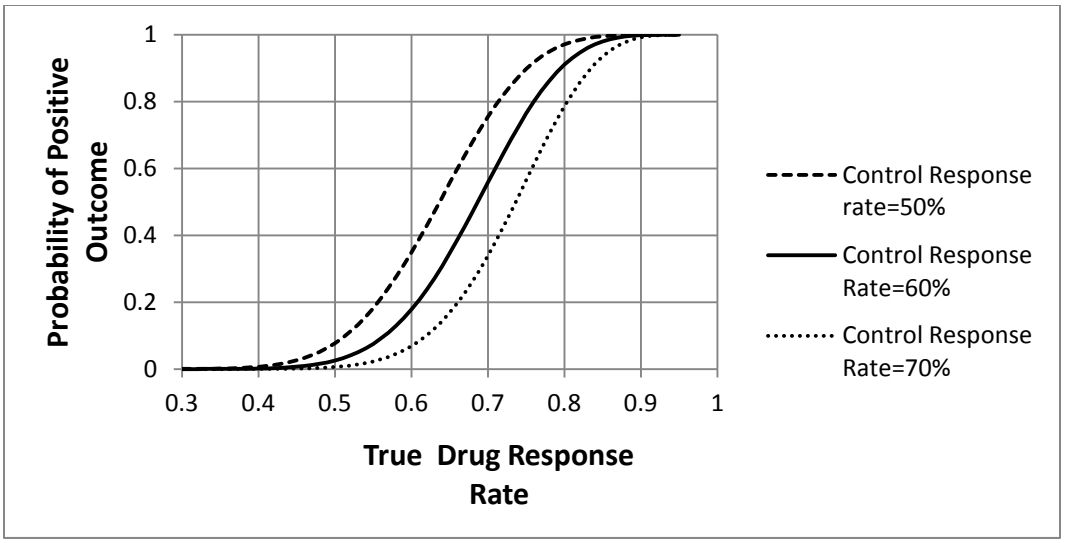


**Figure 3**, the operating characteristic curve for the proposed sample sizes for each treatment arm and prior information.

With a sample size of 20 subjects per Treatment A arm and 20 subjects per Control treatment arm, the likelihood of establishing comparability diminishes substantially. With the smaller sample size, if the Treatment A triple therapy yields an RR rate of 80%, there is only a 75% probability of rejecting the null hypothesis, as compared with a 91% probability with the selected sample size.

### *2.1.2 Bayesian Setting*

For the primary endpoint (RR) analysis, let:
$X_A$ = number of subjects receiving Treatment A triple arm to attain RR, and
$X_{CTR}$ = number of subjects receiving active control to attain RR

The binomial distribution is the assumed likelihood of the RR data, as follows:
$X_A \sim \text{Binomial}(40, p_A)$
$X_{CTR} \sim \text{Binomial}(20, p_{CTR})$

Since the true RR rate is unknown, prior distributions are placed on these parameters of interest to reflect current beliefs and balanced with acceptable decision criteria performance. Conjugate beta densities are assumed. The information pertaining to the control RR is well understood and the prior that was chosen reflects the belief that the RR is between 44% and 76% with 95% certainty. The RR for Treatment A is less well understood, which necessitates a less informed, conservative prior density, where it is assumed that the RR will be between 41% and 85% with 95% certainty.
$P_A \sim \text{Beta}(11, 6)$
$p_{CTR} \sim \text{Beta}(20, 13)$

The posterior probability that the RR for Treatment A triple arm demonstrates the comparability with the Control triple therapy is as follows:
$p_1 = P(p_A > p_{CTR} - 0.08 \mid \text{data})$
A posterior probability of at least 95% (i.e., $p_1 > 0.95$) corresponds to "substantial evidence of positive outcome" and is chosen as the weight of evidence threshold.

## 3. Bayesian Interim analysis

An interim analysis will be included based on the predictive probability of success [3, 4] after half of the subjects have been enrolled (20 each of treatment A triple arm, 10 Control triple arm). During the interim analysis, the probability that the trial will be successful at the completion will be computed. At the same time, determine the probability that the each treatment A dose will be "failure" at the end of the study. If this probability is small (<5%) then stop that dose arm for futility. Otherwise, continue enrolling. For this particular interim analysis, we are not focus on stopping by success (e.g, the predictive probability is high >90%, can stop for success), but the same ideas can be applied to stopping the trial for declaring success. The simulations can help to test the design characteristics [5].

The predictive probability can be calculated by [6]:

$$\sum_{y_1=0}^{N_1-n_1}\sum_{y_2=0}^{N_2-n_2} I(\Pr(p_1 - p_2 > 0) \geq 0.90)\Pr(y_1|x_1)\Pr(y_2|x_2)$$

Here $x_i$= # Complete Responses observed at the interim for arm and $y_i$= # Complete Responses observed between the interim and the final analyses for arm *i*.

The interim analysis can provide a summary table 2 for the predictive probability given different combinations for treatment A and Control arm, the clinicians and study team can very easily to tell the probability of success by looking at this summary table. Due to the double blind randomization setting, this interim look/analysis will be conducted by the Independent Safety Review Committee (ISRC), and all core study team should keep blinded to the randomization code during the interim analysis.

| # Successes in Control Triple Arm (out of 10) \ # Successes in Treatment A triple arm (out of 20) | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.04 | 0.11 | 0.22 | 0.39 | 0.58 | 0.75 | 0.88 | 0.95 | 0.98 |
| 4 | 0.02 | 0.05 | 0.12 | 0.25 | 0.43 | 0.62 | 0.79 | 0.90 | 0.96 |
| 5 | 0.006 | 0.02 | 0.06 | 0.14 | 0.28 | 0.46 | 0.66 | 0.82 | 0.92 |
| 6 | 0.002 | 0.007 | 0.024 | 0.07 | 0.16 | 0.31 | 0.50 | 0.70 | 0.85 |
| 7 | 0.0003 | 0.002 | 0.0082 | 0.03 | 0.08 | 0.19 | 0.35 | 0.55 | 0.74 |
| 8 | 0.0001 | 0.0004 | 0.0023 | 0.01 | 0.03 | 0.09 | 0.21 | 0.38 | 0.59 |

**Table 2**, Predictive probability of success at the interim analysis under various scenarios. Red color represents the predictive probability of success less than 5%, and green color represents the predictive probability of success greater than 90%.

## 4. Conclusion and Discussion

The Bayesian design is a good alternative to traditional designs. It allows for a comparison between the investigate drug with the controls, and also uses information from historical trials to increase the power. Sensitivity analyses may be performed to assess the impact of the informative prior for each treatment arm, though the analysis described above will remain the primary for decision making purposes. An alternative

prior could be to use a non-informative distribution (e.g. uniform (0, 1)) for both treatment arms.

## References

1, Green S, Benedetti J, Crowley J. Clinical trials in oncology, 2nd edn. Boca, Chapman & Hall/CRC, Raton, FL, 2002

2, Palesch YY, Tilley BC, Sackett DL, Johnson KC, Woolson R. Applying a phase Ii futility study design to therapeutic strock trials. Strock 2005; 36:2410-14

3, Spiegelhalter DJ, Freedman LS, Blackburn PR. Monitoring clinical trials: conditional or predictive power? Cont Clin Trials 1986; 7:8-17.

4, Johns D, Andersen JS. Use of predictive probabilities in phase II and Phase III clinical trials. JBiopharm Stat 1999; 9:67-79

5, J Gan, H Zhang, R Best. Mixture of Measurement Errors and their impact on parameter inferences. Journal of Statistical Computation and Simulation, DOI:10.1018/00949655.2011.630001

6, J Lee, D Liu. A predictive probability design for Phase II cancer clinical trials. Clinical Trials 2008;5:93-106