BIAS: Bayesian Interim Analysis Software

Introduction

The BIAS tool was created to allow investigators conducting clinical trials to: (a) make educated decisions about trial design and (b) conduct interim trial analysis.

The intended user of this tool is someone designing one of two types of clinical trials:

1. Independent or cross-trial – the trial is informed by a preceding, independent trial, but the eventual analysis will only be based on the most recent data collected
2. Seamless or within-trial – we are conducting an interim analysis as part of a larger trial, and all previous data will go towards the final analysis.

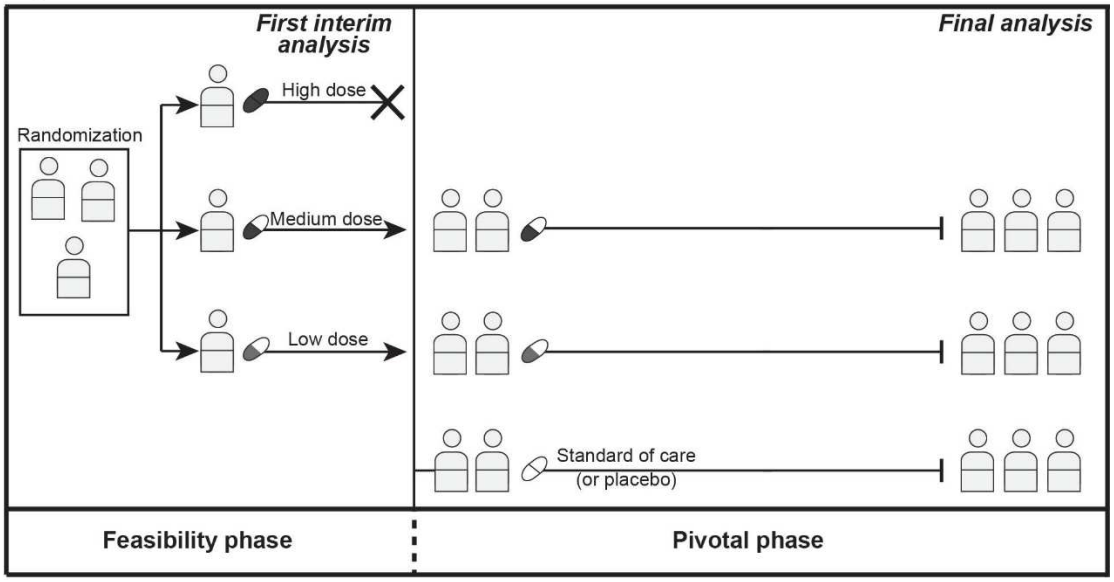
An important question to answer at the outset: why a Bayesian approach? The Bayesian paradigm is particularly appealing and convenient when there is a continual accrual of data and a subsequent adjustment of expectation.

Cautionary note: For proper use of this tool properly, it is recommended that the user be familiar with *both* Bayesian concepts and phase II/III trial design & analysis.

Seamless Trials

Seamless designs are less common than independent trials. Here is an illustration that should help clarify how they work:

Figure 3. Seamless trials



Legend: After first interim analysis, the high dose arm showing serious toxicity could be discontinued from the trial. Thereafter, the trial transitions seamlessly from the feasibility into the pivotal phase with of standard therapy arm being introduced into the trial.

Note that the total number of participants at the end of the seamless trial (phase II and phase II together) is the sum of those that started at the beginning and those that were after the interim analysis.

Bayesian Predictive Power

Before we dive into the functionality of BIAS, we must first understand *Bayesian predictive power (BPP)*. BPP is the probability of rejecting the null hypothesis (H0) when the alternative hypothesis (H1) is true, under a Bayesian framework. An interesting aspect of BPP is that it's the probability of rejecting the null hypothesis with *data that has not been collected yet*, given data that we have already collected. We don't know what future data will be like, but we make educated estimates based on the data that has been collected so far and the mathematical models we chose for it.

BPP is like the idea of power from frequentist statistics and is used in similar ways, the most well-known of which is to choose sample size that ensures studies are adequately powered[1]. In the field of clinical trials, BPP has two main uses:

- Determining sample size per arm required to meet the target 80% BPP in independent trials *and* seamless trials.
- Screening out proposed treatments in multi-arm phase II clinical trials.

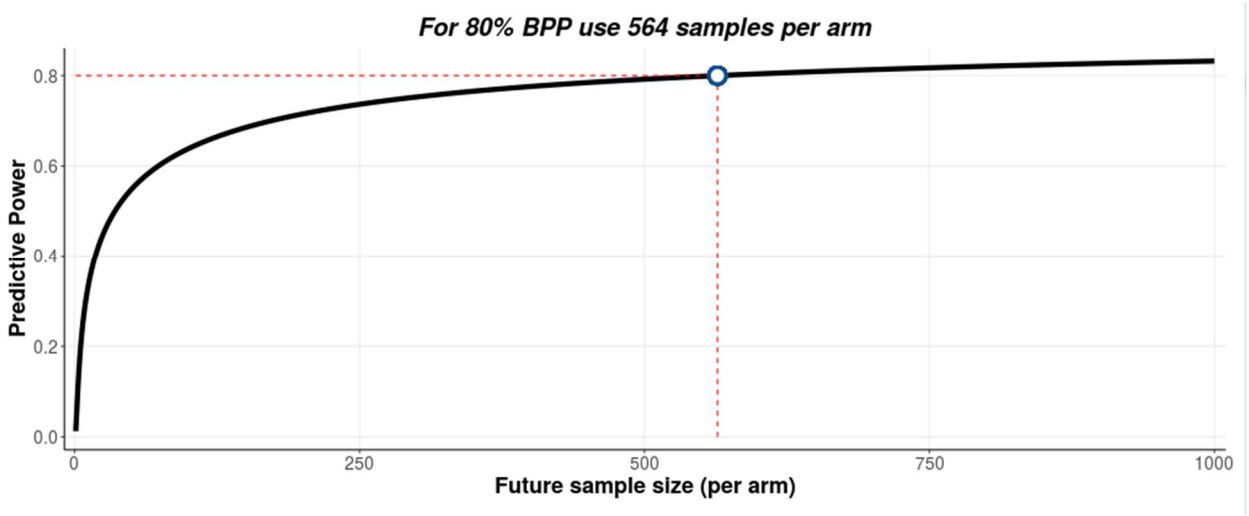
[1] This is conventionally determined by requiring power or BPP to be at least 0.8.

BPP Calculations Tab

The first tab we will cover is the BPP Calculations tab. Here the user can insert data summaries for both a control and treatment arm, and then assess the BPP (probability of rejecting the null hypothesis) in favor of the treatment arm vs the control some number of patients into the future. Ultimately, the user can use this tab to understand how many samples will be needed per arm to meet a specified BPP requirement. Here we explain some of the fields in the BPP Calculations tab:

- Outcome type – Numeric refers to a continuous value like temperature, weight, severity. Dichotomous refers to binary outcomes.
- Predictive power calculation – Within-trial means that the BPP will be calculated by combining "current" samples with the "future" samples, requiring fewer samples to reach a BPP of 80%. Cross-trial means that the BPP will be calculated by only using "future" samples.
- Control/Treatment current sample size – Number of samples that have already been collected in this seamless trial (within-trial) or in a preceding trial (cross-trial).
- Control/Treatment future sample size – Number of samples to be collected as the trial continues (within-trial) or for the newly beginning trial (cross-trial).
- Superiority vs Non-inferiority – Non-inferiority testing is done to show that a treatment is no worse[2] than the control by more than a set clinically important difference.
- Control/Treatment arm average response – Estimated from the data collected thus far (within-trial) or from the preceding phase (cross-trial).
- Control/Treatment arm standard deviation – Also estimated from previous data.
- Type I error rate – The significance level at which the one-sided test will be conducted at the end of the follow-up trial. Usually 0.05.

Once all these values are set, you'll see the BPP displayed in a green box. Clicking on "Calculate Sample Size" will produce a plot like this:



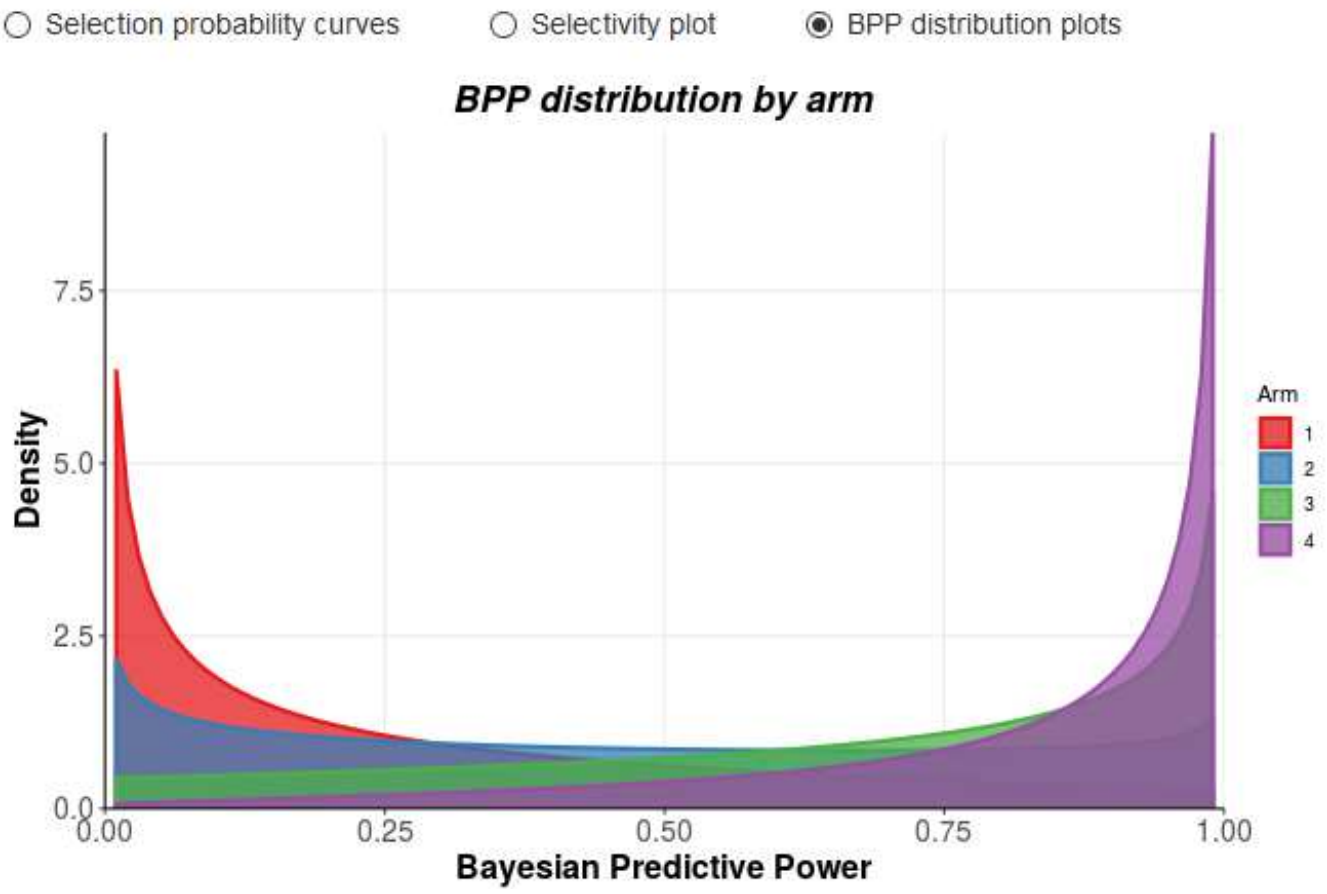
The BPP is on the y-axis and the number of future samples is on the x-axis. The number of future samples required to obtain the target BPP is marked with a red vertical dashed line.

[2] I.e. control and treatment can be equivalent.

BPP Trial Design Tab

Phase III trials are often limited to a fixed number of arms, and the arms themselves are chosen from a superset of arms tested in a phase II trial. In this tab, the user can explore possibilities of arm selection between phase II and phase III, before the phase II trial actually occurs.

The approach taken is to find the subset of treatments using the *selectivity* criterion, which is the cutoff that maximizes the probability of selecting the best treatments[3]. Here "best" means the treatments with the largest BPPs after conducting a phase II trial. We have not run the trial yet, so there is uncertainty in what the BPP will eventually be after the phase II trial, and we thus have a *probability distribution (density)* over future BPP (one per arm), which are plotted when selecting the "BPP distribution plots" option:



We see that in this example arms 3 and 4 have the largest probability of high BPP, while arms 1 and 2 are likely to have low BPP. The expected BPP (reported in the table to the right of the above plots) is the expected value (mean) of these distributions. *Single-value summaries can be misleading, so it's important to look at entire probability distributions.*

To reiterate, the overall goal of this tab is to help a user choose the BPP cutoff above which the arms will advance to the phase II trial. To do so, we start with a specified number of currently active arms (this can be

[3] This is the same fixed number of treatments mentioned at the beginning of the section.

adjusted with the slider near the top of the page). For each arm, the user must enter the initial sample sizes, future sample sizes, (future) mean responses, and (future) standard deviations.

We then move on to the bottom part of the screen, where we are asked to enter “Number of arms to advance” and “Predictive power cutoff”. The selectivity criterion is not dependent on “Predictive power cutoff” – changes in that field only modify the “Selection probability” curves. Each time the “Number of arms to advance” is modified, the user will see a new recommended BPP cutoff in the green box below.

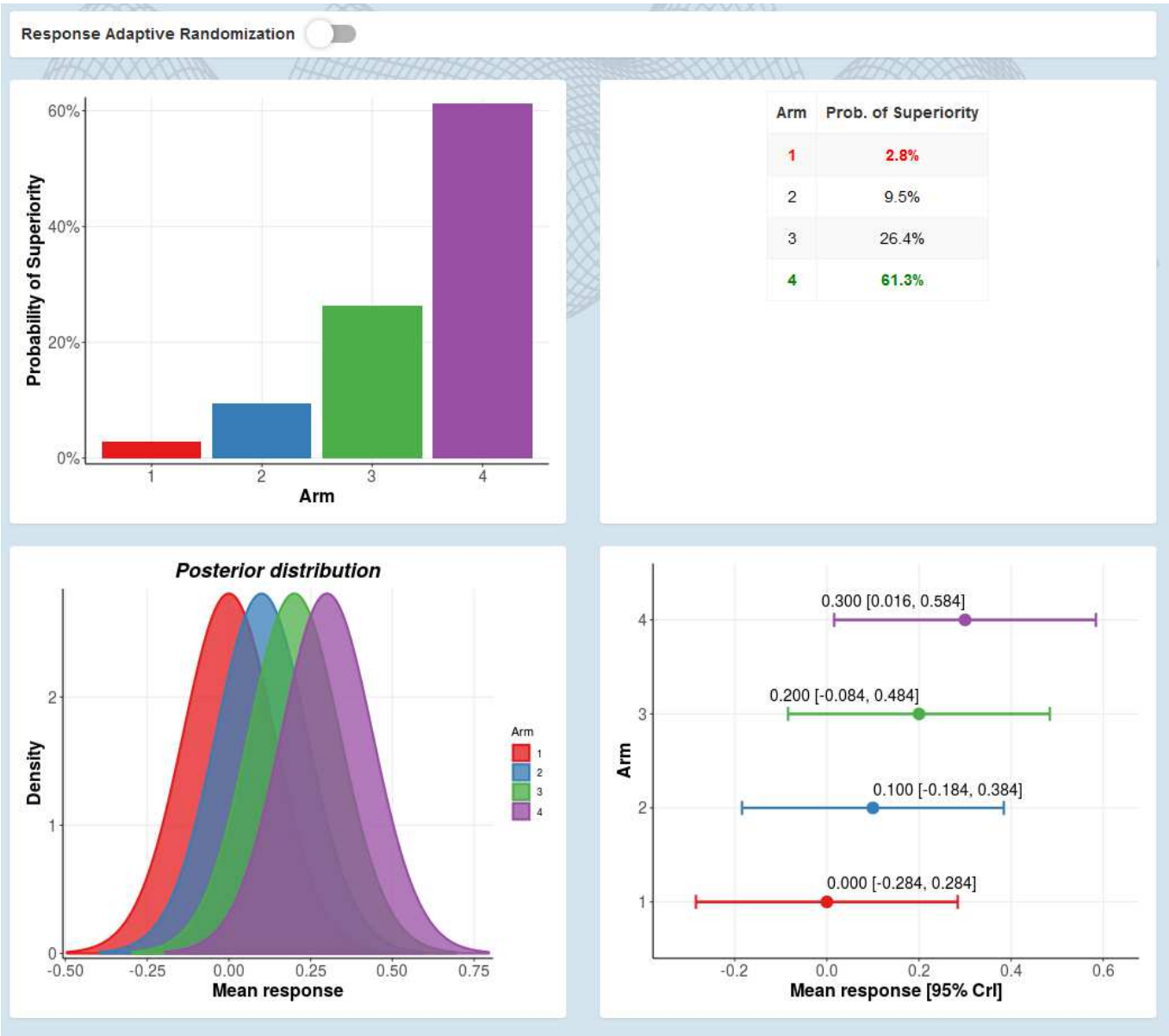
The recommended cut-off is the cut-off that optimizes the likelihood of advancing the both the correct treatments and the desired number of treatments to advance. In practice, clinical trial data are subject to play of chance, and so are the probability of selection and expected BPP. It should therefore be noted that while the recommended cut-off optimized the likelihood of making the correct decision to advance treatments, it inevitably comes with some risk of sub-optimal treatment advancement, as do all cut-offs. The larger the initial sample size of the smaller the likelihood of a sub-optimal recommendation.

In the case of the above figure, we see that arm #1 has a 45.1% chance of being advanced to phase III even though it’s the least effective arm (the control in this case). If one increases the number of future samples, this probability will go down.

Posterior Superiority Tab

This part of the BIAS tool allows the user to apply Bayesian decision rules and response adaptive randomization (RRR) that are based on an interim analysis. The trial's goal is to find the best arm among multiple, and adaptations are made according to the probability of superiority of each arm. That is, the posterior probability that each arm is superior to all other study arms.

The top parts of the tool have the familiar outcome types, number of arms, sample sizes, mean responses, etc. Once these are set, and the “Calculate Probabilities” button is clicked, the user is presented with the following:



The upper-left figure shows the probability of superiority of each arm. The upper-right table shows the same information, but in tabular form. The bottom-left figure shows the full posterior probability distribution of each arm's response (outcome). The bottom-right figure shows the same information, but in 95% confidence interval form.

Once the "Response Adaptive Randomization" switch is triggered above, the top two figures change:

Now in addition to plotting/showing the probability of superiority, we now also see the probability of *allocation* of each arm as the trial continues (within-trial) or in the next trial (cross-trial). The allocation probability to an arm (i.e. the proportion of future patients to be randomized to it) can be calculated by raising its probability of superiority to the "Response adaptive randomization power." For example, if the probability of superiority is 0.613 and the response adaptive randomization power is ½, then the allocation probability is $0.613^{½} = 0.442$ or 44.2%. This randomization power can be adjusted from 0 (no response adaptation) to 1 (strong response adaptation). The investigator may set a threshold on the probability of allocation to decide whether the trial can be stopped early, as well as drop study arms that show little promise along the way.