POLITECNICO DI TORINO
Repository ISTITUZIONALE

Investigating the impact of data monitoring committee recommendations on the probability of trial success

Original
Investigating the impact of data monitoring committee recommendations on the probability of trial success / Rondano, L., Saint-Hilary, G., Gasparini, M., Vezzoli, S.. - In: JOURNAL OF BIOPHARMACEUTICAL STATISTICS. - ISSN 1054-3406. - 36:3(2026), pp. 487-503. [10.1080/10543406.2024.2430308]

Availability:
This version is available at: 11583/2995652 since: 2024-12-19T16:47:39Z

Publisher:
Taylor and Francis

Published
DOI:10.1080/10543406.2024.2430308

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Taylor and Francis postprint/Author's Accepted Manuscript con licenza CC by-nc-nd

This is an Accepted Manuscript version of the following article: Investigating the impact of data monitoring committee recommendations on the probability of trial success / Rondano, L., Saint-Hilary, G., Gasparini, M., Vezzoli, S.. - In: JOURNAL OF BIOPHARMACEUTICAL STATISTICS. - ISSN 1054-3406. - 36:3(2026), pp. 487-503. [10.1080/10543406.2024.2430308]. It is deposited under the terms of the CC BY-

(Article begins on next page)

14 September 2026

Investigating the impact of Data Monitoring Committee recommendations on the probability of trial success

Luca Rondano[1], Gaëlle Saint-Hilary[2,1], Mauro Gasparini[1], Stefano Vezzoli[3]

Abstract

Determining the probability of success of a clinical trial using a prior distribution on the treatment effect can significantly enhance decision-making by the sponsor. In a group sequential design, the probability of success calculated at the design stage can be updated to incorporate the information disclosed by the Data Monitoring Committee (DMC), usually consisting in a simple statement that advises to continue or to stop the trial, either for efficacy or futility, following pre-specified rules defined in the protocol. We define the "probability of success post interim" as the probability of success conditioned on the assumption that the DMC recommends continuing the trial after an interim analysis. A good assessment of this probability helps mitigate the tendency of the study team to express excessive optimism or unwarranted pessimism regarding the trial's ultimate outcome after the DMC recommendation. We explore the relationship between this "probability of success post interim" and the initial probability of success, and we provide an in-depth investigation of how interim boundaries impact these probabilities. This analysis offers valuable insights that can guide the selection of boundaries for both efficacy and futility interim analyses, leading to more informed clinical trial designs.

[1]Politecnico di Torino, Italy
[2]Saryga, France
[3]Chiesi Farmaceutici, Italy

Keywords: assurance; efficacy; futility; interim analysis; group sequential design; hybrid Bayesian/frequentist design; probability of success.

1 Introduction

In a clinical study analysed according to standard frequentist principles, the sample size and the threshold for declaring statistical significance can be determined as a function of the type I and type II error, the treatment effect size of interest (e.g., expressed as mean difference, hazard ratio, etc.) and other relevant parameters (e.g., standard deviation of the response variable, hazard in the control group, etc.). A prior distribution of the treatment effect is not strictly necessary to design a clinical trial, but when such prior information is available, the sponsor may want to use it to make better informed decisions on how to conduct the study. In a frequentsist trial, this prior distribution may play a role in the design, but not in the analysis of the trial. In particular, this prior distribution can be used to determine the *probability of success* (*PoS*) of a study, also known as *assurance*. The concept of *PoS*/*assurance* traces back to O'Hagan, Stevens, and Campbell (2005), and has been studied by many other authors in the following years. Chuang-Stein (2006) introduces the very similar concept of *average power*, and some early works on the properties of *PoS*, to name a few, are Gasparini et al. (2013) and Carrol (2013). Rufibach, Burger, and Abt (2016) gives useful reccomendations on which prior to use when computing *PoS*. Crisp et al. (2018) reports the practical experience of GSK with the use of *PoS*. All these concepts surrounding *PoS* are summarized in Chuang-Stein and Kirby (2017) and a review of the different terminologies can also be found in Kunzmann et al. (2021).

As the prior distribution plays no role in the frequentist analysis of the study, this approach is defined as hybrid Bayesian/frequentist. A thorough review of hybrid Bayesian/frequentist designs can be found in Grieve (2022). If an interim analysis is planned, the *PoS* calculated at the design stage can be updated to incorporate the information disclosed by the Data Monitoring Committee (DMC), usually consisting in a simple statement that advises to continue or to stop the trial (either for efficacy or futility), following the pre-specified rules defined in the protocol.

Our current research is driven by the observation that, in several instances, the study team exhibited either excessive optimism or pessimism regarding the final results following the DMC recommendations to continue the trial. In practice, the impact of the DMC recommendation on the *PoS* for the trial was often found to be relatively minor, but further research was needed to understand their relationship.

In this paper we focus on the relationship between *PoS* and its updated version after the DMC recommendation to continue the trial, extending the work of Temple and Robertson (2021) and Rufibach, Jordan, and Abt (2016). Temple and Robertson (2021) extends the use of *PoS* of a single study to the

conditional *PoS* of a subsequent study given a success in a previous study with the same endpoint (for example a Phase II study with the same treatment). Rufibach, Jordan, and Abt (2016) discusses statistical approaches that can be used to sequentially update *PoS* of a Phase III study, using either external (e.g., coming from another concurrent study) or internal (e.g., coming from the results of an interim analysis) information, in a time-to-event setting. Furthermore, a very recent paper by Grieve (2023) develops the use of *PoS* in group sequential designs for an arbitrary number of interim analyses.

Compared to these previous work, our research explores more specifically the relationship between *PoS* and its updated version, hereafter defined as *PoS post interim*. Moreover, we illustrate in detail the influence of interim boundaries on these probabilities. We believe that the assessment of *PoS* and *PoS post interim* helps inform the choice of the boundaries for efficacy and for futility.

In Section 2, we define *PoS* and *PoS post interim* in the general case and show their relationship, together with other operating characteristics of interest. In addition, we demonstrate the impact of the choice of the boundaries for futility and efficacy on these probabilites.

In Section 3, we show how to compute *PoS* and *PoS post interim* in the case of a two-arm trial where the treatment effect is the mean difference between the two groups, assumed to be normally distributed. Under these conditions, the computations become straightforward.

In Section 4, we provide three examples of a fictional study. In these examples we explore the impact of different futility and efficacy boundaries on *PoS* and *PoS post interim*, as well as providing some recommendations on selecting such boundaries. Furthermore, we analyse the impact of changing the information fraction in the same three examples.

Conclusive remarks are presented in Section 5.

2 *PoS* and *PoS* post interim

Consider a double-blind clinical trial with one interim analysis. Denote the estimators of the true treatment effect θ at the interim and final analysis as $\hat{\theta}_{int}$ and $\hat{\theta}_{fin}$, respectively. The threshold to reach statistical significance at the final analysis is denoted by θ_{suc}. The trial may also be stopped early at the interim analysis, for futility or efficacy, if $\hat{\theta}_{int}$ is lower than a futility boundary θ_{fut} or greater than an efficacy boundary θ_{eff}. We define the (overall) success of the trial as reaching statistical significance at the interim analysis or at the final analysis, i.e., $\hat{\theta}_{int} > \theta_{eff}$ or $\hat{\theta}_{fin} > \theta_{suc}$. For the sake of simplicity, nuisance parameters are assumed known.

Let us assume that a prior distribution for the treatment effect θ is available. For example, it was obtained from historical data or by elicitation from experts (Crisp et al., 2018). Let $q_0(\theta)$ be the probability density function of the prior distribution of θ. Given this prior distribution, we consider the classical definition

of *probability of success* (PoS), with succes defined as reaching statistical significance at the interim or final analysis:

$$\begin{aligned} PoS &= P(\textit{no early stop and success at final analysis}) + P(\textit{early stop for efficacy}) \\ &= \int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff},\, \hat{\theta}_{fin} > \theta_{suc}|\theta)\, q_0(\theta) d\theta + \int P(\hat{\theta}_{int} > \theta_{eff}|\theta)\, q_0(\theta) d\theta. \end{aligned}$$

Since the trial is double-blind, the DMC recommendation to continue the study at the interim analysis only informs the sponsor that $\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}$, but the exact value of the estimated treatment effect obtained at the interim analysis remains unknown. How does this information affect the PoS? Let us define $q_1(\theta)$ as the posterior density of θ when the trial is continued after the interim:

$$q_1(\theta) = q_1(\theta|\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}) = \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta)\, q_0(\theta)}{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta')\, q_0(\theta') d\theta'}.$$

Then, we define the *probability of success post interim* (PoS_{post}) as the probability of reaching statistical significance at the end of the trial, given that the trial was continued after the interim:

$$PoS_{post} = \int P(\hat{\theta}_{fin} > \theta_{suc}|\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}, \theta)\, q_1(\theta) d\theta. \tag{1}$$

PoS_{post} is useful to evaluate the confidence we would have in the success of the trial if it continues after the interim. Prior to the trial start, we can effectively fine-tune such confidence by adjusting the boundaries for futility and efficacy. However, there is a cost in aiming for higher PoS_{post}, as discussed below.

In order to assess the impact of boundary selection, it is first helpful to clarify the relationship between PoS_{post} and PoS. By expanding the conditional probability and the posterior density in equation (1), we get

$$\begin{aligned} PoS_{post} &= \int \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff},\, \hat{\theta}_{fin} > \theta_{suc}|\theta)}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta)} \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta)\, q_0(\theta)}{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta')\, q_0(\theta') d\theta'} d\theta \\ &= \frac{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff},\, \hat{\theta}_{fin} > \theta_{suc}|\theta)\, q_0(\theta) d\theta}{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}|\theta')\, q_0(\theta') d\theta'}. \end{aligned} \tag{2}$$

The numerator is the probability of not stopping early and reaching statistical significance at final analysis. The denominator is the probability of not stopping the trial early, respectively. Hence

$$\begin{aligned} PoS_{post} &= \frac{P(\textit{no early stop and success at final analysis})}{P(\textit{no early stop})} \\ &= \frac{PoS - P(\textit{early stop for efficacy})}{P(\textit{no early stop})}. \end{aligned} \tag{3}$$

The values of PoS_{post}, PoS, $P(\textit{no early stop and success at final analysis})$, $P(\textit{early stop for efficacy})$ and $P(\textit{no early stop})$ depend on the choice of the boundaries θ_{fut} and θ_{eff}.

It should be noted that, in our calculations, we assume that the DMC recommendation after the interim analysis will follow exactly the pre-specified stopping rules. In reality, the futility rule is likely to be non-binding and this might not happen all the times. For example, if the threshold for futility is crossed by a small margin while a number of secondary endpoints show positive results, the DMC may decide to suggest the continuation of the trial, instead of stopping it for futility. For this reason, even though the assumption is that the DMC always follows the futility boundary, θ_{suc} is calculated in a non-binding fashion and thus the type I error is not affected by the futility boundaries. However, we should expect that the DMC will behave at least somewhat consistently with the pre-specified stopping rules. We discuss the potential implications of the non-binding nature of the futility rule in Section 5.

If the interim estimate of a trial does not exceed the efficacy boundary, our confidence in the final success might be reduced. Such loss of confidence is higher for lower (easier to reach) values of θ_{eff}, because failing to reach them suggests that the treatment may not be as good as expected. At the same time, if the interim estimate is above the futility boundary, our confidence in the final success should increase. Such increase is higher for larger values of θ_{fut}, because exceeding them suggests that the treatment has at least a small effect. When computing PoS_{post} we are conditioning on $\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}$, therefore both conditions are fulfilled. This means that PoS_{post} is increasing in both θ_{eff} and θ_{fut} (see proof in appendix).

Therefore PoS_{post} is lower if the futility and efficacy boundaries are small, and higher if the futility and efficacy boundaries are large.

Consider the simpler case where early stopping for efficacy is not planned at all in the protocol, i.e., $\theta_{eff} = +\infty$. Then the expression in equation (3) simplifies to

$$PoS_{post} = \frac{PoS}{P(\textit{no early stop for futility})}.$$

If the trial can only be stopped early for futility, it is intuitive that our confidence in a final success can only increase with respect to PoS if the trial is not stopped at the interim analysis; indeed, $PoS_{post} \geq PoS$ for any choice of θ_{fut} and $PoS_{post} = PoS \Leftrightarrow P(\textit{no early stop for futility}) = 1 \Leftrightarrow \theta_{fut} = -\infty$.

Of note, by inverting the probability of not stopping early, we obtain the relative gain in probability of success after the interim:

$$P(\textit{no early stop for futility})^{-1} = \frac{PoS_{post}}{PoS}.$$

These probabilities are functions of the futility boundary. However, a change in the boundary has opposite

effects on the different probabilities:

$$\theta_{fut} \nearrow \Longrightarrow \begin{array}{l} P(\textit{no early stop for futility}) \searrow \\ PoS \searrow \\ PoS_{post} \nearrow \end{array}$$

Since PoS and PoS_{post} have opposite trends, there is a tradeoff in setting a more (or less) aggressive futility boundary. Computing both probabilities of success for different values of θ_{fut} allows an assessment of this tradeoff and facilitates an informed decision on the choice of a futility boundary.

3 Special case for normally distributed data

We consider a two-arm trial with a sample size of n subjects per group in which θ is the mean treatment difference between a new drug and a control. The sample size is set according to the choice of the type I error rate α, the power $1-\beta$ and a mean treatment effect of interest Δ. The objective of the trial is to demonstrate that the new drug is superior to the control, which is translated in statistical terms as rejecting the null hypothesis H_0: $\theta \leq 0$ in favour of the alternative H_1: $\theta > 0$. The standard deviation of the response variable σ is assumed to be known. An interim analysis is scheduled after n_{int} subjects in each group have completed the study. We set a futility boundary θ_{fut} and an efficacy boundary θ_{eff} so that we may stop either for futility or for efficacy at the interim analysis. The threshold to reach statistical significance at the final analysis is again denoted by θ_{suc}.

In this particular case, we assume that the mean treatment effect θ is normally distributed. We will show that under this assumption PoS and PoS_{post} can be computed using a bivariate normal distribution. Given the mean treatment difference θ, the estimator $\hat{\theta}_{fin}|\theta$ at the final analysis is

$$\hat{\theta}_{fin}|\theta \sim \mathcal{N}\left(\theta, \frac{2\sigma^2}{n}\right).$$

The estimator $\hat{\theta}_{int}|\theta$ at the interim analysis also follows a normal distribution:

$$\hat{\theta}_{int}|\theta \sim \mathcal{N}\left(\theta, \frac{2\sigma^2}{n_{int}}\right)$$

and the conditional bivariate normal distribution of both is

$$\begin{pmatrix} \hat{\theta}_{int} \\ \hat{\theta}_{fin} \end{pmatrix} | \theta \sim N \left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} \frac{2\sigma^2}{n_{int}} & \frac{2\sigma^2}{n} \\ \frac{2\sigma^2}{n} & \frac{2\sigma^2}{n} \end{pmatrix} \right).$$

Let us also assume that the prior for theta is a normal distribution:

$$\theta \sim \mathcal{N} \left(\theta_0, \frac{2\sigma^2}{n_0} \right), \tag{4}$$

where n_0 is a fixed positive real number. For example, n_0 might be set to be equal (or smaller, in case of discounting) to the sample size of a previous study providing an estimate of the mean treatment effect. In general n_0 is not necessarily an integer and it may be chosen freely to obtain any sensible value for the variance of the prior (Grieve, 2022, Chapter 2). Expressing the variance of the prior as $\frac{2\sigma^2}{n_0}$ is convenient when deriving the unconditional distributions of the estimators $\hat{\theta}_{int}$ and $\hat{\theta}_{fin}$. Since both $\hat{\theta}_{fin}|\theta$ and θ are normally distributed, it follows that the unconditional distribution of $\hat{\theta}_{fin}$ is normal too. The same is true for the for unconditional distribution of $\hat{\theta}_{int}$. Their unconditional bivariate distribution is

$$\begin{pmatrix} \hat{\theta}_{int} \\ \hat{\theta}_{fin} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_0 \\ \theta_0 \end{pmatrix}, \begin{pmatrix} 2\sigma^2 \left(\frac{n_{int}+n_0}{n_{int}n_0} \right) & 2\sigma^2 \left(\frac{n+n_0}{nn_0} \right) \\ 2\sigma^2 \left(\frac{n+n_0}{nn_0} \right) & 2\sigma^2 \left(\frac{n+n_0}{nn_0} \right) \end{pmatrix} \right). \tag{5}$$

For the reader's convenience, we provide a proof of (5) in the Appendix. The unconditional multivariate distribution for an arbitrary number of interim analyses can be found in Grieve (2023).

A very similar result can be obtained in a time-to-event setting if the treatment effect is measured as hazard ratio, by making use of the normal approximation of the log hazard ratio, as shown in Rufibach, Burger, and Abt (2016).

When the distribution in (5) holds, the probabilities in equation (3) become easy to compute by making use of the bivariate normal cumulative distribution function of $(\hat{\theta}_{int}, \hat{\theta}_{fin})$. The explicit formulas of these probabilities are given in the Appendix.

4 Examples

4.1 Example 1: No early stop for efficacy

Consider a parallel group trial for a new treatment where the treatment effect is assumed normally distributed and the standard deviation is known ($\sigma = 1$). Assume $\Delta = 0.3$ as the treatment effect size

of interest, expressed in terms of mean difference between treatments. This choice of Δ and σ reflects a median standardized effect size of 0.3 found in pivotal trials (Rothwell, Julious, and Cooper, 2018). The power and the type I error rate are set to $1-\beta = 0.9$ and one-sided $\alpha = 0.025$ respectively. Based on these assumptions, a total of 468 subjects ($n = 234$ per group) are needed to reach the target power. The threshold for statistical significance θ_{suc} is computed as:

$$\theta_{suc} = z_{1-\alpha}\sqrt{\frac{2}{n}}\sigma = 0.181.$$

An interim analysis for futility is scheduled when half of the subjects have completed the study ($n_{int} = 117$ in each group). In this first example, the trial is not allowed to stop early for efficacy, i.e., the efficacy boundary is fixed at $\theta_{eff} = +\infty$.

We consider three normal prior distributions to evaluate PoS and PoS_{post} of this trial. These are defined as in (4) with $n_0 = 10$ and different choices of θ_0. We call the priors *pessimistic*, *realistic* and *optimistic*, as an evocative reflection of their characteristics. The pessimistic prior is centered in a value below the effect size of interest (i.e., $\theta_0 = \Delta - 0.2$), the realistic prior is exactly centered in Δ (i.e., $\theta_0 = \Delta$) and the optimistic prior is centered in a value above Δ (i.e., $\theta_0 = \Delta + 0.2$). The same variance is assumed in all scenarios. These three prior distributions may be seen as plausible posterior distibutions obtained from a small study conducted on a total of $2n_0$ subjects before the start of the new trial. The pessimistic prior reflects a scenario where the treatment effect was below expectation in the small trial. In such a scenario, we may still consider advancing the development program, for example in a disease with a high unmet need. The realistic and optimistic priors correspond to cases with an observed effect in the small trial in line with expectations or above expectations, respectively.

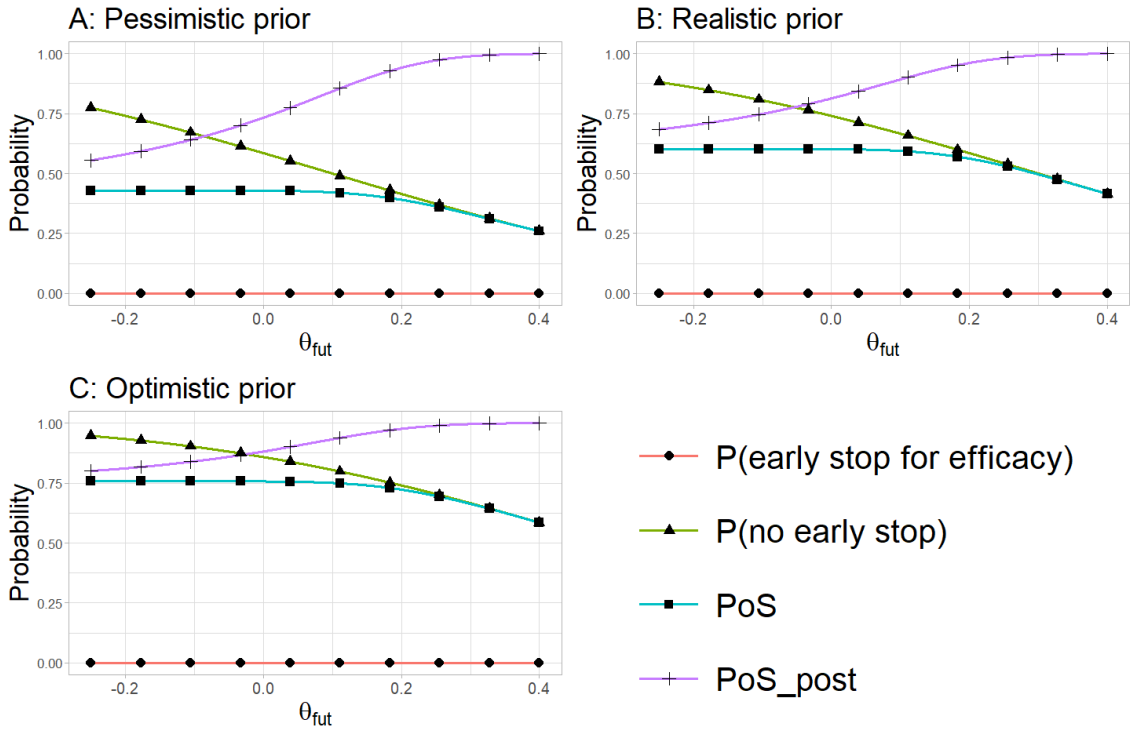
As already noted, if a trial has a stopping rule for futility only, then $PoS_{post} \geq PoS$. Therefore, if the trial is not stopped at the interim look, the probability of reaching significance at the final analysis increases, regardless of the choice of the futility threshold. Relevant gains in PoS_{post} with negligible losses in PoS can ususally be achieved through an appropriate choice of θ_{fut}.

Figure 1 compares PoS_{post} and PoS as functions of θ_{fut}. In all three scenarios, PoS is almost constant for $\theta_{fut} < 0$, as the probability of success is practically unaffected by the definition of a small futility boundary. But even for those extremely cautious futility boundaries, there is an evident gain in terms of PoS_{post}.

To evaluate the gain in PoS_{post} against the loss of PoS, let us define the maximum possible PoS ($MPPoS$ in short) as the upper bound for PoS. $MPPoS$ is PoS in a trial without a stopping rule for futility, i.e., when $\theta_{fut} = -\infty$.

In the investigated scenarios, PoS decreases slowly while PoS_{post} increases relatively fast with increasing θ_{fut}. For this reason, a quite large futility boundary could be set to increase PoS_{post} for a small cost in

Figure 1: PoS_{post}, PoS, $P(\textit{early stop for efficacy})$ and $P(\textit{no early stop})$ as functions of θ_{fut} for three different priors when there is no stopping rule for efficacy. $P(\textit{early stop for efficacy}) = 0$ because early stop for efficacy is not allowed. PoS is always smaller than $P(\textit{no early stop})$ because a success in this case requires that the trial is not stopped for futility at the interim. $PoS_{post} \geq PoS$ because the confidence in a success increases if the trial is not stopped at the interim analysis.



terms of PoS reduction. For example, let us consider the trial design with the realistic prior, in which $MPPoS = 0.60$. Examining PoS_{post} as a function of θ_{fut}, we can identify the futility boundary that gives $PoS = 0.59$; that is $\theta_{fut} = 0.11$. This choice of θ_{fut} leads to $PoS_{post} = 0.90$. In a similar fashion, we can identify the smallest futility boundary that gives $PoS = 0.58$; that is $\theta_{fut} = 0.15$. This choice of θ_{fut} leads to $PoS_{post} = 0.93$. Limited losses of 1% or 2% compared to MPPoS lead to a significantly higher PoS_{post}, which corresponds to a stronger confidence in a final success in case of no early stop for futility. These findings are summarised in Table 1, including also the results for the following other examples.

The difference between the three priors, instead, is shown in Figure 1. As one would expect, the pessimistic prior gives the lowest PoS and PoS_{post}, while the optimistic prior gives the highest ones (θ_{fut} being equal). On the other hand, the largest increase in PoS_{post} compared to PoS is associated with the pessimistic prior, the lowest with the optimistic prior. This is because the optimistic prior assumes that the treatment effect is likely to be above Δ. Therefore, limited evidence of a positive treatment effect does not substantially alter the expected outcome. In contrast, the same evidence has a greater impact when the pessimistic prior is initially assumed.

4.2 Example 2: O'Brien-Fleming efficacy boundary

In the same scenarios previously described, we consider adding the possibility of an early stop for efficacy if the interim estimated effect is above a chosen efficacy boundary θ_{eff}, obtained using an O'Brien-Fleming alpha spending function (DeMets and Lan, 1994)

$$\alpha_{int}(I) = 2 - 2\phi\left(\frac{z_{1-\alpha/2}}{\sqrt{I}}\right).$$

Assuming the information fraction to be $I = 1/2$, i.e., half the subjects are considered for the interim analysis, $\alpha_{int}(1/2) = 0.0015$. The corresponding efficacy boundary at the interim analysis is

$$\theta_{eff} = z_{1-\alpha_{int}(1/2)}\sqrt{\frac{2}{n_{int}}}\sigma = 0.387. \tag{6}$$

Since we are introducing the option to stop early for efficacy, there are now two opportunities to reject the null hypotesis. Therefore, the threshold to reach statistical significance at the final analysis θ_{suc} has to be increased in order to preserve the overall type I error rate. In this setting, $\theta_{suc} = 0.182$ should be set.

As we can see in Figure 2, when adding a stopping rule for efficacy ($\theta_{eff} = 0.387$), $PoS_{post} \geq PoS$ does not hold for every possible choice of θ_{fut}, as was the case in Example 1 ($\theta_{eff} = +\infty$), but only when θ_{fut} is large enough. This is due to the negative impact of a failed interim for efficacy on PoS_{post} (PoS_{post} is increasing in θ_{eff}).

As it was the case in the previous example, PoS and PoS_{post} are highest with the optimistic prior but the difference $PoS_{post} - PoS$ is largest with the pessimistic prior. Indeed, we can achieve $PoS_{post} \geq PoS$ with a much smaller θ_{fut} when using the pessimistic prior, compared to the other scenarios.

Regardless of the prior used, there is always the possibility to select a futility boundary that gives PoS slightly below the $MPPoS$ in order to increase PoS_{post} at a minimal cost (see Table 1 for the detailed results).

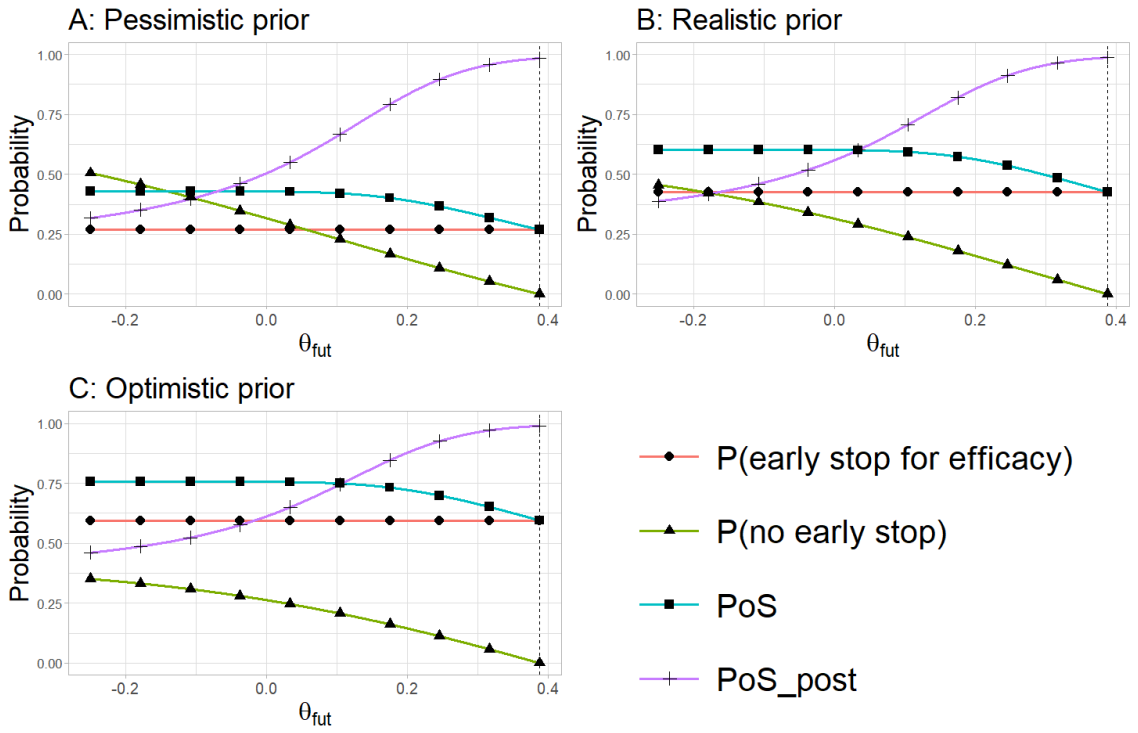
4.3 Example 3: Pocock efficacy boundary

This example is based on the same setting of Example 2, with one interim analysis with stopping rules for futility or efficacy, but using a different efficacy boundary, based on the Pocock alpha spending function (DeMets and Lan, 1994)

$$\alpha_{int}(I) = \alpha \ln\left[1 + (e-1)I\right].$$

Using the formula in (6), we obtain $\theta_{eff} = 0.282$. In order to preserve the overall type I error rate, the

Figure 2: PoS_{post}, PoS, $P(\textit{early stop for efficacy})$ and $P(\textit{no early stop})$ as functions of θ_{fut} for three different priors when an O'Brien-Fleming efficacy boundary (vertical dashed line) is used.

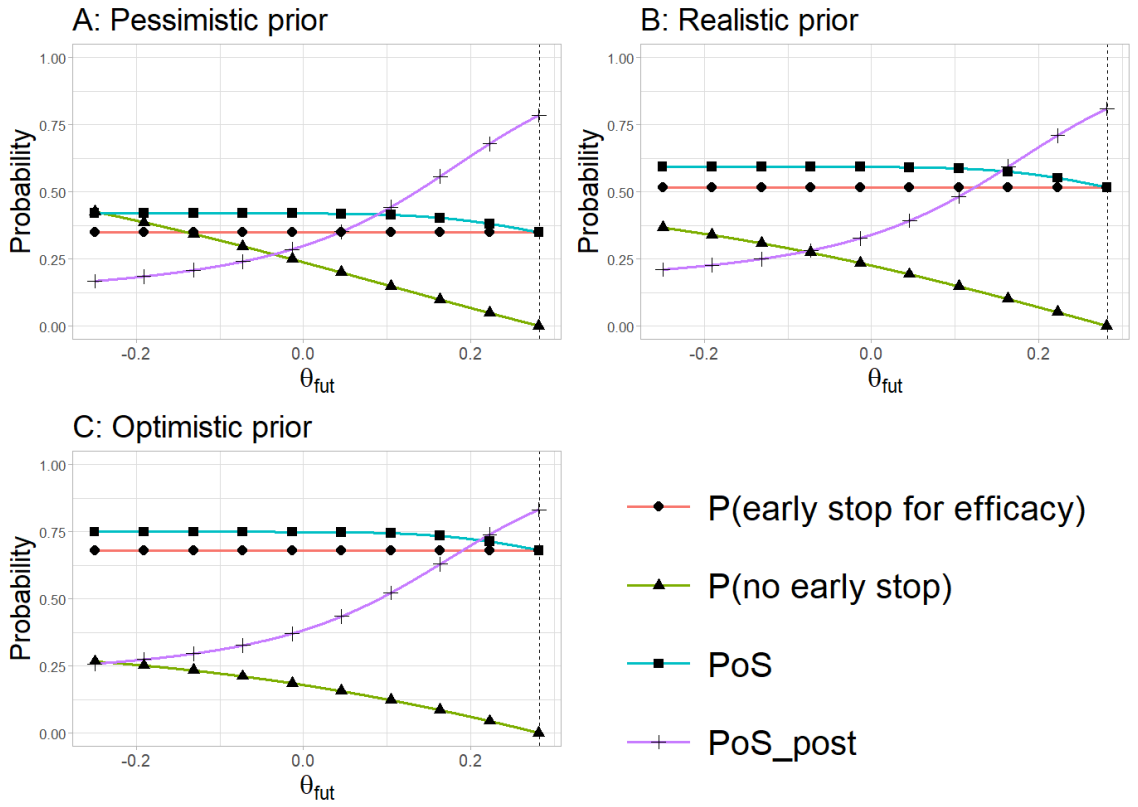


threshold for the final analysis is now $\theta_{suc} = 0.203$

Figure 3 presents the behaviour of PoS and PoS_{post} in this example. Since the Pocock efficacy boundary is smaller than the prior mean of both the realistic and optimistic prior, in both scenarios a very large futility boundary is needed to obtain $PoS_{post} \geq PoS$. Intuitively this is reasonable, because continuing after an interim analysis implies that the observed effect was smaller than the target effect ($\hat{\theta}_{int} \leq 0.282 < \Delta = 0.3$). On the other hand, with the pessimistic prior even a moderate futility boundary can result in a PoS_{post} more favorable than the initial PoS. Indeed, in the pessimistic scenario $PoS_{post} \geq PoS$ for the choice of θ_{fut} that reduces PoS by only 0.01, as shown in Table 1.

In any case, by choosing a futility boundary that gives PoS slightly below the $MPPoS$, we can always increase PoS_{post} (see Table 1 for detailed results), as in the previous examples. Although PoS_{post} will not necessarily exceed PoS, it may still be a valid option to choose a large futility boundary, in order to increase the confidence in a final success of the trial in case it continues after the interim analysis.

Figure 3: PoS_{post}, PoS, $P(\textit{early stop for efficacy})$ and $P(\textit{no early stop})$ as functions of θ_{fut} for three different priors when a Pocock efficacy boundary (vertical dashed line) is used.



4.4 Further comments on the examples

Table 1: PoS and PoS_{post} tradeoff according to futility boundaries. PoS is at its maximum ($MPPoS$) and PoS_{post} at its minimum when there is no stopping rule for futility ($\theta_{fut} = -\infty$). The two columns on the right-hand side show the value of PoS_{post} for the choice of futility boundaries that reduce PoS by 0.01 and 0.02 compared to $MPPoS$, respectively.

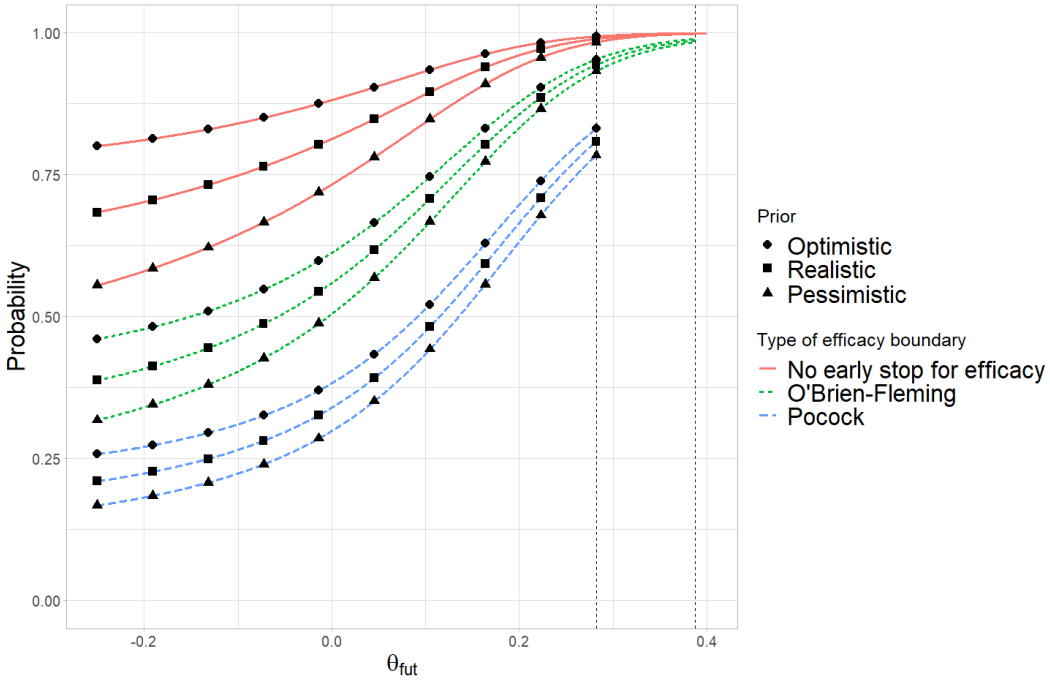
Example	Prior	PoS for $\theta_{fut} = -\infty$ ($MPPoS$)	PoS_{post} for $\theta_{fut} = -\infty$	PoS_{post} when PoS is reduced by 0.01	PoS_{post} when PoS is reduced by 0.02
No early stop for efficacy	Pessimistic	0.43	0.43	0.85	0.90
	Realistic	0.60	0.60	0.90	0.93
	Optimistic	0.76	0.76	0.94	0.96
O'Brien-Fleming boundary	Pessimistic	0.43	0.22	0.68	0.75
	Realistic	0.60	0.31	0.72	0.78
	Optimistic	0.76	0.40	0.77	0.83
Pocock boundary	Pessimistic	0.42	0.11	0.50	0.58
	Realistic	0.60	0.16	0.53	0.61
	Optimistic	0.75	0.21	0.58	0.66

The effect of the chosen efficacy boundary on PoS_{post} can be seen in Figure 4, where the different PoS_{post} functions from the three examples are plotted together. As we have already mentioned, PoS_{post} is lower for smaller efficacy boundaries. Indeed PoS_{post} is lowest when a Pocock efficacy boundary is used, and highest when there is no stopping rule for efficacy.

Table 1 shows that even a quite small futility boundary, leading to a reduction in PoS of only 0.01, yields a PoS_{post} significantly greater than the one obtained without futility stopping rules. However, the benefit of choosing a larger futility boundary may not always outweigh the loss of PoS. For example, it looks reasonable to reduce PoS by 0.02 if an efficacy boundary is set (either the O'Brien-Fleming or the Pocock type), because PoS_{post} would increase further by 0.06 or more. On the other hand, in Example 1 ("No early stop for efficacy"), the gain in PoS_{post} is only 0.05 if we are considering the pessimistic prior, or less otherwise. This phenomenon is graphically represented in Figure 4, where the PoS_{post} function in the "No early stop for efficacy" example is less steep than in the "Pocock" example, meaning that in the latter situation PoS_{post} increases more quickly for larger values of θ_{fut}.

In Figure 4 we can also observe the impact of the three priors (the pessimistic, realistic and optimistic ones) on PoS_{post}. For low values of θ_{eff}, i.e., when a Pocock alpha-spending function is used, the difference of PoS_{post} in the pessimistic and realistic scenarios, as well as the difference between the realistic and the optimistic scenarios, is relatively constant for all values of θ_{fut}. For higher values of θ_{eff} (the extreme case being the "no early stop for efficacy" example), PoS_{post} is quite different between the three scenarios for low values of θ_{fut}, but then quickly converges to the same value with increasing θ_{fut}.

Figure 4: Comparison of PoS_{post} as a function of θ_{fut} in the three discussed examples. From left to right, the vertical dashed lines correspond to the Pocock and the O'Brien-Fleming efficacy boundaries from example 3 and 2 respectively.



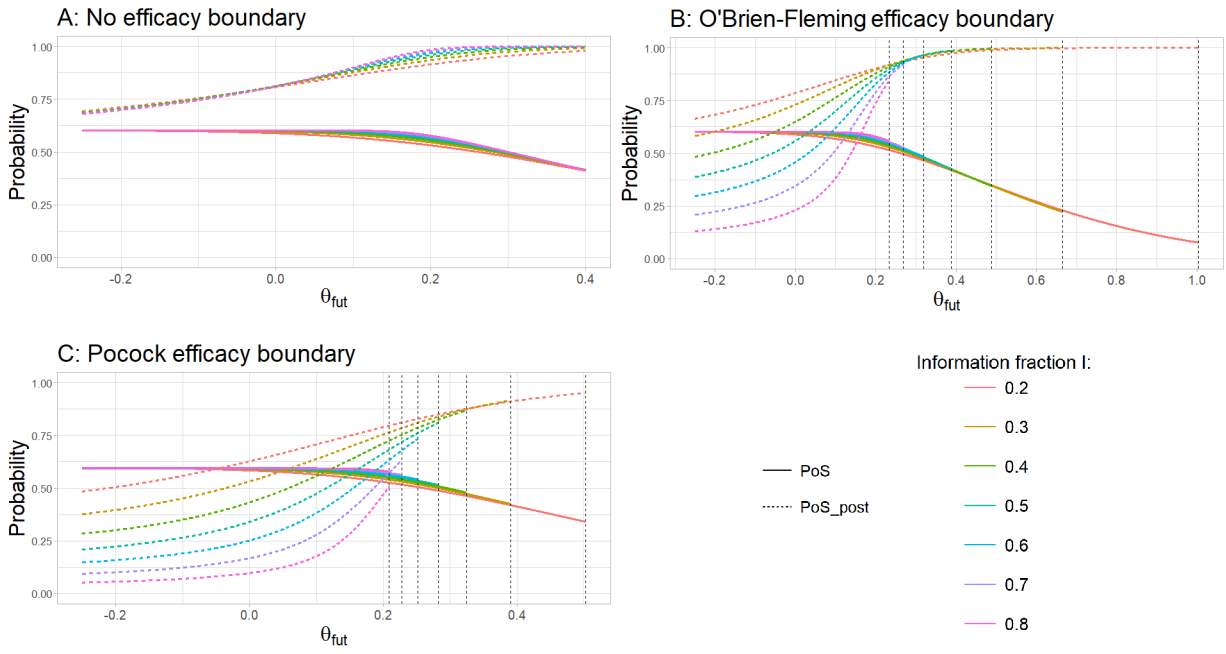
Moreover, focusing on the efficacy boundary, PoS_{post} decreases significantly for smaller θ_{eff}. We can observe a relevant drop in PoS_{post} when moving from the example with no early stop for efficacy to the example with the O'Brien-Fleming boundary and from the O'Brien-Fleming to the Pocock boundary. This does not necessarily mean that a lower efficacy boundary is detrimental to the study in general, but it means that, if the study continues after the interim, the confidence in its success may be reduced.

4.5 Impact of the information fraction

Our previous examples focused on trials with a single interim analysis conducted when half the subjects completed the study, corresponding to an information fraction $I = 0.5$. The theoretical results from Sections 2 and 3 remain valid regardless of the information fraction. Nevertheless, despite the general validity of their described relationship, variations in information fraction may influence PoS and PoS_{post} values. In this section, we investigate this impact within the specific context of the designs previously examined.

Figure 5 compares PoS and PoS_{post} for information fractions from 0.2 to 0.8, assuming the realistic prior. Note that the range of θ_{fut} depends on I, as θ_{fut} must be less than θ_{eff}, which decreases as I increases. Figure 6 and Table 2 provide more details for $I = 0.2$ and $I = 0.8$ using the O'Brien-Fleming alpha spending function under the same prior.

Figure 5: PoS and PoS_{post} under the realistic prior as functions of θ_{fut} for information fractions between $I = 0.2$ and $I = 0.8$. The vertical dashed lines correspond to the efficacy boundaries.



In the trial design with no early stop for efficacy, the information fraction has a limited impact on both PoS and PoS_{post}. PoS is almost not affected by the information fraction. Regarding PoS_{post}, the curve describing its relationship with θ_{fut} becomes flatter as I decreases, because the results informing the interim decision become less predictive of the final outcome.

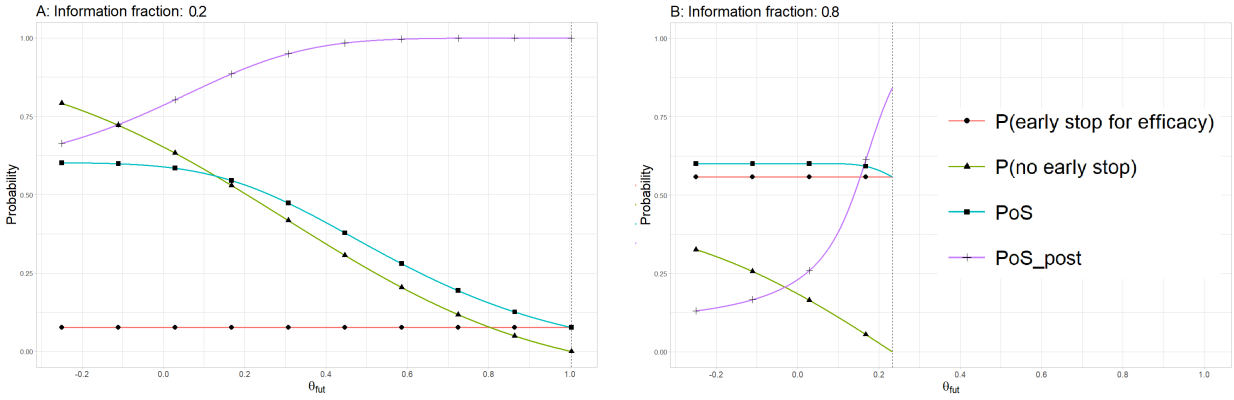
In the two designs allowing early efficacy stop, the information fraction minimally affects PoS overall. Despite its negligible impact on overall PoS, the information fraction significantly influences its components ($PoS = P(\textit{early stop for efficacy}) + P(\textit{no early stop}) \cdot PoS_{post}$), as shown in Table 2: higher I values increase the probability of early efficacy stop, while reducing the relative contribution of final analysis success to PoS. This occurs because larger information fractions lead to greater alpha spending at the interim analysis, resulting in smaller, more easily crossed efficacy boundaries θ_{eff}. Contrary to PoS, PoS_{post} is greatly affected by the information fraction. As in the design without early efficacy stop, the PoS_{post} vs. θ_{fut} curve flatten with lower I, but this trend is significantly more pronounced when an interim efficacy analysis is planned. This is induced by gradually decreasing values of θ_{eff} as information accumulates. Failing to cross a smaller efficacy boundary later in the study has a more substantial negative impact on PoS_{post} compared to not reaching a larger, more challenging boundary earlier.

For all three trial designs, consistent trends are observed when considering pessimistic and optimistic prior distributions.

Table 2: PoS, PoS_{post}, $P(no\ early\ stop)$, $P(early\ stop\ for\ efficacy)$ and $P(early\ stop\ for\ futility)$ as functions of θ_{fut} for information fractions $I = 0.2$ and $I = 0.8$ when an O'Brien-Fleming efficacy boundary is used and the realistic prior is assumed.

	$I = 0.2$					$I = 0.8$				
θ_{fut}	PoS	PoS_{post}	$P(no\ early\ stop)$	$P(stop\ for\ eff.)$	$P(stop\ for\ fut.)$	PoS	PoS_{post}	$P(no\ early\ stop)$	$P(stop\ for\ eff.)$	$P(stop\ for\ fut.)$
-0.20	0.60	0.68	0.77	0.08	0.15	0.60	0.14	0.30	0.56	0.14
-0.15	0.60	0.70	0.74	0.08	0.18	0.60	0.15	0.28	0.56	0.16
-0.10	0.60	0.73	0.72	0.08	0.21	0.60	0.17	0.25	0.56	0.19
-0.05	0.59	0.76	0.68	0.08	0.24	0.60	0.19	0.22	0.56	0.22
0.00	0.59	0.79	0.65	0.08	0.27	0.60	0.23	0.19	0.56	0.26
0.05	0.58	0.82	0.62	0.08	0.31	0.60	0.29	0.15	0.56	0.29
0.10	0.57	0.85	0.58	0.08	0.34	0.60	0.38	0.11	0.56	0.33
0.15	0.55	0.88	0.54	0.08	0.38	0.60	0.54	0.07	0.56	0.37
0.20	0.53	0.90	0.50	0.08	0.42	0.58	0.74	0.03	0.56	0.41

Figure 6: PoS_{post}, PoS, $P(early\ stop\ for\ efficacy)$ and $P(no\ early\ stop)$ as functions of θ_{fut} for information fractions $I = 0.2$ and $I = 0.8$ when an O'Brien-Fleming efficacy boundary (vertical dashed lines) is used and the realistic prior is assumed.



5 Discussion

PoS is the unconditional probability of success of the study, based on a prior distribution on the treatment effect θ. PoS_{post}, on the other hand, is the PoS conditioned on the fact that the trial continues after the interim analysis, i.e., $\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}$.

Our experience has demonstrated that, on many occasions, the study team tends to express excessive optimism regarding the future outcome of the trial upon receiving the news that the trial will continue following the interim phase. Our research objectively examines the relationship between PoS and PoS_{post} to determine if this behavior is warranted, and to provide recommendations for adjusting the interim boundaries to align with the expectations. Moreover, we believe that PoS_{post} offers a straightforward interpretation of the impact of the DMC recommendations to continue the trial after the interim analysis, making it a valuable tool for communication with non-statisticians.

We considered the case where the treatment effect is measured as a mean difference, and provided some examples of a fictive trial with a fixed maximum sample size. In these exmples we explored the use of three different prior distributions and efficacy boundaries, while the futility boundary was allowed to vary on the entire continuous scale.

For a given choice of θ_{eff}, we have shown that PoS_{post} is increasing in θ_{fut}, while PoS is decreasing. PoS_{post} is also increasing in θ_{eff}, hence a higher efficacy boundary will yield a higher PoS_{post}, but at the cost of a smaller chance of an early stop for efficacy.

Our results show that, in certain situations, a somewhat large futility boundary may be chosen to obtain a substantial increase of PoS_{post} at the cost of a slight reduction of PoS. On the other hand, if the futility rule is not aggressive enough, the fact of passing an interim analysis should not be received with too much optimism, since the chances of success may remain quite low. This is particularly true for studies with efficacy boudaries that are quite easy to reach. When instead there are no efficacy stopping rules, the chances of success always increase after passing the futility interim analysis.

For all these reasons, it is important to understand the impact of the interim analysis on the PoS of the study at the design stage, by fine-tuning the interim boundaries, in order to avoid later disappointments. To the best of our knowledge, our research is the first to study the relationship between PoS and PoS_{post} and to connect it with the choice of the interim boundaries.

We based our definitions of PoS and PoS_{post} on the assumption that the DMC would always recommend to continue a trial if the futility rule is fulfilled. Although in reality the futility rule is likely to be non-binding, it is reasonable to assume that the DMC recommendation will be at least somewhat consistent with the pre-specified stopping rules. Incorporating the probability that the DMC follows the futility rule into our calculations would be theoretically possible. However, the parameters of such modeling would certainly be quite subjective and context dependent: this is an area for future research.

Aknowledgements

This publication is part of the project PNRR-NGEU which has received funding from the MUR – DM 352/2022. This publication is under the National Plan for Complementary Investments to the NRRP, project "D34H—Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care" (project code: PNC0000001), Spoke 4, funded by the Italian Ministry of University and Research.

Supplementary material

Tables summarizing the impact of information fraction for other designs and prior distributions and the R codes to produce the results can be found on https://github.com/LucaRondano/Supplementary-Material.

References

Carrol, K.J. (2013). "Decision making from phase II to phase III and the probability of success: reassured by "assurance"?" In: *Journal of Biopharmaceutical Statistics* 23, pp. 1188–1200.

Chuang-Stein, C. (2006). "Sample size and the probability of a successful trial". In: *Pharmaceutical Statistics* 5, pp. 305–309. DOI: https://doi.org/10.1002/pst.232.

Chuang-Stein, C. and S. Kirby (2017). *Quantitative Decisions in Drug Development*. Springer.

Crisp, A. et al. (2018). "Practical experiences of adopting assurance as a quantitative framework to support decision making in drug development". In: *Pharmaceutical Statistics* 17, pp. 317–328.

DeMets, D.L. and K.K.G. Lan (1994). "Interim analysis: the alpha spending function approach". In: *Statistics in Medicine* 13, pp. 1341–1352.

Gasparini, M. et al. (2013). "Predictive probability of success in clinical drug development". In: *Epidemiology, Biostatistics and Public Health* 10.1, e8760-1–e8760-14.

Grieve, A.P. (2022). *Hybrid Frequentist/Bayesian Power and Bayesian Power in Planning Clinical Trials*. CRC press.

Grieve, A.P. (2023). "Probability of success and group sequential designs". In: *Pharmaceutical Statistics*, pp. 1–19. DOI: https://doi.org/10.1002/pst.2346.

Kunzmann, K. et al. (2021). "A Review of Bayesian Perspectives on Sample Size Derivation for Confirmatory Trials". In: *The American Statistician* 75, pp. 424–432.

O'Hagan, A., J.W. Stevens, and M.J. Campbell (2005). "Assurance in clinical trial design". In: *Pharmaceutical Statistics* 4, pp. 187–201. DOI: https://doi.org/10.1002/pst.175.

Rothwell, J.C., S.A. Julious, and C.L. Cooper (2018). "A study of target effect sizes in randomised controlled trials published in the Health Technology Assessment journal". In: *Trials* 19. DOI: https://doi.org/10.1186/s13063-018-2886-y.

Rufibach, K., H. Burger, and M. Abt (2016). "Bayesian predictive power: choice of prior and some recommendations for its use as probability of success in drug development". In: *Pharmaceutical Statistics* 15, pp. 438–446.

Rufibach, K., P. Jordan, and M. Abt (2016). "Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information". In: *Journal of Biopharmaceutical Statistics* 26, pp. 191–201.

Temple, J.R. and J.R. Robertson (2021). "Conditional assurance: the answer to the questions that should be asked within drug development". In: *Pharmaceutical Statistics*, pp. 1–10. DOI: https://doi.org/10.1002/pst.2128.

A Appendix

A.1 Proof that PoS_{post} is increasing in θ_{eff}

Given a futility boundary θ_{fut}, let us explicitly write the dependence of PoS_{post} on the efficacy boundary θ_{eff} as $PoS_{post}(\theta_{eff})$. $PoS_{post}(\theta_{eff})$ is an increasing function of θ_{eff} if

$$PoS_{post}(\theta_{eff1}) < PoS_{post}(\theta_{eff2})$$

for any choice of θ_{eff1}, θ_{eff2} such that $\theta_{fut} < \theta_{eff1} < \theta_{eff2}$.

Let us choose any pair $(\theta_{eff1}, \theta_{eff2})$ satisfying $\theta_{fut} < \theta_{eff1} < \theta_{eff2}$. In a study with an interim analysis

for efficacy, the threshold at the final analysis have to be determined according to the choice of the efficacy boundary in order to preserve the overall type I error rate. Let us call θ_{suc1} and θ_{suc2} the thresholds for significance at the final analysis corresponding to the choice of θ_{eff1} and θ_{eff2} as efficacy boundaries, respectively. Since $\theta_{eff1} < \theta_{eff2}$, it follows that

$$\theta_{suc1} > \theta_{suc2}. \tag{7}$$

Let $q_0(\theta)$ be the probability density function of the prior distribution of the treatment effect θ. For the law of total probability, the unconditional joint distribution of $\hat{\theta}_{int}$ and $\hat{\theta}_{fin}$ is

$$P(\hat{\theta}_{int} \in A,\, \hat{\theta}_{fin} \in B) = \int P(\hat{\theta}_{int} \in A,\, \hat{\theta}_{fin} \in B|\theta) q_0(\theta) d\theta \quad \forall A, B \subset \mathbb{R}.$$

For equation (2),

$$\begin{aligned} PoS_{post}(\theta_{eff1}) &= \frac{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1},\, \hat{\theta}_{fin} > \theta_{suc1}|\theta)\, q_0(\theta) d\theta}{\int P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}|\theta')\, q_0(\theta') d\theta'} \\ &= \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1},\, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1})}. \end{aligned}$$

Let us prove the following inequality first:

$$\frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1},\, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1})} < \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2},\, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2})}. \tag{8}$$

The right-hand side of (8) is equal to

$$\frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1},\, \hat{\theta}_{fin} > \theta_{suc1}) + P(\theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2},\, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}) + P(\theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2})}.$$

Let us call

$$\begin{aligned} a &= P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1},\, \hat{\theta}_{fin} > \theta_{suc1}), \\ b &= P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}), \\ c &= P(\theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2},\, \hat{\theta}_{fin} > \theta_{suc1}), \\ d &= P(\theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2}). \end{aligned}$$

Since $a, b, c, d > 0$,

$$\frac{a}{b} < \frac{a+c}{b+d} \iff a(b+d) < (a+c)b \iff ad < cb \iff \frac{a}{b} < \frac{c}{d}.$$

By definition of conditional probability,

$$\frac{a}{b} = P(\hat{\theta}_{fin} > \theta_{suc1} | \theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}),$$
$$\frac{c}{d} = P(\hat{\theta}_{fin} > \theta_{suc1} | \theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2}).$$

Hence (8) follows from the following inequality:

$$P(\hat{\theta}_{fin} > \theta_{suc1} | \theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}) < P(\hat{\theta}_{fin} > \theta_{suc1} | \theta_{eff1} < \hat{\theta}_{int} \leq \theta_{eff2}).$$

The above inequality is true because $\hat{\theta}_{fin}$ depends linearly on $\hat{\theta}_{int}$. In particular

$$\hat{\theta}_{fin} = \left(\frac{n_{int}}{n}\right) \hat{\theta}_{int} + \left(1 - \frac{n_{int}}{n}\right) \hat{\theta}_{post},$$

where $\hat{\theta}_{post}$ is the estimator of θ in the second part of the trial (Grieve, 2022, Chapter 3).

Using (8), we can complete the proof as follows

$$\begin{aligned}
PoS_{post}(\theta_{eff1}) &\overset{(2)}{=} \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1}, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff1})} \\
&\overset{(8)}{<} \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2}, \hat{\theta}_{fin} > \theta_{suc1})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2})} \\
&\overset{(7)}{<} \frac{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2}, \hat{\theta}_{fin} > \theta_{suc2})}{P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff2})} \\
&\overset{(2)}{=} PoS_{post}(\theta_{eff2}).
\end{aligned}$$

The proof that PoS_{post} is increasing in θ_{fut} is analogous to this one.

A.2 Proof of (5)

From the law of total expectation

$$\mathbb{E}[\hat{\theta}_{int}] = \mathbb{E}\left[\mathbb{E}[\hat{\theta}_{int}|\theta]\right] = \mathbb{E}[\theta] = \theta_0.$$

From the law of total variance

$$var[\hat{\theta}_{int}] = \mathbb{E}\left[var[\hat{\theta}_{int}|\theta]\right] + var\left[\mathbb{E}[\hat{\theta}_{int}|\theta]\right] = \mathbb{E}\left[\frac{2\sigma^2}{n_{int}}\right] + var\left[\theta\right] = \frac{2\sigma^2}{n_{int}} + \frac{2\sigma^2}{n_0} = 2\sigma^2\left(\frac{n_{int} + n_0}{n_{int} n_0}\right).$$

Similarly, $\mathbb{E}[\hat{\theta}_{fin}] = \theta_0$ and $var[\hat{\theta}_{fin}] = 2\sigma^2\left(\frac{n + n_0}{n n_0}\right)$.

Let us define $\hat{\theta}_{post}$ as the estimator of the mean treatment effect in the second part of the trial. Notice that,

conditionally on θ, $\hat{\theta}_{int}|\theta$ and $\hat{\theta}_{post}|\theta$ are independent and $\hat{\theta}_{fin}|\theta = \left[\left(\frac{n_{int}}{n}\right)\hat{\theta}_{int} + \left(1 - \frac{n_{int}}{n}\right)\hat{\theta}_{post}\right]|\theta$. Moreover, from the law of total covariance

$$cov(\hat{\theta}_{int}, \hat{\theta}_{fin}) = \mathbb{E}\left(cov(\hat{\theta}_{int}|\theta, \hat{\theta}_{fin}|\theta)\right) + cov\left(\mathbb{E}(\hat{\theta}_{int}|\theta), \mathbb{E}(\hat{\theta}_{fin}|\theta)\right).$$

Let us compute the two terms in the right-hand side:

$$\begin{aligned}
\mathbb{E}\left(cov(\hat{\theta}_{int}|\theta, \hat{\theta}_{fin}|\theta)\right) &= \mathbb{E}\left(cov(\hat{\theta}_{int}|\theta, \left[\left(\frac{n_{int}}{n}\right)\hat{\theta}_{int} + \left(1 - \frac{n_{int}}{n}\right)\hat{\theta}_{post}\right]|\theta)\right) \\
&= \mathbb{E}\left(\frac{n_{int}}{n} var(\hat{\theta}_{int}|\theta) + 0\right) \\
&= \mathbb{E}\left(\frac{n_{int}}{n}\frac{2\sigma^2}{n_{int}}\right) = \frac{2\sigma^2}{n},
\end{aligned}$$

$$cov\left(\mathbb{E}(\hat{\theta}_{int}|\theta), \mathbb{E}(\hat{\theta}_{fin}|\theta)\right) = var(\theta) = \frac{2\sigma^2}{n_0}.$$

It immediately follows

$$cov(\hat{\theta}_{int}, \hat{\theta}_{fin}) = \frac{2\sigma^2}{n} + \frac{2\sigma^2}{n_0} = 2\sigma^2\left(\frac{n + n_0}{n n_0}\right).$$

A.3 Probabilities in (3) for normally distributed data

$$\begin{aligned}
&P(\textit{early stop for efficacy}) = P(\hat{\theta}_{int} > \theta_{eff}) \\
&= 1 - \phi\left(\frac{\theta_{eff} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int} n_0}{n_{int} + n_0}}\right)
\end{aligned}$$

$$\begin{aligned}
&P(\textit{no early stop}) = P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}) \\
&= \phi\left(\frac{\theta_{eff} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int} n_0}{n_{int} + n_0}}\right) - \phi\left(\frac{\theta_{fut} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int} n_0}{n_{int} + n_0}}\right)
\end{aligned}$$

Let us define the bivariate normal cumulative distribution function of $(\hat{\theta}_{int}, \hat{\theta}_{fin})$ as $\phi_{biv}(\cdot, \cdot)$.

$$\begin{aligned}
&PoS = P(\hat{\theta}_{int} > \theta_{eff}) + P(\theta_{fut} \leq \hat{\theta}_{int} \leq \theta_{eff}, \hat{\theta}_{fin} > \theta_{suc}) \\
&= 1 - \left[\phi\left(\frac{\theta_{fut} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int} n_0}{n_{int} + n_0}}\right) + \phi_{biv}(\theta_{eff}, \theta_{suc}) - \phi_{biv}(\theta_{fut}, \theta_{suc})\right]
\end{aligned}$$

$$PoS_{post} = \frac{PoS - \ P(\textit{early stop for efficacy})}{P(\textit{no early stop})}$$
$$= 1 - \frac{\phi_{biv}(\theta_{eff}, \theta_{suc}) - \phi_{biv}(\theta_{fut}, \theta_{suc})}{\phi\left(\frac{\theta_{eff} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int}n_0}{n_{int}+n_0}}\right) - \phi\left(\frac{\theta_{fut} - \theta_0}{\sqrt{2}\sigma}\sqrt{\frac{n_{int}n_0}{n_{int}+n_0}}\right)}$$