Error Rates of Adaptive Power Priors

Master Thesis in Biostatistics (STA495)

by

Natalia Popova

18-720-623

supervised by

Prof. Dr. Leonhard Held

Department of Biostatistics

University of Zurich

Zurich, May 2020

Abstract

The appropriate use of historical information can lead to more ethical and powerful trial designs, however, with a subsequent risk of a type I error inflation.

This master thesis looks at the power prior Bayesian approach applied to a normally distributed outcome and explores operating characteristics and related rejection ratio. The critical factor, which was considered in this thesis, influencing power and type I error, is a bias between historical and current data. Three methods determine the influence of historical information in a power prior design: a conditional power prior, an empirical Bayesian and a fully Bayesian approaches.

We considered two common scenarios of borrowing information: one arm and two arm settings. In the one arm setting historical data is available for the treatment effect. In the two arm setting it is available for a control arm only. The influence of historical estimate on the posterior distribution of a treatment effect differs between these two scenarios. As a result, the dependence of operating characteristics on the bias between historical and current data is also different. In addition, we considered both types of historical estimates: a fixed and a randomly distributed around a true value.

Further, we inspected two posterior distribution critical values used in Bayesian testing. A critical value, which depends only on the data from the current experiment, can lead to a type I error inflation. Another critical value, which depends on historical data and the data from the current experiment, allows strict control of type I error, however with a cost of no power gain.

We concluded that integration of historical information could give a simultaneous gain in power and type I error under the three critical conditions. First, it can happen only in the two arm setting. Second, the bias between a historical and current experiment should be minimal. Third, the critical value of a decision rule should depend only on the data from the current experiment.

Acknowledgements

I want to thank Prof. Dr. Leonhard Held for a constant support and advice while writing this thesis, along with offering me such an exciting topic. I am also grateful to all members of the biostatistics master program for all the interesting and informative modules they offer. I have learned so much during this study, and I would always feel grateful for this opportunity.

My special thanks go to my fellow students Chiara Vanetta, Lucas Kook, Seraphina Kissling, Katrin Petermann, Uriah Daugaard, Xijin Chen, and Mei-Yee Ng, who made my studying time very fun.

I am immensely grateful to Dr. Eva Furrer, who made this wonderful journey possible for me and who shows constant support to all master students of this program.

Finally, I would like to thank my husband, Benjamin, who constantly supports and encourages me to be the better version of myself. I am unusually lucky to have such an amazing person by my side.

Natalia Popova
May 2020

Contents

Chapter 1

Introduction

Interest in incorporating historical data into an analysis of a new experiment in clinical trials is growing year by year. The expression "historical data" is used in this master thesis to refer to any extra information available outside of the current experiment. Usually, it is information from the previous studies, or can also be from a different sub-population. For example, it often happens that multiple randomized control trials (RCT) are available for a particular disease, where control arms are represented by identical "standards of care" treatment. All this information for a control arm can be combined and used as prior knowledge for a current experiment. Another example might be an implementation of adult sub-population data into an analysis of pediatric trials.

The appropriate use of historical data can lead to a reduction in sample size or improved operating characteristics. At the same time, disregard of extra information can cause the conduct of an unethical trial. For a highly effective drug, it might be unethical to have an equal number of patients in a control arm and a treated arm. With the integration of historical controls, the sample size of a control arm in a new experiment might be reduced, and more patients might be allocated to a treatment group.

However, one should be aware of possible bias between new and historical information, which can lead to the loss of all advantages of using historical data or even to a wrong conclusion from a new experiment. A combination of historical information with a randomized clinical trial is a very complex task, and currently, there is no uniform solution. Multiple ways were invented to adapt the use of historical information according to the bias, which are called dynamic borrowing approaches.

The frequentist approach *Test-then-pool* combines historical and randomized controls if the difference between groups is not significant. *Robust MAP* uses an extra parameter to adapt the variance of the prior depending on the bias Neuenschwander *et al.* (2010), Schmidli *et al.* (2014). In the *Commensurate prior* approach a parameter of interest from historical data is connected to a new experiment through a link function Hobbs *et al.* (2012). In *Power prior* extra parameter related to the likelihood of the historical information defines the influence of historical information Gravestock and Held (2017), Gravestock and Held (2018).

There are numerous papers that compare different methods. In the paper Viele *et al.* (2014) authors considered "dynamic" and "static" borrowing designs for a binary endpoint. They examined the influence of the bias on mean squared error, power, and type I error. They emphasized the importance of the analysis of determining where historical data is similar to the current experiment. In the paper Dejardin *et al.* (2018), authors also compared different dynamic borrowing approaches, such as test and pool, power prior, commensurate prior and robust mixture prior for a binary endpoint with a single historical dataset. They concluded that methods are comparable, but the robust mixture prior is the easiest to implement. Similarly, in the paper van Rosmalen *et al.* (2018) authors also investigated different methods for multiple historical studies for a time to event endpoint.

In my master thesis, we focused on the power prior approach and a single historical study. The following research questions were explored:

Question 1: Explore the influence of bias between historical and current experiments on different operating characteristics in the power prior setting for a normal outcome. A similar analysis was performed for a binary outcome in Gravestock and Held (2017).

There are different ways to integrate historical information, and we explored one arm and two arm scenarios. Historical data in the one arm case influence a treatment effect directly. In the two arm case, it also influences a treatment effect, but through a change in the posterior distribution of an outcome in a control arm (Fig. 1.1). Therefore, operating characteristics for these two scenarios change differently depending on the bias between historical and current experiments. The impact of the type of a historical estimate, fixed or random, was also investigated. First, we examined the situation where we have one fixed estimate from one historical study, and, second, we considered if estimates would be randomly normally distributed around a true value.

Question 2: Is it true that borrowing information cannot lead to an increased power while strictly controlling type I error? This question was inspired by the paper Kopp-Schneider *et al.* (2019), where the authors concluded based on the uniformly most powerful (UMP) test that no power gain is possible when strictly controlling type I error.

Power and type I error strongly depend on the chosen hypothesis test and related critical value of a decision rule. In Bayesian testing, we considered two different critical values. First, we examined the situation when the critical value is a function of current data. In this case, using historical information can cause type I error inflation. Second, we examined the situation when the critical value is a function of historical and current data, such that, it gives a fixed value of type I error. Further, we explored how these critical values influenced power of a related tests.

One arm and two arm setting have different structures of hierarchical models and related parameters. Related mathematical derivations were divided into two separate chapters (chapter 2 and chapter 3). Derivations for a random effect of historical data were added after each derivation for a fixed case. Three main ways to deal with a nuisance power parameter in Bayesian setting were considered: fixed values of a parameter, an empirical Bayesian and a fully Bayesian approaches for each setting. The main function of interest was H_0 rejection rate given effect. It was derived mathematically or using simulation techniques for each scenario. Resulting type I error, power, and rejection ratio plots are shown in chapter 4.1. The UMP and Bayesian testing are considered in chapter 5. The conclusion for each research question is given in chapter 6.

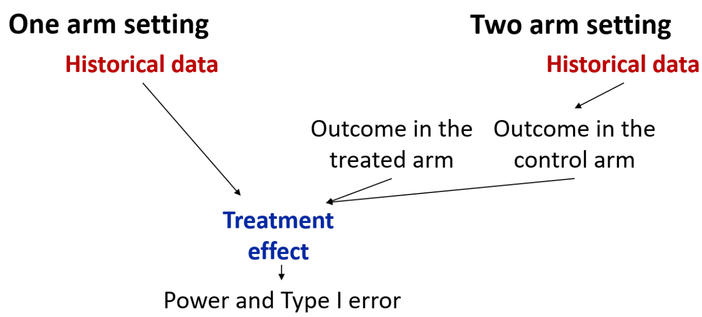

Figure 1.1: One arm and two arm setting

Chapter 2

One arm setting

In this chapter, we were looking at the scenario when historical information is available directly for a treatment effect. For example, when we have an estimate of a treatment effect from an adult trial, and we would like to use it as prior information for our new experiment for pediatric patients. The outcome estimate, we are referring to in this chapter, is the difference between sample means of a normally distributed relevant measure in a treatment group and a control group from related experiments. The relevant measure is a patient-level biological measure of the disease progression, for example, blood pressure, weight, or others.

In this scenario, the next parameters influence H_0 rejection rate: true treatment effect in historical and current experiments, a sample size in a historical trial, a sample size of a new experiment, the bias between historical and current data, and variance of the relevant measure.

Since we are interested in the influence of bias on operating characteristics, we made some simplifying assumptions for other parameters. We set identical variances of the outcome measure and identical sample sizes in historical and current experiments.

- Historical data: $Y_0 \sim \mathrm{N}\left(\Delta_0, \frac{\sigma^2}{m_0}\right)$, where y_0 is effect estimate, m_0 is the parameter related to the number of patients and Δ_0 is a true treatment effect from historical experiment

- New data: $Y_\star \sim \mathrm{N}\left(\Delta_\star, \frac{\sigma^2}{m_\star}\right)$, where $y_\star$ is effect estimate, $m_\star$ is the parameter related to the number of patients and $\Delta_\star$ is a true treatment effect from a new experiment

- For our simulation study we used $m_0 = 50$, $m_\star = 50$ and variance as $\sigma = 1$

- With mentioned above sample sizes we reach a power of 80% with a significance level of 5% for an one-sided test when the treatment effect under H_1 is 0.35 in RCT setting

- We varied $\Delta_\star$ from -0.2 to 0.45 to see the influence of the treatment effect on H_0 rejection rate

- We varied $y_0 =$ from -0.2 to 0.44 to see the influence of bias on power and type I error

The difference between estimates y_0 and $y_\star$ is the measure for the bias between historical and current experiments. In the situation, when historical estimate is fixed, the bias depends only on one random variable: $y_0 - Y_\star \sim \mathrm{N}(y_0 - \Delta_\star, \frac{\sigma^2}{n_\star})$. When random historical estimate is random, the bias estimate depends on two random variables: $Y_0 - Y_\star \sim \mathrm{N}(\Delta_0 - \Delta_\star, \frac{\sigma^2}{n_\star} + \frac{\sigma^2}{n_0})$.

2.1 Power parameter in a power prior

We use power prior to include historical data in the analysis:

$$p(\Delta \mid y_0) \propto L(y_0 \mid \Delta)^{\delta} p(\Delta) \tag{2.1}$$

$$\Delta \mid y_0, \delta \sim \mathrm{N}\left(y_0, \frac{\sigma^2}{\delta m_0}\right), \tag{2.2}$$

where Δ is the parameter of interest and δ is a nuisance power parameter, which determines the weight of the prior information.

For further analysis of the operating characteristics, we need to compute the marginal posterior of the treatment effect $p(\Delta \mid y_0, y_\star)$. There are three ways to deal with the nuisance power parameter δ: to fix δ based on an available assumption (a conditional power prior), the empirical Bayes and full Bayes approaches. The posterior distributions for the treatment effect, which are derived by using each method and related H_0 rejection rates, are summarized in the Table 2.1.

2.1.1 Conditional power prior

In the conditional power prior approach, δ is considered as a fixed value, determined by an existing belief about the compatibility of the historical and the new experiments. In our analysis, we considered the next δ: 0, 0.2, 0.4, 0.6, 0.8, 1 for estimation of operation characteristics.

The posterior distribution of the treatment effect Δ, which combines the historical power prior with the current likelihood is:

$$p\left(\Delta \mid y_0, y_\star\right) \propto p(\Delta \mid y_0, \delta) L(y_\star \mid \Delta) \tag{2.3}$$

$$p(\Delta \mid y_0, y_*) \propto p(\Delta)\left(\frac{1}{\sqrt{2\pi\sigma^2/m_0}} \exp\left(-\frac{m_0}{2}\frac{(\Delta - y_0)^2}{\sigma^2}\right)\right)^{\delta}\left(\frac{1}{\sqrt{2\pi\sigma^2/m_\star}} \exp\left(-\frac{m_\star}{2}\frac{(y_\star - \Delta)^2}{\sigma^2}\right)\right)$$

Using quadratic formula available in the App. 1.

$$p(\Delta \mid y_0, y_\star, \delta) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\delta m_0 + m_\star} \exp\left(-\frac{1}{2\sigma^2}\left(\delta m_0 + m_\star\right)\left(\Delta - \frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}\right)^2\right)$$

Thus, in the case of a fixed power parameter, the posterior distribution of a treatment effect for the one arm case has a normal shape:

$$\Delta \mid y_0, y_\star, \delta \sim \mathrm{N}\left(\frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}, \frac{\sigma^2}{\delta m_0 + m_\star}\right) \tag{2.4}$$

2.1.2 Empirical Bayes approach

In this approach, the nuisance power parameter δ is estimated based on the data, using maximum likelihood estimate (MLE) of the marginal likelihood of $L(y_0, y_\star \mid \delta)$.

$$\hat{\delta}(y_o, y_*) = \underset{\delta \in [0,1]}{\operatorname{argmax}} L(y_\star, y_0 \mid \delta) \tag{2.5}$$

By the law of total probability:

$$\begin{aligned}
L(y_\star \mid \delta, y_0) &= \int L(y_\star \mid \Delta) p(\Delta \mid \delta, y_0) d\Delta \\
&= \int \mathrm{N}\left(y_\star \mid \Delta, \frac{\sigma^2}{m_\star}\right) \mathrm{N}\left(\Delta \mid y_0, \frac{\sigma^2}{\delta m_0}\right) d\Delta \\
&= \frac{\sqrt{m_\star \delta m_0}}{2\pi\sigma^2} \int \exp\left(-\frac{1}{2\sigma^2}\left(m_\star (y_\star - \Delta)^2 + \delta m_0 (\Delta - y_0)^2\right)\right) d\Delta \\
&\propto \int \exp\left(-\frac{1}{2\sigma^2}(m_\star + \delta m_0)\left(\Delta - \frac{m_\star y_\star + \delta m_0 y_0}{m_\star + \delta m_0}\right)\right) \exp\left(-\frac{1}{2\sigma^2}\frac{\delta m_0 m_\star}{m_\star + \delta m_0}(y_\star - y_0)^2\right) d\Delta
\end{aligned}$$

Since this has the form of a normal likelihood, invariance property of the maximal likelihood estimate $(y_\star - y_0)^2$ can be applied.

$$\sigma^2 \left(\frac{1}{\delta m_0} + \frac{1}{m_\star}\right) = (y_\star - y_0)^2 \tag{2.6}$$

$$\hat{\delta} = \frac{1}{m_0} \cdot \frac{\sigma^2 m_\star}{m_\star (y_\star - y_0)^2 - \sigma^2} \tag{2.7}$$

We need to respect the condition that $0 \leq \delta \leq 1$, what leads to $(y_\star - y_0)^2 > \frac{\sigma^2}{m_\star}$ and when this condition is not met $\hat{\delta} = 1$.

```
fun.delta.EB <- function(x_star, x_0, sigma, n_star, n_0) {
  if ( (x_star - x_0)^2 >=  sigma^2/n_star + sigma^2/n_0) {
    delta.hat <- (1/n_0) * (sigma^2) * n_star / (n_star*(x_star - x_0)^2 - sigma
   ^2)
  }  else {delta.hat <- 1}
  return(delta.hat)
}
```

Thus, the bigger the difference between the historical and the current estimate $(y_\star - y_0)^2$, the smaller the estimate of the power parameter $\hat{\delta}$, and therefore the smaller the influence of historical data. The left plot of the Figure 2.1 shows the variability of estimation of $\hat{\delta}$ depending on the true value of effect $\Delta_\star$, when y_0 is fixed and equal to 0. The higher the value of $\Delta_\star$, the higher the related simulated value $y_\star \sim \mathrm{N}\left(\Delta_\star, \frac{\sigma^2}{m_\star}\right)$, and therefore, the bias is higher too. The right plot shows an average over multiple simulations.

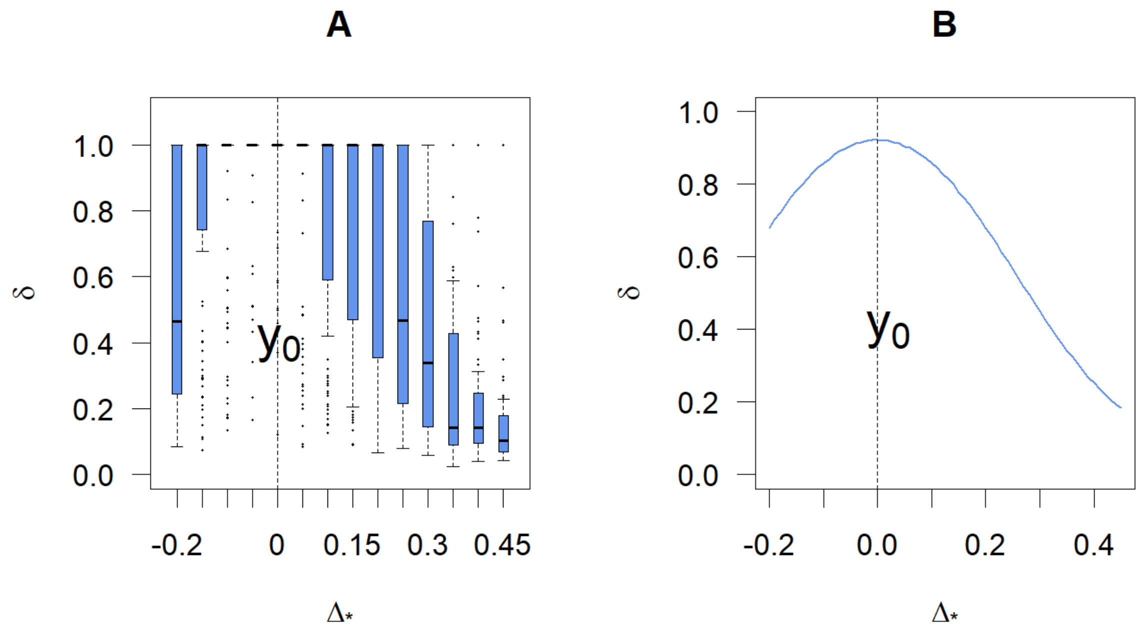

Figure 2.1: A power parameter $\hat{\delta}$ based on the empirical Bayes approach. A) Box plots for 100 simulations of $y_\star$ for each value of $\Delta_\star$ and fixed $y_0 = 0$ B) $\hat{\delta}$ averaged over multiple simulations

If we substitute the derived estimate of $\hat{\delta}$ to the posterior distribution for fixed δ, we obtain the next posterior distribution for treatment effect using the empirical Bayes approach:

$$\Delta \,|\, y_\star, y_0, \hat{\delta} = \begin{cases} \mathrm{N}\left(\frac{\sigma^2}{m_\star(y_0-y_\star)} + y_\star, \frac{\sigma^2}{m_\star}\left(1 - \frac{\sigma^2}{m_\star(y_\star-y_0)^2}\right)\right) & \text{if } (y_\star - y_0)^2 > \frac{\sigma^2}{m_\star} \\ \mathrm{N}\left(\frac{m_0 y_0 + m_\star y_\star}{m_0+m_\star}, \frac{\sigma^2}{m_0+m_\star}\right) & \text{if } (y_\star - y_0)^2 \leq \frac{\sigma^2}{m_\star} \end{cases} \tag{2.8}$$

2.1.3 Normalized power prior

In the full Bayes approach, we treat δ as an extra parameter with its own prior distribution: $\delta \sim \mathrm{Be}(\alpha, \beta)$, where α, β are hyper-parameters. For our analysis we have used $\alpha = 0.5$ and $\beta = 0.5$.

The joint power prior based on the definition of the conditional density is:

$$p(\Delta, \delta \,|\, y_0) = p(\Delta \,|\, y_0, \delta) p(\delta \,|\, y_0) \tag{2.9}$$

where $p(\delta \,|\, y_0) = p(\delta)$ since a prior distribution $\delta \sim \mathrm{Be}(\alpha, \beta)$.
Using Bayesian theorem we can write the the next formula for $p(\Delta \,|\, y_0, \delta)$:

$$p(\Delta \,|\, y_0, \delta) = \frac{L(y_0 \,|\, \Delta)^\delta p(\Delta)}{\int L(y_0 \,|\, \Delta)^\delta p(\Delta) d\Delta}$$

Thus, considering $p(\Delta)$ as non-informative the joint posterior is:

$$
\begin{aligned}
p(\Delta, \delta \,|\, y_0, y_\star) &\propto L(y_\star \,|\, \Delta) p(\Delta, \delta \,|\, y_0) \\
p(\Delta, \delta \,|\, y_0, y_\star) &\propto L(y_\star \,|\, \Delta) p(\Delta \,|\, y_0, \delta) p(\delta) \\
p(\Delta, \delta \,|\, y_0, y_\star) &\propto L(y_\star \,|\, \Delta) L(y_0 \,|\, \Delta)^\delta p(\Delta) p(\delta) \\
p(\Delta, \delta \,|\, y_0, y_\star) &\propto L(y_\star \,|\, \Delta) L(y_0 \,|\, \Delta)^\delta p(\delta) \\
p(\Delta, \delta \,|\, y_0, y_\star) &\propto \mathrm{N}_\Delta\left(y_0, \frac{\sigma^2}{\delta m_0}\right) \mathrm{N}_\Delta\left(y_\star, \frac{\sigma^2}{m_\star}\right) \mathrm{Be}(\alpha, \beta)
\end{aligned}
$$

and marginal posterior for Δ is:

$$
p(\Delta \,|\, y_0, y_\star) \propto \int_0^1 p(\Delta, \delta \,|\, y_0, y_\star) d\delta \propto \int_0^1 L(y_\star \,|\, \Delta) L(y_0 \,|\, \Delta)^\delta p(\delta) d\delta
$$

Thus, the marginal posterior for the treatment effect is $p(\Delta \,|\, y_0, y_\star)$ in this case is:

$$
p(\Delta \,|\, y_0, y_\star) \propto \int_0^1 \mathrm{N}_\Delta\left(y_0, \frac{\sigma^2}{\delta m_0}\right) \mathrm{N}_\Delta\left(y_\star, \frac{\sigma^2}{m_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\delta \tag{2.10}
$$

```
poster.control.FB <- function (theta, x_0, x_star, alpha, beta, n_star,
                                 n_0, sigma) {
  # d - delta(power parameter) over what we integrate
  # y - theta(effect) parameter of our interest
  dens<- function (d,y) {
    dnorm(y, mean = x_0,sd = sqrt(sigma^2/(n_0*d)))*dnorm(y, mean = x_star,sd =
   sqrt(sigma^2/(n_star)))*dbeta(d, alpha, beta)}
  return(sapply(theta, function(z){integrate(dens,0,1, y = z)$value}))
  #as return we have function (theta)
  #for each theta we integrate over delta
}
```

And marginal posterior for the power parameter $p(\delta \,|\, y_0, y_\star)$ is

$$
p(\delta \,|\, y_0, y_\star) \propto \int_\infty^\infty \mathrm{N}_\Delta\left(y_0, \frac{\sigma^2}{\delta m_0}\right) \mathrm{N}_\Delta\left(y_\star, \frac{\sigma^2}{m_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\Delta \tag{2.11}
$$

```
#Marginal for delta one arm
margin.poster.FB.delta <- function (delta, x_0, x_star,
                                     alpha, beta, n_star,
                                     n_0, sigma) {
  # d - delta(power parameter)
  # y - theta(effect) parameter of our interest
  dens<- function (d,y) {
    dnorm(y, mean = x_0,sd = sqrt(sigma^2/(n_0*d)))*dnorm(y, mean = x_star,sd =
   sqrt(sigma^2/(n_star)))*dbeta(d, alpha, beta)}
  return(sapply(delta, function(z){integrate(dens,-Inf,Inf, d = z)$value}))
  #as return we have function (delta)
  #for each delta we integrate over theta/y
}
```

The marginal distribution for $p(\delta \,|\, y_0, y_\star)$ depends on the difference between historical and new experiment estimates $(y_0 - y_\star)^2$, thus the full Bayes approach is also an adaptable, dynamic borrowing approach. Figure 2.2 shows the marginal posterior distribution for the treatment effect for large (A) and small (B) bias values. When the bias is large, the marginal posterior distribution has a bigger variance and is shifted towards a new experiment.

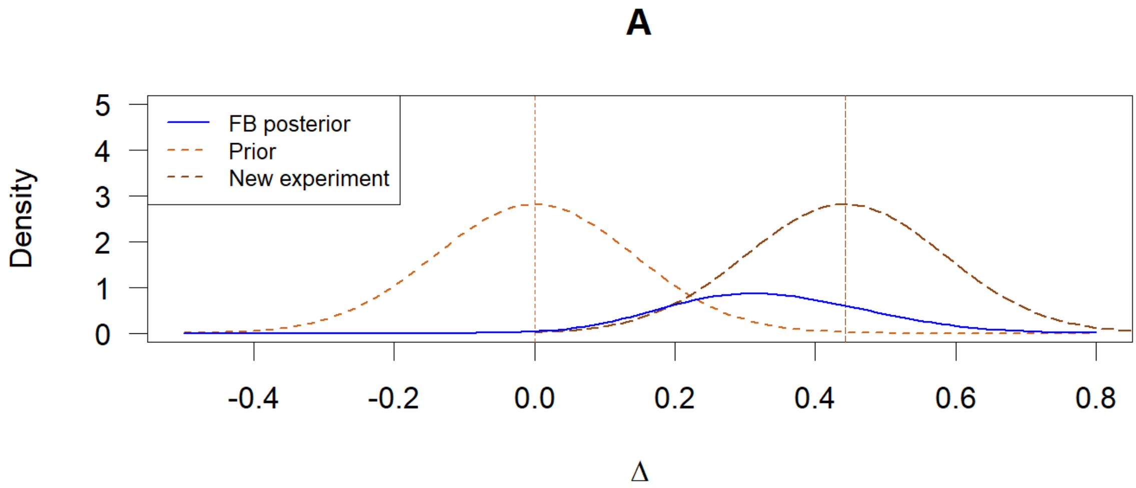

Figure 2.2: The marginal posterior distribution for the treatment effect under using the full Bayes approach: A) With a large bias B) With a small bias

Adaptability, according to the bias, can also be seen on the marginal posterior of the power parameter $p(\delta \,|\, y_0, y_\star)$. Figure 2.3 shows the prior distribution for $\delta \sim \mathrm{Be}(0.5, 0.5)$ and how the shape of the marginal posterior distribution is changed according to the difference between historical and new experiment estimates $(y_0 - y_\star)^2$. When the difference is large, higher values of the density are concentrated in the region $0 < \delta < 0.5$.

Figure 2.4 shows the change in the median point of the marginal posterior distribution for the power parameter δ according to the bias. The higher the bias, the bigger the shift of a median point towards lower values.

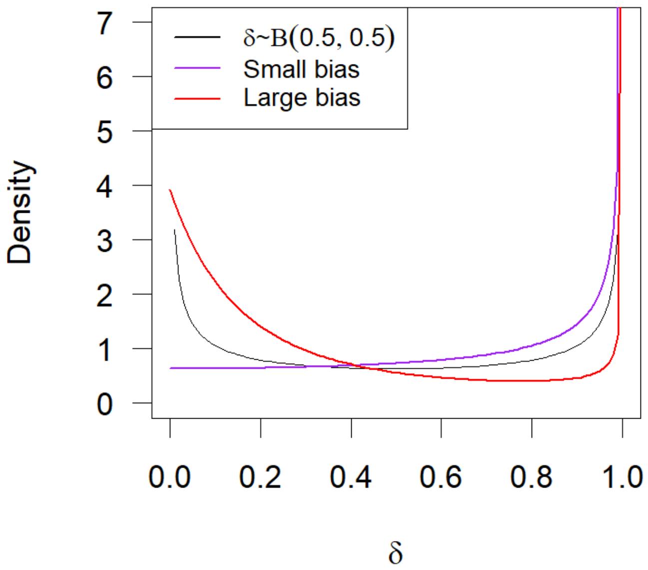

Figure 2.3: The marginal posterior distribution for the power parameter δ using the full Bayes approach under a large (red line) and a small (pink line) bias

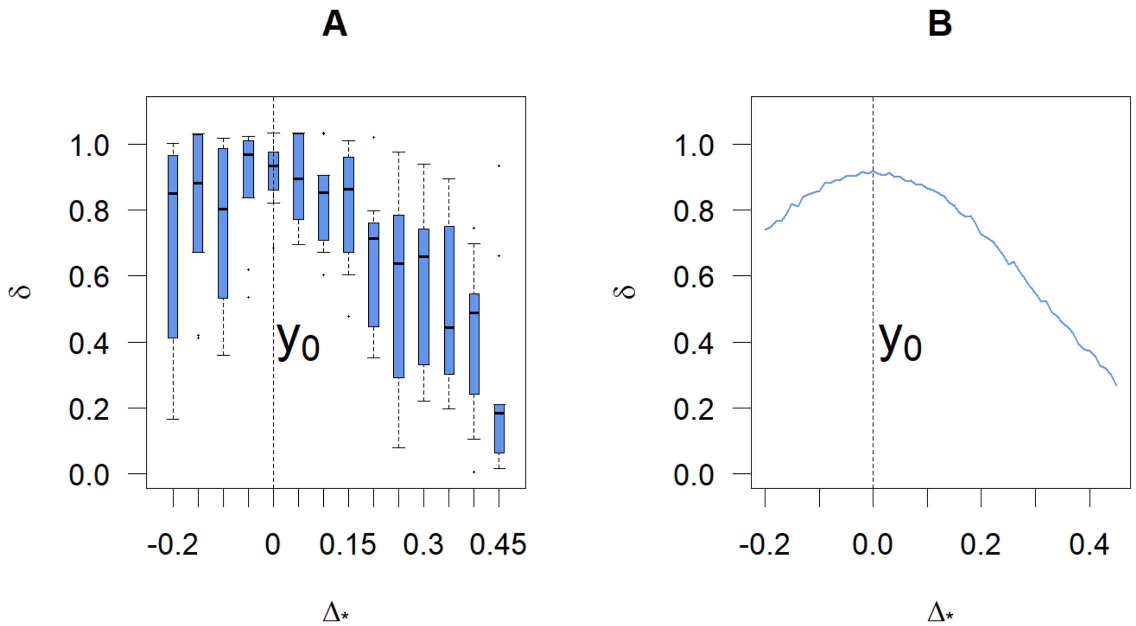

Figure 2.4: Median estimates for the marginal posterior distribution of δ. A) Boxplots for 100 simulations for different values of $\Delta_\star$ and fixed $y_0 = 0$ B) Averaged over multiple simulations

Posterior distributions of a treatment effect for one simulation of $y_\star$

Figure 2.5 shows the posterior distributions of a treatment effect using different approaches for one simulation of $y_\star$ under $\Delta = \Delta_{H_1}$ and a fixed value $y_0 = 0$. Different shadows of grey lines demonstrate how the posterior distribution shifts towards the likelihood of a new experiment when δ is getting smaller. The red line is for posterior distribution with the power parameter estimated with the empirical Bayes approach. For this simulation, the difference between y_0 and $y_\star$ is large and consequently, a the value of $\hat{\delta}$ is small. The blue line shows the posterior under the full Bayes approach, which is located between the prior distribution and the current likelihood with a higher variance also due to a high difference between historical and new experiment estimates.

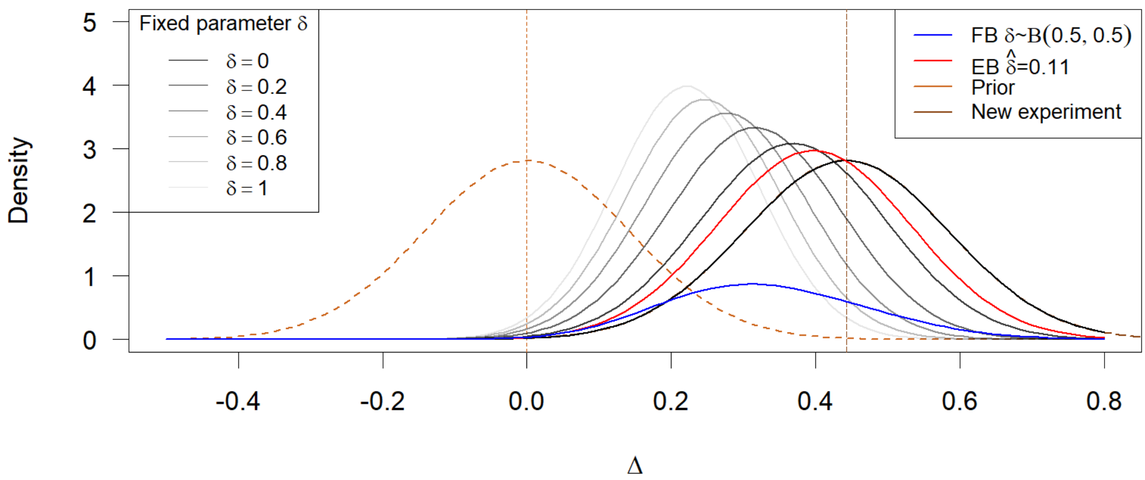

Figure 2.5: The posterior distributions of a treatment effect for one simulation of $y_\star$ under $\Delta = \Delta_{H_1}$ and a fixed value $y_0 = 0$

2.2 H_0 rejection rate

For estimation of operating characteristics we need to derive the H_0 rejection rate conditional on a treatment effect Δ. We test the following hypotheses $H_0 : \Delta \leq 0$ and $H_1 : \Delta > 0$.

Even though we apply a Bayesian approach, the frequentist's operating characteristics are required to answer our research questions: power and type I error. Hence Δ_{H_0} and Δ_{H_1} should be established. Based on the established hypotheses Δ_{H_0} is 0. For determining Δ_{H_1} we fixed the target power of 80% with significance level of 5%. Thus, for the one-sided test in RCT setting Δ_{H_1} is 0.35 conditional on $m_0 = 50$, $m_\star = 50$ and variance $\sigma = 1$.

- **Type I error**: H_0 rejection rate (how often we reject H_0) when H_0 is true.
  In a Bayesian setting how often $\mathsf{Pr}(\Delta > 0 \,|\, \text{data}) > 1 - \alpha$ or $\mathsf{Pr}(\Delta \leq 0 \,|\, \text{data}) \leq \alpha$ under the condition $\Delta = \Delta_{H_0}$
- **Power**: H_0 rejection rate (how often we reject H_0) when H_1 is true.
  In a Bayesian setting how often $\mathsf{Pr}(\Delta > 0 \,|\, \text{data}) > 1 - \alpha$ or $\mathsf{Pr}(\Delta \leq 0 \,|\, \text{data}) \leq \alpha$ under the condition $\Delta = \Delta_{H_1}$

where we fixed the significance level α as 5% for our analysis.

2.2.1 Classical approach

Power estimation in classical RCT setting was taken from Spiegelhalter *et al.* (2004). Under classical analysis, H_0 is rejected if a sample value $Y_\star$ is below a specific critical value.

$$Y_\star > -z_\alpha \sigma \sqrt{\frac{1}{m_\star}}, \tag{2.12}$$

where z_α is quantile of a prescpecified significance level (Type I error).

$$\begin{aligned}
\mathsf{Pr}_C(\text{reject } H_0 \,|\, \Delta) &= \mathsf{Pr}\left(Y_\star > -z_\alpha \sigma \sqrt{\frac{1}{m_\star}} \,\middle|\, \Delta\right) \\
&= 1 - \Pr\left(Y_\star \leq -\sqrt{\frac{1}{m_\star}} z_\alpha \sigma \,\middle|\, \Delta\right) \\
&= \Phi\left[\frac{\Delta}{\sigma}\sqrt{m_\star} + z_\alpha\right]
\end{aligned} \tag{2.13}$$

```
#Conditional classical power for 1 arm
one.arm.cond.class.power <- function (n_new, ze, effect, sigma){
  out <- pnorm(effect*sqrt(n_new)/sigma + ze)
  return (out)
}
```

2.2.2 Bayesian approach

In a Bayesian approach, the goal is to incorporate prior (historical) information to make a better decision. We are interested in how often we reject H_0 based on a Bayesian decision rule when testing the null hypothesis : $\Delta \leq 0$ against an alternative $H_A : \Delta > 0$. In the Bayesian setting, the decision depends on the posterior distribution of the parameter of interest, which is a treatment effect inferred from historical and current experiments $p(\Delta \mid y_\star, y_0)$. The resulting H_0 rejection rate and related operating characteristics strongly depend on the approach chosen for the derivation of the posterior distribution of the treatment effect, which we considered in section 2.1.

Fixed power parameter

The derivation of the H_0 rejection rate for a normal prior in the one arm case was shown in Spiegelhalter *et al.* (2004). We adapted this derivation for a power prior with a fixed power parameter.

As it was derived above (2.4) the posterior distribution of the treatment effect is in conditional power prior setting:

$$\Delta \mid y_\star, y_0, \delta \sim \mathrm{N}\left(\frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}, \frac{\sigma^2}{\delta m_0 + m_\star}\right)$$

In order to estimate the H_0 rejection rate, we need to see how often $\mathsf{Pr}(\Delta \leq 0 \mid \text{data}) \leq \alpha$, where by data we mean a historical outcome estimate y_0 and a current experiment estimate $y_\star$ or $\mathsf{Pr}(\Delta > 0 \mid \text{data}) > 1 - \alpha$.

$$\begin{aligned} \mathsf{Pr}(\Delta \leq 0 \mid y_0, y_\star, \delta) &\leq \alpha \\ \Phi\left[\frac{-\frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}}{\sigma\sqrt{1/(\delta m_0 + m_\star)}}\right] &\leq \alpha \\ \frac{-\frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}}{\sigma\sqrt{1/(\delta m_0 + m_\star)}} &\leq z_\alpha \end{aligned}$$

Thus, the next condition for $Y_\star$ can be derived, which causes a significant result, namely H_0 rejection:

$$Y_\star \geq \frac{1}{m_\star}\left(-z_\alpha \sigma \sqrt{\delta m_0 + m_\star} - \delta m_0 y_0\right) \tag{2.14}$$

Since $Y_\star \sim \mathrm{N}\left(\Delta_\star, \frac{\sigma^2}{m_\star}\right)$ we can derive the conditional H_0 rejection rate for a fixed historical outcome estimate y_0.

$$\begin{aligned} \mathsf{Pr}_B(\text{reject} H_0 \mid \Delta_\star) &= 1 - \Phi\left[\frac{\frac{1}{m_\star}\left(-z_\alpha \sigma \sqrt{\delta m_0 + m_\star} - \delta m_0 y_0\right) - \Delta_\star}{\sigma/\sqrt{m_\star}}\right] \\ &= \Phi\left[\frac{\frac{1}{m_\star}\left(z_\alpha \sigma \sqrt{\delta m_0 + m_\star} + \delta m_0 y_0\right) + \Delta_\star}{\sigma/\sqrt{m_\star}}\right] \end{aligned} \tag{2.15}$$

```
#Conditional bayesian power for 1 arm
one.arm.cond.bayes.power <- function (delta, n_0, y_0, n_new, ze, effect, sigma)
  {
  out <- pnorm(sqrt(n_new) * effect / sigma + delta * n_0 * y_0 / (sigma * sqrt(
    n_new)) +sqrt((delta * n_0  + n_new) / n_new) * ze
  )
  return (out)
}
```

Using this formula (2.15) we estimated the H_0 rejection rate depending on the true treatment effect in a new experiment $\Delta_\star$, the bias, the value of y_0, and the value of fixed δ.

```
Pr.B.rej.H0.given.effect.f1 <-
  function (x, y, z) {
    one.arm.cond.bayes.power(
      delta = z,
      n_0 = m_0,
      y_0 = y,
      n_new = m_star,
      ze = qnorm(0.05),
      effect = x,
      sigma = sd_0
    )
  }
one.arm.H0.rej.cond.bayes <-
  array(do.call(Pr.B.rej.H0.given.effect.f1, c(expand.grid(
    list(x = effect_star, y = y_0, z = deltas.fixed)
  ))),
  dim = c(length(effect_star), length(y_0), length(deltas.fixed)))
```

In the case when y_0 is a random variable, instead of one fixed value y_0, we explore what happens if $Y_0 \sim \mathrm{N}\left(\Delta_0, \frac{\sigma^2}{m_0}\right)$. Similar to a fixed scenario, different Δ_0 are considered to be able to see the influence of the bias $\Delta_0 - \Delta_\star$. An estimate of the bias for each simulation is $y_0 - y_\star$.

$$\mathsf{Pr}_B(\text{reject}H_0 \,|\, \Delta) = \int \Phi\left[\frac{\frac{1}{m_\star}\left(z_\alpha\sigma\sqrt{\delta m_0 + m_\star} + \delta m_0 y_0\right) + \Delta}{\sigma/\sqrt{m_\star}}\right]\phi_{y_0}\left(y_0, \frac{\sigma}{\sqrt{m_0}}\right)dy_0 \qquad (2.16)$$

We estimated H_0 rejection rate depending on the true treatment effect in a new experiment $\Delta_\star$ for each value of Δ_0 and fixed δ, using the formula (2.16),

```
#we estimate P(rejec H0) for a range of y0, effects and fixed delta values
Pr.B.rej.H0.given.effect.f1.y0rv <-
  function (x, y, z) {
    Pow.y_0 <- function(y0) {
      one.arm.cond.bayes.power(
        y_0 = y0,
        delta = z,
        n_0 = m_0,
        n_new = m_star,
        ze = qnorm(0.05),
        effect = x,
        sigma = sd_0
      ) * dnorm(y0, mean = y, sd = sd_0 / sqrt(m_0))
    }
    out <-
      integrate(Vectorize(Pow.y_0), -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
par <- expand.grid(x = effect_star, y = y_0, z = deltas.fixed)
one.arm.H0.rej.cond.bayes.y0rv <- array(
  mapply(Pr.B.rej.H0.given.effect.f1.y0rv, par$x, par$y, par$z),
  dim = c(length(effect_star), length(y_0), length(deltas.fixed))
)
```

The resulting plots, which show H_0 rejection rate and related operating characteristics, can be found in chapter 4.1.

Empirical Bayes

Similarly to the conditional power prior case, we need to find which $Y_\star$ leads to H_0 rejection : $\mathsf{Pr}(\Delta \leq 0 \,|\, y_\star, y_0) \leq \alpha$. First, we consider the situation when $(y_\star - y_0)^2 > \frac{\sigma^2}{m_\star}$:

$$
\begin{aligned}
\mathsf{Pr}(\Delta \leq 0 \,|\, \text{data}) &\leq \alpha \\
\Phi\left[\frac{0 - \frac{\sigma^2}{m_\star(y_0 - y_\star)} - y_\star}{\sqrt{\frac{\sigma^2}{m_\star}\left(1 - \frac{\sigma^2}{m_\star(y_\star - y_0)^2}\right)}}\right] &\leq \alpha \\
\frac{\frac{\sigma^2}{m_\star(y_0 - y_\star)} + y_\star}{\sqrt{\frac{\sigma^2}{m_\star}\left(1 - \frac{\sigma^2}{m_\star(y_\star - y_0)^2}\right)}} &\leq z_\alpha \\
\frac{\sigma^2 m_\star + y_\star m_\star^2 (y_0 - y_\star)}{\sqrt{\sigma^2 m_\star (y_0 - y_\star)^2 - \sigma^4}} &\leq z_\alpha
\end{aligned}
$$

To have a closed-form solution for the H_0 rejection rate, we need to solve this inequality in terms of $y_\star$. Since there is no simple way to do that, we applied simulation techniques: simulations for $y_\star$ with the following estimation of $\hat{\delta}$ and averaging over simulations to estimate H_0 rejection rate.

Similarly to the conditional power prior case, we considered different values for the true treatment effect in a new experiment $\Delta_\star$ and different values of historical estimate y_0. The bias for this scenario is $y_0 - \Delta_\star$, with a bias estimate, which depends only on one random variable, $y_0 - y_\star$.

```r
for (n.sim in c(1:sim.N)) {
  print(n.sim)
  for (n.effect_star in c(1:length(effect_star))) {
    #1.simulate y_star
    y_star_sim <-
      rnorm(1, effect_star[n.effect_star], sd_0 / sqrt(m_star))
    for (n.y_0 in c(1:length(y_0))) {
      #2.estimate delta.hat
      delta.hat <- fun.delta.EB(
        x_star = y_star_sim,
        x_0 = y_0[n.y_0],
        sigma = sd_0,
        n_star = m_star,
        n_0 = m_0
      )
      one.arm.delta.EB[n.sim, n.effect_star, n.y_0] <- delta.hat
      #3.Derive mean and variance of EB posterior of p(effect)
      post.mean.fixed <-
        (delta.hat * m_0 * y_0[n.y_0] + m_star * y_star_sim) / (delta.hat * m_0
    + m_star)
      post.sd.fixed <- sd_0 / sqrt(delta.hat * m_0 + m_star)
      #P(Delta>0)
      One.arm.P.EB[n.sim, n.effect_star, n.y_0] <-
        1 - pnorm(0, mean = post.mean.fixed, sd = post.sd.fixed)
    }
  }
}
```

```
#4.Estimate H0 rejection rate
One.arm.H0.rej.rate.EB <-
  array(apply(One.arm.P.EB > 0.95, c(2,3), FUN = mean), dim = c(length(effect_
    star),length(y_0)))
```

In the case when y_0 is random, an additional simulation of y_0 is required. y_0 was simulated from $\mathrm{N}\left(\Delta_0, \frac{\sigma^2}{m_0}\right)$.

```
for (n.sim in c(1:sim.N)) {
  print(n.sim)
  for (n.effect_star in c(1:length(effect_star))) {
    for (n.y_0 in c(1:length(y_0))) {
      #1.simulate y_star and y_0
      y_star_sim <-
        rnorm(1, effect_star[n.effect_star], sd_0 / sqrt(m_star))
      y_0_sim <- rnorm(1, y_0[n.y_0], sd_0 / sqrt(m_0))
      #2.estimate delta.hat
      delta.hat <- fun.delta.EB(
        x_star = y_star_sim,
        x_0 = y_0_sim,
        sigma = sd_0,
        n_star = m_star,
        n_0 = m_0
      )
      one.arm.delta.EB.y0rv[n.sim, n.effect_star, n.y_0] <-
        delta.hat
      #3.Derive mean and variance of EB posterior of p(effect)
      post.mean.fixed <-
        (delta.hat * m_0 * y_0_sim + m_star * y_star_sim) / (delta.hat * m_0 + m
    _star)
      post.sd.fixed <- sd_0 / sqrt(delta.hat * m_0 + m_star)
      #P(delta>0)
      One.arm.P.EB.y0rv[n.sim, n.effect_star, n.y_0] <- +
        1-pnorm(0, mean = post.mean.fixed, sd = post.sd.fixed)
    }
  }
}
#4.Estimate H0 rejection rate
One.arm.H0.rej.rate.EB.y0rv <-
  array(apply(One.arm.P.EB.y0rv > 0.95, c(2,3), FUN = mean),
        dim = c(length(effect_star),length(y_0)))
```

Full Bayes

As it was derived above (2.11), the posterior distribution for a treatment effect for the full Bayes approach:

$$p(\Delta \,|\, y_0, y_\star) \propto \int_0^1 \mathrm{N}_\Delta\left(y_0, \frac{\sigma^2}{\delta m_0}\right) \mathrm{N}_\Delta\left(y_\star, \frac{\sigma^2}{m_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\delta$$

For calculation of H_0 rejection rate, we simulated $y_\star$ with the following derivation of the marginal posterior distribution of the treatment effect $p(\Delta \,|\, y_0, y_\star)$. Further, we calculated how often $p(\Delta > 0 \,|\, y_0, y_\star) > 1 - \alpha$.

```
foreach(n.sim = c(1:sim.N)) %do% {
  print(n.sim)
```

```
  one.arm.Pr.rej.H0 <- function(x, y) {
    #1.simulate y_star
    y_star <-
      rnorm(1, x, sd_0 / sqrt(m_star))
    #2. Derive marginal FB posterior(effect)
    poster.effect.FB <-
      function (eff) {
        poster.control.FB (
          theta = eff,
          x_0 = y,
          x_star = y_star,
          alpha = 0.5,
          beta = 0.5,
          n_star = m_star,
          n_0 = m_0,
          sigma = sd_0
        )
      }
    out <-
      integrate(Vectorize(poster.effect.FB), 0, Inf, abs.tol = 0)$value/
   integrate(Vectorize(poster.effect.FB), -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
  #3.P(delta>0)
  one.arm.P.FB[n.sim, ,] <-
    (outer(effect_star2, y_0, FUN = Vectorize(one.arm.Pr.rej.H0)))
}
 #4.H0 rejection rate: how often P(delta>0) > 95%
one.arm.rej.H0.FB <-
  array(apply(one.arm.P.FB > 0.95, c(2,3), FUN = mean, na.rm = TRUE),
        dim = c(length(effect_star2),length(y_0)))
```

For random historical data we also simulated different values of $y_0 \sim \mathrm{N}\left(\Delta_0, \frac{\sigma^2}{m_0}\right)$.

```
foreach(n.sim = c(1:sim.N)) %do% {
  print(n.sim)
  one.arm.Pr.rej.H0.y0rv <- function(x, y) {
    #1.simulate y_star and y_0
    y_star <-
      rnorm(1, x, sd_0 / sqrt(m_star))
    y_0_sim <-
      rnorm(1, y, sd_0 / sqrt(m_0))
    #2. Derive marginal FB posterior(effect)
    poster.effect.FB.y0rv <-
      function (eff) {
        poster.control.FB (
          theta = eff,
          x_0 = y_0_sim,
          x_star = y_star,
          alpha = 0.5,
          beta = 0.5,
          n_star = m_star,
          n_0 = m_0,
          sigma = sd_0
        )
      }
    out <-
      integrate(Vectorize(poster.effect.FB.y0rv), 0, Inf, abs.tol = 0)$value/
```

```
    integrate(Vectorize(poster.effect.FB.y0rv), -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
  #3.P(delta>0)
  one.arm.P.FB.y0rv[n.sim, ,] <-
    (outer(effect_star2, y_0, FUN = Vectorize(one.arm.Pr.rej.H0.y0rv)))
}
#4.H0 rejection rate: how often P(delta>0) > 95%
one.arm.rej.H0.FB.y0rv <-
  array(apply(one.arm.P.FB.y0rv > 0.95, c(2,3), FUN = mean, na.rm = TRUE),
        dim = c(length(effect_star2),length(y_0)))
```

Summary

The resulting posterior distributions and H_0 rejection rates are shown in the Table 2.1. Related operating characteristics are shown on the plots in chapter 4.1.

Type	The posterior distribution $p(\Delta \mid y_0, y_\star)$	Calculation of the H_0 Rejection rate (fixed y_0) conditional on the treatment effect Δ	Calculation of the H_0 Rejection rate (random y_0) conditional on the treatment effect Δ
Fixed	$\mathrm{N}\left(\frac{\delta m_0 y_0+m_\star y_\star}{\delta m_0+m_\star}, \frac{\sigma^2}{\delta m_0+m_\star}\right)$	$\Phi\left[\frac{\frac{1}{m_\star}\left(z_\alpha \sigma\sqrt{\delta m_0+m_\star}+\delta m_0 y_0\right)+\Delta}{\sigma/\sqrt{m_\star}}\right]$	$\int \Phi\left[\frac{\frac{1}{m_\star}\left(z_\alpha \sigma\sqrt{\delta m_0+m_\star}+\delta m_0 y_0\right)+\Delta}{\sigma/\sqrt{m_\star}}\right] \phi_{y_0}\left(y_0, \frac{\sigma}{\sqrt{m_0}}\right) dy_0$
EB	$\begin{cases} \mathrm{N}\left(\frac{\sigma^2}{m_\star(y_0-y_\star)}+y_\star, \frac{\sigma^2}{m_\star}\left(1-\frac{\sigma^2}{m_\star(y_\star-y_0)^2}\right)\right), \\ \text{if } (y_\star-y_0)^2 > \frac{\sigma^2}{m_\star} \\ \mathrm{N}\left(\frac{m_0 y_0+m_\star y_\star}{m_0+m_\star}, \frac{\sigma^2}{m_0+m_\star}\right), \\ \text{if } (y_\star-y_0)^2 \leq \frac{\sigma^2}{m_\star} \end{cases}$	1. Simulate $y_\star$ 2. Estimate $\hat{\delta}$ 3. Derive mean and variance of Normal EB posterior $p(\Delta \mid y_0, y_\star)$ 4. Estimate H_0 rejection rate $\mid \Delta$: Check how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ over simulations	1. Simulate $y_\star$ and y_0 2. Estimate $\hat{\delta}$ 3. Derive mean and variance of Normal EB posterior $p(\Delta > 0 \mid y_0, y_\star)$ 4. Estimate H_0 rejection rate $\mid \Delta$: Check how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ over simulations
FB	$\int_0^1 \mathrm{N}_\Delta\left(y_0, \frac{\sigma^2}{\delta m_0}\right) \mathrm{N}_\Delta\left(y_\star, \frac{\sigma^2}{m_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\delta$	1. Simulate $y_\star$ 2. Derive marginal FB posterior $p(\Delta \mid y_0, y_\star)$ 3. $\mathrm{Pr}(\Delta > 0 \mid y_0, y_\star) = \int_0^\infty p(\Delta \mid y_0, y_\star) d\Delta$ 4. Power: how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ under Δ_{H_a} and Type I error: how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ under Δ_{H_0}	1. Simulate $y_\star$ and y_0 2. Derive marginal FB posterior $p(\Delta \mid y_0, y_\star)$ 3. $\mathrm{Pr}(\Delta > 0 \mid y_0, y_\star) = \int_0^\infty p(\Delta \mid y_0, y_\star) d\Delta$ 4. Power: how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ under Δ_{H_a} and Type I error: how often $\mathsf{Pr}(\Delta > 0 \mid y_0, y_\star) > 1-\alpha$ under Δ_{H_0}

Table 2.1: Posterior distributions and H_0 rejection rates for the one arm setting

Chapter 3

Two arm setting

In the two arm case, we are interested in the scenario of a new clinical study, which aims to compare normally distributed outcomes for control and treatment arms, however additional historical information is available only for a control arm. For example, such a situation can happen when there are multiple RCT for a particular disease with a control arm as standards of care. Available information for the control arm from different studies can be analyzed together.

Thus, in a new experiment, a prior distribution for an outcome in a control arm can be based on historical information. For a treatment arm, a non-informative prior can be applied. In this chapter, the outcome estimate is the sample mean of a normally distributed relevant measure in a group of interest.

Similarly, like in the one arm scenario 2, we established an identical variance of the outcome measure for both arms in a new experiment and in a control arm of the historical experiment. Besides, we also fixed the same sample sizes of the control and the treated groups of a new study and in a historical control group.

- Historical data: $X_0 \sim \mathrm{N}\left(\theta_0, \frac{\sigma^2}{n_0}\right)$, where x_0 is the outcome estimate, n_0 is the sample size and θ_0 is the true outcome in historical control group
- New control data: $X_\star \sim \mathrm{N}\left(\theta_\star, \frac{\sigma^2}{n_\star}\right)$, where $x_\star$ is the outcome estimate, $n_\star$ is the sample size and $\theta_\star$ is the true value of the outcome in a control group of a new experiment
- We use the notation θ_c for the true treatment effect in the control group, when we did the inference using both, historical and new experiment data together
- New treated data: $X_t \sim \mathrm{N}\left(\theta_t, \frac{\sigma^2}{n_t}\right)$, $\theta_t = \theta_\star + \Delta$, where x_t is the outcome estimate in a current treated group, n_t is the sample size in the current treated group
- For our simulation study we fixed $n_\star = 100$, $n_0 = 100$, $n_t = 100$ and variance as $\sigma = 1$
- Fixed historical estimate was $x_0 = 10$ and for the case when historical data is random, we used: $x_0 \sim \mathrm{N}(\theta_0, \frac{\sigma^2}{n_0}), \theta_0 = 10$
- We varied $\theta_\star$ for a new experiment from 9.5 to 10.5 to see the influence of the bias in a fixed historical estimate case: $x_0 - \theta_\star$ and in random: $\theta_0 - \theta_\star$ on operating characteristics
- We varied Δ from -0.2 to 0.45 to see the influence of the treatment effect on the H_0 rejection rate and have estimates of type I error and a power for a particular values of Δ
- With the mentioned above sample sizes we reach power of 80% with a significance level of 5% for one-sided test when a treatment effect under H_1 is 0.35 in the RCT setting

3.1 Power parameter in a power prior

3.1.1 Conditional power prior

To derive the posterior distribution for a treatment effect, first, we need to have a posterior distribution for an outcome measure for each arm separately. For a treated group, a non-informative prior was used, and the posterior distribution of an outcome was derived using Bayes theorem.

$$p(\theta_t \,|\, x_t) = \frac{L(x_t \,|\, \theta_t) p(\theta_t)}{\int L(x_t \,|\, \theta_t) p(\theta_t) d\theta_t}$$

$$p(\theta_t \,|\, x_t) \sim \mathrm{N}\left(x_t, \frac{\sigma^2}{n_t}\right) \tag{3.1}$$

Similarly to the derivation of (2.3) we have the next posterior distribution of the outcome in a control group:

$$p(\theta_c \,|\, x_0, x_\star, \delta) \sim \mathrm{N}\left(\frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n\star}, \frac{\sigma^2}{\delta n_0 + n\star}\right)$$

The posterior distribution of a treatment effect: $\Delta = \theta_t - \theta_c$. Thus, it can be derived as the difference of two normal distributions:

$$\Delta \,|\, x_t, x_0, x_\star = \theta_t \,|\, x_t - \theta_c \,|\, x_0, x_\star$$

$$\Delta \,|\, x_t, x_\star, x_0, \delta \sim \mathrm{N}\left(x_t - \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{\delta n_0 + n_\star}\right) \tag{3.2}$$

3.1.2 Empirical Bayes approach

Similar to the derivation in the one arm case (2.5), the posterior distribution of the outcome in a control group is:

$$\theta_c \,|\, x_0, x_\star \sim \begin{cases} \mathrm{N}\left(\frac{\sigma^2}{n_\star(x_0 - x_\star)} + x_\star, \frac{\sigma^2}{n_\star}\left(1 - \frac{\sigma^2}{n_\star(x_\star - x_0)^2}\right)\right), \text{if } (x_\star - x_0)^2 > \frac{\sigma^2}{n_\star} \\ \mathrm{N}\left(\frac{n_0 x_0 + n_\star x_\star}{n_0 + n_\star}, \frac{\sigma^2}{n_0 + n_\star}\right), \text{if } (x_\star - x_0)^2 \le \frac{\sigma^2}{n_\star} \end{cases}$$

And, consequently, the posterior distribution of the treatment effect $\Delta = \theta_t - \theta_c$ again as the difference of the two normal distributions can be easily derived:

$$\Delta \,|\, x_t, x_\star, x_0, \hat{\hat{\delta}} \sim \begin{cases} \mathrm{N}\left(x_t - \frac{\sigma^2}{n_\star(x_0 - x_\star)} - x_\star, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{n_\star}\left(1 - \frac{\sigma^2}{n_\star(x_\star - x_0)^2}\right)\right), \text{if } (x_\star - x_0)^2 > \frac{\sigma^2}{n_\star} \\ \mathrm{N}\left(x_t - \frac{n_0 x_0 + n_\star x_\star}{n_0 + n_\star}, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{n_0 + n_\star}\right), \text{if } (x_\star - x_0)^2 \leq \frac{\sigma^2}{n_\star} \end{cases} \tag{3.3}$$

3.1.3 Normalized power prior

Similar to the derivation for one arm case (2.9), the posterior distribution of an outcome in the control group is:

$$p(\theta_c \,|\, x_0, x_\star) \propto \int_0^1 \mathrm{N}_{\theta_c}\left(x_0, \frac{\sigma^2}{\delta n_0}\right) \mathrm{N}_{\theta_c}\left(x_\star, \frac{\sigma^2}{n_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\delta \tag{3.4}$$

Posterior distribution of the treatment effect is now the difference of two random variables $\Delta \,|\, x_t, x_0, x_\star = \theta_t \,|\, x_t - \theta_c \,|\, x_0, x_\star$, one of them is normal, and the other has a more complex shape (3.4). Using the formula of the difference between two random variables available in the App. 2, we derive the next formula for the posterior distribution of the treatment effect:

$$p(\Delta \,|\, x_t, x_0, x_\star) = \int_\infty^\infty \mathrm{N}_{\theta_t}\left(x_t, \frac{\sigma^2}{n_t}\right) p_{\theta_c \,|\, x_0, x_\star}(\theta_t - \Delta) d\theta_t$$

3.2 H_0 rejection rate

Analogously to the one arm scenario, we test the following hypotheses $H_0 : \Delta \leq 0$ and $H_1 : \Delta > 0$. The treatment effect under the null hypothesis is $\Delta_{H_0} = 0$. The treatment effect under the alternative hypothesis is $\Delta_{H_1} = 0.35$, as estimated for a target power of 80% with significance level 5% in one-sided test in the RCT setting conditional on $n_0 = 100$, $n_\star = 100$, $n_t = 100$ and variance as $\sigma = 1$.

Operating characteristics were defined in the same way as in chapter 2.2. For the estimation of operating characteristics, we need to derive H_0 rejection rate conditional on a treatment effect $\Delta = \theta_t - \theta_c$.

3.2.1 Classical approach

Similar to the derivation for the one arm scenario (2.13), a classical power for two arm test is:

$$x_t - x_\star > -z_\alpha \sigma \sqrt{\frac{n_\star + n_t}{n_\star n_t}} \tag{3.5}$$

where z_α is a quantile of a prescpecified significance level (Type I error)

Thus, H_0 rejection rate in the classical RCT approach without the use of historical data is:

$$
\begin{aligned}
\Pr_C(\text{reject } H_0 \,|\, \Delta_\star) &= \Pr(x_t - x_\star > -z_\alpha \sigma \sqrt{\frac{n_\star + n_t}{n_\star n_t}} \,\Big|\, \Delta) \\
&= 1 - \Pr(x_t - x_\star \le -\sqrt{\frac{n_\star + n_t}{n_\star n_t}} z_\alpha \sigma \,\Big|\, \Delta) \\
&= 1 - \Phi\left[\frac{-z_\alpha \sigma / \sqrt{\frac{n_\star + n_t}{n_\star n_t}} - \Delta}{\sigma / \sqrt{\frac{n_\star + n_t}{n_\star n_t}}}\right] \\
&= \Phi\left[\frac{z_\epsilon \sigma / \sqrt{\frac{n_\star + n_t}{n_\star n_t}} + \Delta}{\sigma / \sqrt{\frac{n_\star + n_t}{n_\star n_t}}}\right] \\
&= \Phi\left[\frac{\Delta}{\sigma}\sqrt{\frac{n_\star n_t}{n_\star + n_t}} + z_\alpha\right] \qquad (3.6)
\end{aligned}
$$

```
#Conditional classical power for 2 arms
two.arm.cond.class.power <- function (n_c,n_t, ze, effect, sigma){
  out <- pnorm((effect/sigma)*sqrt((n_c *  n_t)/(n_c + n_t)) + ze)
  return (out)
}
```

```
#Classical approach
two.arm.H0.rej.cond.class <-
  sapply(effects, two.arm.cond.class.power.effect)
```

3.2.2 Bayesian approach

In a Bayesian approach, we are integrating historical information into the analysis. We derive a H_0 rejection rate function for each approach of dealing with a nuisance power parameter: a conditional power prior, the empirical Bayes, and the full Bayes approaches.

Conditional power prior

The posterior for the effect was derived in (3.1.1):

$$
\Delta \,|\, x_t, x_\star, x_0, \delta \sim \mathrm{N}\left(x_t - \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{\delta n_0 + n_\star}\right)
$$

Similar to the one arm scenario (2.2.2), we need to find for which $x_\star$ the result is significant:

$$
\begin{aligned}
\Pr(\Delta < 0 \,|\, x_\star, x_0, x_t, \delta) &\le \alpha \\
\Phi\left[\frac{0 - x_t + \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}}{\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}}\right] &\le \alpha \\
\frac{0 - x_t + \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}}{\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}} &\le z_\alpha
\end{aligned}
$$

Thus:

$$X_\star \leq \frac{1}{n_\star}\left\{(\delta n_0 + n_\star)\left(x_t + z_\alpha \sigma \sqrt{\frac{1}{\delta n_0 + n_\star} + \frac{1}{n_t}}\right) - \delta n_0 x_0\right\} \tag{3.7}$$

Since $X_\star \sim \mathrm{N}\left(\theta_\star, \frac{\sigma^2}{n_\star}\right)$, the H_0 rejection rate for a fixed historical estimate is:

$$\begin{aligned}\mathsf{Pr}(\text{reject} H_0 \,|\, \theta_\star, \Delta, x_t) &= \Phi\left[\frac{\frac{1}{n_\star}\left\{(\delta n_0 + n_\star)\left(x_t + z\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}\right) - \delta n_0 x_0\right\} - \theta_\star}{\sigma/\sqrt{n_\star}}\right] \\ &= \Phi\left[\frac{\delta n_0 + n_\star}{\sigma\sqrt{n_\star}}\left(x_t + z_\alpha\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}\right) - \frac{\delta n_0 x_0}{\sigma\sqrt{n_\star}} - \frac{\theta_\star\sqrt{n_\star}}{\sigma}\right]\end{aligned}$$

which is still dependent on x_t and we know that $X_t \sim \mathrm{N}\left(\theta_\star + \Delta, \frac{\sigma^2}{n_t}\right)$. So, we can apply the law of the total probability and integrate over the space of x_t :

$$\mathsf{Pr}(\text{reject } H_0 \,|\, \theta_\star, \Delta) = \int \Phi\left[\frac{\delta n_0 + n_\star}{\sigma\sqrt{n_\star}}\left(x_t + z_\alpha\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}\right) - \frac{\delta n_0 x_0}{\sigma\sqrt{n_\star}} - \frac{\theta_\star\sqrt{n_\star}}{\sigma}\right]\phi_{x_t}\left(\theta_\star + \Delta, \frac{\sigma^2}{n_t}\right) dx_t$$

```
#Bayesian H0 rejection rate given effect
two.arm.bayes.power.given.effect.xt.var <- function (delta, x_0, x_t, n_t, n_0,
    n_c, ze, sigma, effect, theta_star){
  out <- pnorm( ((delta*n_0 + n_c)/(sigma*sqrt(n_c)))*(x_t + ze*sigma*sqrt(1/(
    delta*n_0+n_c) + 1/n_t)) - delta*n_0*x_0/(sigma*sqrt(n_c)) - theta_star *
    sqrt(n_star)/sigma  ) * dnorm(x_t, mean = theta_star + effect, sd = sigma/
    sqrt(n_t))
  return (out)
}
```

We applied this formula to different values of Δ to investigate the H_0 rejection rate depending on a treatment effect. We also explored different values of $\theta_\star$ to examine the influence of the bias between historical and current experiments. Besides, we considered different values of δ to investigate the influence of a power parameter in a conditional power prior approach.

```
#Bayesian approach with fixed power parameter
B.Pr.rej.Ho.f1 <-
  function (x, y, z) {
    Pow.x_t_delta <- function(xt) {
      two.arm.bayes.power.given.effect.xt.var(
        x_t = xt,
        delta = z,
        x_0 = x_0,
        n_0 = n_0,
        n_c = n_star,
        n_t = n_t,
        ze = qnorm(0.05),
        theta_star = y,
        sigma = sd_0,
        effect = x
      )
    }
```

```
    out <- integrate(Pow.x_t_delta, -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
par <- expand.grid(x = effects, y = theta_star, z = deltas.fixed)
two.arm.H0.rej.cond.bayes <- array(mapply(B.Pr.rej.Ho.f1, par$x, par$y, par$z),
                                   dim = c(length(effects), length(theta_star),
    length(deltas.fixed)))
```

In the case of random historical data, we again can apply the law of total probability and update the H_0 rejection function, considering the randomness of $X_0 \sim (\theta_0, \frac{\sigma^2}{n_0})$:

$$\mathsf{Pr}(\text{reject } H_0 \,|\, \Delta) = \int\int \Phi\left[\frac{\delta n_0 + n_\star}{\sigma\sqrt{n_\star}}\left(x_t + z_\alpha\sigma\sqrt{1/(\delta n_0 + n_\star) + 1/n_t}\right) - \frac{\delta n_0 x_0}{\sigma\sqrt{n_\star}} - \frac{\theta_c\sqrt{n_\star}}{\sigma}\right]$$
$$\phi\left(\theta_0, \frac{\sigma^2}{n_0}\right)\phi\left(\theta_c + \Delta, \frac{\sigma^2}{n_t}\right)dx_0 dx_t$$

```
f1 <-
  function (x, y, z) {
    #H0 rejection rate depending on two variables xt and xo
    Pow.x_t_x_0 <- function(xt1, xo1) {
      two.arm.bayes.power.given.effect.xt.var(
        x_t = xt1,
        delta = z,
        x_0 = xo1,
        n_0 = n_0,
        n_c = n_star[1],
        n_t = n_t,
        ze = qnorm(0.05),
        theta_star = y,
        sigma = sd_0,
        effect = x
      )
    }
    #Now we perform a double integration over xo and xt :
    #First, we integrate over xt
    # xt random
    # xo random
    Pow.x_t <- function (x0) {
      # d delta over what we integrate
      # y theta parameter of interest
      dens <- function (x0, xt) {
        Pow.x_t_x_0(xt1 = xt, xo1 = x0) * dnorm(x0, mean = x_0, sd = sd_0 / sqrt
    (n_0))
      }
      return(sapply(x0, function(z) {
        integrate(dens, 5, Inf, abs.tol = 0, xt = z)$value
      }))
      #for each x0 we integrate over xt
    }
    #Second, we integrate over x0
    out <- integrate(Pow.x_t, 5, Inf, abs.tol = 0)$value
    return(out)
  }
par <- expand.grid(x = effects, y = theta_star, z = deltas.fixed)
two.arm.H0.rej.cond.bayes.xorv <-
  array(mapply(f1, par$x, par$y, par$z),
```

```
      dim = c(length(effects), length(theta_star), length(deltas.fixed)))
```

Empirical Bayes

Similar to the one arm case (2.2.2), the closed-form solution for the H_0 rejection rate is not easily obtained, so we used simulations techniques to asses operating characteristics.

When historical data x_0 is fixed, we simulated $x_\star$ and x_t, derived the posterior distribution for a treatment effect using the formula from above (3.3), estimated $\Pr(\Delta > 0 \,|\, x_0, x_\star, x_t)$, and calculated how often H_0 is rejected : $\Pr(\Delta > 0 \,|\, x_0, x_\star, x_t) > 0.95$

```
######################################## For fixed and EB:
for (n.sim in c(1:sim.N)) {
  print(n.sim)
  for (n.theta_star in c(1:length(theta_star))) {
    for (n.effect in c(1:length(effects))) {
      #1.simulate x_star
      x_star1 <- rnorm (1, theta_star[n.theta_star] ,
                        sd_0 / sqrt(n_star[n.n_star]))
      #2.simulate x_t
      x_new_t <-
        rnorm (1, theta_star[n.theta_star] + effects[n.effect],
               sd_0 / sqrt(n_t))

      #3.estimate delta.hat
      delta.hat <-
        fun.delta.EB(
          x_star = x_star1,
          x_0 = x_0,
          sigma = sd_0,
          n_star = n_star,
          n_0 = n_0
        )
      #4.Derive mean and variance of EB posterior of p(effect)
      mean_post.2arm.simul.EB <-
        (delta.hat * x_0 * n_0 + x_star1 * n_star) / (n_star + delta.hat *
                                                          n_0)
      sd_post.2arm.simul.EB <-
        sqrt(sd_0 ^ 2 / (n_star + delta.hat * n_0))
      #P(Delta >0)
      P.EB[n.sim, n.theta_star, n.effect] <- +
        1 - pnorm(0, x_new_t - mean_post.2arm.simul.EB,
                  sqrt(sd_0 ^ 2 / (n_t) + sd_post.2arm.simul.EB ^ 2))

    }
  }
}
Reject.H0.EB <-
  array(apply(P.EB > 0.95, c(2,3), FUN = mean, na.rm = TRUE),
        dim = c(length(theta_star),length(effects)))
```

In the case with random historical data $x_0 \sim \mathrm{N}\left(\theta_0, \frac{\sigma^2}{n_0}\right)$, we additionally simulated a historical estimate.

```
for (n.sim in c(1:sim.N)) {
  print(n.sim)
  #1.simulate x_0
  x_0_sim <- rnorm(1, x_0, sd = sd_0 / sqrt(n_0))
```

```
  for (n.theta_star in c(1:length(theta_star))) {
    for (n.effect in c(1:length(effects))) {
      #2.simulate x_t and x_star
      x_new_t <-
        rnorm (1, theta_star[n.theta_star] + effects[n.effect], sd_0 / sqrt(n_t)
    )
      x_star1 <-
        rnorm (1, theta_star[n.theta_star] , sd_0 / sqrt(n_star)) #New control
      #3.estimate delta.hat
      delta.hat <-
        fun.delta.EB(
          x_star = x_star1,
          x_0 = x_0_sim,
          sigma = sd_0,
          n_star = n_star,
          n_0 = n_0
        )
      #4.Derive mean and variance of EB posterior of p(effect)
      mean_post.2arm.simul.EB.xorv <-
        (delta.hat * x_0_sim * n_0 + x_star1 * n_star) / (n_star + delta.hat *
                                                          n_0)
      sd_post.2arm.simul.EB.xorv <-
        sqrt(sd_0 ^ 2 / (n_star + delta.hat * n_0))
      #P(Delta>0)
      P.EB.xorv[n.sim, n.theta_star, n.effect] <- 1 - pnorm(
        0,
        x_new_t - mean_post.2arm.simul.EB.xorv,
        sqrt(sd_0 ^ 2 / (n_t) + sd_post.2arm.simul.EB.xorv ^ 2)
      )
    }
  }
}
#5.how often we reject H0: P(Delta>0)>95%
Reject.H0.EB.xorv <-
  array(apply(P.EB.xorv > 0.95, c(2,3), FUN = mean, na.rm = TRUE),
        dim = c(length(theta_star),length(effects)))
```

Full Bayes

As it was shown before, the posterior distribution of a treatment effect has a complex shape (3.5). Thus, a two-step approach: first, derive the posterior distribution of a treatment effect and then estimate $P(\Delta > 0 \,|\, x_t, x_0, x_\star)$ would require numerous numerical integrations and consequently, a lot of time. Therefore, we focused on the direct derivation of $P(\Delta > 0 \,|\, x_t, x_0, x_\star)$ using the normality of the posterior distribution of an outcome in a treated group.

$$
\begin{aligned}
P(\Delta > 0 \,|\, x_t, x_0, x_\star) &= \\
P((\theta_t \,|\, x_t - \theta_c \,|\, x_0, x_\star) > 0) &= \\
P(\theta_t \,|\, x_t > \theta_c \,|\, x_0, x_\star) &= \\
1 - P(\theta_t \,|\, x_t \leq \theta_c \,|\, x_0, x_\star) &= \\
1 - \Phi\left[\frac{\theta_c \,|\, x_0, x_\star - x_t}{\sqrt{\sigma^2/n_t}}\right] &= \\
\Phi\left[\frac{x_t - \theta_c \,|\, x_0, x_\star}{\sqrt{\sigma^2/n_t}}\right] &
\end{aligned}
$$

Now, applying the law of total probability, we need to integrate over the space of the parameter θ_c :

$$
\Pr(\Delta > 0 \,|\, x_t, x_0, x_\star) = \int \Phi\left[\frac{x_t - \theta_c}{\sqrt{\sigma^2/n_t}}\right] p(\theta_c \,|\, x_0, x_\star) d\theta_c
$$

In order to get the H_0 rejection rate, we needed to see how often $\Pr(\Delta > 0 \,|\, x_t, x_0, x_\star) > 0.95$, depending on different simulations of x_t and $x_\star$. This approach was also computationally time-consuming, so we looked only at the $\Delta = \Delta_{H_0}$ and $\Delta = \Delta_{H_1}$, since our primary interest is power and type I error. We also considered different values for $\theta_\star$ to investigate the influence of the bias $x_0 - \theta_\star$.

```
foreach(n.sim = c(1:sim.N)) %do% {
  print(n.sim)
  #Under H0
  f1.H0 <- function (t) {
    #1.simulate x_star
    x_star1 <-
      rnorm (1, t , sd_0 / sqrt(n_star))
    #2.Derive marginal posterior p(theta_c given x_star, x_0)
    post.theta.control <- function(theta) {
      poster.control.FB(
        theta = theta,
        x_0 = x_0,
        x_star = x_star1,
        alpha = 0.5,
        beta = 0.5,
        n_star = n_star,
        n_0 = n_0,
        sigma = sd_0
      )
    }
    K<-1/integrate(Vectorize(post.theta.control), -Inf, Inf, abs.tol = 0)$value
    #3.simulate x_t given H0
    x_new_t.H0 <- rnorm (1, t + effect.H0,
                         sd_0 / sqrt(n_t))

    #4. Derive P(delta>0) under H0
    ## First we derive p(effect>0) as function of theta_c under H0
    r.v.diff.H0.FB <- function(theta_c) {
      pnorm((x_new_t.H0 - theta_c) / sqrt(sd_0 ^ 2 / (n_t)), 0, 1) * post.theta.
```

```
    control(theta_c)*K
    }
    out <- integrate(r.v.diff.H0.FB, -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
  P.under.H0.FB[n.sim, ] <- sapply(theta_star, Vectorize(f1.H0))
  #Under Ha
  f1.Ha <- function (t) {
    #1.simulate x_star
    x_star1 <-
      rnorm (1, t , sd_0 / sqrt(n_star))
    #2.Derive marginal posterior p(theta_c given x_star, x_0)
    post.theta.control <- function(theta) {
      poster.control.FB(
        theta = theta,
        x_0 = x_0,
        x_star = x_star1,
        alpha = 0.5,
        beta = 0.5,
        n_star = n_star,
        n_0 = n_0,
        sigma = sd_0
      )
    }
    K<-1/integrate(Vectorize(post.theta.control), -Inf, Inf, abs.tol = 0)$value
    #3.simulate x_t given H1
    x_new_t.Ha <- rnorm (1, t + effect.Ha,
                         sd_0 / sqrt(n_t))
    #4. Derive P(delta>0) under H1
    ## First we derive p(effect>0) as function of theta_c under H1
    r.v.diff.Ha.FB <- function(theta_c) {
      pnorm((x_new_t.Ha - theta_c) / sqrt(sd_0 ^ 2 / (n_t)), 0, 1) *post.theta.
    control(theta_c)*K
    }
    ## Then we integrate over the space of theta_c
    out <-
      integrate(r.v.diff.Ha.FB, -Inf, Inf, abs.tol = 0)$value
    return(out)
  }
  P.under.Ha.FB[n.sim, ] <- sapply(theta_star, Vectorize(f1.Ha))
}
TPR.FB <- apply (P.under.Ha.FB > 0.95,2,mean, na.rm = TRUE)
FPR.FB <- apply (P.under.H0.FB > 0.95,2,mean, na.rm = TRUE)
```

For the case with random historical estimate, we just added an extra line of code allowing random simulation of a historical estimate $x_0 \sim \mathrm{N}(\theta_0, \frac{\sigma^2}{n_0})$. Here we also considered different values of $\theta_\star$ to investigate the bias $\theta_0 - \theta_\star$.

Summary

The resulting posterior distributions and the H_0 rejection rates are shown in the tables 3.1 and 3.2. Related operating characteristics are shown on the plots in chapter 4.2.

Type	Posterior of an outcome in control arm $p(\theta_c \mid x_0, x_\star)$	Posterior of an effect $p(\Delta \mid x_0, x_\star, x_t)$
Fixed	$\mathrm{N}\left(\frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}, \frac{\sigma^2}{\delta n_0 + n_\star}\right)$	$\mathrm{N}\left(x_t - \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{\delta n_0 + n_\star}\right)$
EB	$\mathrm{N}\left(\frac{\sigma^2}{n_\star(x_0 - x_\star)} + x_\star, \frac{\sigma^2}{n_\star}\left(1 - \frac{\sigma^2}{n_\star(x_\star - x_0)^2}\right)\right)$, when $(x_\star - x_0)^2 > \frac{\sigma^2}{n_\star}$	$\mathrm{N}\left(x_t - \frac{\sigma^2}{n_\star(x_0 - x_\star)} + x_\star, \frac{\sigma^2}{n_t} + \frac{\sigma^2}{n_\star}\left(1 - \frac{\sigma^2}{n_\star(x_\star - x_0)^2}\right)\right)$, when $(x_\star - x_0)^2 > \frac{\sigma^2}{n_\star}$
FB	$\int_0^1 \mathrm{N}_{\theta_c}\left(x_0, \frac{\sigma^2}{\delta n_0}\right) \mathrm{N}_{\theta_c}\left(x_\star, \frac{\sigma^2}{n_\star}\right) \mathrm{Be}_\delta(\alpha, \beta) d\delta$	$p(\Delta \mid x_t, x_0, x_\star) = \int_{\infty}^{\infty} \mathrm{N}_{\theta_t}\left(x_t, \frac{\sigma^2}{n_t}\right) p_{\theta_c \mid x_0, x_\star}(\theta_t - \Delta) d\theta_t$

Table 3.1: Posterior distributions for the two arm setting

Type	Calculation of H_0 rejection rate (fixed x_0)	Calculation of H_0 rejection rate (random x_0)
Fixed	$\int \Phi \left[\frac{\delta n_0 + n_\star}{\sigma \sqrt{n_\star}} \left(x_t + z_\alpha \sigma \sqrt{1/(\delta n_0 + n_\star) + 1/n_t} \right) - \frac{\delta n_0 x_0}{\sigma \sqrt{n_\star}} - \frac{\theta_\star \sqrt{n_\star}}{\sigma} \right] * \phi_{x_t} \left(\theta_\star + \Delta, \frac{\sigma^2}{n_t} \right) dx_t$	$\int \int \Phi \left[\frac{\delta n_0 + n_\star}{\sigma \sqrt{n_\star}} \left(x_t + z_\alpha \sigma \sqrt{1/(\delta n_0 + n_\star) + 1/n_t} \right) - \frac{\delta n_0 x_0}{\sigma \sqrt{n_\star}} - \frac{\theta_\star \sqrt{n_\star}}{\sigma} \right] * \phi_{x_0} \left(x_0, \frac{\sigma^2}{n_0} \right) \phi_{x_t} \left(\theta_\star + \Delta, \frac{\sigma^2}{n_t} \right) dx_0 dx_t$
EB	1. Simulate x_t and $x_\star$ 2. Estimate $\hat{\delta}$ 3. Derive mean and variance of Normal EB posterior $\mathsf{Pr}(\Delta \mid x_0, x_\star, x_t)$ 4. Estimate H_0 rejection rate: Check how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$	1. Simulate x_0 2. Simulate x_t and $x_\star$ 3. Estimate $\hat{\delta}$ 4. Derive mean and variance of Normal EB posterior $\mathsf{Pr}(\Delta \mid x_0, x_\star, x_t)$ 5. Estimate H_0 rejection rate: Check how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$
FB	1. Simulate $x_\star$ 2. Derive the marginal posterior $p(\theta_c \mid x_0, x_\star)$ 3. Simulate $x_t \mid \Delta_{H_a}$ and $x_t \mid \Delta_{H_0}$ 4. Derive $\mathsf{Pr}(\Delta > 0 \mid x_t, x_0, x_\star) = \int \Phi \left[\frac{x_t - \theta_c}{\sqrt{(\sigma^2/(n_t)}} \right] p(\theta_c \mid x_0, x_\star) d\theta_c$ for each $x_t \mid \Delta_{H_a}$ and $x_t \mid \Delta_{H_0}$ 5. Power: how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$ under Δ_{H_a} and Type I error: how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$ under Δ_{H_0}	1. Simulate $x_\star$ and x_0 2. Derive the marginal posterior $p(\theta_c \mid x_0, x_\star)$ 3. Simulate $x_t \mid \Delta_{H_a}$ and $x_t \mid \Delta_{H_0}$ 4. Derive $\mathsf{Pr}(\Delta > 0 \mid x_t, x_0, x_\star) = \int \Phi \left[\frac{x_t - \theta_c}{\sqrt{(\sigma^2/(n_t)}} \right] p(\theta_c \mid x_0, x_\star) d\theta_c$ for each $x_t \mid \Delta_{H_a}$ and $x_t \mid \Delta_{H_0}$ 5. Power: how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$ under Δ_{H_a} and Type I error: how often $\mathsf{Pr}(\Delta > 0 \mid x_0, x_\star, x_t) > 1 - \alpha$ under Δ_{H_0}

Table 3.2: H_0 rejection rates for two arm setting

Additional characteristics for the two arm case

We analysed a few additional characteristics in order to explore the influence of a bias in the two arm scenario.

MSE

The MSE is the expected value of the squared error of a point estimate of θ_c compared to the true value in a control group.

$$\mathrm{E}(\hat{\theta}_c - \theta_\star)^2$$

where θ_c is the posterior mean estimate of an outcome in a control group and $\theta_\star$ is a true value of an outcome in a control group of a new experiment.

Coverage

We estimated 95% equi-tailed credible intervals of the posterior distribution of θ_c, to judge about the influence of historical estimate on the shape of the distribution.

Bias of an estimate of an outcome in a control group

We analysed the influence of historical data on an estimate of an outcome in a control group, derived as a mean estimate from posterior distribution. Bias is calculated as the expected difference between this posterior mean and the true outcome in a control group.

$$\mathrm{E}(\hat{\theta}_c - \theta_\star)$$

The results of simulation and estimation of these characteristics are shown in the figure 4.15.

Chapter 4

Results

4.1 One arm setting

Figure 4.1 shows the dependence of the H_0 rejection rate on the treatment effect $\Delta_\star$ in a new experiment. The black line is for the RCT setting when no historical data was used, and expectedly, if $\Delta_\star = 0$ the H_0 rejection rate is 5% and if $\Delta_\star = 0.35$ the H_0 rejection rate is 80%. The purple lines highlight $\Delta_{H_0} = 0$ and $\Delta_{H_1} = 0.35$. Different shades of grey lines show the H_0 rejection rate using a conditional power prior approach for different values of a fixed δ. The red line is for an empirical Bayes approach.

The two upper plots present the case when historical data is in favor of a negative effect, namely the null hypothesis. Thus, using historical data leads to a decrease in type I error with a simultaneous decrease in power depending on the impact of historical data, which is determined by the power parameter δ. The smaller the value of δ, the closer the H_0 rejection rate to the RCT values.

The lower plots show the situation when historical data shows evidence of a treatment effect. In this case, the use of historical data leads to an increase in power with a simultaneous increase in a type I error. The influence of historical data is similarly determined by the power parameter δ.

An important difference between a fixed and a random case can be observed on these plots. In a fixed case scenario, on the left plots, the H_0 rejection functions change similarly to the RCT setting, the increase in power is similar to the increase in type I error. However, for a random case scenario, the H_0 rejection functions change differently. The increase in power is significantly smaller than the increase in a type I error in the lower plot for a random case.

Figure 4.2 shows the dependence of a power and a type I error on historical data also for fixed and random cases. In both cases, historical estimates, which are in favor of H_1, increase power and type I error. Historical estimates, which are in favor of H_0, decrease power and type I error. Adaptable methods as the empirical Bayes and full Bayes approaches show similar dependence. However, the influence of historical data decreases with the growth of the bias for these methods, and consequently, operating characteristics converge to RCT values.

In short, if we have historical information that favors the null hypothesis, and we integrate this information in the analysis, we decrease the posterior value of a treatment effect and, consequently, decrease type I error and power. And the other way around, if historical information favors alternative hypothesis, integration of this information shifts the posterior estimate of the effect towards bigger values, which leads to an increase in power and type I error. This behavior is well known as power and type I error trade-off.

On figure 4.4, on the upper two plots, power and a type I error are shown together, where the trade-off is clearly seen. The critical y_0 value where the power of the Bayesian approach exceeds the RCT power is very close to the critical y_0 value of the type I error when the RCT type I error exceeds the Bayesian type I error. For a random case, these critical values are further apart, type I error rises above the RCT value of 5% faster than power reach 80%. Overall, for both cases, there is no set of historical estimates where they both would outperform RCT.

Lower two plots of figure 4.4 show the rejection ratio. It is a summary measure for operating characteristics, which shows the ratio between power and type I error. Naturally, the higher the value of the rejection ratio, the better the trade-off between power and type I error in the test. However, very high values of the ratio, when historical estimates favor the null hypothesis, are caused by very low values of type I error, which are accompanied by the lower values of power. Hence, when comparing different tests and methods, analysis of only rejection ratio is not enough, and power and type I error should be analyzed additionally. A conditional power prior method shows a strong dependence on bias and does not have good operating characteristics when the bias is too large. The rejection ratio of an empirical Bayesian and a fully Bayesian methods do not migrate far from the RCT level, which proves their adaptability. A Fully Bayesian method has a slightly better type I error and power trade-off for both fixed and random y_0 case. Both adaptable methods outperform RCT in terms of rejection ratio and type I error when historical estimates favor the null hypothesis; however, with a cost of a strong decline in power.

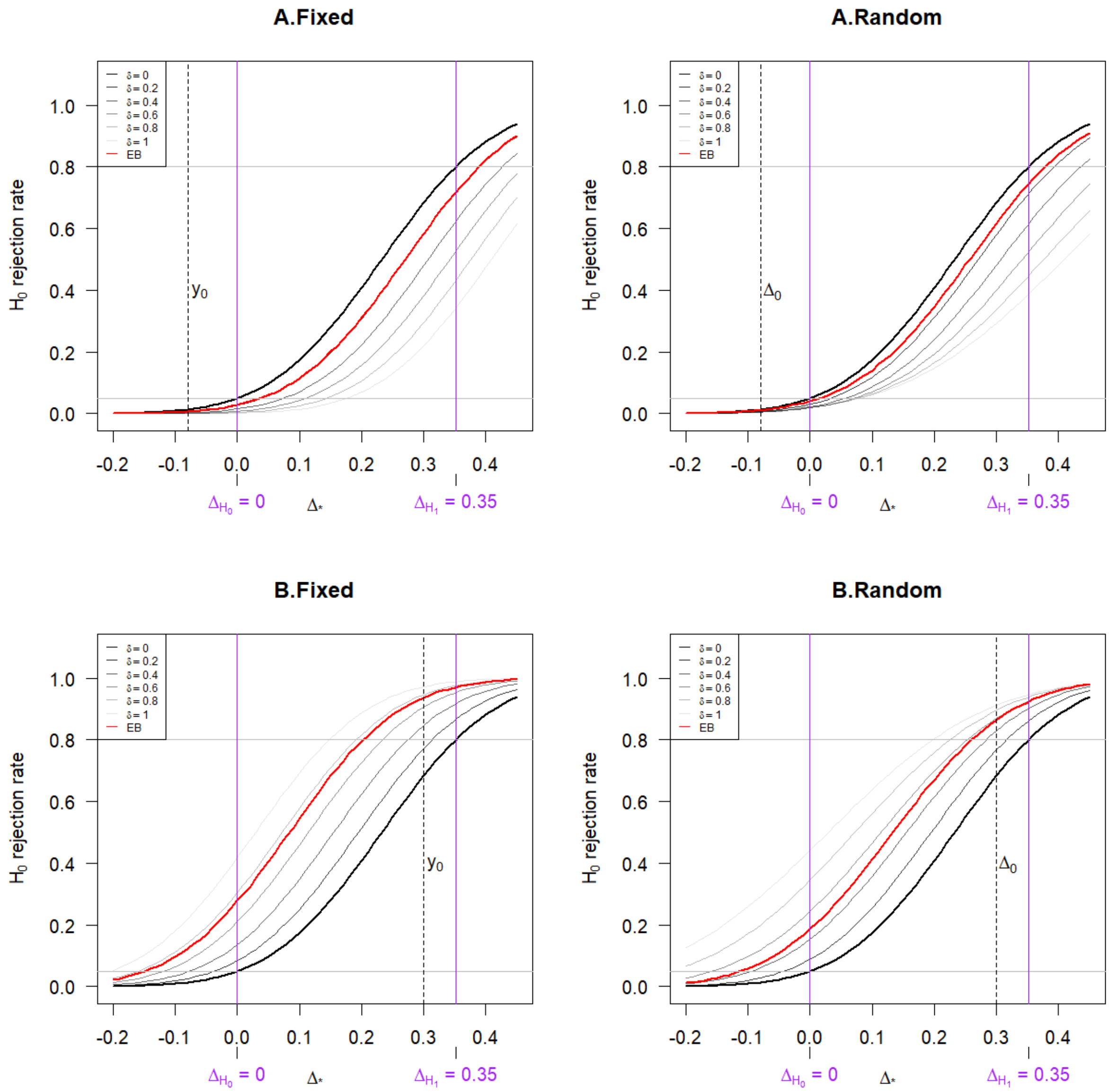

Figure 4.1: The H_0 rejection rate depending on the true effect of a new experiment when **A.** a historical estimate in H_1 region **B.** a historical estimate in H_0 region. **Fixed**: for a fixed historical estimate y_0. **Random**: for a randomly simulated historical estimate $y_0 \sim \mathrm{N}(\Delta_0, \frac{\sigma^2}{m_0})$

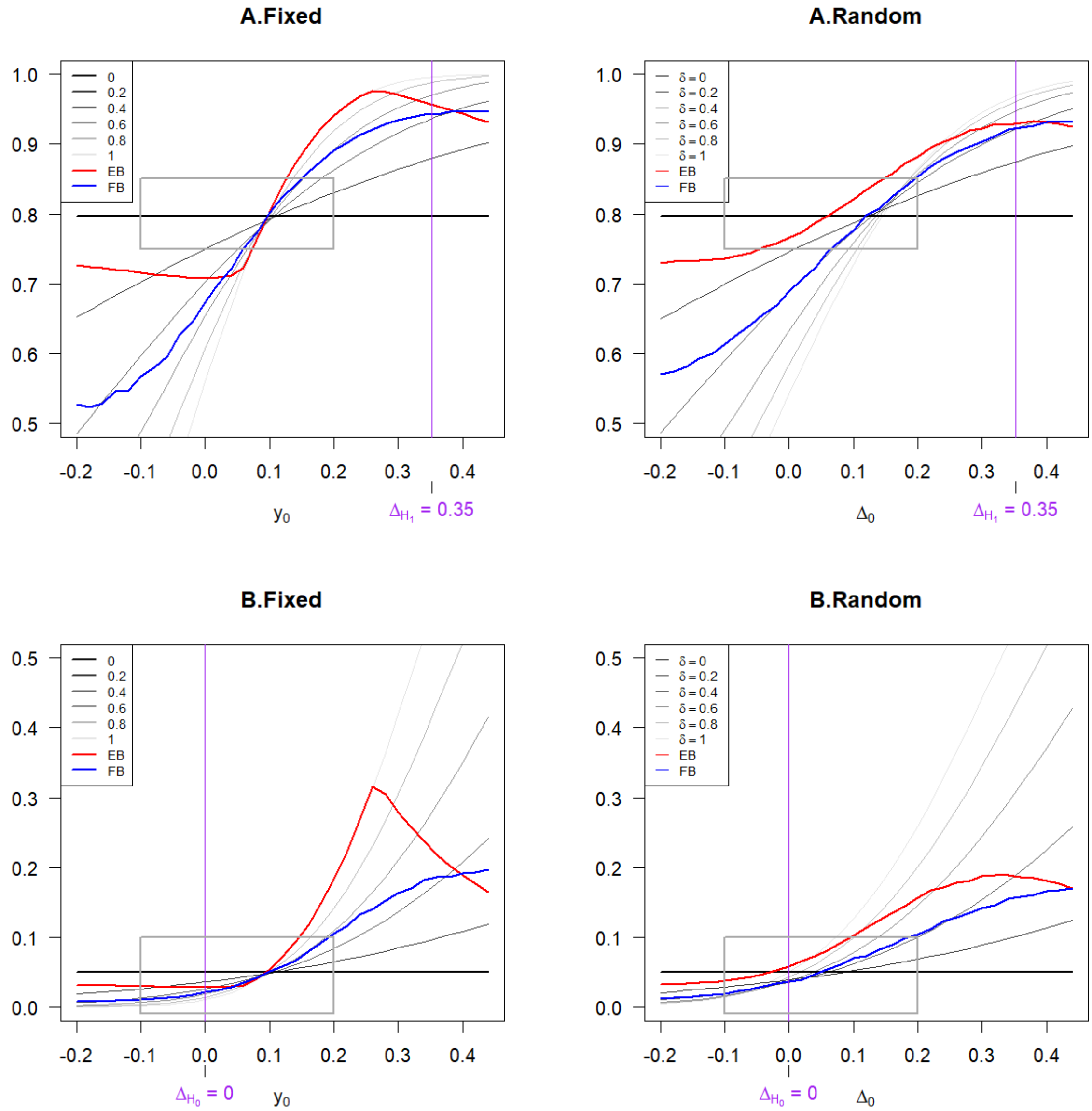

Figure 4.2: **A.** Power and **B.** type I error depending on the value of historical data. **Fixed**: for a fixed historical estimate y_0. **Random**: for randomly simulated historical estimate $y_0 \sim \mathrm{N}(\Delta_0, \frac{\sigma^2}{m_0})$. Grey squares show the area, which is shown on figure 4.3

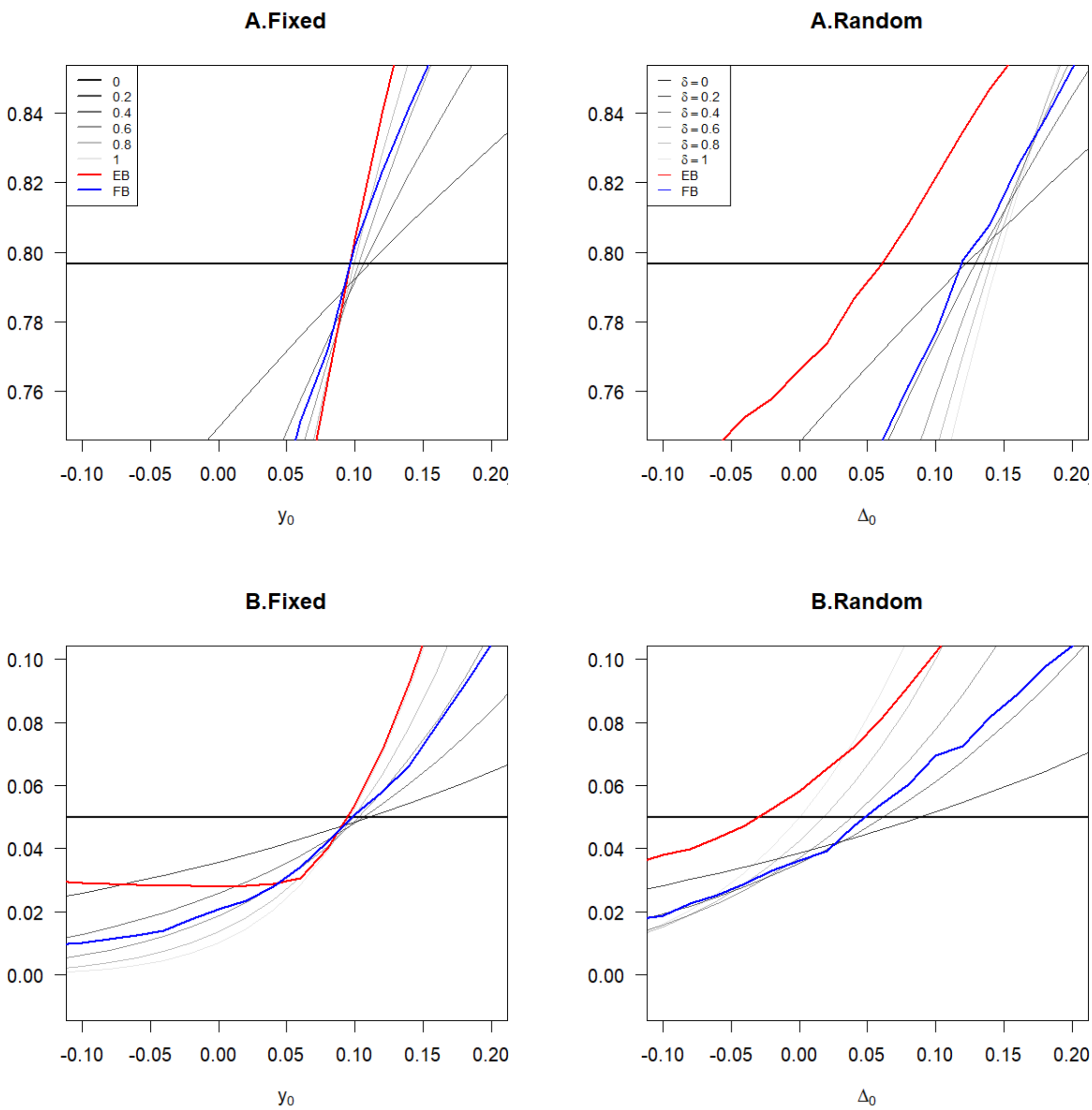

Figure 4.3: "Zoomed in" part of the figure 4.2. **A.** Power and **B.** type I error depending on the value of historical data. **Fixed**: for a fixed historical estimate y_0. **Random**: for randomly simulated historical estimate $y_0 \sim \mathrm{N}(\Delta_0, \frac{\sigma^2}{m_0})$

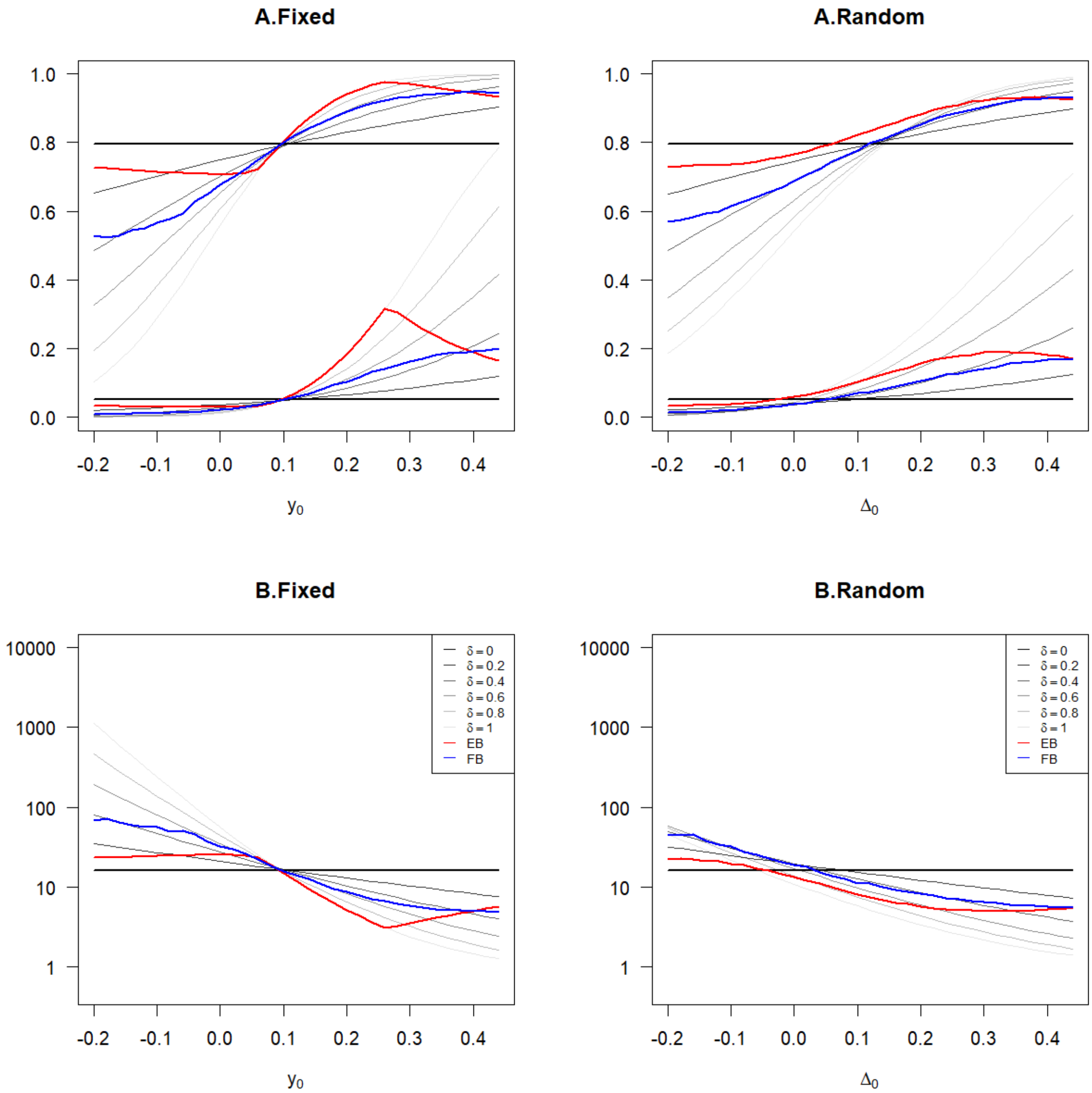

Figure 4.4: A. Power and type I error trade-off and **B.** Rejection ratio depending on historical data. **Fixed**: for fixed historical estimate y_0. **Random**: for randomly simulated historical estimate $y_0 \sim \mathrm{N}(\Delta_0, \frac{\sigma^2}{m_0})$

4.2 Two arm setting

The same plots as for the one arm setting were created for the two arm setting, and are presented on the figures on figures 4.5, 4.6, 4.8. However, there is an important difference between these two scenarios, which is reflected in the H_0 rejection rate function. In two arm setting, bias influences the posterior distribution of an outcome in a control arm and further on the posterior distribution of a treatment effect. In the one arm setting, bias influences directly on the posterior distribution of a treatment effect.

Figure 4.5 shows the H_0 rejection rate depending on the treatment effect $\Delta = \theta_c - \theta_\star$. Upper plots show the situation when the historical outcome estimate in the control arm is smaller than in a new experiment. The posterior distribution is shifted towards smaller values, and it leads to a shift of the posterior distribution for treatment effect towards larger values. It causes an increase in power accompanied by an increase in type I error. The lower plot shows an inverse situation when the outcome estimate in the control group from historical data is higher than from a new experiment. Then the posterior distribution of a treatment effect is shifted towards smaller values, which causes a decrease in power and type I error. In addition, power increases faster on the upper plots than type I error, and slower on the lower plots.

Figure 4.6 shows the dependence of power and type I error on the bias between historical and control data. Empirical Bayes and the full Bayes methods show adaptability in the case of high values of bias similar to the one arm case: the higher the bias, the less influence of historical information and power and type I error of these methods converge to the RCT values.

Figure 4.8 shows power and type I error trade-off on the upper plots and rejection ratio on the lower plots. Opposite to the one arm scenario, there is a little region around historical estimate where the integration of historical information increases power, while type I error is still below 5%. Thus, the Bayesian methods outperform RCT for these conditions. This region still exists when a historical estimate is generated randomly; however, it is smaller than for a fixed value. Empirical Bayes and the full Bayes approaches show better rejection ratio than RCT when the true value of a control group in a new experiment $\theta_\star$ is below the historical true value $\theta_0 = 10$ for a random case. Thus, power decreases slower than type I error, what could be an advantage for some studies.

In conclusion, in the one arm setting, there is no region of bias between historical and current experiment, where Bayesian methods would outperform RCT methods in power and type I error, however, in the two arm case, such a region can exist.

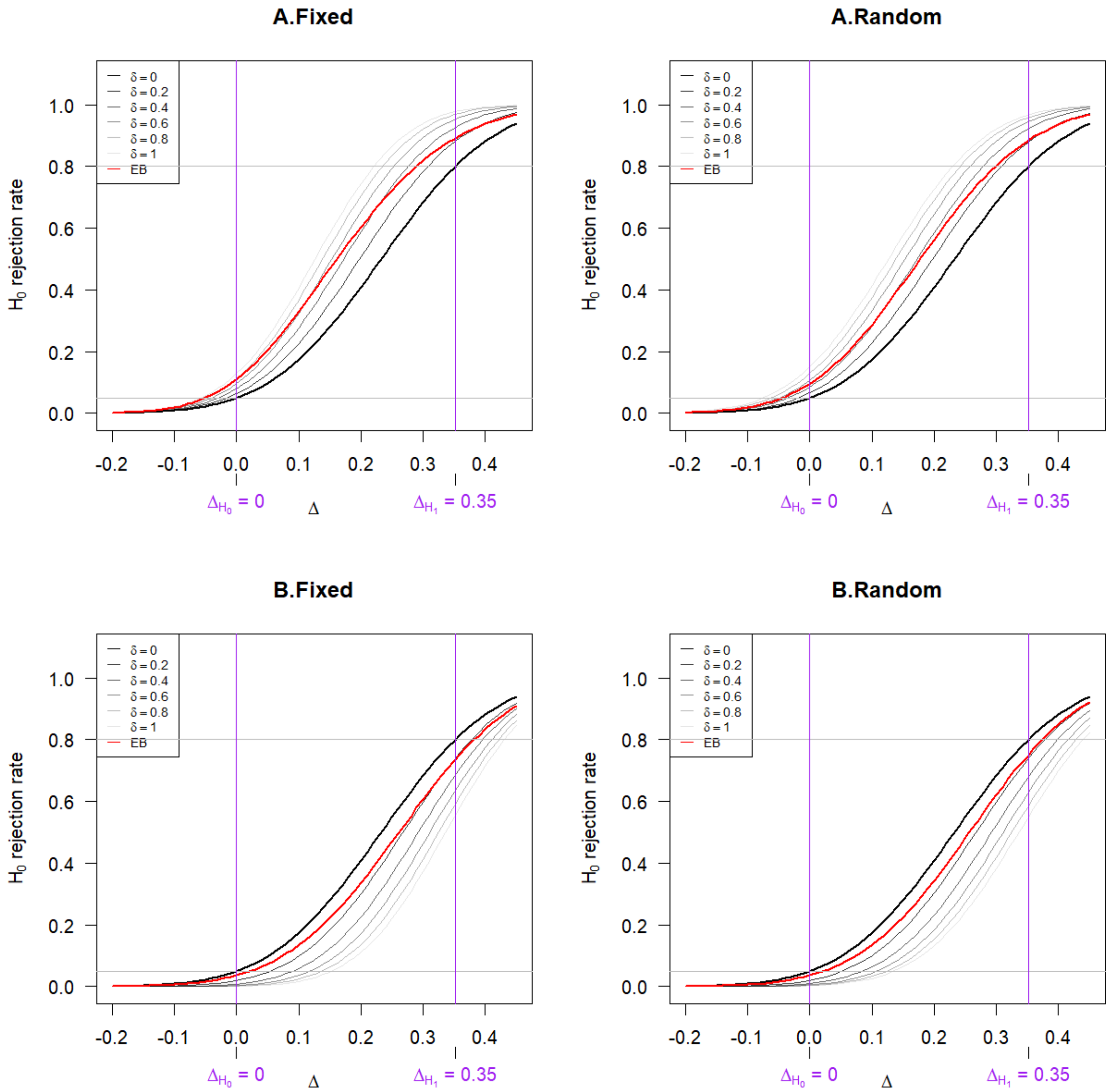

Figure 4.5: The H_0 rejection rate when historical estimate is **A.** lower than new true value in a control group: $\theta_\star = 10.15$ and $x_0 = 10$ **B.** higher than new true value in a control group: $\theta_\star = 9.73$ and $x_0 = 10$. **Fixed**: for fixed historical estimate x_0. **Random**: for randomly simulated historical estimate $x_0 \sim \mathrm{N}(10, \frac{\sigma^2}{n_0})$

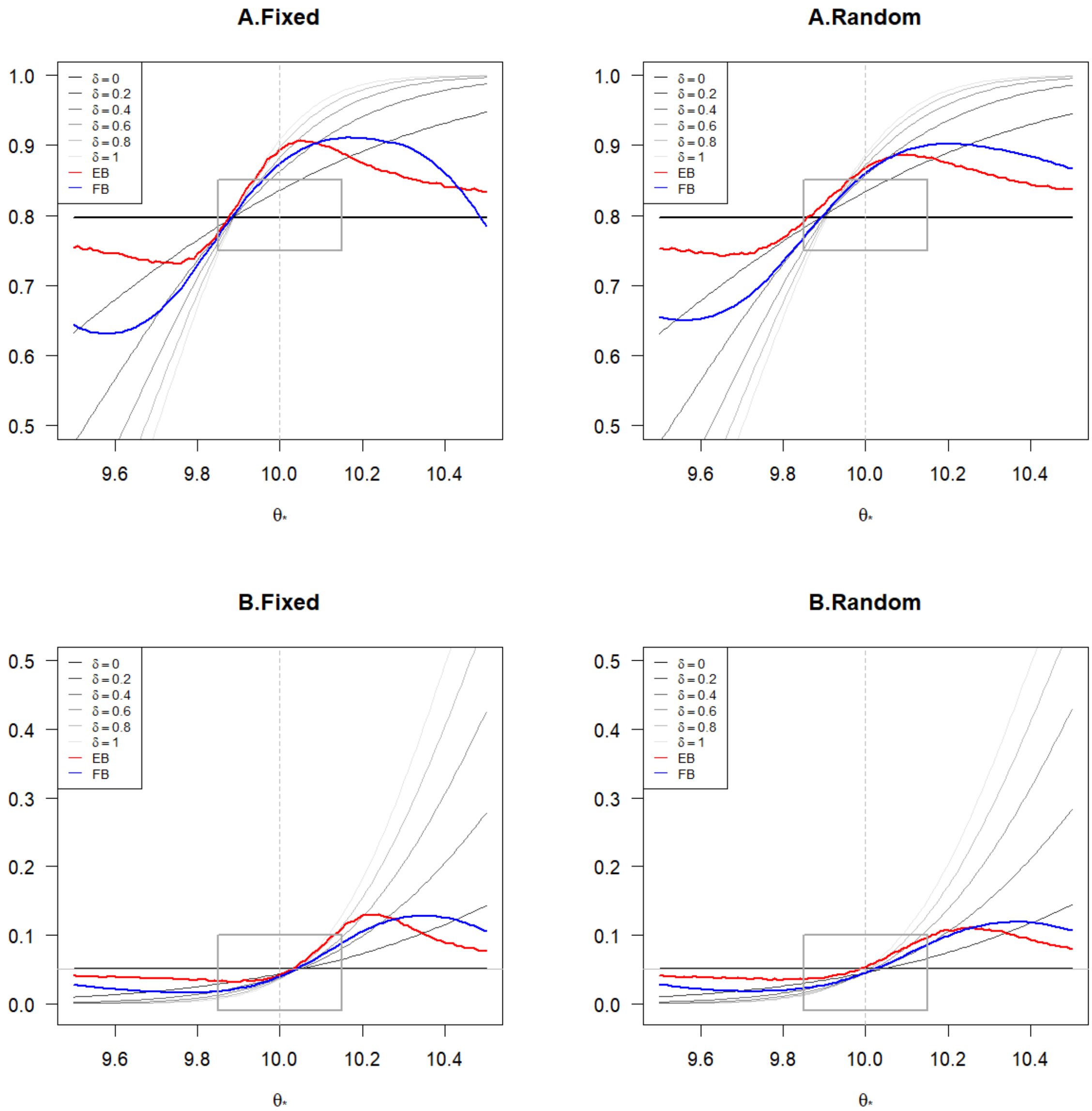

Figure 4.6: **A.** Power and **B.** Type I error depending on bias between historical estimate and new true value in a control group. **Fixed**: for fixed historical estimate. **Random**: for randomly simulated historical estimate. Grey squares show the area, which is shown on figure 4.7

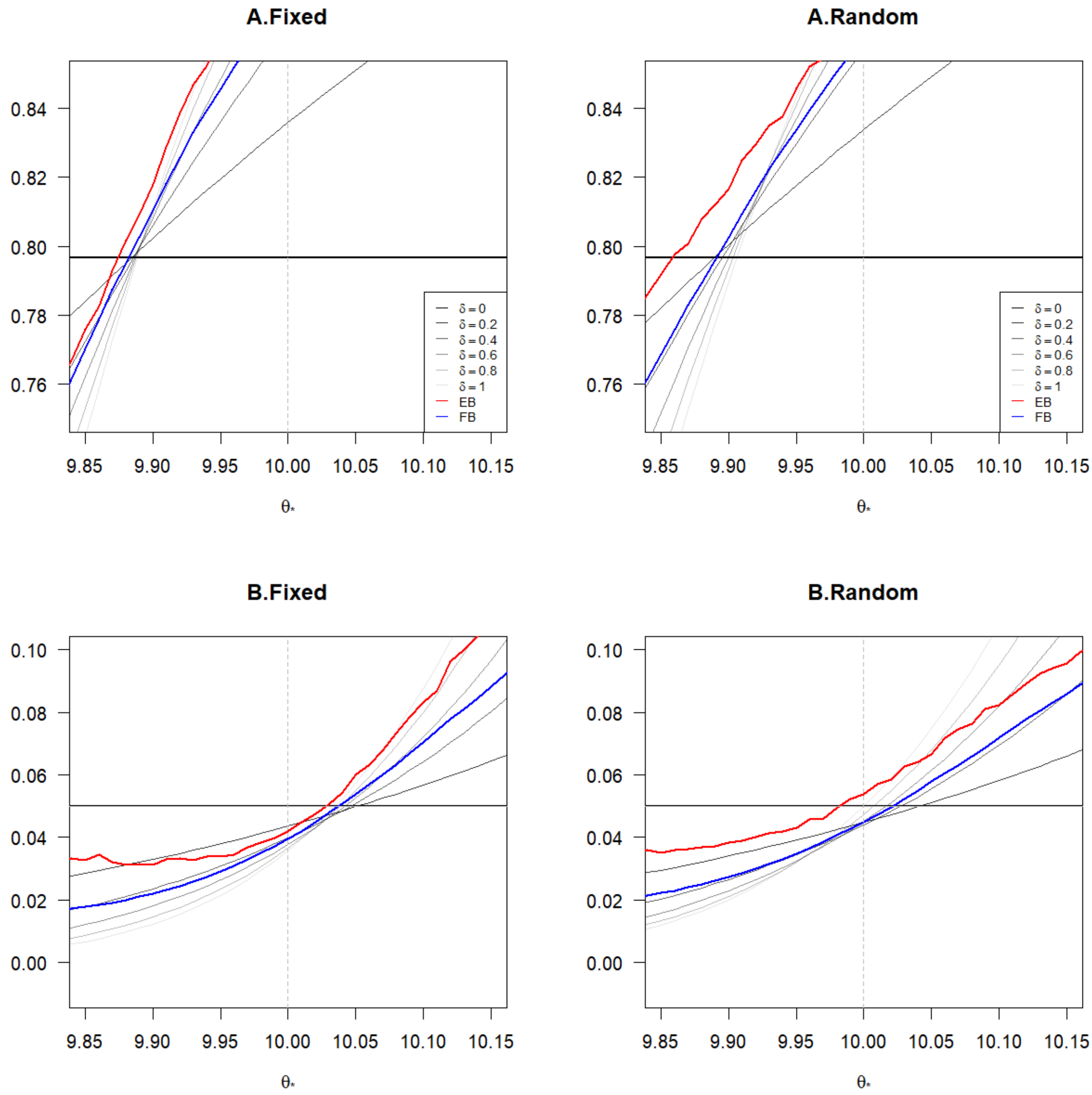

Figure 4.7: "Zoomed in" part of the figure 4.6. **A.** Power and **B.** Type I error depending on bias between historical estimate and new true value in a control group. **Fixed**: for fixed historical estimate. **Random**: for randomly simulated historical estimate.

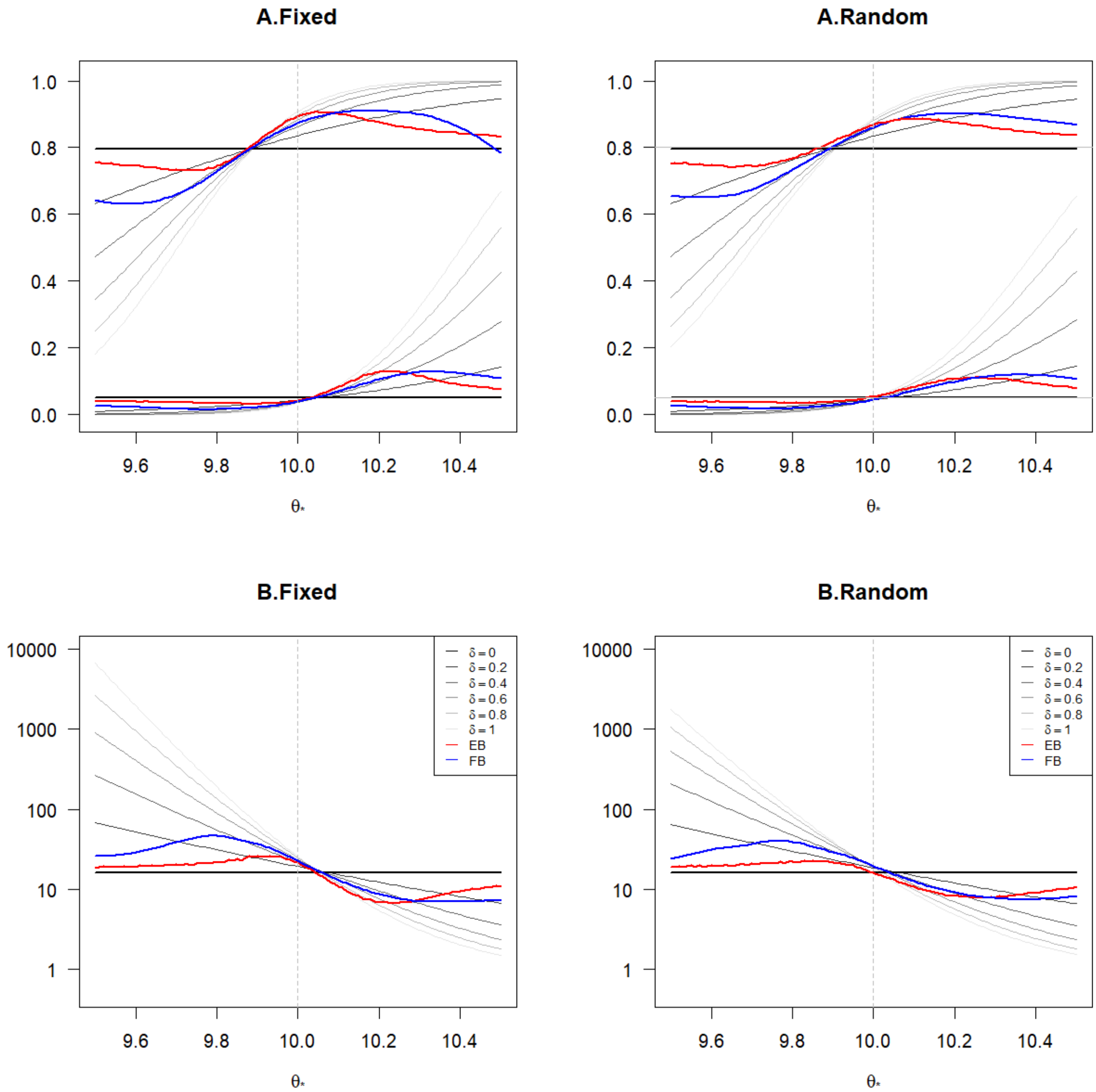

Figure 4.8: A. Power and type I error trade-off and **B.** Rejection ratio depending on bias between a historical estimate and a new true value in a control group. **Fixed**: for fixed historical estimate. **Random**: for randomly simulated historical estimate

4.3 Difference between the one arm and two arm settings

Different results for the trade-off between power and type I error in the one arm and two arm settings is connected to the way historical information influence the posterior distribution of a treatment effect. In this section we explain this difference for the conditional power prior method.

Figures 4.9, 4.10, 4.11, 4.12 show the results of multiple simulations of posterior distributions of a treatment effect in the one arm and two arm settings under null and alternative hypotheses. Based on Bayesian approach, we would reject H_0 hypothesis, where probability to have positive treatment effect is more than 95%. Since the variance of these posterior distributions stays constant for each simulation, it is enough to look at the distributions of the posterior means to understand the H_0 rejection rate of these simulations. In other words, if a posterior mean is above a test critical value, then H_0 is rejected.

Taking into account results of the derivation for the distributions of the posterior treatment effect from chapters 2 and 3, we can derive the distribution for the posterior means for the conditional power prior approach:

One arm :

$$E(\hat{\Delta}) \sim \mathrm{N}\left(\frac{m_\star}{\delta m_0 + m_\star}\Delta\star + \frac{\delta m_0}{\delta m_0 + m_\star} y_0, \frac{\sigma^2 m_\star}{(\delta m_0 + m_\star)^2}\right),$$

explanations for the notations can be found in section 2.

Two arm :

$$E(\hat{\Delta}) \sim \mathrm{N}\left(\Delta + \frac{\delta n_0}{\delta n_0 + n_\star}(\theta_\star - x_0), \frac{\sigma^2 n_\star}{(\delta n_0 + n_\star)^2} + \frac{\sigma^2}{n_t}\right),$$

explanations for the notations can be found in section 3.

Figure 4.13 shows the distribution of posterior means for the one arm and two arm settings. In one arm scenario, historical estimate exist directly for an effect and consequently can favor null or an alternative hypothesis but never both simultaneously. Thus, if the power increases, type I error increases too, and the other way around, if power decreases, type I error decreases. In the two arm settings, the historical data influence on the posterior distribution of the outcome in the control arm. When the bias is small, the variance of the outcome in the control arm is reduced, what leads to a smaller variance of the posterior distribution of the treatment effect under both, null and alternative hypotheses. Thus, in this situation, simultaneous improvement of power and type I error becomes possible. However, when the bias is large, posterior mean of the outcome shifts far from the true value and it leads to simultaneous increase (or decrease) in type I error and power.

Figure 4.14 shows the change of the H_0 rejection function when historical data is integrated. In one arm setting, a historical estimate can favor or null or alternative hypothesis, which causes a shift of the H_0 rejection function upwards or downwards from the RCT line. In the two arm setting, under the condition of a small bias between current and historical experiments, the H_0 rejection function changes its shape, which can lead to a simultaneous gain in power and type I error.

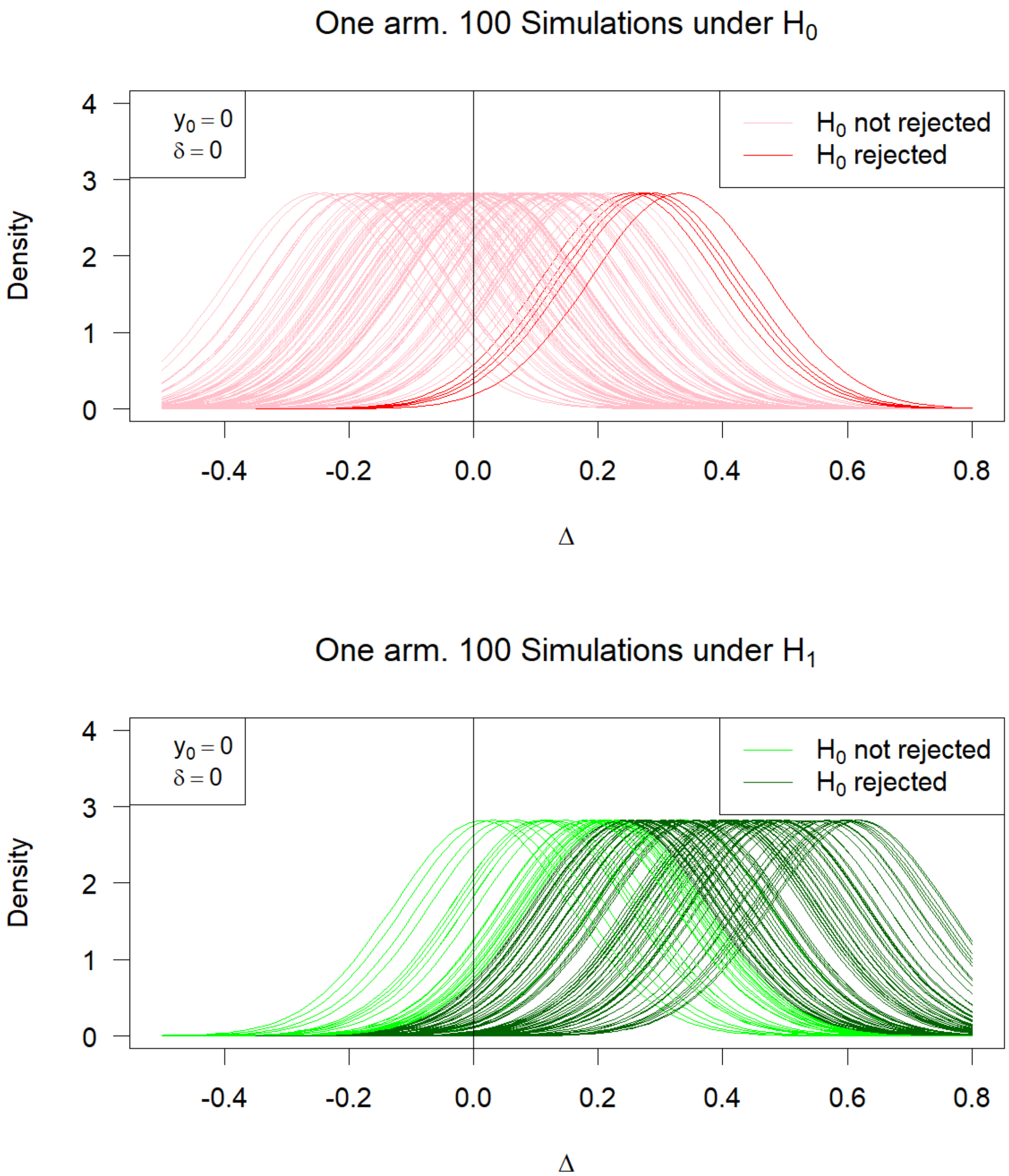

Figure 4.9: Posterior distributions of the treatment effect for 100 simulations of $y_\star$ without taking into account historical data.

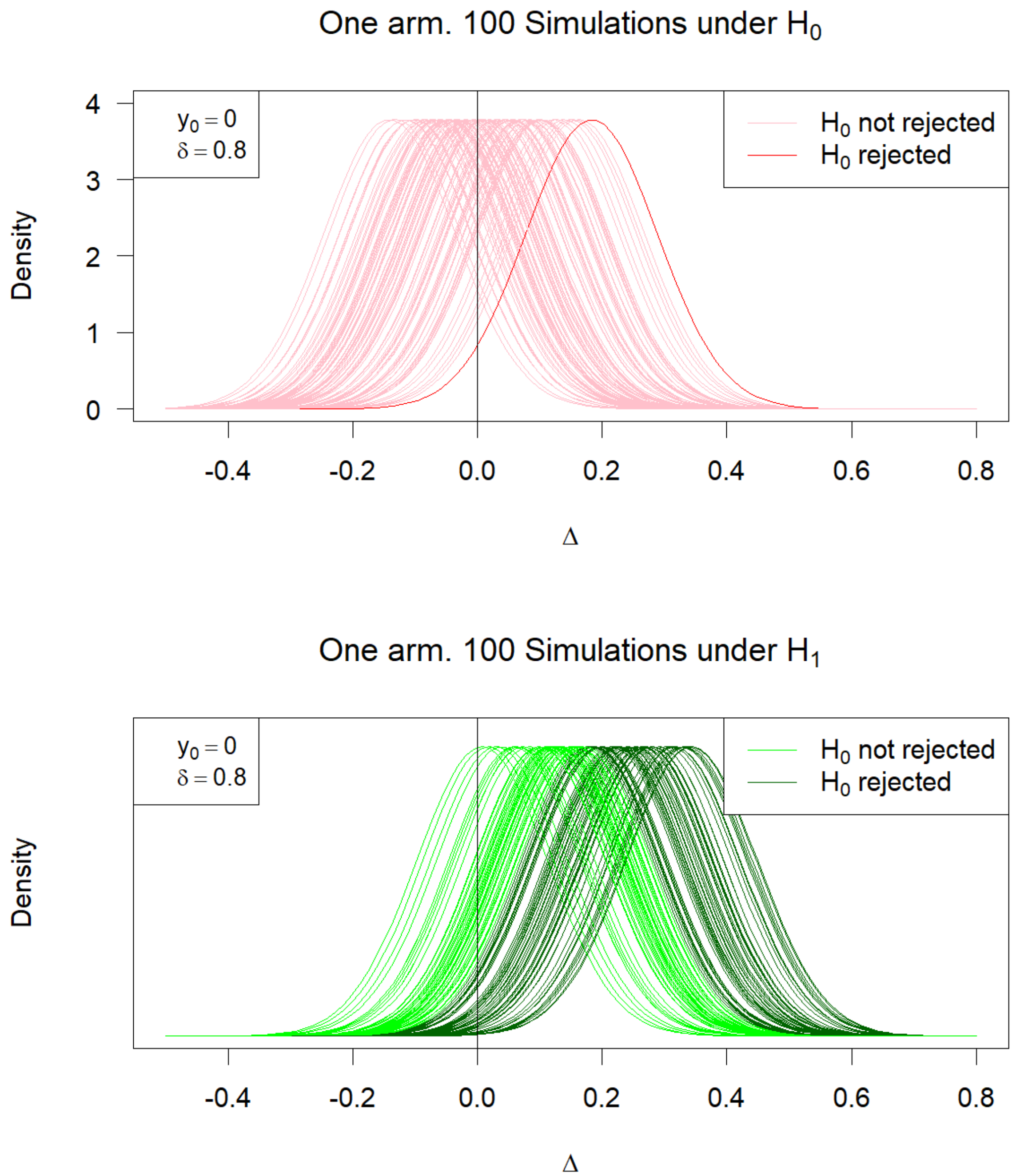

Figure 4.10: Posterior distributions of the treatment effect for 100 simulations of $y_\star$ with taking into account historical data using conditional power prior method

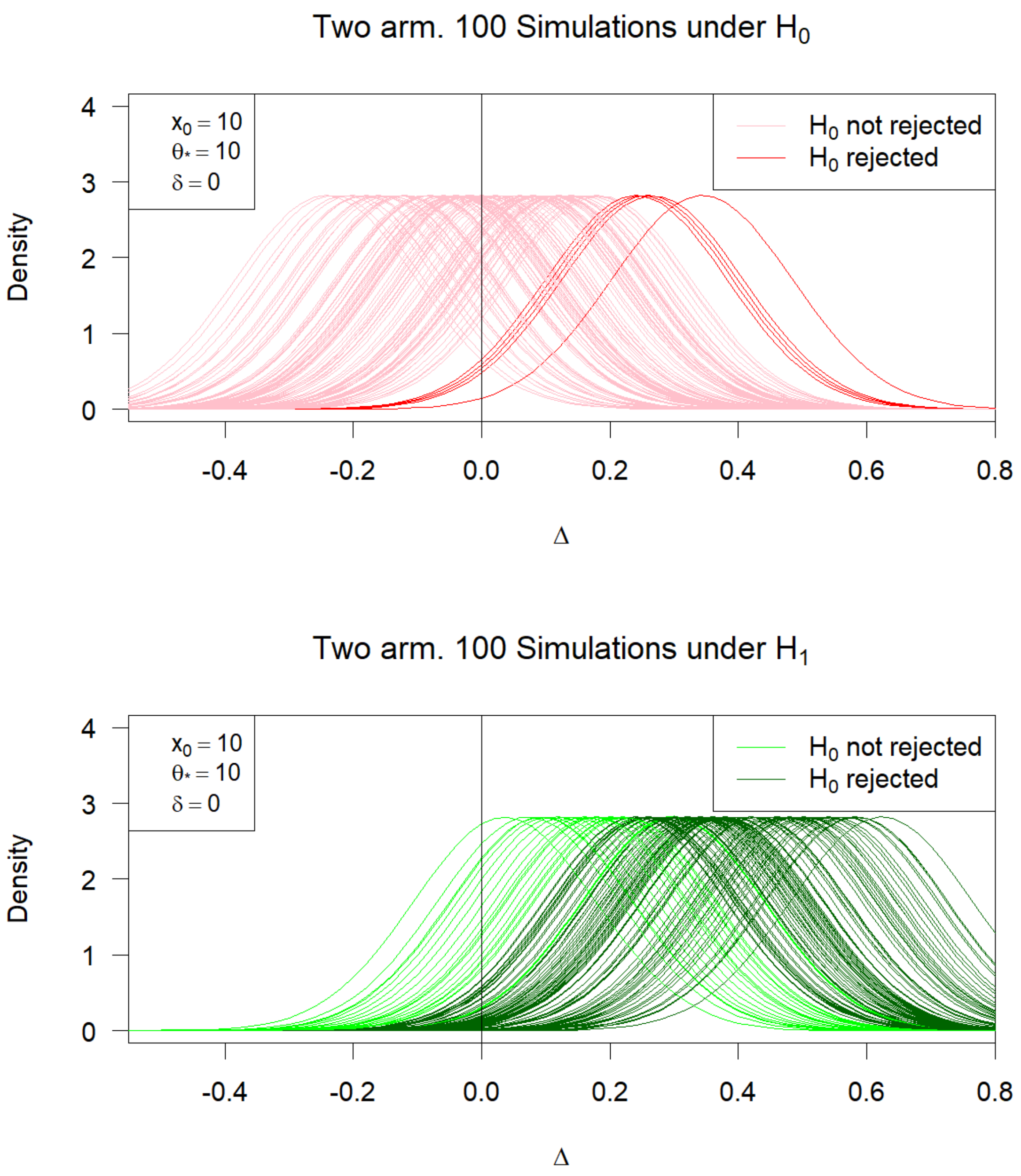

Figure 4.11: Posterior distributions of the treatment effect for 100 simulations of $x_\star$ and x_t without taking into account historical data.

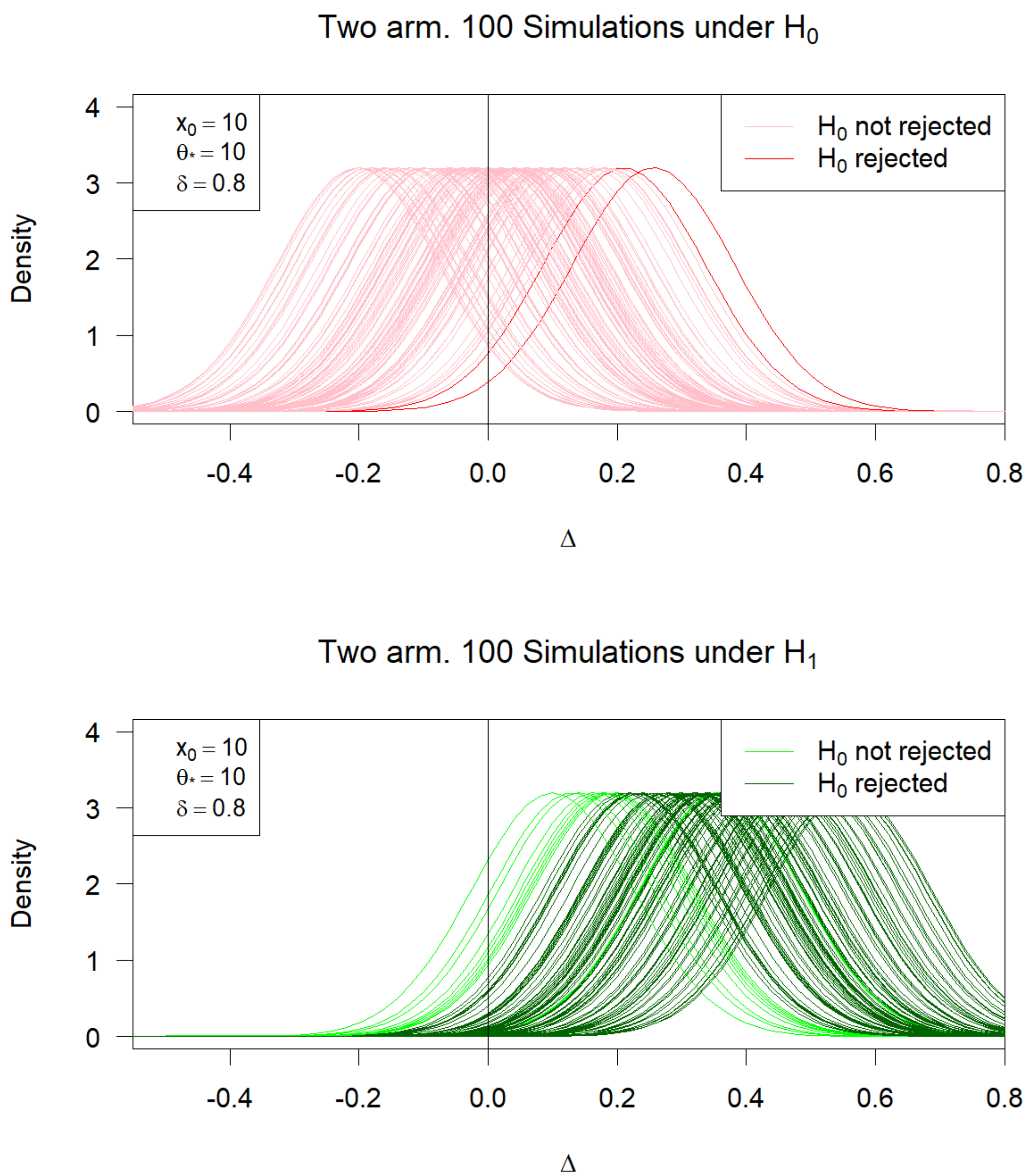

Figure 4.12: Posterior distributions of the treatment effect for 100 simulations of $x_\star$ and x_t with taking into account historical data using conditional power prior method. No bias between historical data and true value of current control was taken for this simulations.

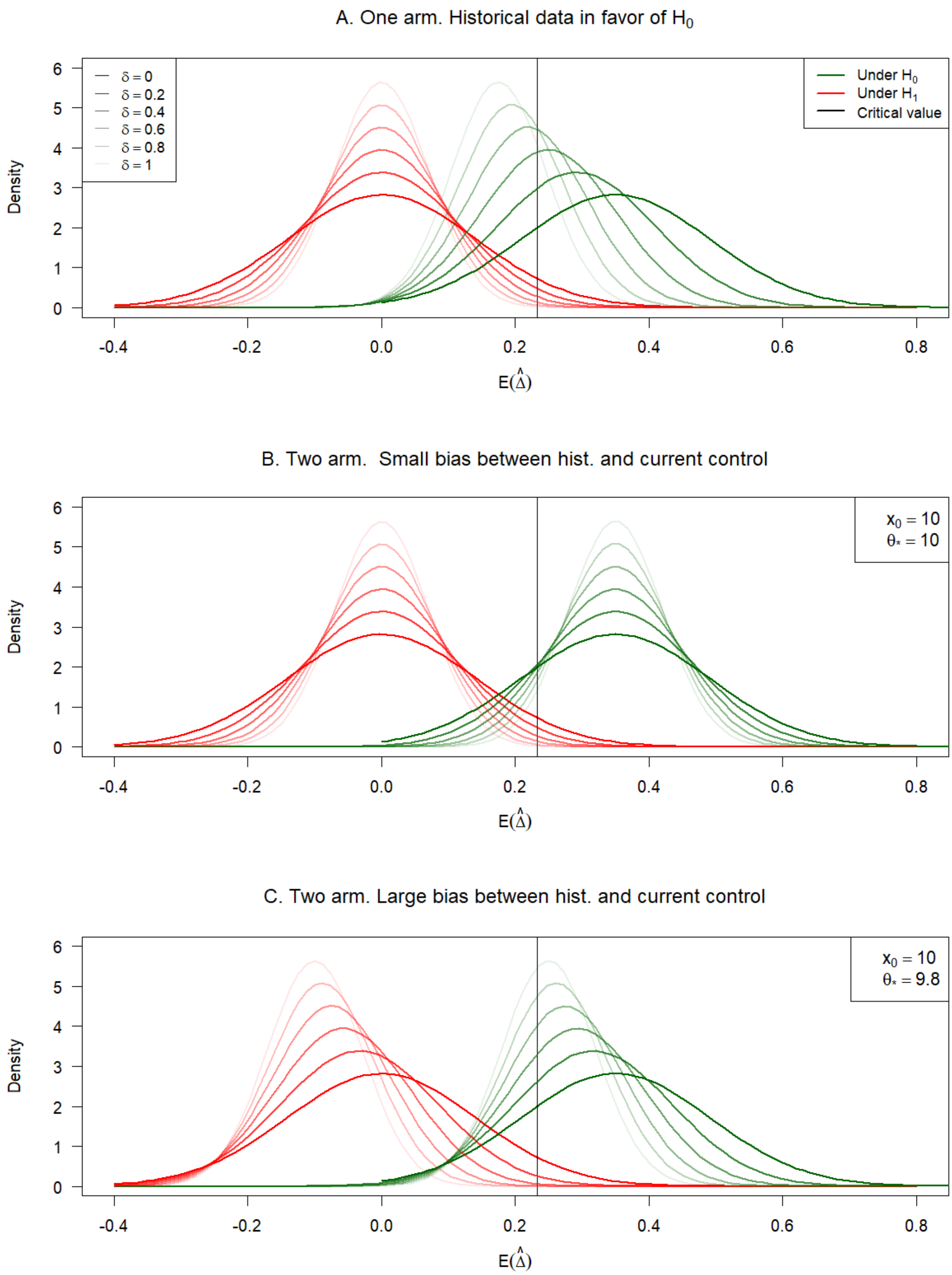

Figure 4.13: The distribution of posterior means of the treatment effect in the one arm and two arm settings

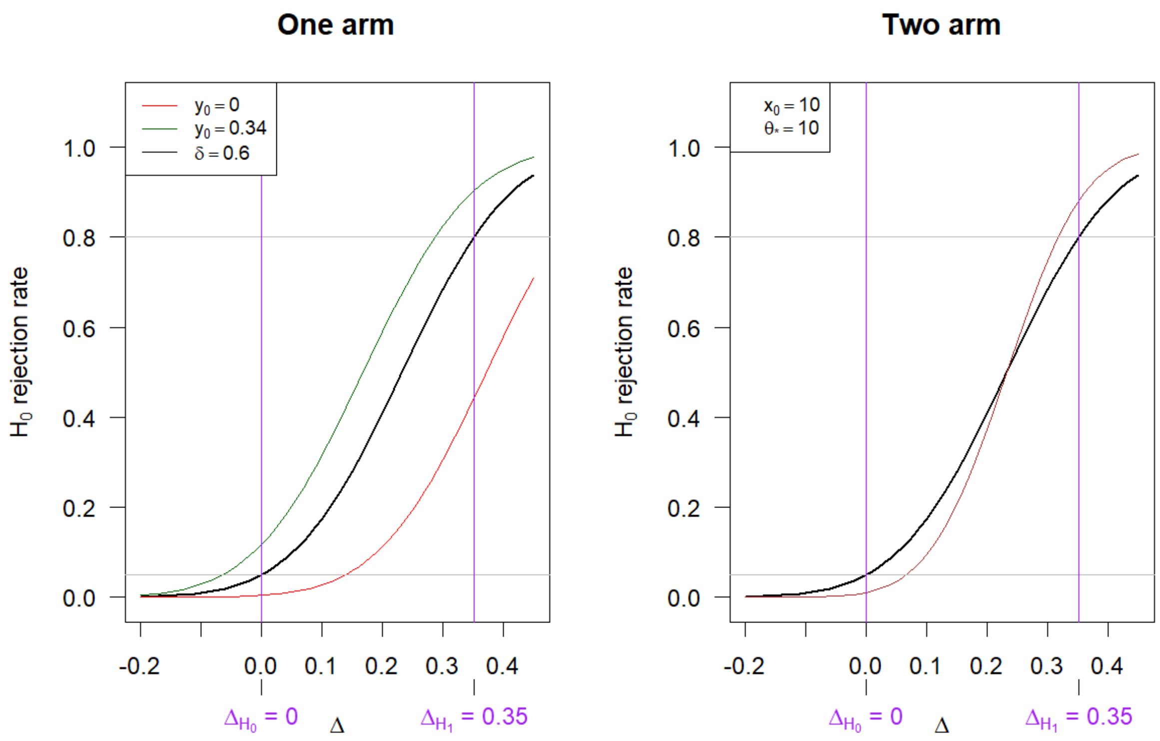

Figure 4.14: The difference in the influence of historical information on posterior distributions for one and two arm case. Historical information in one arm settings is available for the effect and can favor only the null or alternative hypothesis. In the two arm scenario, historical information can reduce the variance of the outcome in the control arm under the condition of the small bias, which leads to a simultaneous improvement of type I error and power

Additional characteristics for the two arm case

All approaches that incorporate historical data shift the posterior distribution towards the historical estimate, which causes adverse effects when the bias between a historical and a current experiment is high.

Figure 4.15 shows additional performance characteristics for the fixed power parameter and empirical Bayes approaches. When the bias between a historical estimate and a true value of a new experiment is small, the incorporation of historical information improves additional characteristics: bias between a true value and an inferred value (posterior mean), mean squared error (MSE) and coverage. However, when the bias is increasing, only an adaptive empirical Bayes method reduces the weight of historical data and mitigates the adverse effects of the bias.

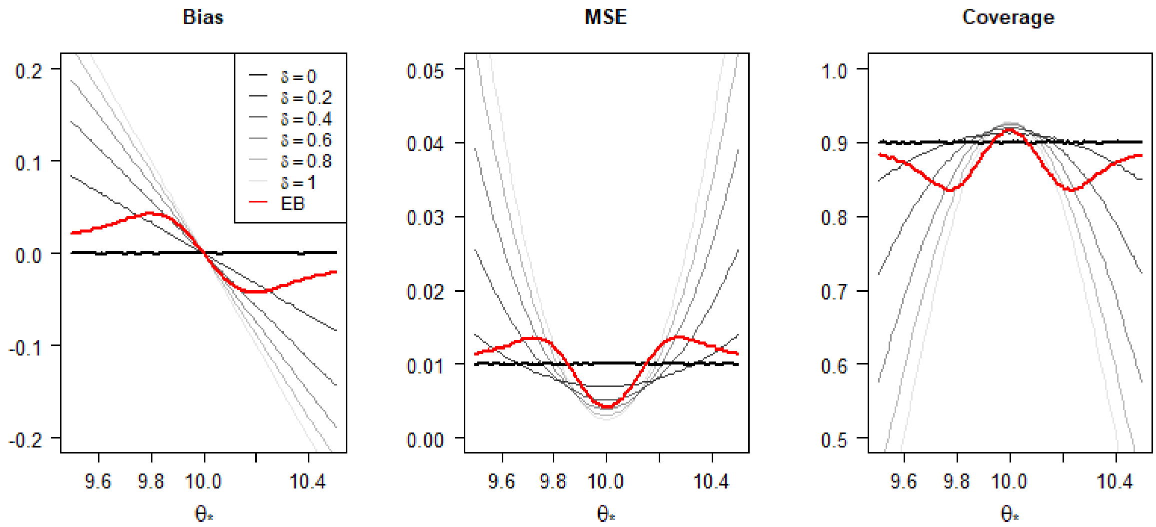

Figure 4.15: Additional performance characteristics of the Bayesian approaches

Chapter 5

Hypotheses testing with conditional power prior

In this chapter, we would like to explore the importance of a decision rule applied in our hypotheses testing. The results shown in the chapter 4 were created using the decision rule, where critical value depends on a new experiment only, what can lead to the inflation of type I error when historical information is used. However, it is possible to have a strict control of type I error if the decision rule is adapted for the historical estimate.

The decision rule of the test is defined by a value C, which is a critical value of a test statistic. For the hypotheses $H_0 : \Delta \leq 0$ and $H_1 : \Delta > 0$, if the test statistic $T(y_\star)$ of an experiment exceeds a critical value then H_0 is rejected. The probability of a type I error, rejection of true H_0 is the (nominal) size of the test. The power of the test is the probability of the rejection of a false H_0.

$$\begin{aligned} \alpha &= \mathsf{Pr}(\text{reject } H_0 : T(y_\star) > C \mid H_0) \\ 1 - \beta &= \mathsf{Pr}(\text{reject } H_0 : T(y_\star) \leq C \mid H_1) \end{aligned}$$

Type I error and power are connected through a critical value C. Thus, decrease in the value C causes decrease in α and increase in β. A usual solution is to fix α and then to find related C and β.

In order to visually understand the difference between two decision rules considered in this master thesis, we need to look again at the distribution of the posterior means described in the section 4.3. Figure 5.1 shows the distribution for the posterior means for the one arm settings, when historical estimate favors H_1 hypothesis. Blue line shows the critical value, which depends only on the current experiment. If we continue to use this critical value and integrate historical information, then we allow inflation of type I error. The second decision rule, which is adaptable for a historical estimate, allows a strict control of type I error. Related critical values are shown with different shadow of grey.

5.1 One arm case

5.1.1 UMP test

A uniformly most powerful (UMP) test is a hypothesis test that has the greatest power $1 - \beta$ among all possible tests of a given size α.

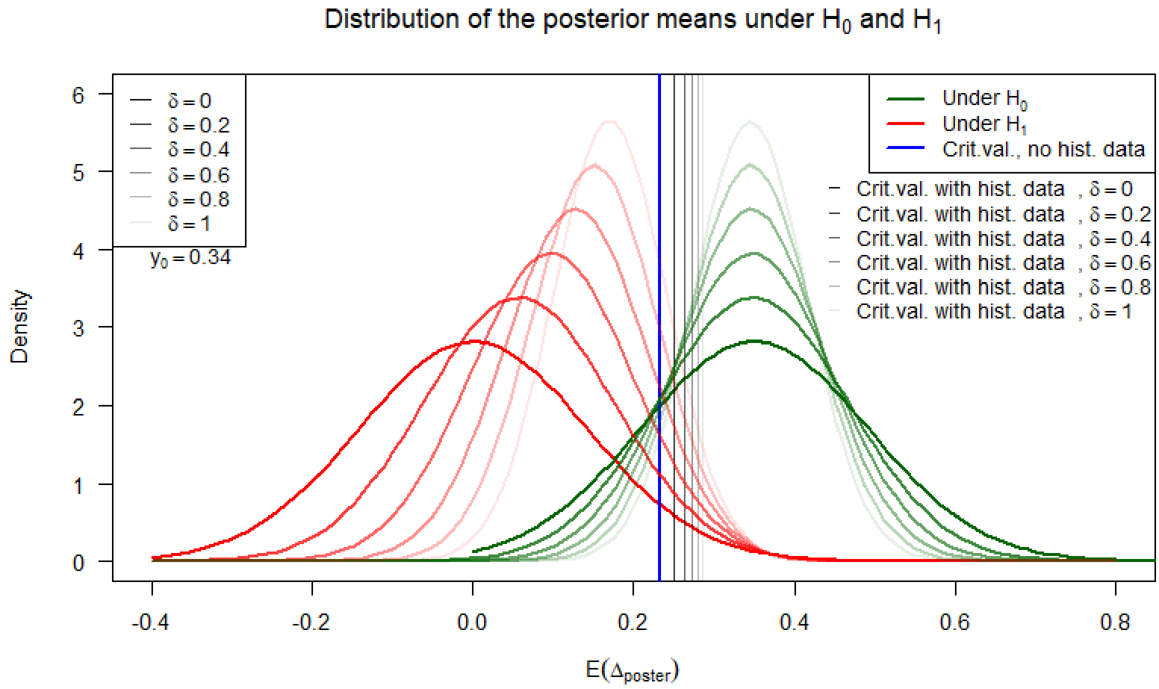

Figure 5.1: The distribution of posterior means of the treatment effect in the one arm settings and two different types of the decision rules: dependent and independent of historical information

$$\phi_{UMP}(y_\star) = \begin{cases} 1 & \text{if } Y_\star > C \\ 0 & \text{if } Y_\star \leq C \end{cases} \tag{5.1}$$

where 1 is for a decision to reject H_0 and 0 is for a decision not to reject H_0, $Y_\star$ is a sufficient statistics. In the one arm scenario, where $Y_\star$ is the effect estimate from a new experiment, C can be defined such that this test controls type I error at the level α.

$$\begin{aligned}\alpha &= \Pr(Y_\star > C \mid H_0) \\ &= 1 - \Pr\left(\frac{Y_\star - \Delta_\star}{\sigma/\sqrt{m_\star}} < \frac{C - \Delta_\star}{\sigma/\sqrt{m_\star}} \mid H_0\right) \\ &= 1 - \Phi\left[\frac{C}{\sigma/\sqrt{m_\star}}\right]\end{aligned}$$

Thus,

$$C = -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{m_\star}} \tag{5.2}$$

The UMP test decision rule for the one arm case is :

$$\phi_{UMP}(y_\star) = \begin{cases} 1 & \text{if } Y_\star > -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{m_\star}} \\ 0 & \text{if } Y_\star \leq -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{m_\star}} \end{cases} \tag{5.3}$$

where α is a prescpecified significance level of the test.

$$\phi_{UMP}(y_\star) = \begin{cases} 1 & \Phi\left[\frac{y_\star\sqrt{m_\star}}{\sigma}\right] > 1-\alpha \\ \\ 0 & \Phi\left[\frac{y_\star\sqrt{m_\star}}{\sigma}\right] \leq 1-\alpha \end{cases} \tag{5.4}$$

5.1.2 Bayesian test with no historical data

In the supplementary materials of the paper Psioda and Ibrahim (2019), it is shown that in the one arm scenario, in order to ensure a strict frequentist type I error control prior information should not be used. We repeat here their proof for a fixed δ, adapting it for our notations.

The decision rule of the Bayesian approach depends on the posterior distribution of the treatment effect:

$$\phi_B(y_\star) = \begin{cases} 1 & \text{if } P(\Delta > 0 \,|\, y_\star) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \,|\, y_\star) \leq \phi_0 \end{cases} \tag{5.5}$$

where ϕ_0 is the posterior probability critical value, which is independent of historical information.

First, authors Psioda and Ibrahim (2019) proved that: "Inference based on the Bayesian rejection rule is identical to the frequentist one sided, level α UMP test when $\phi_0 = 1-\alpha$".

The type I error of the test is $\mathrm{E}[1\{P(\Delta > 0 \,|\, y_\star) \geq \phi_0 \,|\, H_0\}]$. We need to find ϕ_0 such that type I error of the test $= \alpha$.

$$\begin{aligned} P(\Delta > 0 \,|\, y_\star) \geq \phi_0 &\Leftrightarrow P\left(\frac{\Delta - y_\star}{\sigma/\sqrt{m_\star}} > \frac{0 - y_\star}{\sigma/\sqrt{m_\star}} \,\Big|\, y_\star\right) \geq \phi_0 \\ &\Leftrightarrow P\left(Z_1 > \frac{0 - y_\star}{\sigma/\sqrt{m_\star}} \,\Big|\, y_\star\right) \geq \phi_0 \\ &\Leftrightarrow P\left(Z_2 < \frac{y_\star}{\sigma/\sqrt{m_\star}} \,\Big|\, y_\star\right) \geq \phi_0 \\ &\Phi\left[\frac{y_\star\sqrt{m_\star}}{\sigma}\right] \geq \phi_0, \end{aligned}$$

where Z_1 and Z_2 are standard normal random variables.

The last equality is the same as in the one sided normal UMP test (5.4). Thus, the authors Psioda and Ibrahim (2019) concluded that : "optimal choice of ϕ_0 that satisfies $\Phi^{-1}(\phi_0) = Z_{1-\alpha}$, which implies $\phi_0 = 1-\alpha$".

In short, it was shown that when no historical information is used, $\delta = 0$, then the most unbiased Bayesian test procedure requires $\phi_0 = 1-\alpha$.

5.1.3 Bayesian test with historical data

The Bayesian decision rule, which incorporates historical data, is :

$$\phi_{B_\delta}(y_\star, y_0) = \begin{cases} 1 & \text{if } P(\Delta > 0 \,|\, y_\star, y_0, \delta) > \phi_\delta \\ 0 & \text{if } P(\Delta > 0 \,|\, y_\star, y_0, \delta) \leq \phi_\delta \end{cases} \tag{5.6}$$

where $\Delta \,|\, y_0, y_\star, \delta \sim \mathrm{N}(\mu_{0\star}, \sigma_{0\star})$ with $\mu_{0\star} = \frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star}$ and $\sigma_{0\star} = \frac{\sigma^2}{\delta m_0 + m_\star}$; ϕ_δ is the posterior probability critical value dependent on the historical information.

Authors Psioda and Ibrahim (2019) showed that, for a strict control of type I error at level α in a Bayesian testing, a critical value ϕ_δ is $\Phi\left[\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{m_\star}}{\sigma} + \frac{y_0 \delta m_0}{\sigma^2}\right)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}}\right]$.

Again, we fix type I error of the test as α:

$$\mathrm{E}[1\{P(\Delta > 0 \,|\, y_\star, y_0, \delta) \geq \phi_\delta \,|\, \Delta_{\text{true}} = 0, y_0, \delta\}] = \alpha$$

$$\begin{aligned}
P(\Delta > 0 \,|\, y_\star, y_0, \delta \geq \phi_\delta) &\Leftrightarrow P\left(\frac{\Delta - \mu_{0\star}}{\sigma_{0\star}} > \frac{0 - \mu_{0\star}}{\sigma_{0\star}} \,\middle|\, y_\star, y_0, \delta\right) \geq \phi_\delta \\
&\Leftrightarrow P\left(Z_1 > \frac{0 - \mu_{0\star}}{\sigma_{0\star}} \,\middle|\, y_\star, y_0, \delta\right) \geq \phi_\delta \\
&\Leftrightarrow P\left(Z_2 < \frac{\mu_{0\star}}{\sigma_{0\star}} \,\middle|\, y_\star, y_0, \delta\right) \geq \phi_\delta \\
&\Leftrightarrow \Phi\left[\frac{\mu_{0\star}}{\sigma_{0\star}}\right] \geq \phi_\delta \\
&\Leftrightarrow \frac{\mu_{0\star}}{\sigma_{0\star}} \geq \Phi^{-1}(\phi_\delta) \\
&\Leftrightarrow \frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star} \geq \Phi^{-1}(\phi_\delta)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}} \\
&\Leftrightarrow \delta m_0 y_0 + m_\star y_\star \geq \Phi^{-1}(\phi_\delta)\sigma\sqrt{\delta m_0 + m_\star} \\
&\Leftrightarrow \frac{y_\star\sqrt{m_\star}}{\sigma} \geq \frac{1}{\sigma\sqrt{m_\star}}\left(\Phi^{-1}(\phi_\delta)\sigma\sqrt{\delta m_0 + m_\star} - \delta m_0 y_0\right) \\
&\Leftrightarrow \frac{y_\star\sqrt{m_\star}}{\sigma} \geq \frac{\sigma}{\sqrt{m_\star}}\left(\frac{\Phi^{-1}(\phi_\delta)}{\sigma/\sqrt{\delta m_0 + m_\star}} - \frac{y_0}{\sigma^2/\delta m_0}\right)
\end{aligned}$$

This inequality is also a rejection rule for the one-sided frequentist UMP test. Thus:

$$Z_{1-\alpha} = \Phi^{-1}(1-\alpha) = \frac{\sigma}{\sqrt{m_\star}}\left(\frac{\Phi^{-1}(\phi_\delta)}{\sigma/\sqrt{\delta m_0 + m_\star}} - \frac{y_0}{\sigma^2/\delta m_0}\right)$$

We can derive an updated posterior probability critical value ϕ_δ :

$$\begin{aligned}
\phi_\delta &= \Phi\left[\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{m_\star}}{\sigma} + \frac{y_0 \delta m_0}{\sigma^2}\right)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}}\right] \\
&= \Phi\left[\Phi^{-1}(1-\alpha)\left(\frac{\sqrt{m_\star}}{\sigma} + \frac{y_0 \delta m_0}{\sigma^2 \Phi^{-1}(1-\alpha)}\right)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}}\right]
\end{aligned}$$

Naturally, the power function also changes when the posterior probability critical value is changed from ϕ_0 to ϕ_δ. Authors Psioda and Ibrahim (2019) showed, that: "the power functions for the set of Bayesian hypothesis tests based on the rejection rules $P(\Delta > 0 \,|\, y_\star, y_0, \delta) \geq \phi_\delta, \delta \in [0,1]$ are identical". It means that power function for any arbitrary δ is identical to the power function with $\delta = 0$ if the posterior probability critical value ϕ_δ, which is dependent on historical data, is used. Thus, historical information is not used when the rejection rule is adapted to ensure a strict control of type I error for any $\delta \in [0,1]$.

$$
\begin{aligned}
P(\Delta > 0 \,|\, y_\star, y_0, \delta) \geq \phi_\delta &\Leftrightarrow \Phi\left[\frac{\mu_{0\star}}{\sigma_{0\star}}\right] \geq \Phi\left[\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{m_\star}}{\sigma} + \frac{y_0\delta m_0}{\sigma^2}\right)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}}\right] \\
&\Leftrightarrow \frac{\delta m_0 y_0 + m_\star y_\star}{\delta m_0 + m_\star} \geq \left(\frac{\Phi^{-1}(1-\alpha)\sqrt{m_\star}}{\sigma} + \frac{y_0\delta m_0}{\sigma^2}\right)\frac{\sigma^2}{\delta m_0 + m_\star} \\
&\Leftrightarrow \frac{y_\star\sqrt{m_\star}}{\sigma} \geq \Phi^{-1}(1-\alpha) \\
&\Leftrightarrow P(\Delta > 0 \,|\, y_\star) \geq \phi_0
\end{aligned}
$$

In short, if type I error has to be strictly fixed as α (for example 5%), then the Bayesian decision rule should be updated according to historical information. However it leads to a simultaneous update of the power function, which, in turn, will be equal to the power function without any extra information.

Summary table

Table 5.1 shows different approach to hypotheses testing in the one arm scenario.

UMP test	Bayesian decision rule with no hist. data	Bayesian decision rule with hist.data and fixed δ
$\phi_{UMP}(y_\star) = \begin{cases} 1 & \Phi\left[\frac{Y_\star\sqrt{m_\star}}{\sigma}\right] > 1-\alpha \\ 0 & \Phi\left[\frac{Y_\star\sqrt{m_\star}}{\sigma}\right] \leq 1-\alpha \end{cases}$	$\phi_B(y_\star) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid y_\star) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \mid y_\star) \leq \phi_0 \end{cases}$ where $\phi_0 = 1-\alpha$	If the critical value of a decision rule is independent of hist. data $\phi_{B_0}(y_\star, y_0) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid y_\star, y_0, \delta) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \mid y_\star, y_0, \delta) \leq \phi_0 \end{cases}$ where $\phi_0 = 1-\alpha$ If the decision rule is adapted according to hist. data to have a strict control of type I error $\phi_{B_\delta}(y_\star, y_0) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid y_\star, y_0, \delta) > \phi_\delta \\ 0 & \text{if } P(\Delta > 0 \mid y_\star, y_0, \delta) \leq \phi_\delta \end{cases}$ where $\phi_\delta = \Phi\left[\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{m_\star}}{\sigma} + \frac{y_0\delta m_0}{\sigma^2}\right)\frac{\sigma}{\sqrt{\delta m_0 + m_\star}}\right]$

Table 5.1: Hypotheses testing for one arm case

Figure 5.2 shows the results of the simulation. Left plots show the case, when the posterior probability critical value is adapted for historical information, thus the decision rule strictly controls type I error. Right plots show the situation, when the posterior probability critical value depends only on a current experiment, thus type I error is inflated when the bias is large between historical and current experiment.

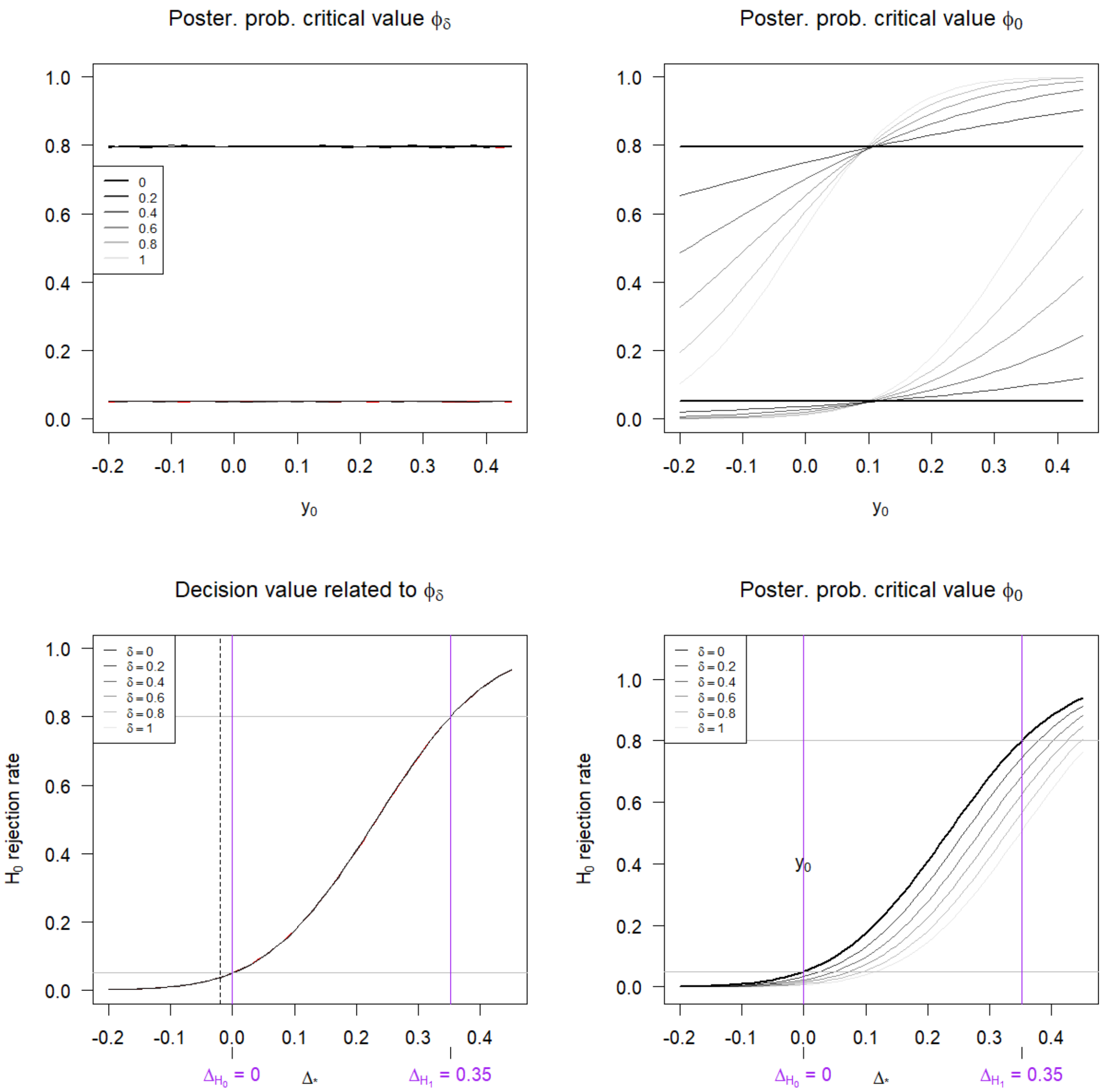

Figure 5.2: Change of power and type I error for different values of historical data depending on decision rule for the one arm setting. Upper plots show power and type I error trade off. Lower plots show the H_0 rejection function

5.2 Two arm case

5.2.1 UMP test

Similarly to the one arm scenario the decision rule for UMP test was derived for two-arm case, using the condition that the test controls type I error at the level α.

$$\begin{aligned}
\alpha &= \Pr\left(\frac{(x_t - x_\star)}{\sigma\sqrt{(n_t+n_\star)/n_t n_\star}} > C \,\middle|\, H_0\right)\\
&= 1 - \Pr\left(\frac{(x_t - x_\star) - \Delta_\star}{\sigma\sqrt{(n_t+n_\star)/n_t n_\star}} < \frac{C - \Delta_\star}{\sigma\sqrt{(n_t+n_\star)/n_t n_\star}} \,\middle|\, H_0\right)\\
&= 1 - \Phi\left[\frac{C}{\sigma\sqrt{(n_t+n_\star)/n_t n_\star}}\right]
\end{aligned}$$

Thus,

$$C = -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n_t n_\star/(n_t+n_\star)}}$$

and the rule can be rewritten :

$$\phi_{UMP}(y_\star) = \begin{cases} 1 & x_t - x_\star > -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n_t n_\star/(n_t+n_\star)}} \\ 0 & x_t - x_\star \le -\Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n_t n_\star/(n_t+n_\star)}} \end{cases} \tag{5.7}$$

$$\phi_{UMP}(y_\star) = \begin{cases} 1 & \Phi\left[\frac{x_t - x_\star}{\sigma}\sqrt{\frac{n_t n_\star}{n_t+n_\star)}}\right] > 1-\alpha \\ \\ 0 & \Phi\left[\frac{x_t - x_\star}{\sigma}\sqrt{\frac{n_t n_\star}{n_t+n_\star)}}\right] \le 1-\alpha \end{cases} \tag{5.8}$$

5.2.2 Bayesian test with no historical data

Since only the variance has changed for the two arm case, and we consider it as fixed value, the decision rule is as in the one arm case.

$$\phi_B(y_\star) = \begin{cases} 1 & \text{if } P(\Delta > 0 \,|\, x_t, x_\star) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \,|\, x_t, x_\star) \le \phi_0 \end{cases},$$

where $\phi_0 = 1 - \alpha$

5.2.3 Bayesian test with historical data

We adapted the derivation used in the one arm case, which were based on the supplementary materials of the paper Psioda and Ibrahim (2019) for the two arm case.

The Bayesian decision rule for the analysis with historical information is:

$$\phi_{B_\delta}(x_t, x_\star, x_0, \delta) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) > \phi_\delta \\ 0 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) \leq \phi_\delta \end{cases}, \tag{5.9}$$

where $\Delta \mid x_t, x_\star, x_0, \delta \sim \mathrm{N}(\mu_{0\star}, \sigma_{0\star})$ with $\mu_{0\star} = x_t - \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star}$ and $\sigma_{0\star} = \frac{\sigma^2}{n_t} + \frac{\sigma^2}{\delta n_0 + n_\star}$, ϕ_δ is the posterior probability critical value established to control type I error at the level α.

To find ϕ_δ we must solve this equation:

$$\mathrm{E}[1\{P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) \geq \phi_\delta \mid \Delta_{\text{true}} = 0, x_0, \delta\}] = \alpha \tag{5.10}$$

$$\begin{aligned} P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) \geq \phi_\delta &\Leftrightarrow P\left(\frac{\Delta - \mu_{0\star}}{\sigma_{0\star}} > \frac{0 - \mu_{0\star}}{\sigma_{0\star}} \;\middle|\; x_t, x_\star, x_0, \delta\right) \geq \phi_\delta \\ &\Leftrightarrow P\left(Z_1 > \frac{0 - \mu_{0\star}}{\sigma_{0\star}} \;\middle|\; x_t, x_\star, x_0, \delta\right) \geq \phi_\delta \\ &\Leftrightarrow P\left(Z_2 < \frac{\mu_{0\star}}{\sigma_{0\star}} \;\middle|\; x_t, x_\star, x_0, \delta\right) \geq \phi_\delta \\ &\Leftrightarrow \Phi\left[\frac{\mu_{0\star}}{\sigma_{0\star}}\right] \geq \phi_\delta \\ &\Leftrightarrow \frac{\mu_{0\star}}{\sigma_{0\star}} \geq \Phi^{-1}(\phi_\delta) \\ &\Leftrightarrow x_t - \frac{\delta n_0 x_0 + n_\star x_\star}{\delta n_0 + n_\star} \geq \Phi^{-1}(\phi_\delta)\sigma\sqrt{\frac{1}{n_t} + \frac{1}{\delta n_0 + n_\star}} \end{aligned}$$

If we try to revise this formula into the decision rule for the UMP test for two arm scenario (5.8):

$$\frac{x_t - x_\star}{\sigma}\sqrt{\frac{n_t n_\star}{n_t + n_\star}} \geq \sqrt{\frac{n_t n_\star}{n_t + n_\star}}\left(\Phi^{-1}(\phi_\delta)\left(\frac{\delta n_0}{n_\star} + 1\right)\sqrt{\frac{1}{n_t} + \frac{1}{\delta n_0 + n_\star}} - \frac{\delta n_0 (x_t - x_0)}{\sigma n_\star}\right)$$

We see, that the right side still contains x_t. Technically, we could still derive ϕ_δ, which would give a strict control of type I error, but would still depend on x_t. Thus, this calculated decision rule can be developed only after conducting an experiment.

$$\Phi^{-1}(1 - \alpha) = \sqrt{\frac{n_t n_\star}{n_t + n_\star}}\left(\Phi^{-1}(\phi_\delta)\left(\frac{\delta n_0}{n_\star} + 1\right)\sqrt{\frac{1}{n_t} + \frac{1}{\delta n_0 + n_\star}} - \frac{\delta n_0 (x_t - x_0)}{\sigma n_\star}\right)$$

$$\phi_\delta = \Phi\left[\sqrt{\frac{n_\star(n_t + n_\star)}{(\delta n_0 + n_\star)(\delta n_0 + n_\star + n_t)}}\left(\Phi^{-1}(1 - \alpha) + \frac{x_t - x_0}{\sigma}\frac{\delta n_0 \sqrt{n_t}}{\sqrt{n_\star(n_t + n_\star)}}\right)\right]$$

The attempt to derive the UMP test analogy in the Bayesian settings for two arm scenario with a conditional power prior was not successful in this master thesis. However, the general case for all settings was considered in the paper Kopp-Schneider *et al.* (2019), where it was shown that gain in power is not possible if the decision rule is dependent on the historical information and a strict control of type I error is required, meaning the decision rule with the posterior probability critical value ϕ_δ is applied.

Summary table

Table 5.2 shows the different approach to hypotheses testing in the one arm scenario.

UMP test	Bayesian decision rule with no hist. data	Bayesian decision rule with hist.data and fixed δ
$\phi(x_t, x_\star) = \begin{cases} 1 & \text{if } \Phi\left[\frac{x_t - x_\star}{\sigma}\sqrt{\frac{n_t n_\star}{n_t + n_\star)}}\right] > 1-\alpha \\ 0 & \text{if } \Phi\left[\frac{x_t - x_\star}{\sigma}\sqrt{\frac{n_t n_\star}{n_t + n_\star)}}\right] \leq 1-\alpha \end{cases}$	$\phi_B(x_t, x_\star) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid x_t, x_\star) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \mid x_t, x_\star) \leq \phi_0 \end{cases}$ where $\phi_0 = 1 - \alpha$	If the critical value of a decision rule is independent of hist. data $\phi_{B_0}(x_t, x_\star, x_0, \delta) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) > \phi_0 \\ 0 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) \leq \phi_0 \end{cases}$ where $\phi_0 = 1 - \alpha$ If the decision rule is adapted according to hist. data to have a strict control of type I error $\phi_{B_\delta}(x_t, x_\star, x_0, \delta) = \begin{cases} 1 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) > \phi_\delta \\ 0 & \text{if } P(\Delta > 0 \mid x_t, x_\star, x_0, \delta) \leq \phi_\delta \end{cases}$ $\phi_\delta = \Phi\left[\sqrt{\frac{n_\star(n_t+n_\star)}{(\delta n_0+n_\star)(\delta n_0+n_\star+n_t)}}\left(\Phi^{-1}(1-\alpha) + \frac{x_t - x_0}{\sigma}\frac{\delta n_0\sqrt{n_t}}{\sqrt{n_\star(n_t+n_\star)}}\right)\right],$ which depends on x_t and can be defined only after conducting an experiment

Table 5.2: Hypotheses testing for the two arm case

Chapter 6

Conclusions

Question 1: Explore the influence of the bias between historical and new data on the different operating characteristics in a power prior settings for a normal outcome.

The dependence of operating characteristics on the bias between a historical estimate and a new true value was shown in chapter 4. In both one-arm and two-arm settings, there is a trade-off between type I error and power: the higher the power, the higher type I error. Adaptive methods, such as empirical Bayes and full Bayes reduce the influence of historical data when the bias is large, discarding historical information, what causes type I error and power convergence to the RCT values.

Question 2: Is it true that borrowing of information cannot lead to an increased power while strictly controlling type I error?

Answer to this question depends on the prior knowledge about the bias between a historical and a current experiment, since it determines the decision rule and related type I error. If we do not have a prior knowledge of the bias and we want a strict control of type I error, then the posterior probability critical value should be adapted for a historical estimate. Although, it would lead to the power of the test being determined only by the data from the current experiment.

However, if we know, that the bias is minimal and historical information is available for the control arm only, then we can expect that type I error does not go above a specified significance level. In this case, we can apply the posterior probability critical value, which does not depend on a historical estimate. Under these conditions, integration of historical information would lead to a simultaneous gain of type I error and power.

Appendix

Combining quadratic forms

$$A(x-a)^2 + B(x-b)^2 = C(x-c)^2 + \frac{AB}{C}(a-b)^2, c = \frac{Aa+Bb}{C}, C = A+B \tag{1}$$

Formula of the density of the difference between two random variables

$$Z = X - Y$$
$$f_z(z) = \int_{\infty}^{\infty} f_x(x) f_y(x-z) dx \tag{2}$$

Bibliography

Dejardin, D., Delmar, P., Warne, C., Patel, K., van Rosmalen, J., and Lesaffre, E. (2018). Use of a historical control group in a noninferiority trial assessing a new antibacterial treatment: a case study and discussion of practical implementation aspects. *Pharmaceutical Statistics*, **17**, 169–181. 3

Gravestock, I. and Held, L. (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics*, **16**, 349–360. 3, 4

Gravestock, I. and Held, L. (2018). Power priors based on multiple historical studies for binary outcomes. *Biometrical Journal*, **61**, 1201–1218. 3

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, **7**, 639–674. 3

Kopp-Schneider, A., Calderazzo, S., and Wiesenfarth, M. (2019). Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. *Biometrical Journal*, **62**, 361–374. 4, 63

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, **7**, 5–18. 3

Psioda, M. A. and Ibrahim, J. G. (2019). Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics*, **20**, 400–415. 57, 58, 59, 62

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, **70**, 1023–1032. 3

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley and Sons. 13, 14

van Rosmalen, J., Dejardin, D., van Norden, Y., Löwenberg, B., and Lesaffre, E. (2018). Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical methods in medical research*, **27**, 3167–3182. 4

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S.,*et al.* (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, **13**, 41–54. 3