Historical borrowing for clinical trial design

Johnson&Johnson
Innovative Medicine

Madhawa Saranadasa
Senior Data Scientist

Disclaimer

The views expressed here are my own and not necessarily that of Johnson & Johnson and its subsidiaries.

Clinical trial design

Consider the following use case:

- Planning a clinical trial
- Reduce placebo sample size and use a **power prior** to achieve sufficient power

Example scenario (hypothetical)

There is a desire to decrease placebo sample size by 10 patients, however this will negatively impact the power of the study

Original design		
	Drug A	Placebo
Sample size	**30**	**30**
Assumed change from baseline in disease activity score	-3.0 (1.5)	-2.0 (1.5)
Type I error	5%	
Power to detect difference of 1.0	**81.7%**	

Proposed design		
	Drug A	Placebo
Sample size	**30**	**20**
Assumed change from baseline in disease activity score	-3.0 (1.5)	-2.0 (1.5)
Type I error	5%	
Power to detect difference of 1.0	**73.6%**	

Historical data in a comparable population	
	Placebo
Sample size	**100**
Reported change from baseline in disease activity score	-1.5 (1.5)

Questions to address at the design phase:

- What is the **potential** difference in response rate between historical control and current control results?
- How should we specify alpha?
- **Given these moving parts, how to quantify the operating characteristics of a trial?**

Operating characteristics

To understand feasibility of borrowing at the design phase, we can define a framework to assess operating characteristics of the trial:

1. Set up scenarios for potential difference between historical data (known) and where the current trial will land (unknown)

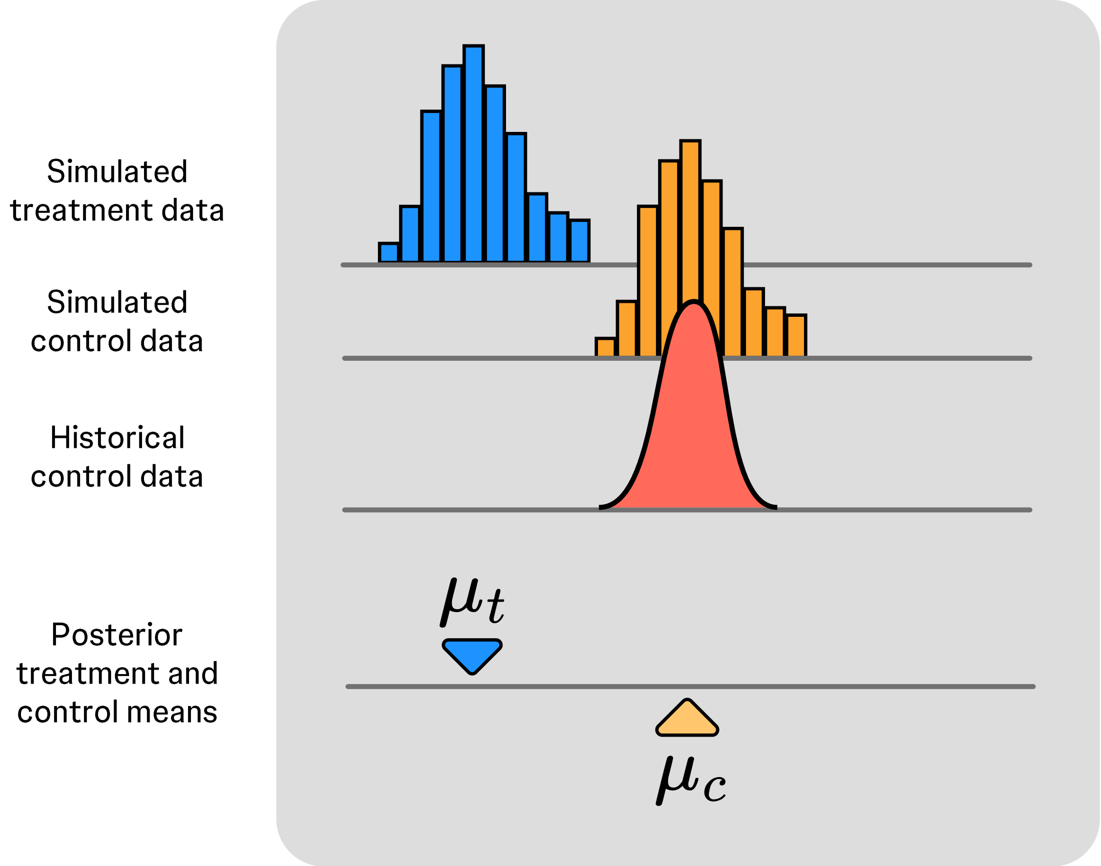
2. See how alpha responds in these scenarios with different specifications
   - Fixed power prior - Alpha is chosen as a fixed value, regardless of discrepancy
   - Empirical Bayes power prior - Alpha is inversely proportional to the discrepancy through an explicit function
   - Many others as well…

3. Quantify **Bayesian power and type I error** under these scenarios

Statistical power with a power prior

Probability of correctly rejecting the null hypothesis

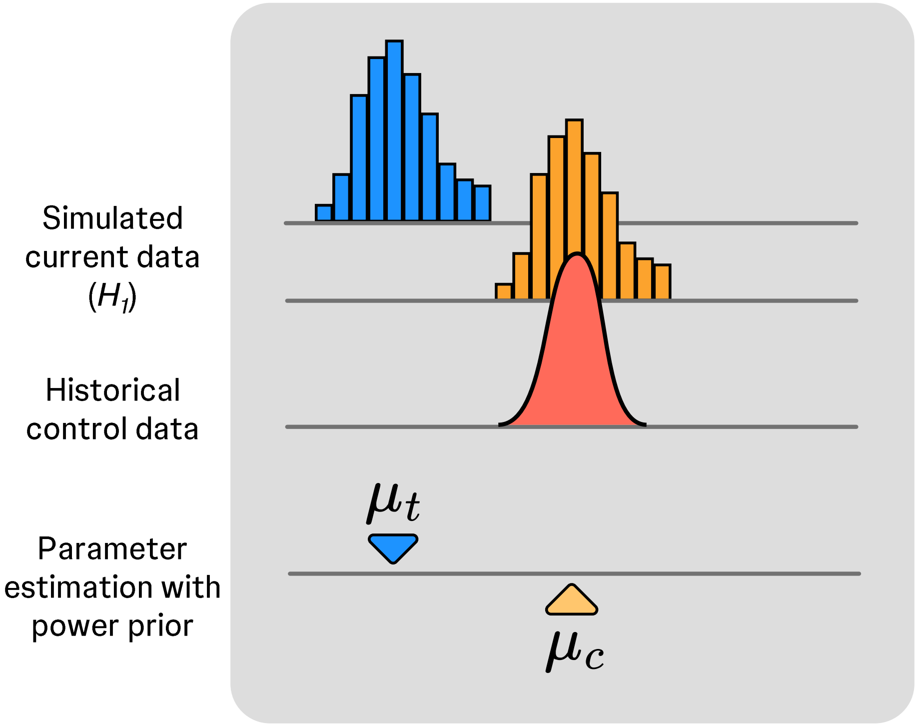
For each iteration:



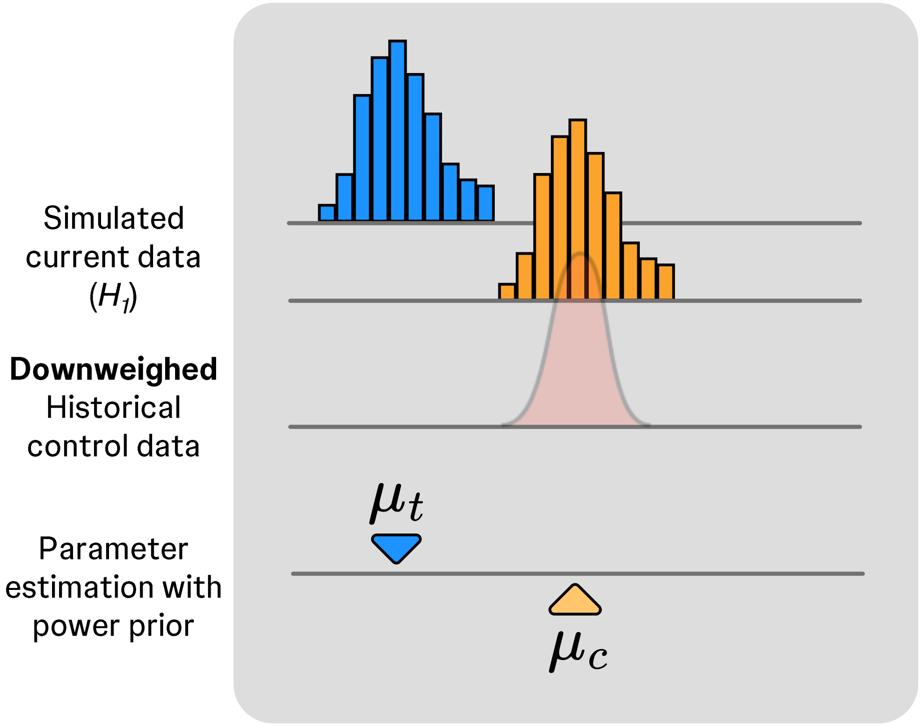
- Generate current trial data consistent with alternative hypothesis
- Incorporate historical trial data (control)
- Estimate parameters using Bayesian power prior model
- Accept or reject null hypothesis using threshold criteria (P > 0.95)

Power estimated as the proportion of iterations where null hypothesis is correctly rejected

Alignment between historical and current control data

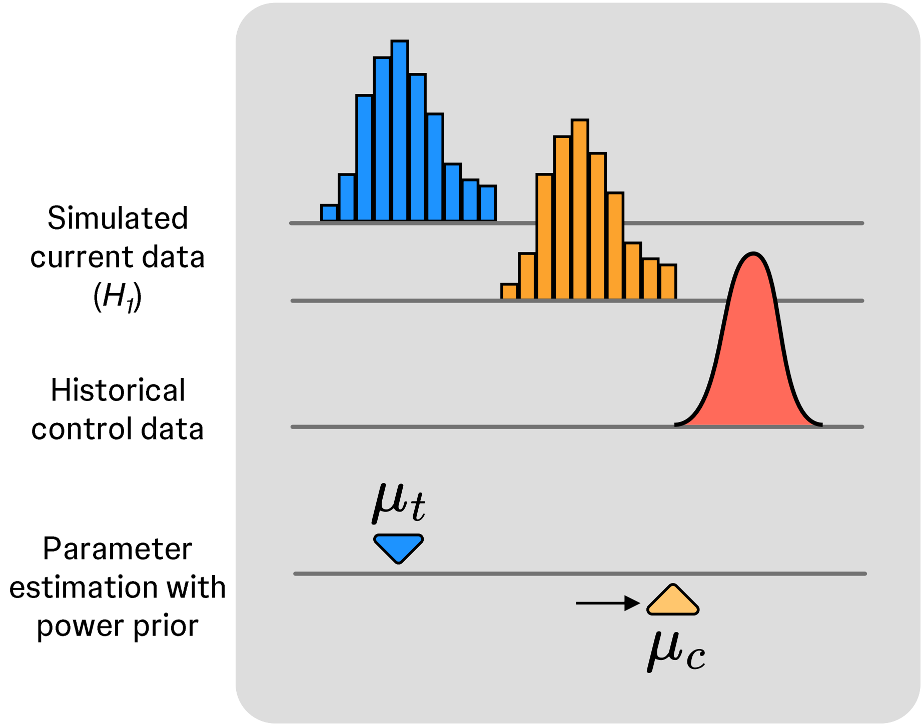


High alpha: Upweight aligned historical data for substantial gain in power

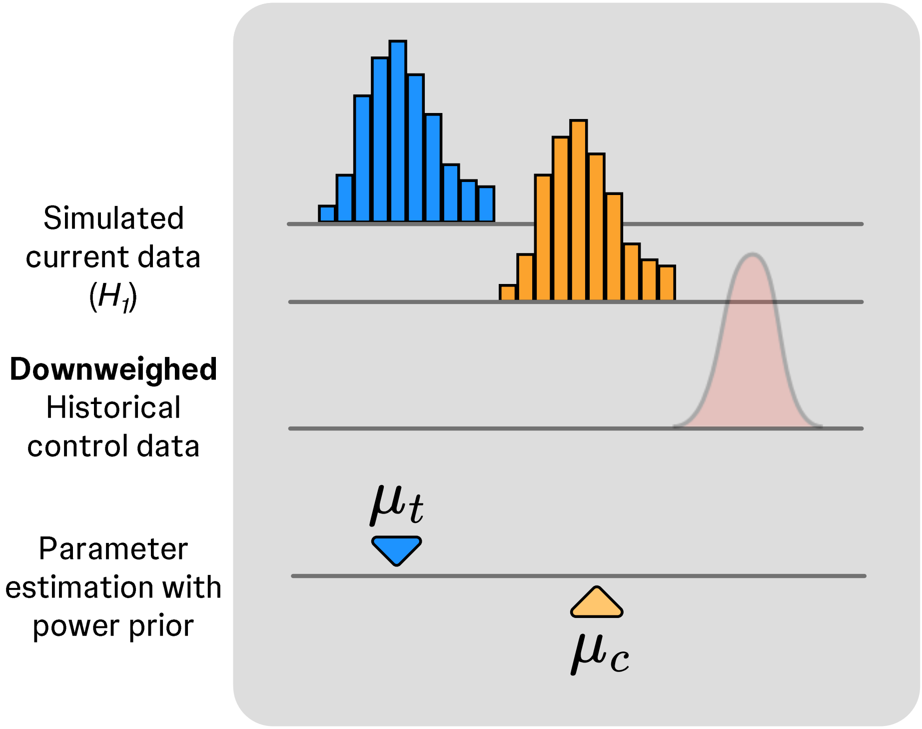


Low alpha: Downweigh aligned historical data for modest gain in power

Discrepancy between historical and current data control data

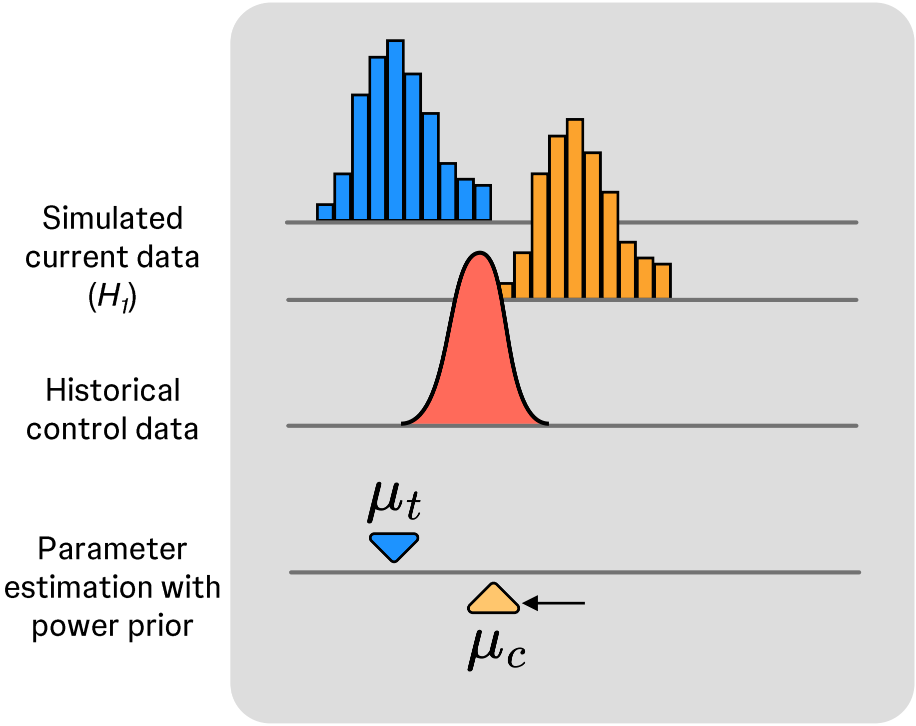


High alpha: Upweight discrepant historical data will result in substantial gain in power

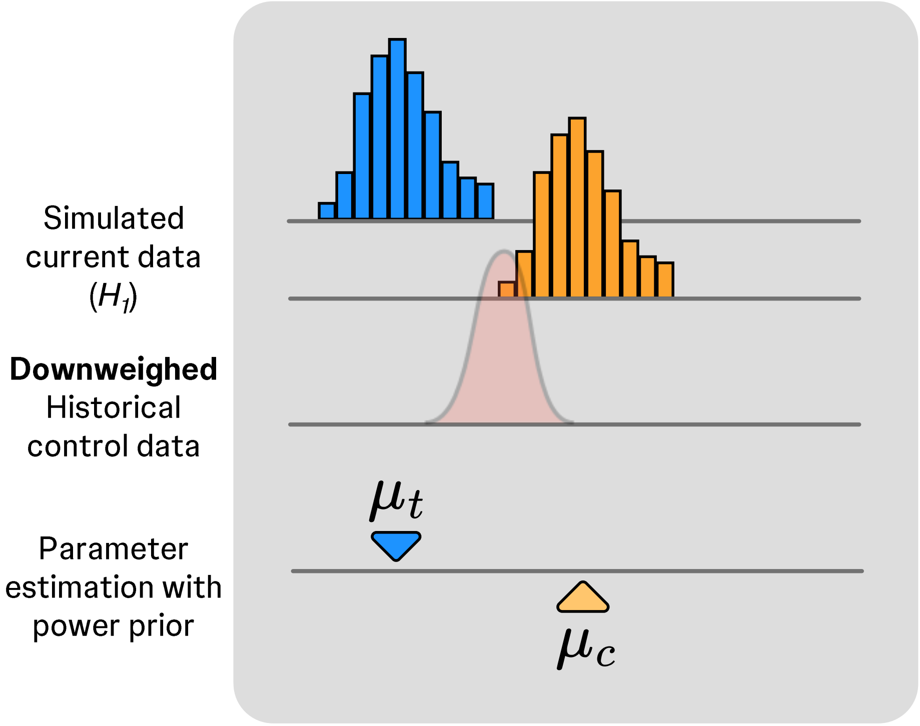


Low alpha: In this case, downweighing would mitigate the increase in power

Discrepancy between historical and current data control data



High alpha: Upweight discrepant historical data will reduce power

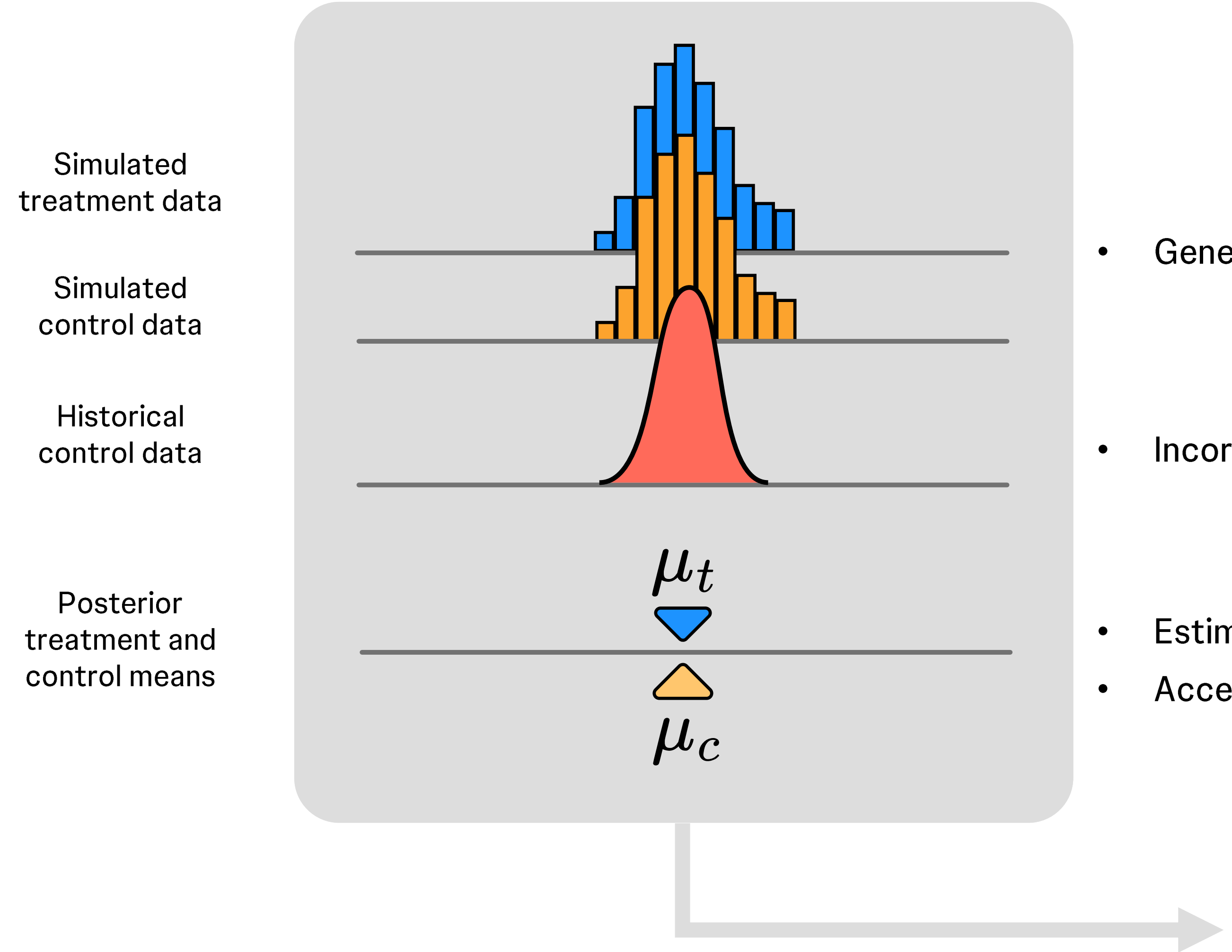


Low alpha: In this case, downweighing would mitigate the reduction in power

Type I error with a power prior

Probability of incorrectly rejecting the null hypothesis

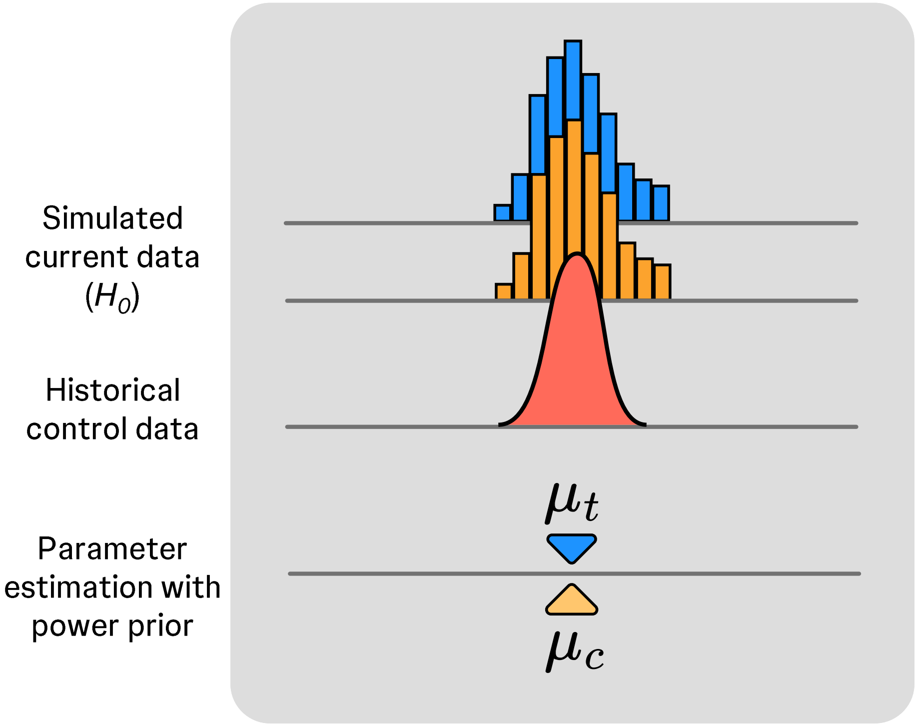
For each iteration:



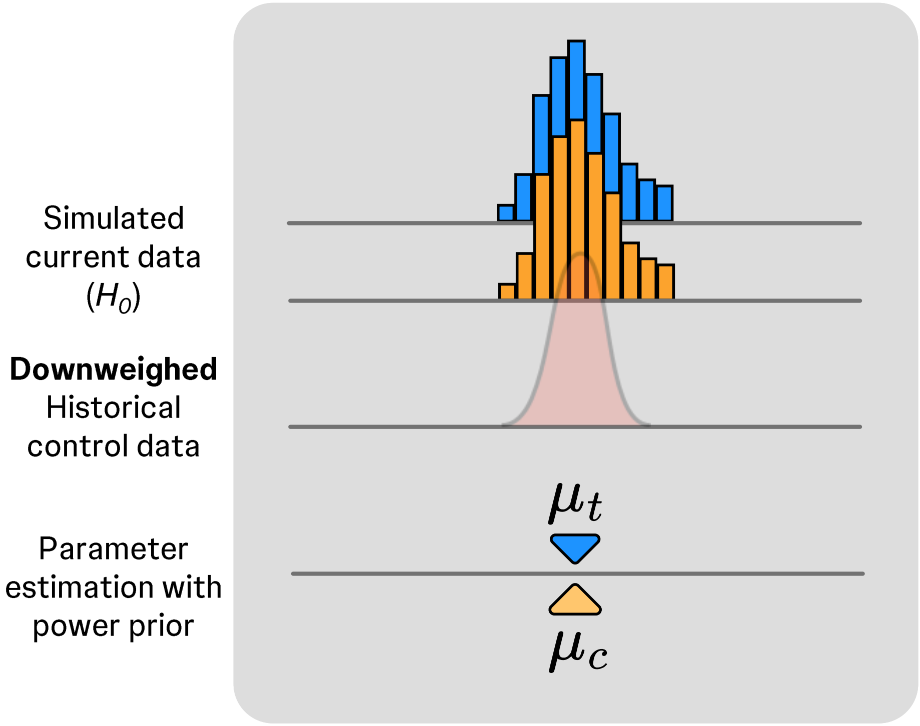
- Generate current trial data consistent with null hypothesis
- Incorporate historical trial data (control)
- Estimate parameters using Bayesian power prior model
- Accept or reject null hypothesis using threshold criteria (P > 0.95)

Type I error estimated as the proportion of iterations where null hypothesis is incorrectly rejected

Alignment between historical and current control data

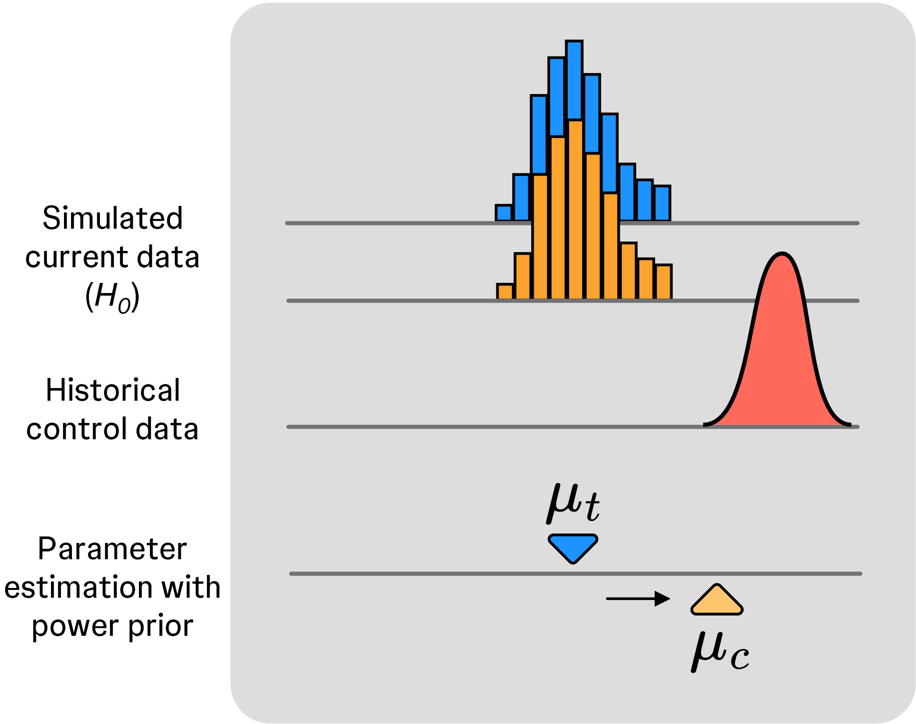


High alpha: Upweight discrepant historical data will result in reduction in type I error

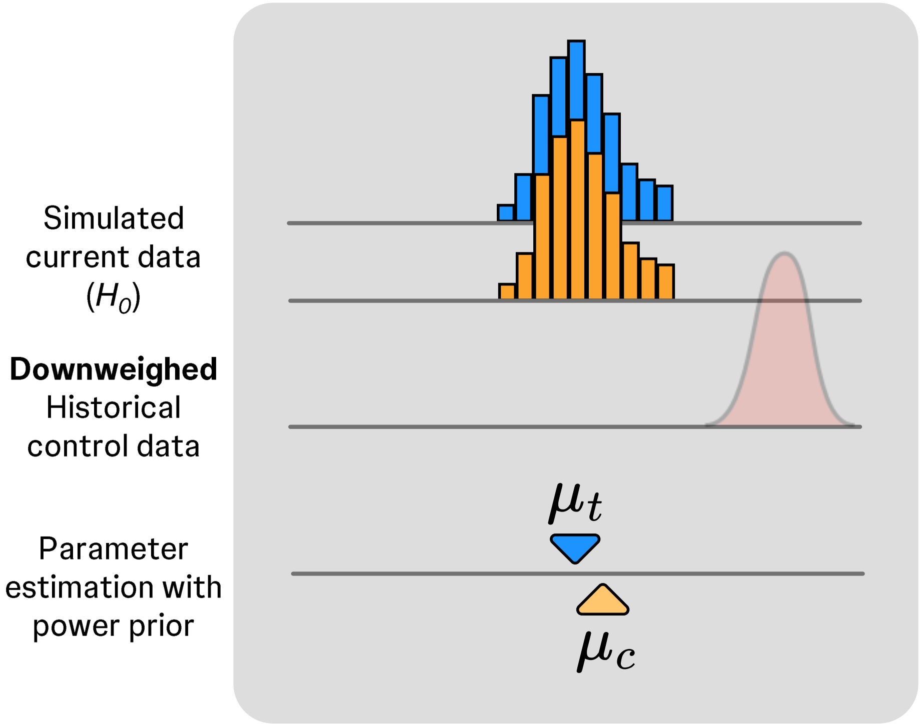


Low alpha: In this case, downweighing would results in a modest reduction in type I error

Discrepancy between historical and current data



High alpha: Upweight discrepant historical data will increase type I error



Low alpha: In this case, downweighing would mitigate the increase in type I error

Example scenario (hypothetical)

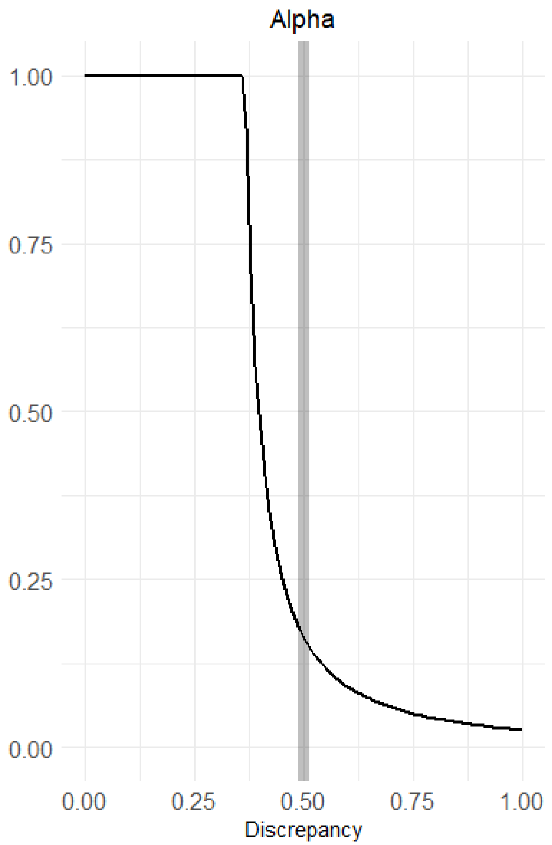
Original operating characteristics		
Arm	Drug A	Placebo
Sample size	**30**	**30**
Assumed change from baseline in disease activity score	-3.0 (1.5)	-2.0 (1.5)
Type I error	5%	
Power to detect difference of 1.0	**81.7%**	

Proposed operating characteristics		
Arm	Drug A	Placebo
Sample size	**30**	**20**
Assumed change from baseline in disease activity score	-3.0 (1.5)	-2.0 (1.5)
Type I error	5%	
Power to detect difference of 1.0	**73.6%**	

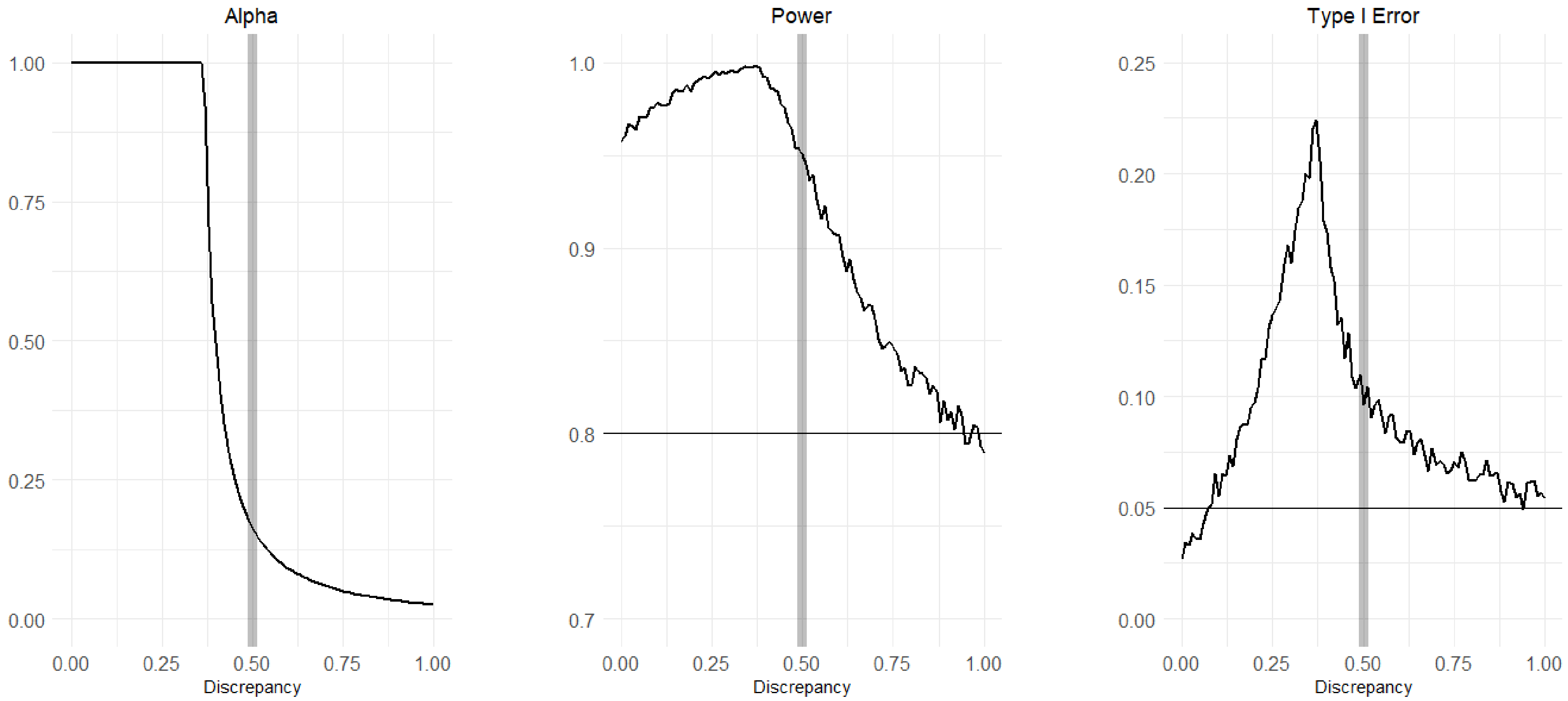
Historical data	
Arm	Placebo
Sample size	**100**
Assumed change from baseline in disease activity score	-1.5 (1.5)

- Scenarios: Current placebo ranges from -1.5 (discrepancy = 0) to -2.5 (discrepancy = 1)
- Use **Empirical Bayes** to define alpha
  - *Alpha is inversely proportional to the discrepancy through an explicit function*
- Quantify Bayesian type I error and power

Empirical Bayes power prior, applied to our example



Empirical Bayes power prior, applied to our example



Given the initial trial assumptions, the Empirical Bayes power prior specified a “cliff” in the behavior of delta. This will lead to full borrowing of discrepant data and an inflation in type I error in some scenarios.

Key Learnings

Why use historical borrowing?

- Reduce sample size at design phase and use borrowing to achieve sufficient power

Communicate benefits/risk

- Given all the moving parts, what is the appropriate way to specify a power prior?
- Framework to assess operating characteristics over plausible scenarios
- Identify problematic scenarios