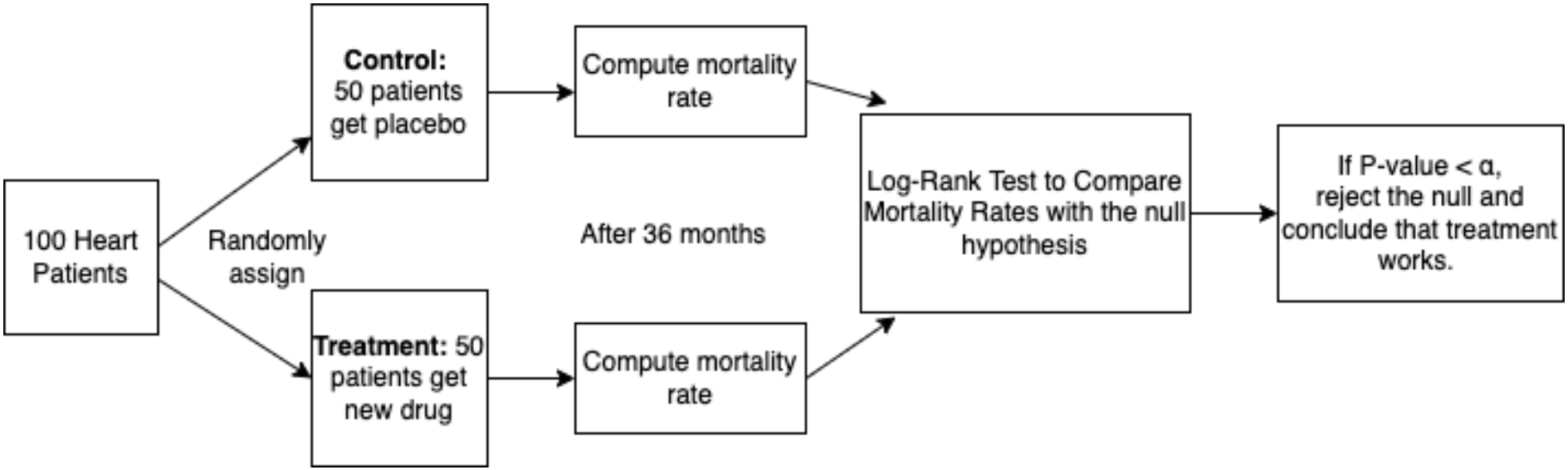
Clinical Trials DRP: Group Sequential Methods in the Beta-blocker Heart Attack Trial

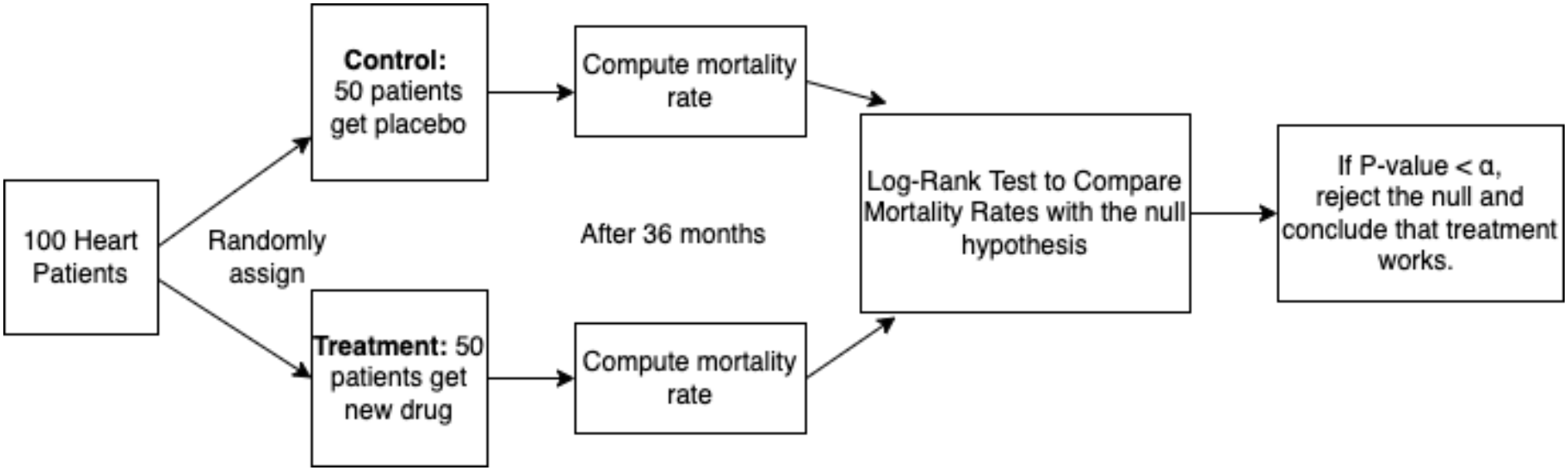
Hisham Bhatti

March 13, 2022

A simple example of a clinical trial

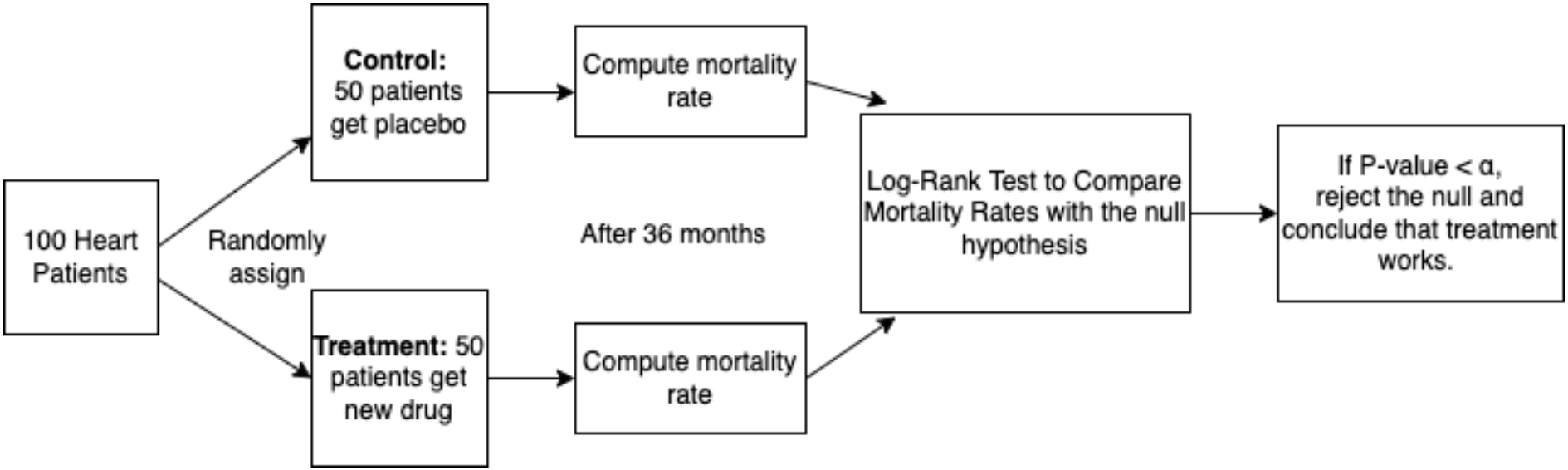


A simple example of a clinical trial



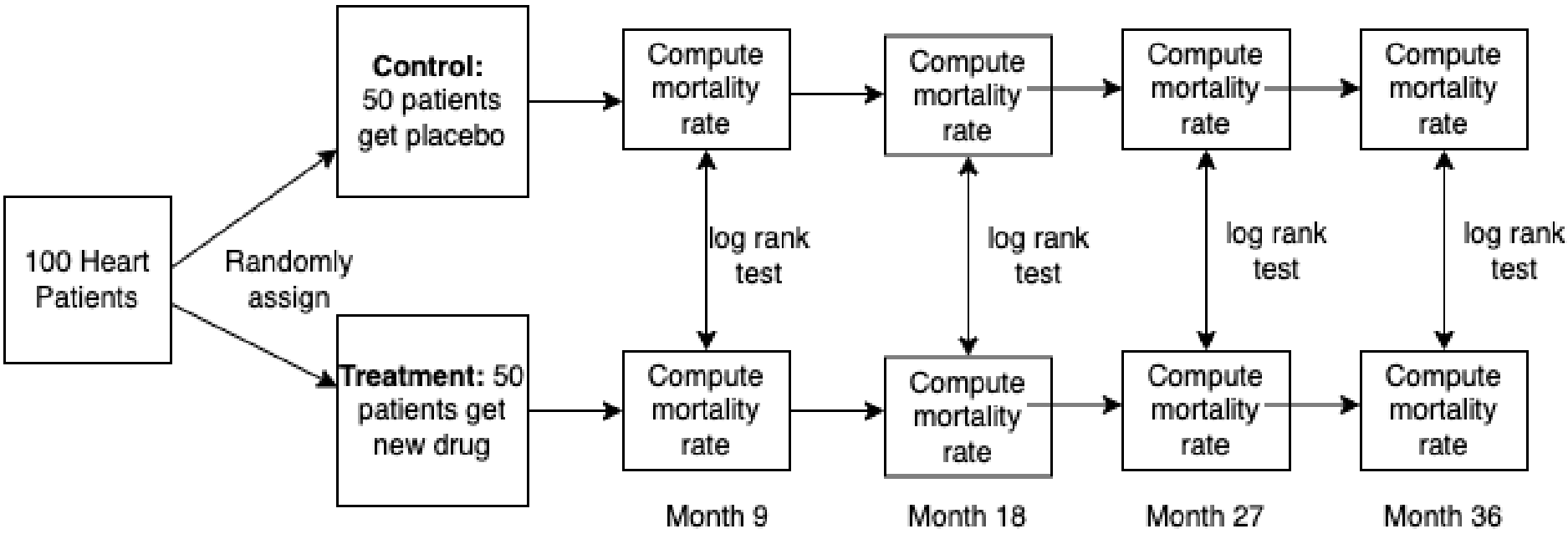
- α is the probability that we reject the null when the null is true.

A simple example of a clinical trial



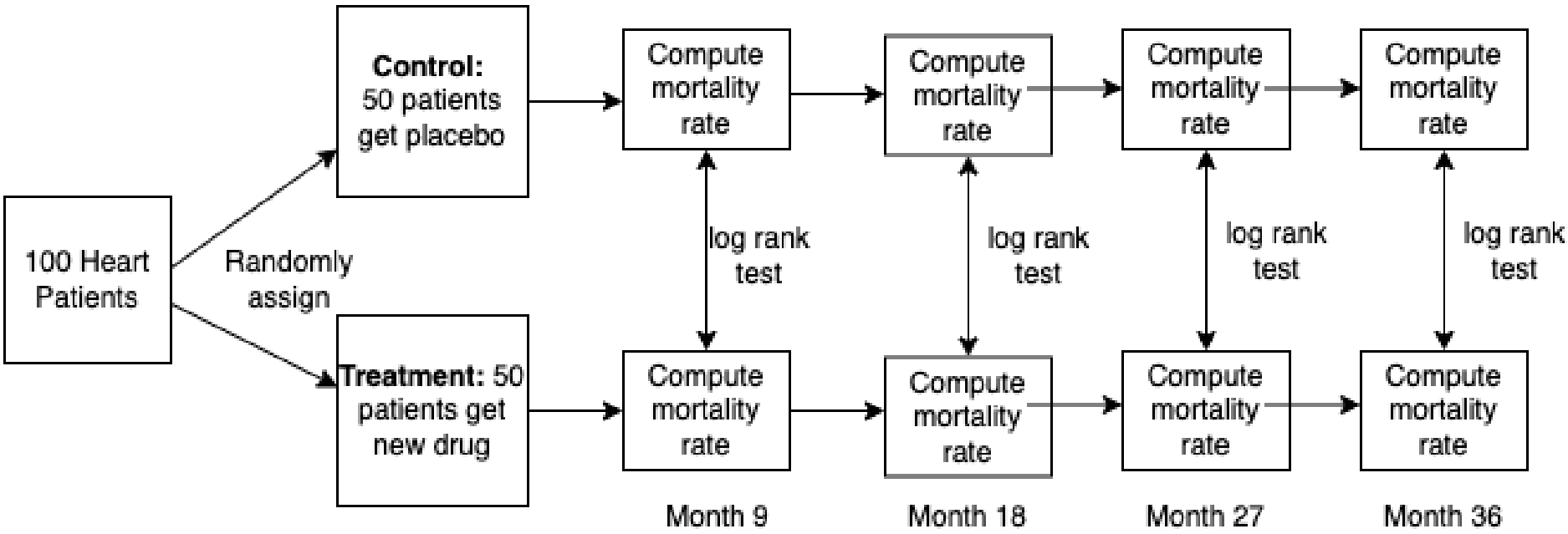
- α is the probability that we reject the null when the null is true.
- **Ethical concern:** If we see early evidence that the treatment is beneficial, we should give the treatment to the people in the control group!!!

Interim analyses



- Can we compute a p-value at each interim analysis and stop as soon as we see a p-value $< \alpha$?

Interim analyses



- Can we compute a p-value at each interim analysis and stop as soon as we see a p-value $< \alpha$?
- If we do this, our probability of falsely rejecting the null is BIGGER than α (multiple testing)

Group Sequential Methods

- If we want overall probability of a mistake (when the null is true) to equal $\alpha = 0.05$, we cannot use $\alpha = 0.05$ as our “boundary" for all 4 tests.
- There are different methods for “spending" our α across the four tests.

Interim #	Bonferroni (α)	Pocock (α)	O'Brien Fleming (α)
1	0.0125	0.0182	0.00005
2	0.0125	0.0182	0.0039
3	0.0125	0.0182	0.0184
4	0.0125	0.0182	0.0412

How do we pick a boundary function"

Interim #	Bonferroni (α)	Pocock (α)	O'Brien Fleming (α)
1	0.0125	0.0182	0.00005
2	0.0125	0.0182	0.0039
3	0.0125	0.0182	0.0184
4	0.0125	0.0182	0.0412

- **Power:** Probability of rejecting the null when the null is false (i.e. saying that the treatment works when it really does!)
- **Stopping time:** How early do we stop the trial and conclude the treatment works?

How do we pick a boundary function"

Interim #	Bonferroni (α)	Pocock (α)	O'Brien Fleming (α)
1	0.0125	0.0182	0.00005
2	0.0125	0.0182	0.0039
3	0.0125	0.0182	0.0184
4	0.0125	0.0182	0.0412

- **Power:** Probability of rejecting the null when the null is false (i.e. saying that the treatment works when it really does!)
- **Stopping time:** How early do we stop the trial and conclude the treatment works?
- Different boundaries = tradeoff between power and stopping time.

The BHAT trial

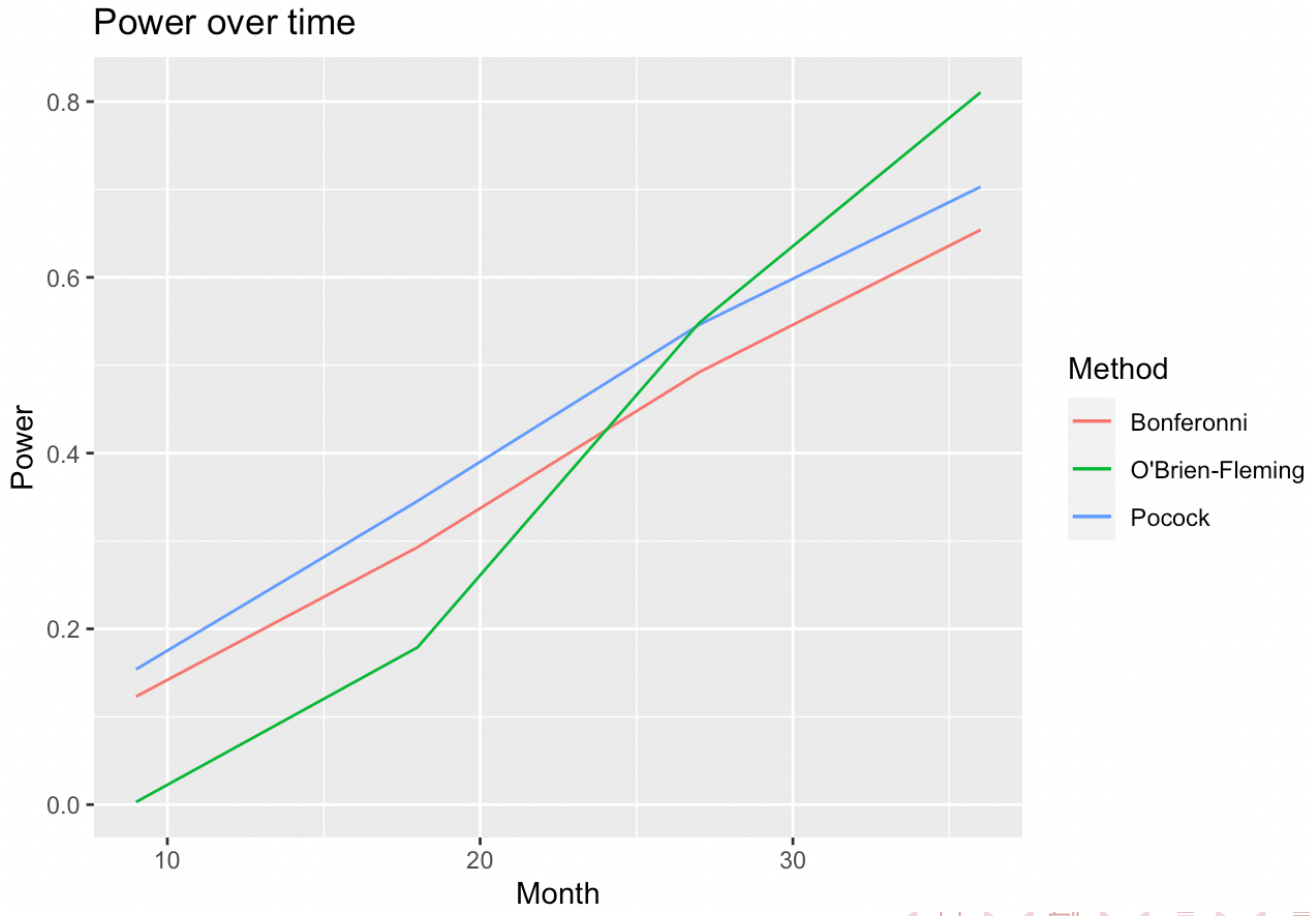
- **Beta-Blocker Heart Attack Trial (BHAT):** Does propranolol hydrochloride cause a significant reduction in mortality to those who recently had a heart attack?
- One of the first clinical trials to use the O'Brien-Fleming boundary
- 3837 patients randomized to propranolol (1916) or placebo (1921).
- Used log-rank test to compare mortality rate at 4 interim analyses

BHAT Results

- At early interim analyses, log-rank test p-values were small enough to end trial with a Pocock boundary.
- But their analysis plan said they would use O'Brien-Fleming boundary!
- Ended up stopping at 3rd interim analysis (9 months early!!) when they crossed O'Brien-Fleming boundary!
- Did they make a mistake by choosing O'Brien-Fleming?

Simulation study

- Goal: Investigate if OF was appropriate boundary, and likelihood of stopping 9 months early
- Used real sample size/mortality rates from BHAT trial (2000 trials)



Conclusions

- **Which sequential method should you use?** It depends
- How willing are you to forgo power for newer treatments?
- Stop EARLY = Pocock (high initial power)
  1. Ex: Cancer Treatments
- Stop LATER with more data = OF (high final power)
  1. Ex: Stomachaches
- For BHAT Trial, I'd guess Pocock would be best
  1. Later studies confirmed effectiveness of β-Blockers