Adaptive Trials and Trial Simulation

Kert Viele

CTMC 2019

Disclosures

- Berry Consultants
  - Consulting firm specializing in Bayesian adaptive clinical trial design
  - Multiple clients

- No off label use of specific drugs discussed

Adaptive Trial Design

- Choices are made at the beginning of every trial based on incomplete information.
  - don't know dose (may know range)
  - don't know treatment effect
  - don't know control information
  - don't know population
  - don't know drug combinations
  - etc.

Driving with your eyes open

- Drug development is ALWAYS adaptive
  - we just typically only adapt between trials
- Prespecified adaptations change trial characteristics mid trial
- Imagine driving to work, do you only open your eyes at intersections, or all the time?

Adaptive Trial Design

- Typically as the trial continues you learn valuable information.
  - this drug doesn't work….
  - these 2 doses/treatments are promising, but another dose/treatment shows nothing…
  - the treatment works quite well!
  - this group of subjects doesn't benefit…
- Some questions are answered before others

Adaptive Trial Design

- Adaptive trials use the accumulating information to change the design of the trial
  - drop doses/treatments mid trial
  - add combinations of treatments.
  - stop for futility (or success)
  - stop enrolling certain subpopulations.
  - seamlessly shift phases of development

Different adaptations

- Futility stopping (very important…let the subjects go to another trial…)
- Success stopping
- Arm dropping/adding
- Adaptive Randomization
  - "softer" form of arm dropping, enroll more subjects to treatments that are performing well
- Enrichment
  - enroll more subjects in populations that seem to benefit from the treatment, potentially drop groups of subjects.

Main idea

- Modern trials have lots of questions….

- As you answer your questions, focus resources on the things you don’t know.

Ebola

- During outbreak, many different treatments proposed for Ebola.
- Many can be given in combination.
- For simplicity, suppose there were four treatments A, B, C, D
  - combination of any 2 allowed
  - (in reality somewhat complex structure of combinations allowable)

Ebola

- How to examine 4 treatments?
- Could sequentially test one at a time.
  - Each experiment requires a fixed number of subjects, provides no information about the other treatments.
  - Unclear how to add/subtract combinations.
  - inefficient UNLESS you can do a good job of picking the best treatment to investigate first.

Ebola

- Could examine multiple treatments at once, N subjects per treatment/combination.
  - lots of subjects placed on ineffective arms.
  - effective arms may not have enough data.
- Any way to bridge the gap?

Ebola

- Adaptively randomize.
  - Start with subjects on all treatments.
  - Look at mortality rates every few subjects.
  - Adjust randomization at looks. More to arms doing well, less to those doing poorly
    - prespecified mathematical formula estimating the chance each treatment/combination is the best
  - Drugs/Combinations may be added freely as trial continues (not considered here)

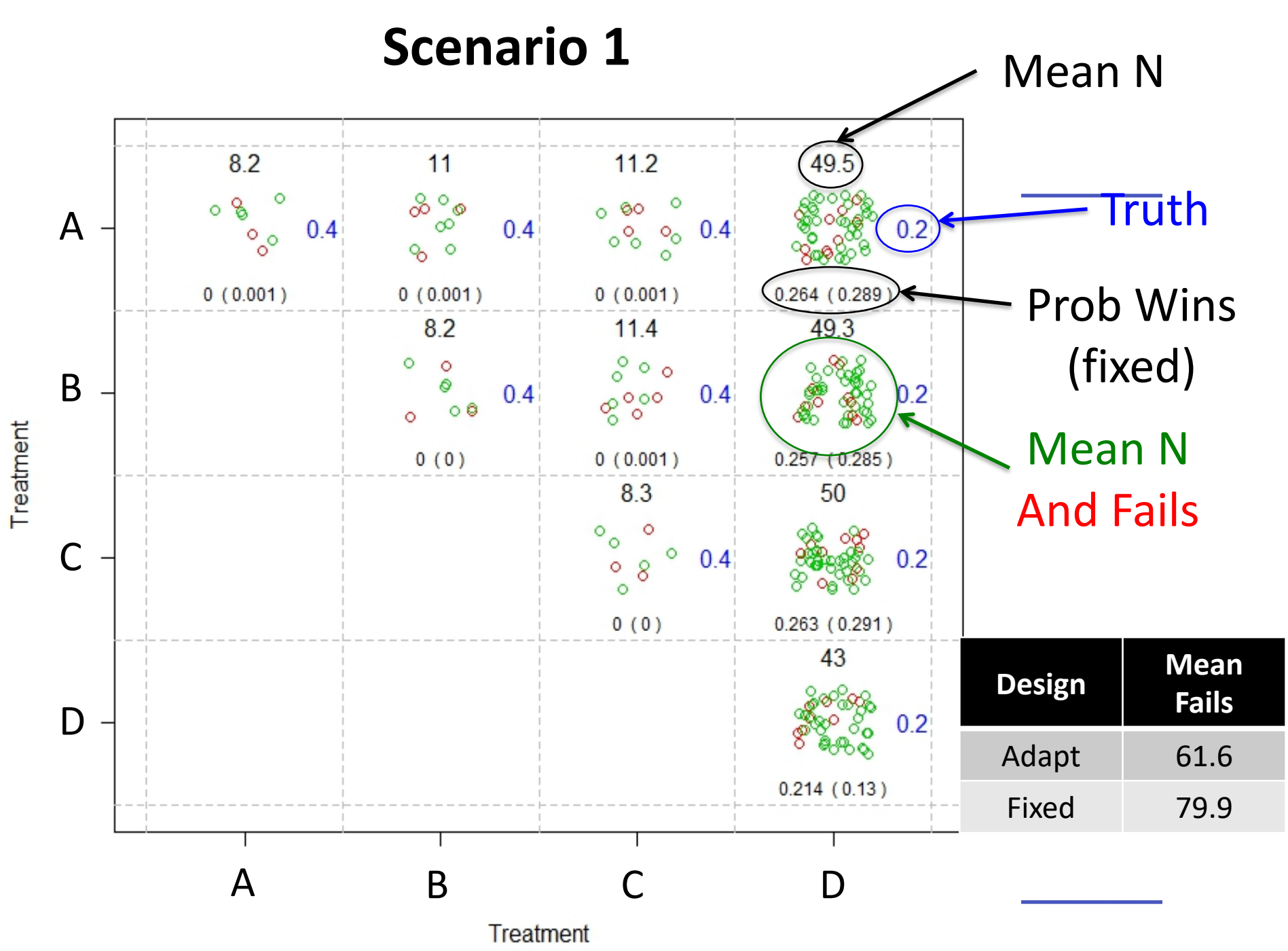
Ebola

- Key things to focus on
  - Mortality rate in study (treatment of patients in trial, always important but potentially more important in rare diseases)
  - Chance of picking the right treatment at the end (treatment of patients outside trial)

Example 1 - Ebola

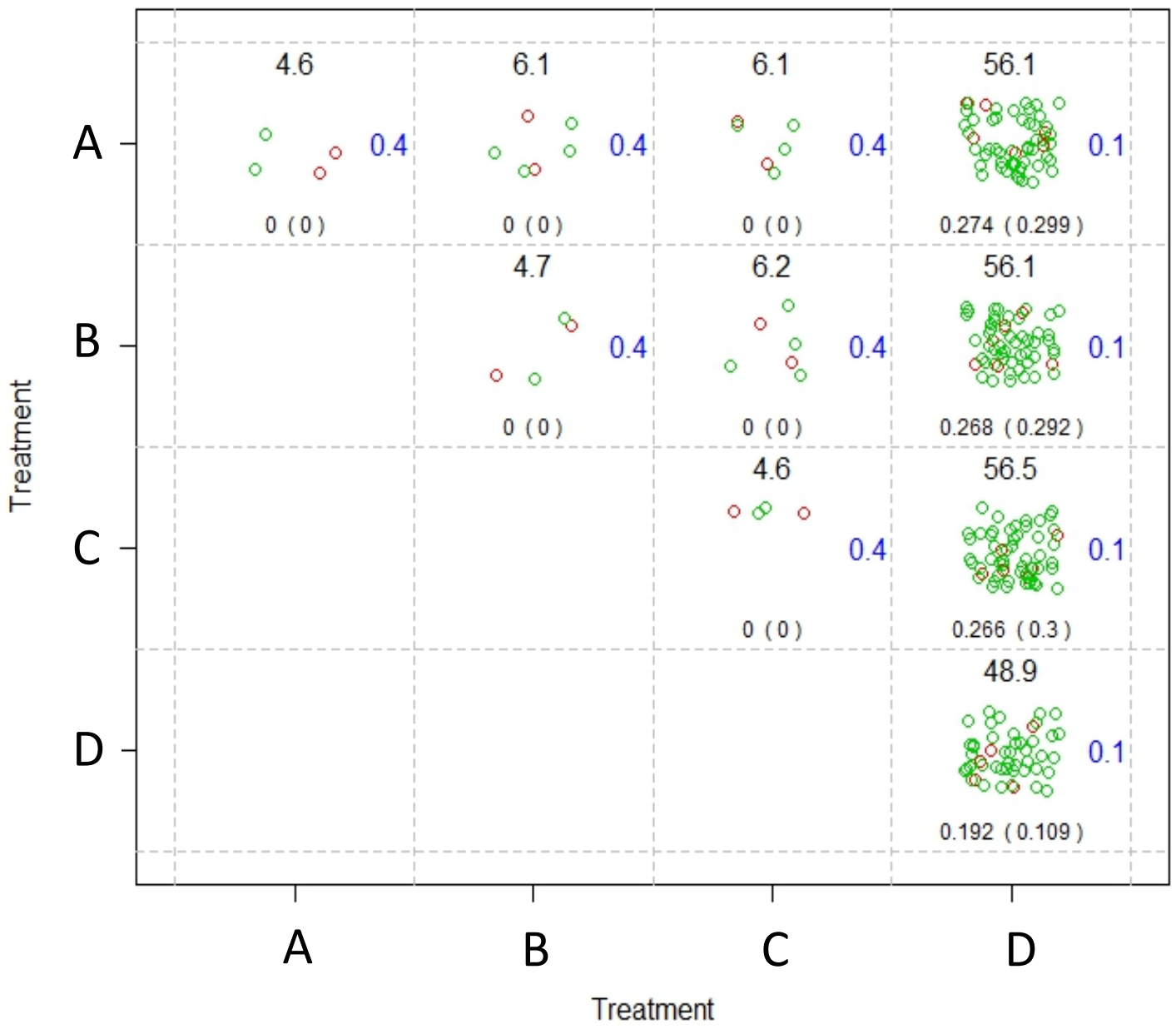
- N=250 subjects
  - "burn in" 3 subjects per combination
  - fit generalized linear model across combinations.
  - change allocation…allocate more to well performing arms.
- Trial can run perpetually.

Scenario 1



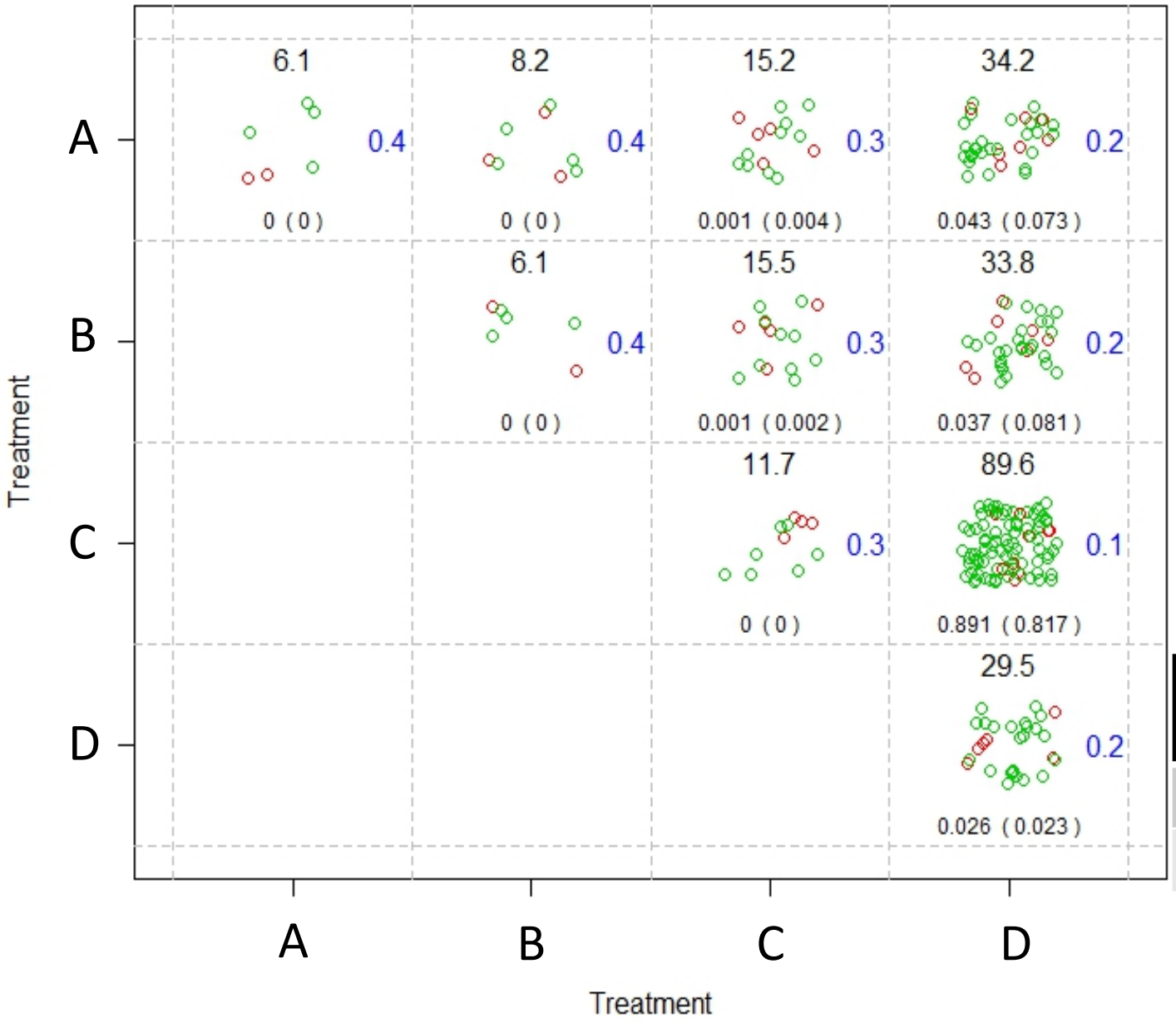
Design	Mean Fails
Adapt	61.6
Fixed	79.9

Scenario 2



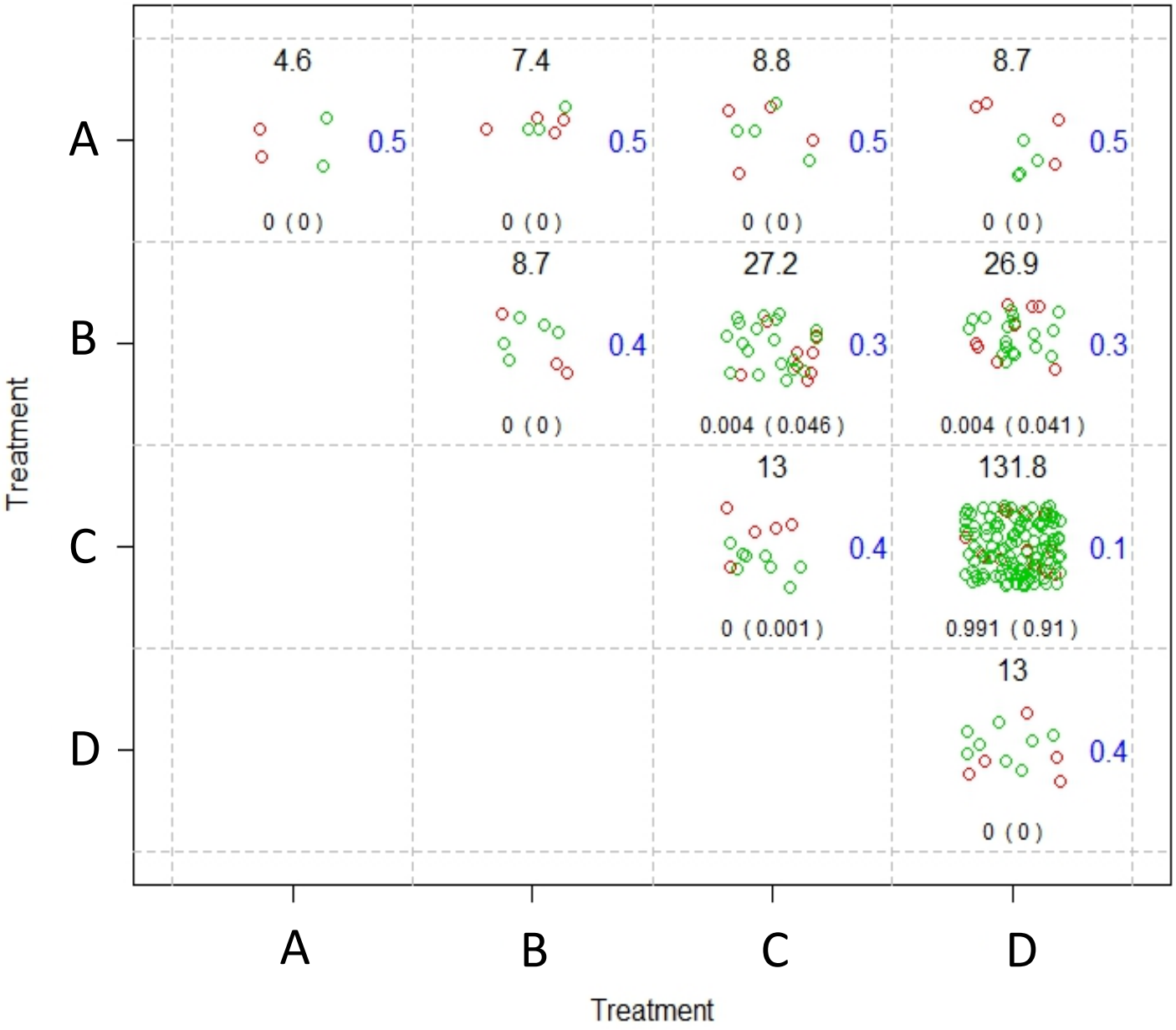
Design	Mean Fails
Adapt	34.7
Fixed	69.9

Scenario 3



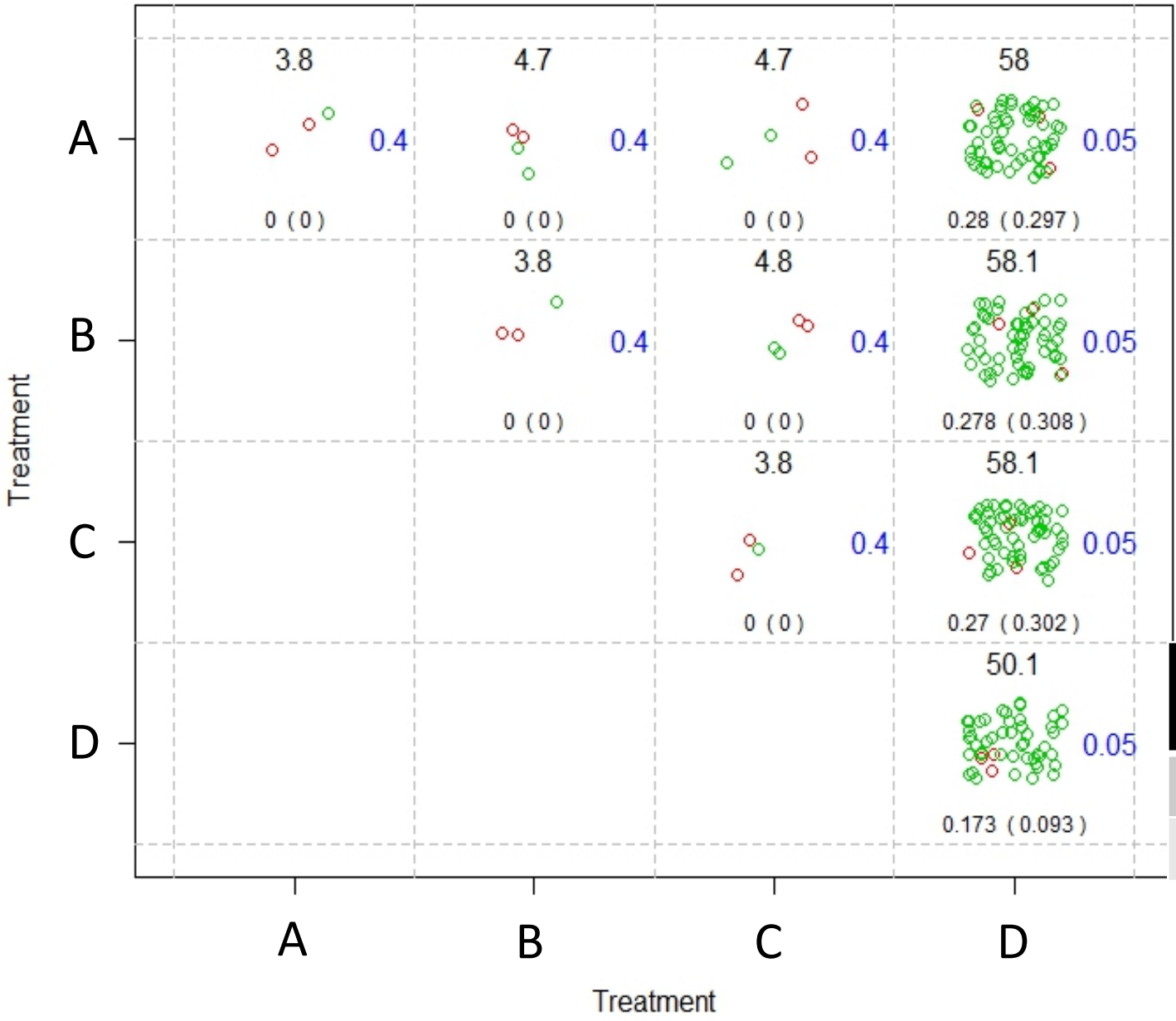
Design	Mean Fails
Adapt	49.4
Fixed	69.9

Scenario 4



Design	Mean Fails
Adapt	58.0
Fixed	97.3

Scenario 5



Design	Mean Fails
Adapt	21.5
Fixed	64.9

Ebola

- Trial treats patients within the study better (less mortality).
- Trial more likely to choose the correct treatment/combination. Treats patients outside the trial better
  - better prepared for next outbreak.

Simulation

- "Complicated" question
  - I flip a fair coin 10 times, what is the probability of getting a streak of 4 heads in a row?

Simulation

- "Complicated" question
  - I flip a fair coin 10 times, what is the probability of getting a streak of 4 heads in a row?
- Answer
  - with some "complicated" math 24.5%
  - can also simulate…meaning let a computer run the experiment LOTS of times.
    - sim1, THTHHTTHTH…..No
    - sim2, HTTHHHHTTH….Yes
    - sim3, HHTTTTTTHH….No
    - sim4, sim5, sim6,…..

Simulation

- The law of large numbers says that if you run enough simulations you get very close to the right answer
  - computers can run a LOT of simulations
- I ran 100,000 simulations
  - after 1,000….rate was 24.2%
  - after 10,000….rate was 24.7%
  - after 100,000….rate was 24.5%

Simulation

- Another complicated question
  - I flip coin N times and look for 4-head streak
  - What N gives me 90% of chance of streak?
    - (same kind of question as power calculation)

Simulation

- Another complicated question
  - I flip coin N times and look for 4-head streak
  - What N gives me 90% of chance of streak?
    - (same kind of question as power calculation)
- By simulation
  - N=10 gives 24.5%
  - N=50 gives 82.9%
  - N=70 gives 91.7%
  - N=65 gives 90.1%

Simulation

- Ok, back to “reality”
- In “simple” clinical trial designs, we can do the math directly to get power, sample size, etc.
- In complex trials (many/most adaptive trials), we have to simulate to get these quantities.
  - basic idea is the same, have computer randomly generate the trial MANY times.

Simulation Example

- Trial with 2 doses (low, high)
  - N=216 total, enroll 36 patients per month
  - simplifying to deterministic enrollment with instant endpoint. Can account for in practice.
- Endpoint is composite event
  - low can be better than high
- Increases of 2 units considered valuable

Fixed Trial

- Enroll 72 subjects per dose (control, low, high)
- Have to adjust for multiplicities
  - use alpha/2=0.025 and test each dose
- Trial always enrolls N=216
- Suppose drug doesn't work
  - true effect in low = 0, true effect in high = 0
  - Pr(success) = 2.3% (type I error rate)
  - essentially always enroll full trial and fail
- Suppose drug does work
  - true effect in low = 1, true effect in high = 3
  - Pr(success) = 77.1%

Naïve adaptive trial

- Interim Analyses
  - After one month stop the trial if neither dose achieves 2 unit increase
  - After three months, choose the dose with the higher observed mean.
- At end of trial perform t-test with selected dose
  - alpha=0.05/2=0.025 significance level (account for two doses in study)

Naïve trial

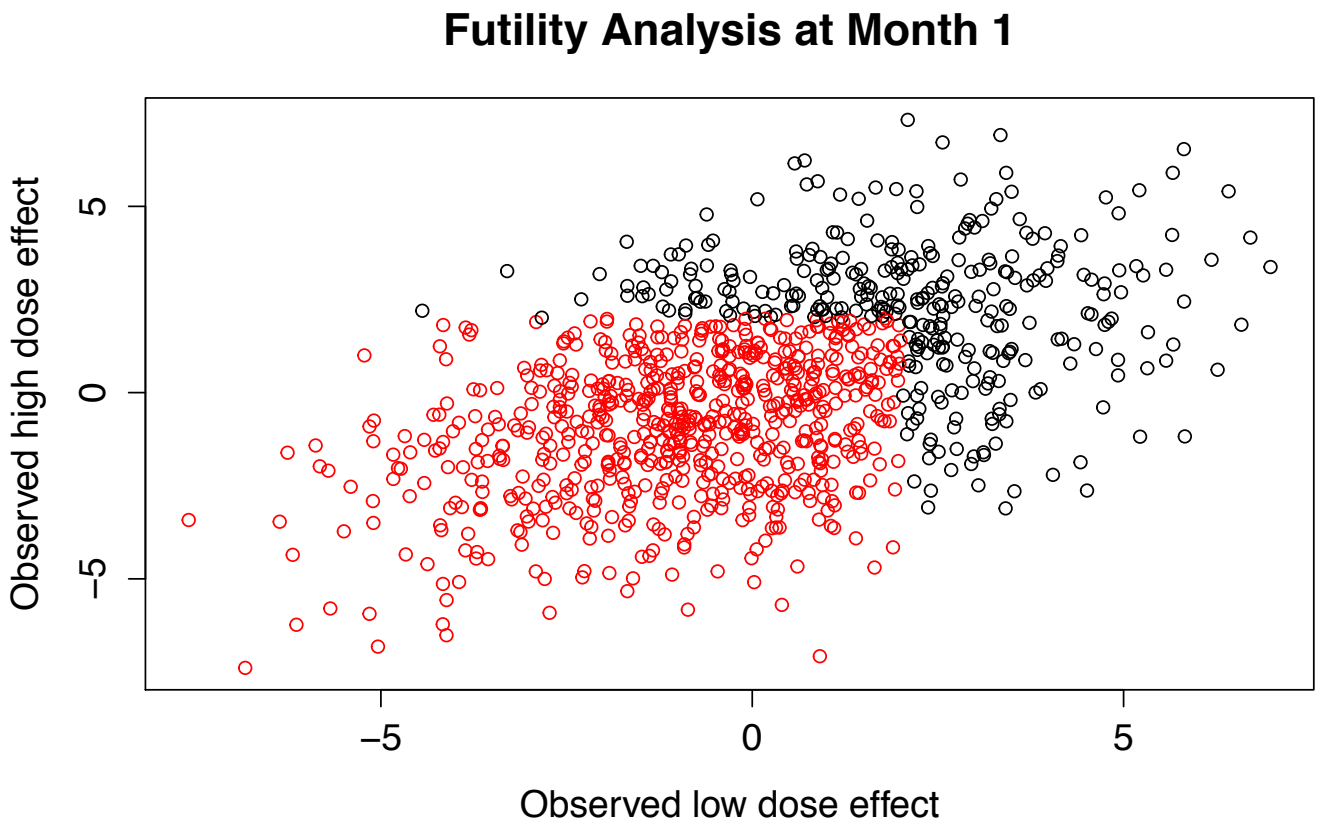
EXAMPLE 1	Placebo Mean	Low dose Mean	High dose Mean	Action
Month 1 Futility	6.4	10.9	9.1	Continue
Month 3 Dose selection	5.6	8.3	7.4	Choose low dose
Month 6 Final Analysis	5.2	7.9	NA	Success, p=0.001

EXAMPLE 2	Placebo Mean	Low dose Mean	High dose Mean	Action
Month 1 Futility	6.8	6.8	7.8	Futility
Month 3 Dose selection				
Month 6 Final Analysis				

Simulate 1000 trials when the drug doesn't work

- We want the trial to declare futility if the drug doesn't work.
  - So let's assume no effect of the drug, and see how often it declare futility
  - like the "streak" example, can do the math here, but we are focused on simulation
  - Can simulate 1000 trials.

Simulate 1000 trials



679 of 1000 trials are futile GOOD…

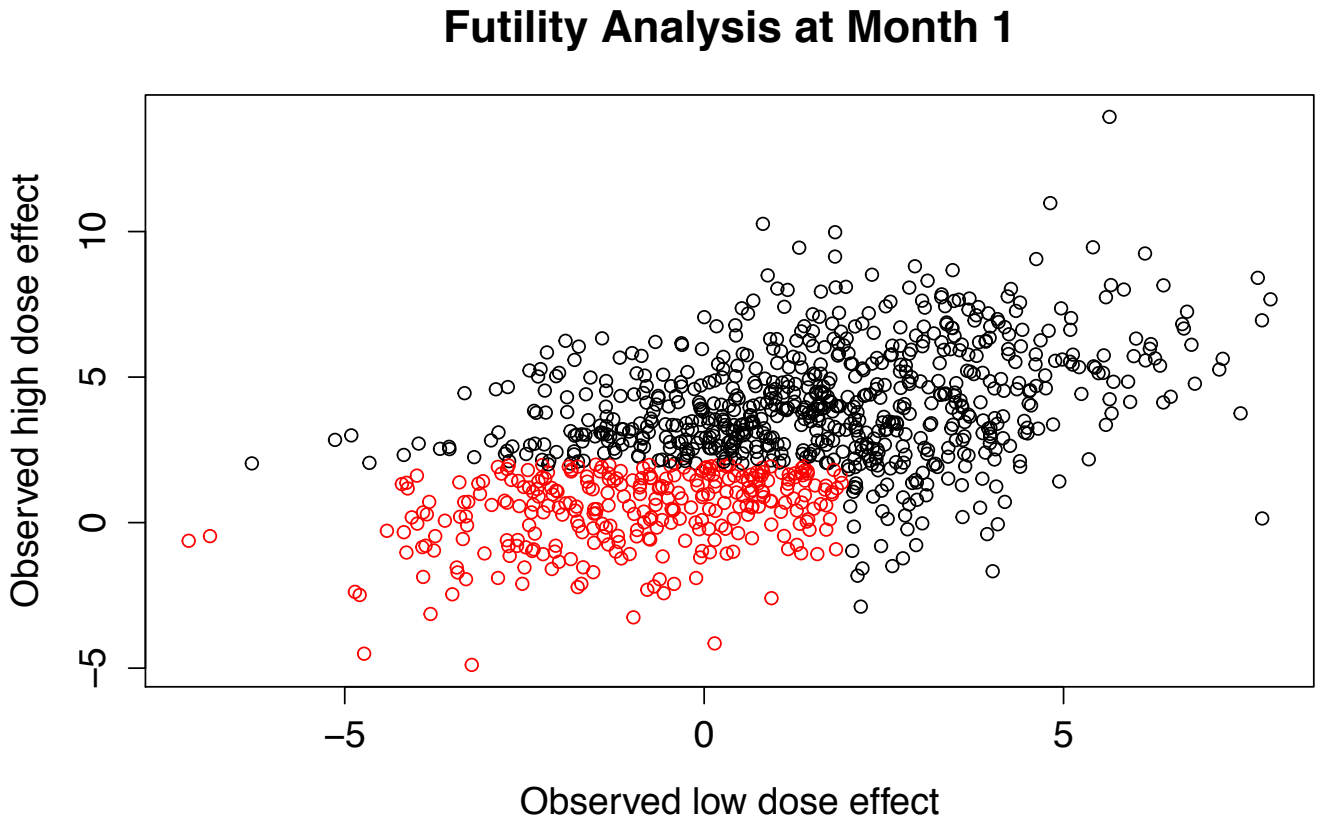
High variation here! Lots of trials that continue have observed treatment effect on the order of 3, 4, 5

Expected sample size
679 trials N=36
321 trials N=216
Average N=93.8

Simulate 1000 trials when the drug does work

- While stopping a lot of bad drugs is good, we do NOT want to stop good drugs
  - Suppose it works
    - low effect 1, high effect 3 (so high is good)
  - Now simulate 1000 trials under this condition

Simulate 1000 trials when the drug DOES work (1,3)

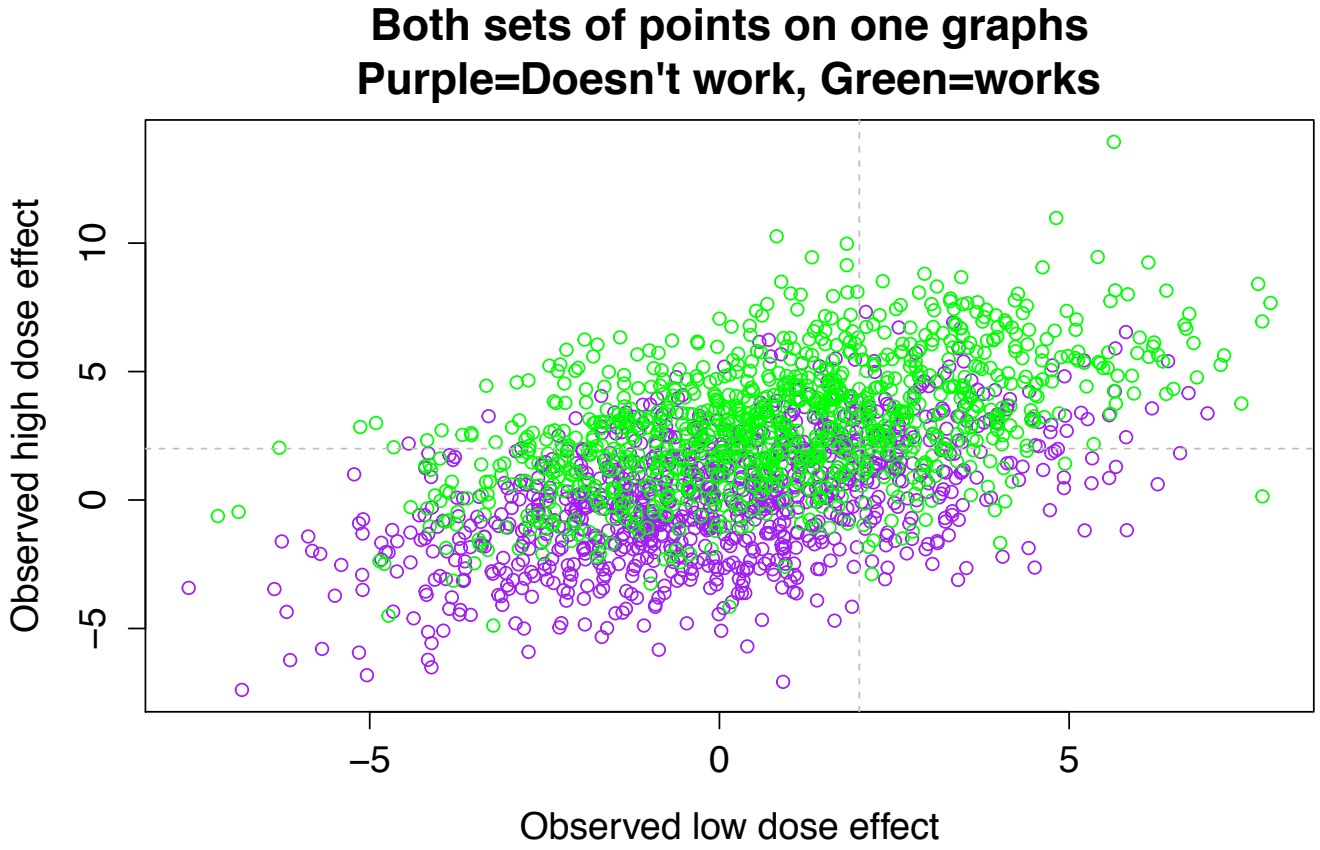


305 trials are stopped for futility UHOH

Even with a good effect on the high dose, the high dose gets unlucky a LOT.

These early futilities directly lower power.

Simulate 1000 trials under both conditions



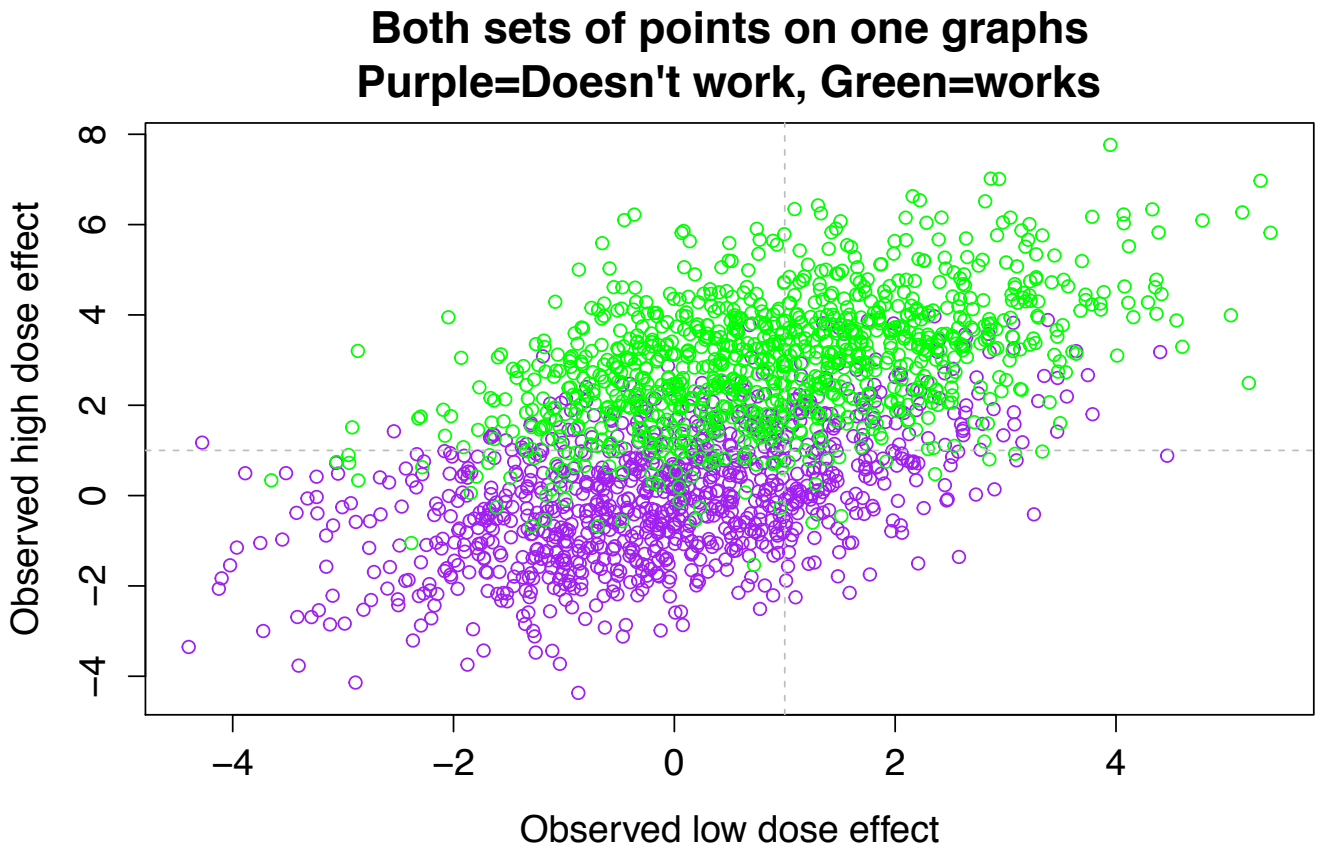
These distributions overlap a lot.
any futility rule which removes a lot of the purple will also remove a lot of green.

If we don't want to eliminate good drugs, need MUCH less restrictive cutoffs

New futility rule

- Need to avoid stopping drugs that work
- Look later (month 3 with dose selection)
  - better discrimination between drugs that work and drugs that don't
- Change form of rule to something more "statistical"
  - neither dose has $p<0.25$ compared to placebo
  - accounts for variation, scales with sample size

Simulate 1000 trials
3 month futility look



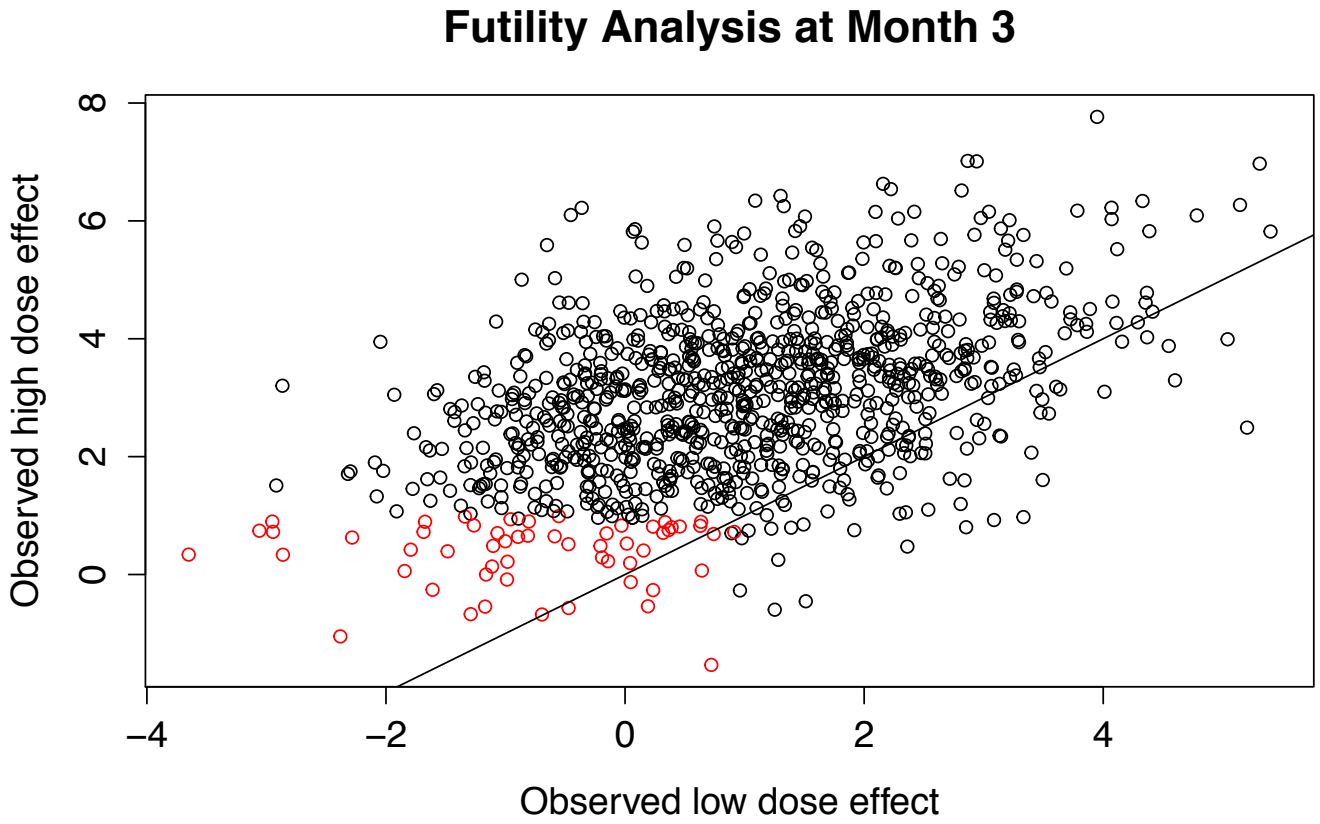
These distributions overlap less.
More discrimination

p<0.25 at 3 month corresponds to about a difference of 1.0

Declares only 5.6% of effective drugs futile
Declares 62.6% of ineffective drugs futile.

Expected N=148.4 when ineffective

Simulate 1000 trials (truth 1,3)
3 months including dose selection



56 Red points futile
944 Black points continue

Below the line you pick the low dose (73 times)

Above the line you pick the high dose (871 times)

Dose selection works pretty well.

Total results for 1000 simulations

For ineffective drugs
(true effects 0,0)

Dose Selection	Successes	Failures
None (futility)	0	626
Low dose	11	187
High Dose	9	167

11+9 = 20 successes = 2%
these are type I errors

For effective drugs
(true effects 1,3)

Dose Selection	Successes	Failures
None (futility)	0	56
Low dose	25	48
High Dose	795	76

795+25 = 820 successes = 82%
this is the power
(although the 25 low dose successes are “type 3 errors”?)

Comparison to fixed trials

	Fixed	Adaptive
Type I error rate	2.3%	2.0%
Power	77.1%	82.0%
Futility savings when drug doesn't work	None	Save half the study 63% of the time
Sample size on selected dose	72	90

Summary

- Adaptive trials allow you to prospectively change the trial based on incoming information.
- Avoids inefficiency due to uncertainty prior to trial start.
- Complex adaptive trials require simulation to assess operating characteristics
- Simulations can guide better decisions.