Bayesian Adaptive Clinical Trial Design

Jason Connor
ConfluenceStat

Jason@ConfluenceStat.com
412-860-3113

Day 2

1

NOTE TO SELF:
START THE RECORDING

2

Phase 3 / Confirmatory Trials

- CDER/CBER Phase 3
- CDRH Confirmatory
- The final test before market
- Control of Type I error rate very important
- Tend not to adaptively randomize
  - Fear of drift
  - Usually two arm
  - No power benefit with adaptive rand. in 2-arm trial

3

3

What is Different About Confirmatory Trials

- Type I error is a dominant factor
- Adjusting the design (goal) in order to accommodate adaptive aspects must still control type I error
- Predictive probabilities much more relevant than posterior probabilities
- Very well-defined goal.
  - A "game" you win or lose

4

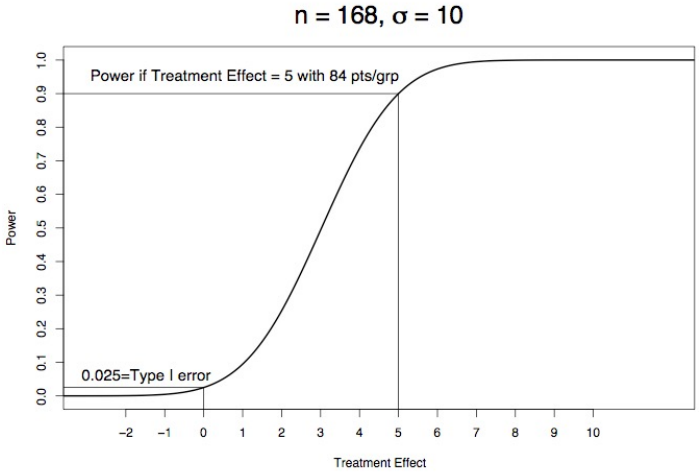
4

Power vs. Prob of Success

- Doctor comes to you.
- Claims her treatment increases IQ by 5 points
- SD = 10
- "How many patients do I need to have 90% power to demonstrate superiority?"

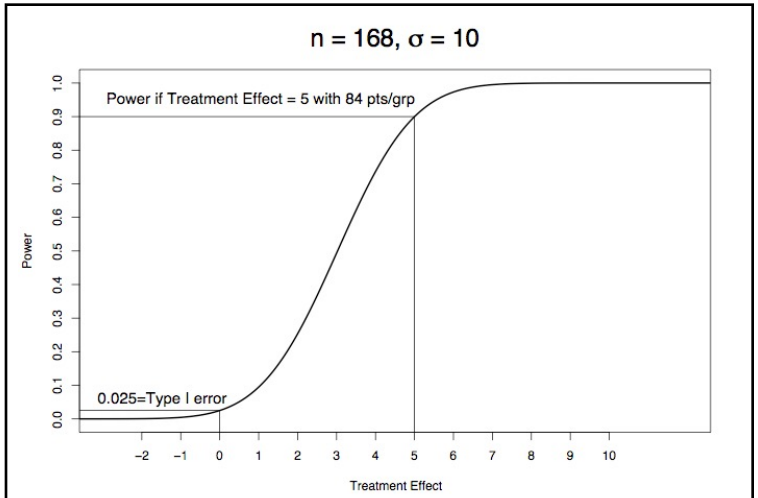
5

5



6

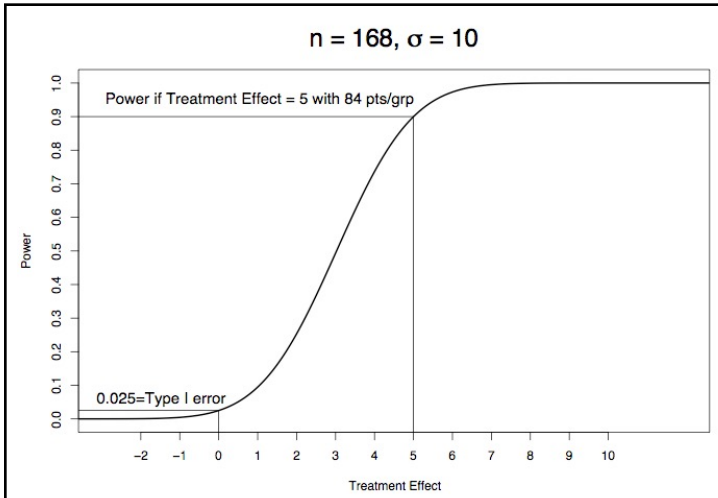
6



Δ = 6 → Power = 97% (↑ 7%)
Δ = 4 → Power = 73% (↓ 17%)

7

7

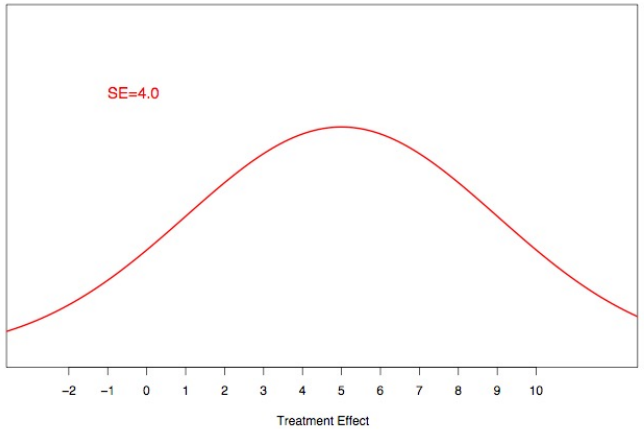


Δ = 7 → Power = 99% (↑ 9%)
Δ = 3 → Power = 49% (↓ 41%)

8

8

We've ignored the error in the pilot data

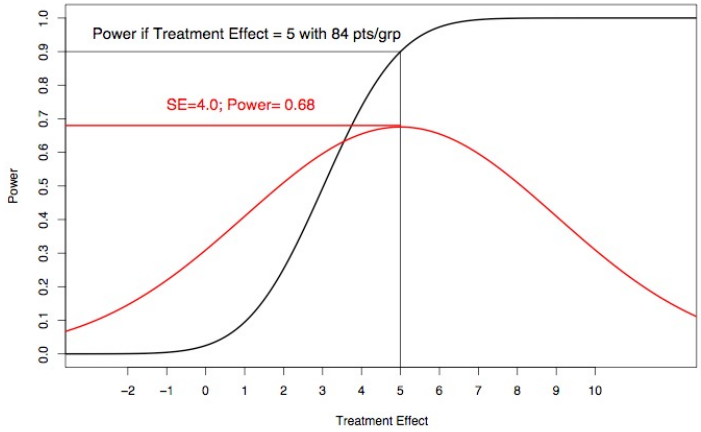


Estimate 5.0 (95% CI -3 to 13)

9

9

n = 168, σ = 10



Probability of success < Power due to Jensen's inequality

10

10

Quick Bayesian Introduction

11

11

Three people get a positive pregnancy test

- My sister with 4 kids who I know wants more
- You or your wife/gf. Using oral contraception
- Me

- What is the probability each person is pregnant?

12

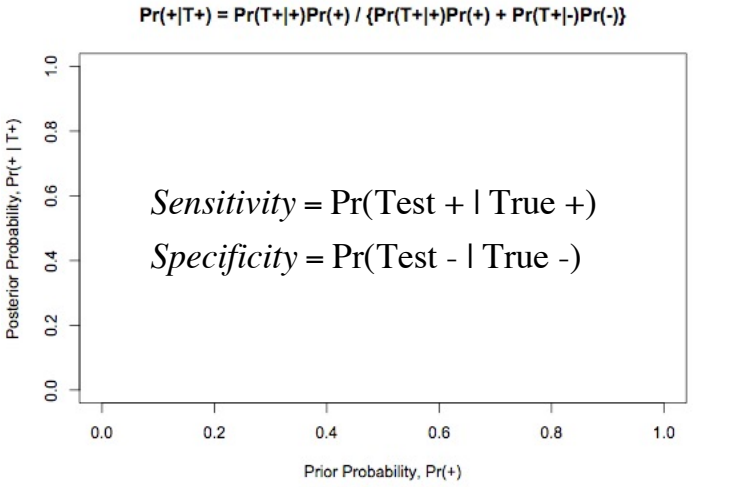
12

Three people get a positive pregnancy test

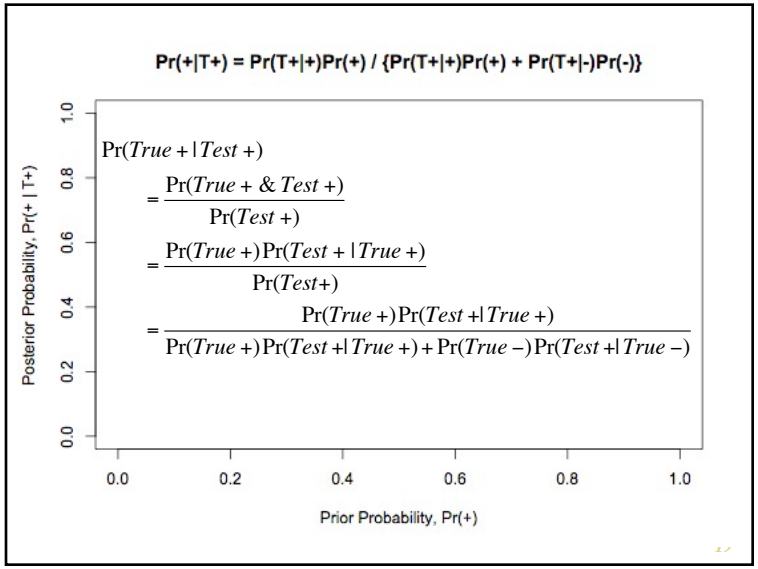
- My sister with 4 kids who I know wants more
- You or your wife/gf. Using oral contraception
- Me
  - Sensitivity 100%, Specificity 95%
- What is the probability each person is pregnant?

13

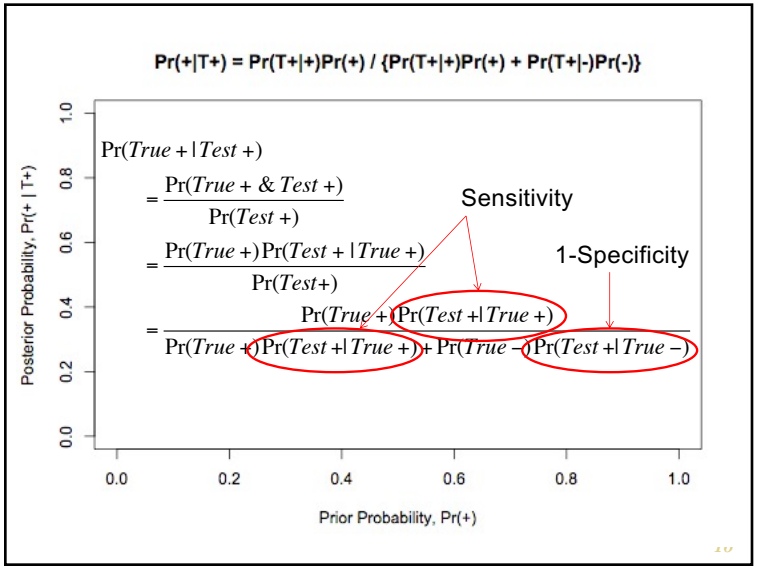
13



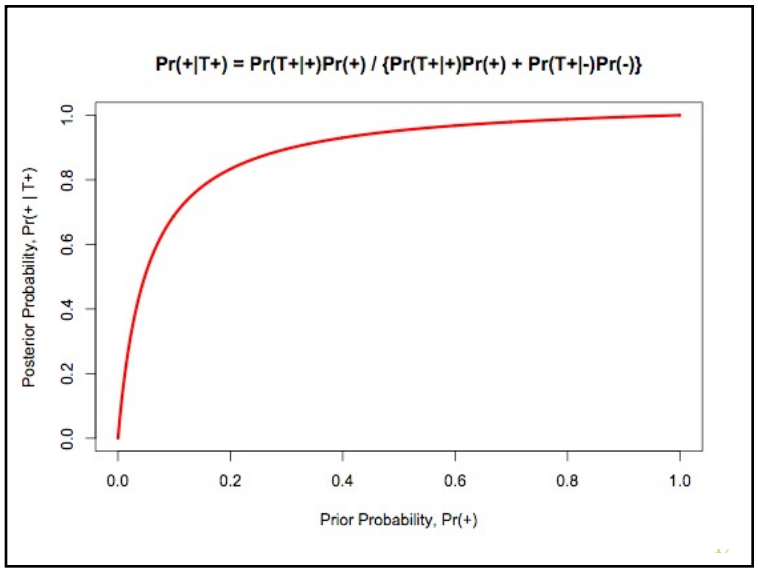
14



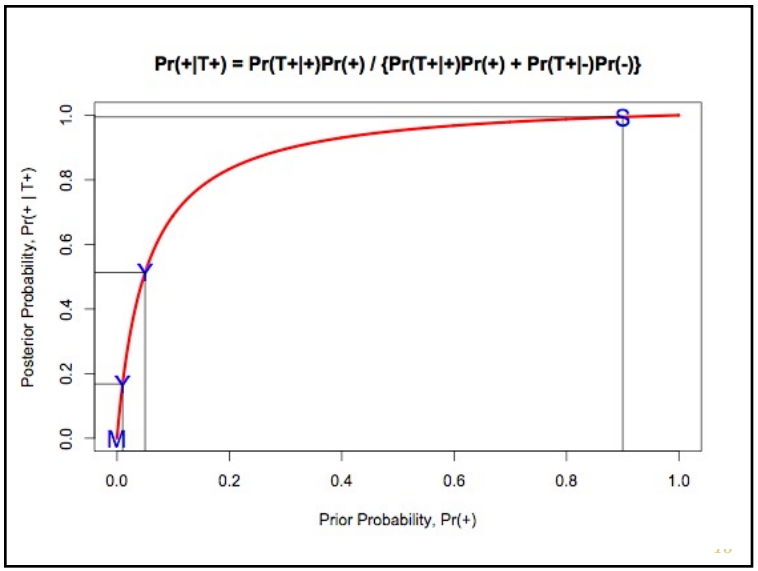
15



16



17



18

Bayes Theorem

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

$$= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid A^C)\Pr(A^C)}$$

$$\Pr(True+ \mid Test+) = \frac{\Pr(Test+ \mid True+)\Pr(True+)}{\Pr(Test+ \mid True+)\Pr(True+) + \Pr(Test+ \mid True-)\Pr(True-)}$$

$$PPV = \frac{Sensitivity \Pr(True+)}{Sensitivity \Pr(True+) + (1 - Specificity)\Pr(True-)}$$

19

19

Bayes Theorem

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

$$= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid A^C)\Pr(A^C)}$$

$$\Pr(Hypothesis \mid Data) = \frac{\Pr(Data \mid Hypothesis)\Pr(Hypothesis)}{\Pr(Data)}$$

$$= \frac{\Pr(Data \mid Hypothesis)\Pr(Hypothesis)}{\int_{\text{All possible hypotheses}} \Pr(Data \mid Hypothesis)\Pr(Hypothesis)}$$

20

20

Bayes Theorem

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

$$= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid A^C)\Pr(A^C)}$$

$$\Pr(Hypothesis \mid Data) = \frac{\Pr(Data \mid Hypothesis)\Pr(Hypothesis)}{\Pr(Data)}$$

$$= \frac{\Pr(Data \mid Hypothesis)\Pr(Hypothesis)}{\int_{\text{All possible hypotheses}} \Pr(Data \mid Hypothesis)\Pr(Hypothesis)}$$

21

21

Common prior for Binomial Outcome

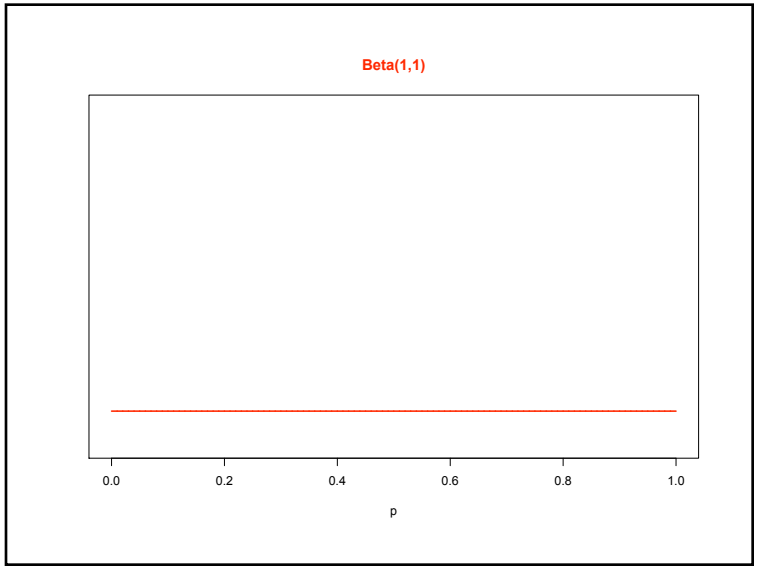
- Beta
  - used for event probabilities
  - conjugate with binomial
  - $x \sim$ Binomial(N, p)
  - $p \sim$ Beta(α,β), $p \in [0,1]$
  - $p \mid x \sim$ Beta(α+x, β+N-x)

$$E(p) = \frac{\alpha}{\alpha+\beta}; \qquad V(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

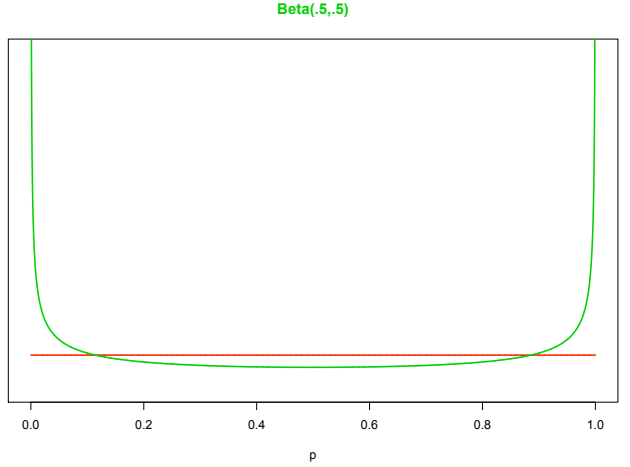
$$E(p \mid x) = \frac{\alpha+x}{\alpha+\beta+N}; \qquad V(p) = \frac{(\alpha+x)(\beta+N-x)}{(\alpha+\beta+N)^2(\alpha+\beta+N+1)}$$

22

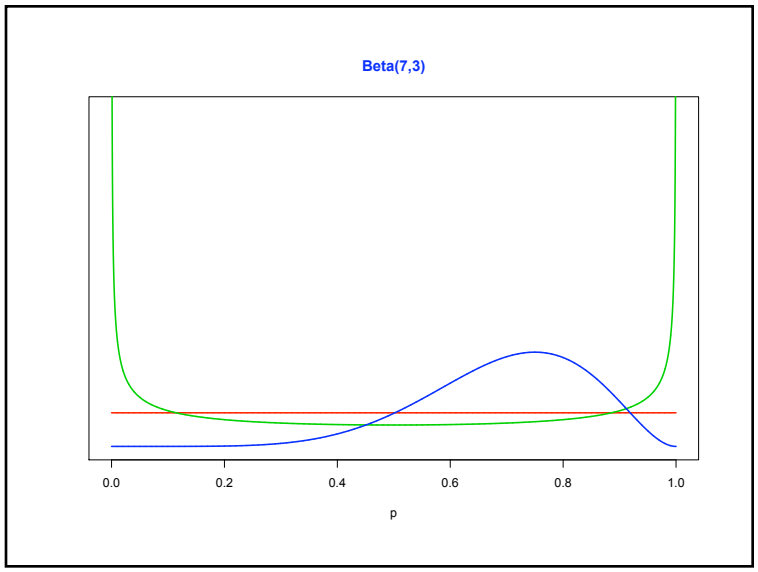
22



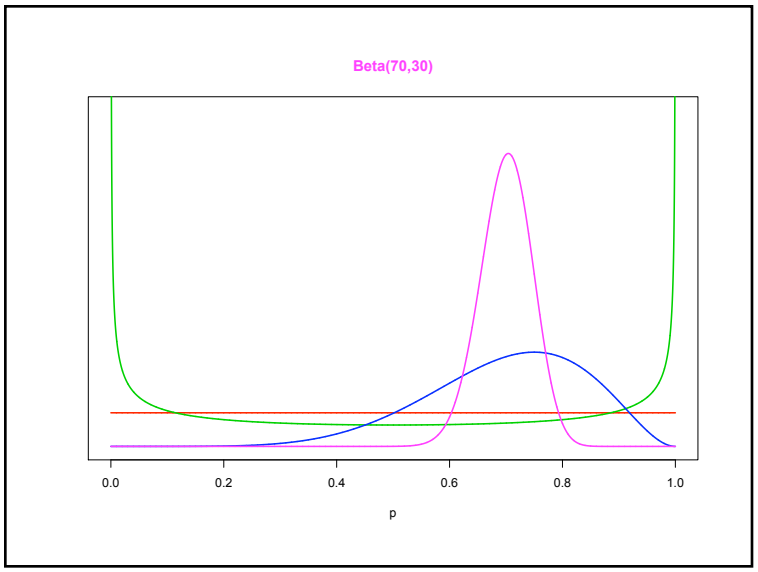
23



24



25



26

Beta Distribution

- $p \sim \text{Beta}(\alpha,\beta)$
  - α is like 'prior' number of successes
  - β is like 'prior' number of failures
  - $\alpha+\beta$ is the prior sample size (or amount of info)
- $p \mid N,x \sim \text{Beta}(\alpha+x, \beta+N-x)$
  - $\alpha+x$ is posterior number of successes
  - $\beta+N-x$ is posterior number of failures
- Posterior mean = $\alpha+x \,/\, (\alpha+\beta+N)$

27

27

Confirmatory Trials & Bayes

- You can't have an informative prior and control Type I error
  - Assuming the informative prior claims the treatment starts off better than the control

28

28

Simple Trial

- Binomial data
- One-armed trial
- n = 100
- Need to show p > 0.5
- H_o: p ≤ 0.5
- H_a: p > 0.5

- FYI: 59/100 → Frequentist p-value = 0.044
  & 1-sided 95% CI (0.503 – 1.00)

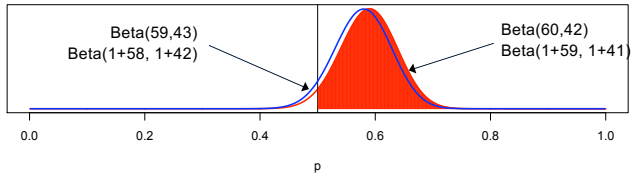
29

29

Phase 3 & Priors

- Simple Trial:
  - Binary data. Observe x ~ Bin(100,p)
  - Need to show Pr(p > 0.5 | x out of 100) > 0.95
  - Assume p ~ Beta(1,1) prior
  - Pr(p > 0.5 | 59 out of 100) = 0.963
  - Pr(P > 0.5 | 58 out of 100) = 0.944

1-sided p-value < 0.05
approx posterior > 0.95



30

30

Phase 3 & Priors

- Simple Trial:
  - Binary data. Observe x ~ Bin(100,p)
  - Need to show Pr(p > 0.5 | x out of 100) > 0.95
  - Assume p ~ Beta(1,1) prior
  - Pr(p > 0.5 | 59 out of 100) = 0.963
  - Pr(P > 0.5 | 58 out of 100) = 0.944
- Pr(X≥59 | p = 0.50) = 0.044
  - Simple binomial calculation
  - This is Type I error and is < 5%
  - Bayesian trial
  - Good frequentist properties

31

31

Phase 3 & Priors

- Now we have "pure" information
  9 of 10 successes in European trial
- Instead of Beta(1,1) prior use Beta(1+9, 1+1)
  = Beta(10,2) prior
- Regulatory agrees it is reasonable to use this as the prior
- Fixed design: for Pr[p > 0.5 | data] ≥ 0.95
  - Pr(p > 0.5 | 55 out of 100, α=10,β=2) = 0.956
  - Pr(P > 0.5 | 54 out of 100, α=10,β=2) = 0.936
- Pr(X≥55 | p = 0.50) = 0.184
  - Type I error is inflated

32

32

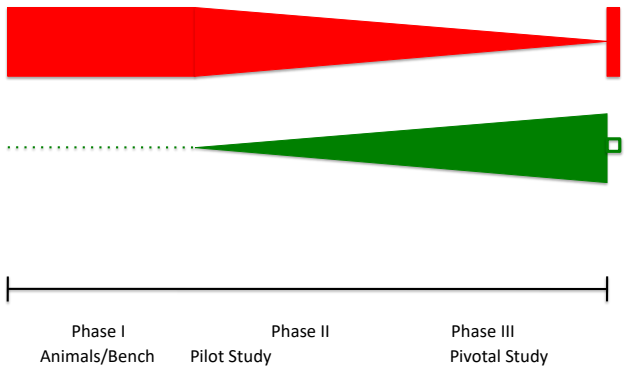
Phase 3 & Priors

- Solution to control Type I error
  - Raise the post probability threshold from 0.95 bar to 0.99 (like decreasing critical level)
  - Pr(p > 0.5 | 59 out of 100, α=10,β=2) = 0.993
  - Pr(P > 0.5 | 58 out of 100, α=10,β=2) = 0.989
  - Pr(X≥59 | p = 0.50) = 0.044
- Need a Beta(59+10,41+2) for a win…59 is back!!!
- The type I error "restriction" forces 59/100 regardless of prior…
- Can't allow beneficial priors AND force Type I of "new" experiment!

33

33

Posterior/Predictive



34

34

Predictive Probabilities

- Simple Trial:
  - Binary data. Observe $x \sim \text{Bin}(100, p)$
  - Need to show $\Pr(p > 0.5 \mid x \text{ out of } 100) > 0.95$
  - Assume $p \sim \text{Beta}(1,1)$ prior
  - $\Pr(p > 0.5 \mid 59 \text{ out of } 100) = 0.963$
  - $\Pr(p > 0.5 \mid 58 \text{ out of } 100) = 0.944$

35

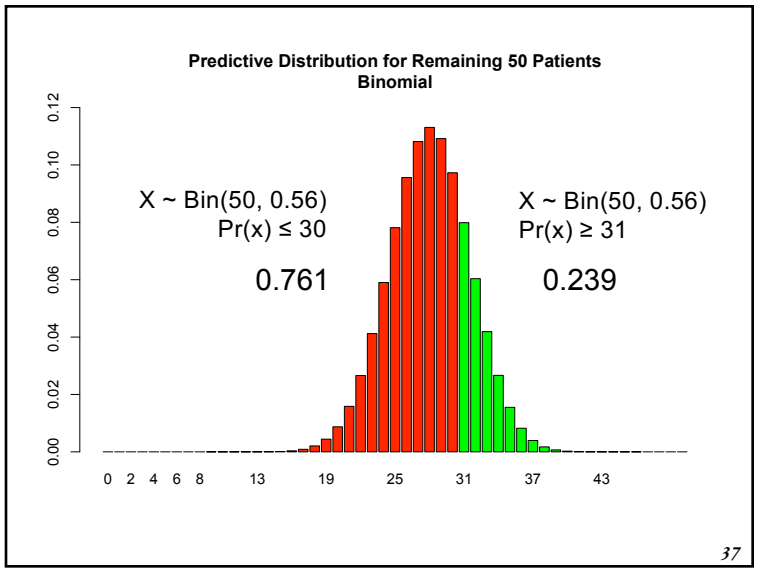
35

Predictive Probabilities

- Simple Trial:
  - Binary data. Observe $x \sim \text{Bin}(100, p)$
  - Need to show $\Pr(p > 0.5 \mid x \text{ out of } 100) > 0.95$
  - Assume $p \sim \text{Beta}(1,1)$ prior
  - $\Pr(p > 0.5 \mid 59 \text{ out of } 100) = 0.963$
  - $\Pr(p > 0.5 \mid 58 \text{ out of } 100) = 0.944$
- Observe data half way through
  - See 28/50 successes
  - Need to see 31/50 to meet threshold
  - What is predictive probability of trial success?

36

36



37

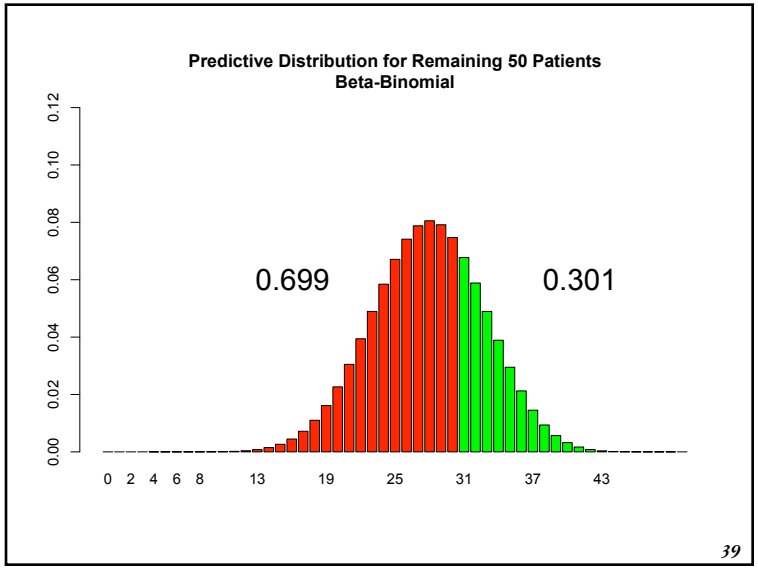
Predictive Probabilities

- Know we need $x \geq 59$ at trial's end
- Have $x_1 = 28$
- Need $x_2 \geq 31$
- $p \sim$ Beta(1+28, 1+22)
- $x_2 \sim$ Binomial(50, p)
- $x_2 \sim$ Beta-binomial(50, α=29, β=23)

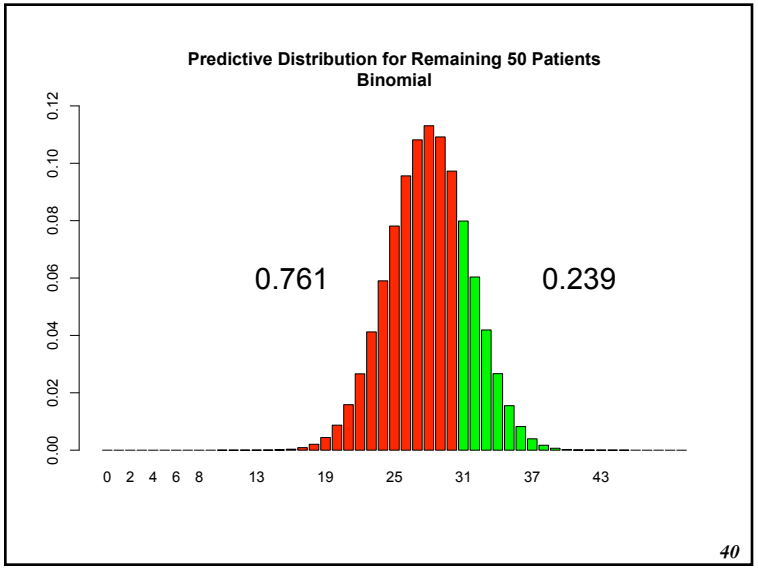
$$\Pr(\text{Win Trial}) = \sum_{x_2=31}^{50} \left\{ \binom{50}{x_2} \frac{B(x_2 + 29, 50 - x_2 + 23)}{B(29,22)} \right\} = 0.301$$

38

38



39



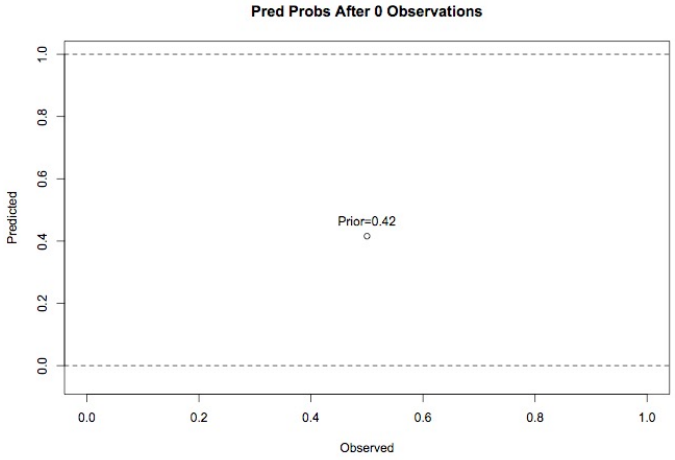
40

R code for predictive probability

```
> ### VIA SIMULATION
> alpha <- 1; beta <- 1
> x <- 28; N <- 50
>
> p <- rbeta(1000000, alpha+x, beta+N-x)
> x.new <- rbinom(1000000, 50, p)
>
> mean(x.new >= 31)
[1] 0.301132
>
>
>
> ### VIA DIRECT CALCULATION
> N.new <- 50
> x.new <- 0:50
> prob <- choose(N.new,x.new) *
+       beta(alpha+x+x.new,(beta+N-x)+(N-x.new)) /
+          beta(alpha+x,      (beta+N-x))
> sum(prob)
[1] 1
> sum(prob[x.new >= 31])
[1] 0.3010906
> barplot(prob, names.arg=0:50, col=c(rep(2,31), rep(3,20)),
+        main="Predictive Distribution for Remaining 50 patients")
```

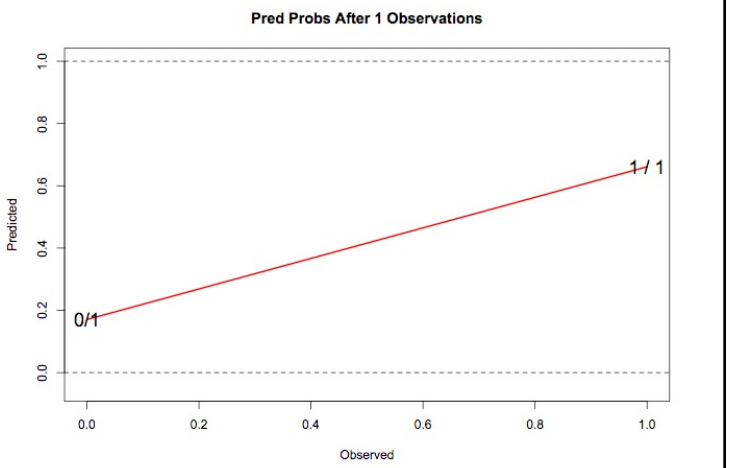
41

41



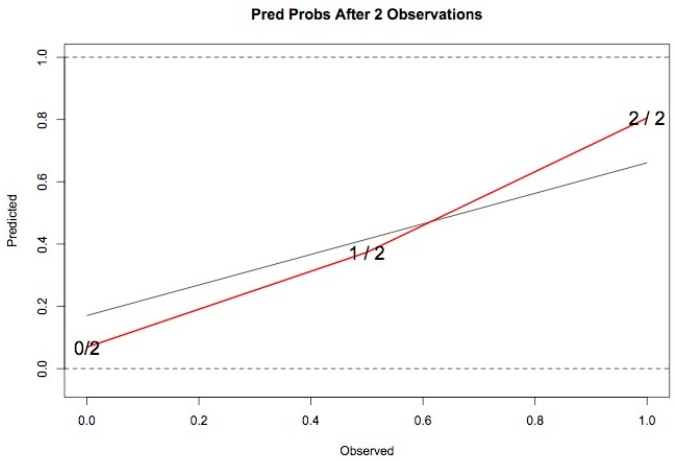
42

42



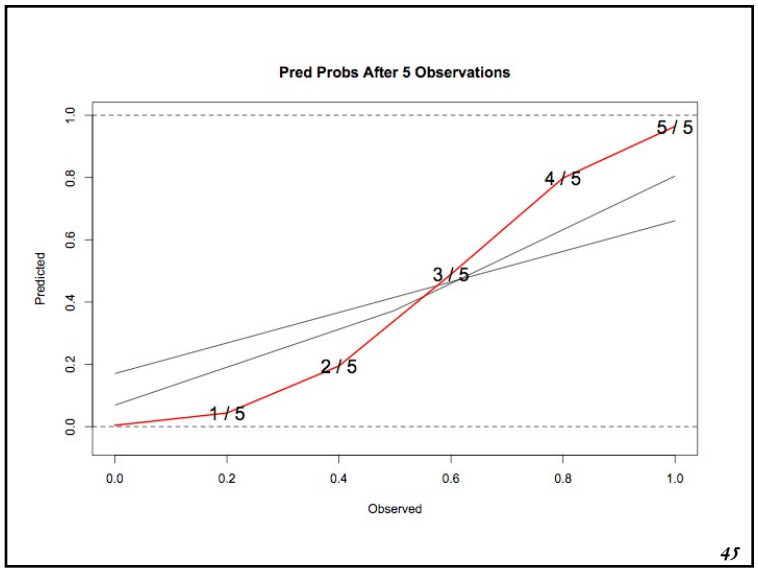
43

43

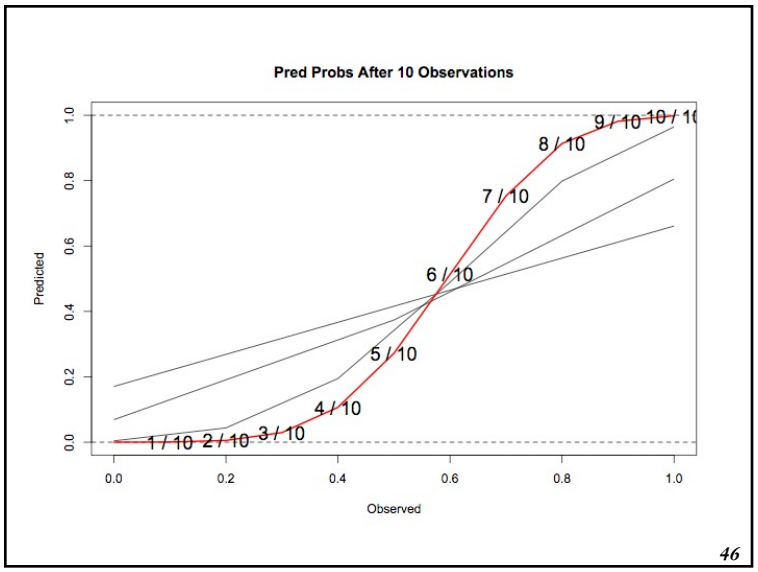


44

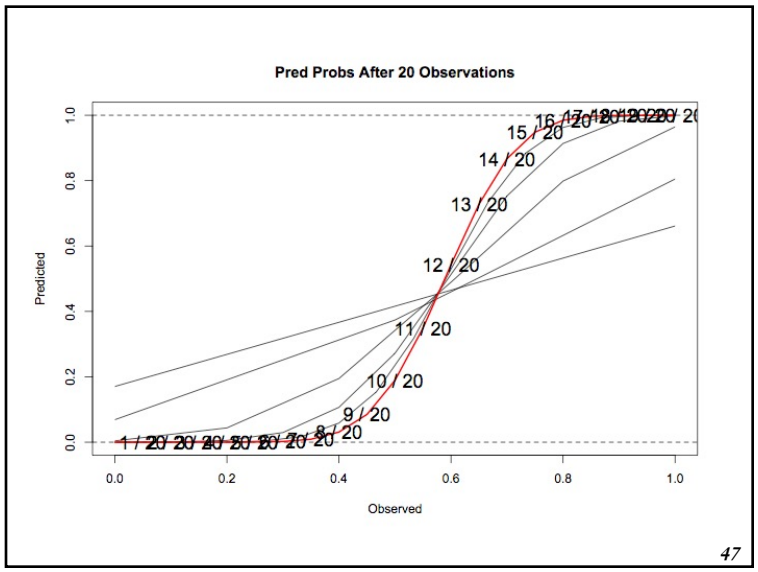
44


Pred Probs After 5 Observations
Predicted
Observed
0.0
0.2
0.4
0.6
0.8
1.0
1 / 5
2 / 5
3 / 5
4 / 5
5 / 5
45

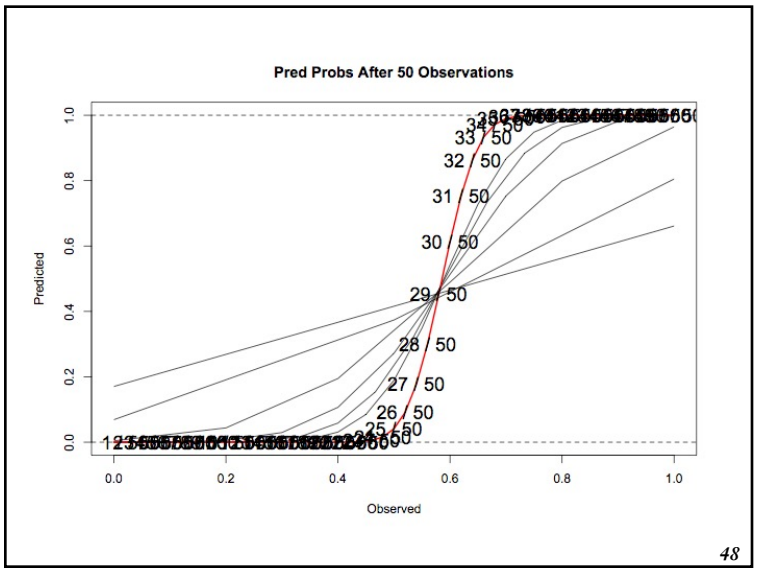
45


Pred Probs After 10 Observations
Predicted
Observed
0.0
0.2
0.4
0.6
0.8
1.0
1 / 10
2 / 10
3 / 10
4 / 10
5 / 10
6 / 10
7 / 10
8 / 10
9 / 10
46

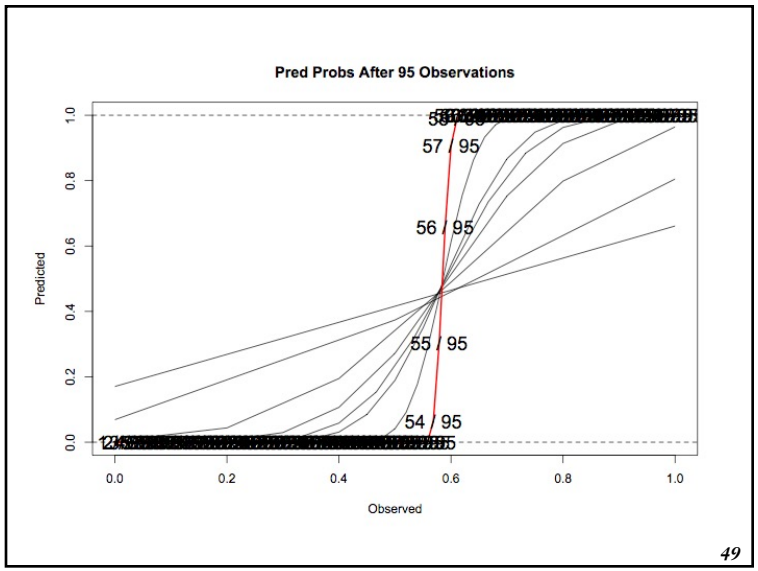
46


Pred Probs After 20 Observations
Predicted
Observed
0.0
0.2
0.4
0.6
0.8
1.0
9 / 20
10 / 20
11 / 20
12 / 20
13 / 20
14 / 20
15 / 20
47

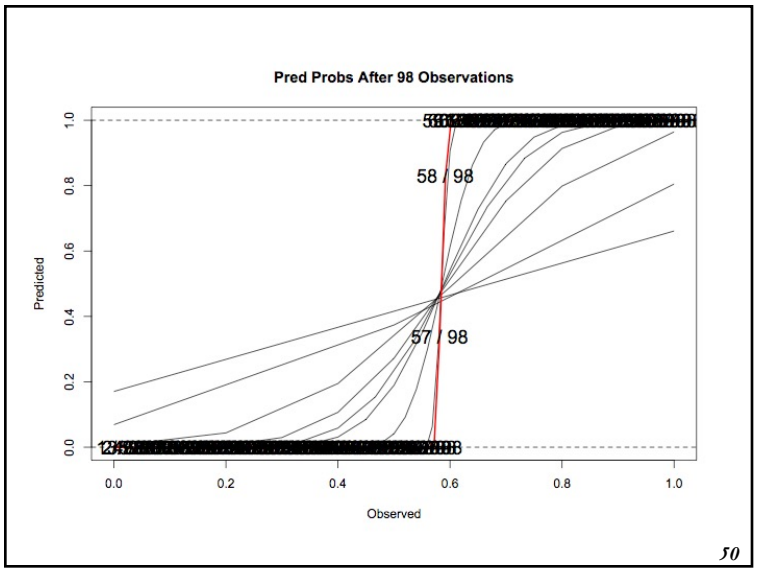
47


Pred Probs After 50 Observations
Predicted
Observed
0.0
0.2
0.4
0.6
0.8
1.0
26 / 50
27 / 50
28 / 50
29 / 50
30 / 50
31 / 50
32 / 50
33 / 50
48

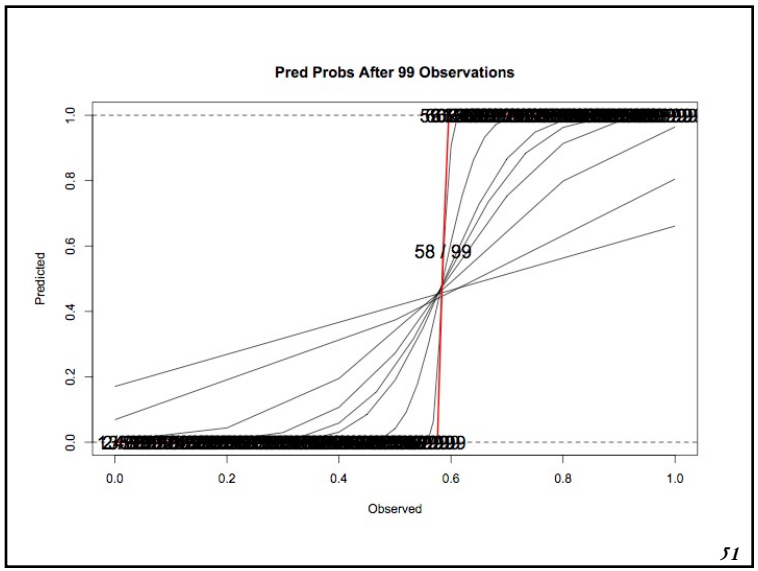
48



49



50



51

Predictive Probabilities

- Observe 12 / 20 (60%)
  - Need 47 / 80 successes; 59% or better rest of way
  - *p-value* = 0.25, Pr(p>0.5) = 0.81
  - Predictive probability of success @ 100 = 0.54

52

52

Predictive Probabilities

- Observe 12 / 20 (60%)
  - Need 47 / 80 successes; 59% or better rest of way
  - *p-value* = 0.25, Pr(p>0.5) = 0.81
  - Predictive probability of success @ 100 = 0.54
- Observe 28 / 50 (56%)
  - Need 31/50 successes; 62% or better rest of way
  - *p-value* = 0.24, Pr(p>0.5) = 0.80
  - Predictive probability of success @ 100 = 0.30

53

53

Predictive Probabilities

- Observe 12 / 20 (60%)
  - Need 47 / 80 successes; 59% or better rest of way
  - *p-value* = 0.25, Pr(p>0.5) = 0.81
  - Predictive probability of success @ 100 = 0.54
- Observe 28 / 50 (56%)
  - Need 31/50 successes; 62% or better rest of way
  - *p-value* = 0.24, Pr(p>0.5) = 0.80
  - Predictive probability of success @ 100 = 0.30
- Observe 41 / 75 (54.7%)
  - Need 18/25 successes; 72% or better rest of way
  - *p-value* = 0.24, Pr(p>0.5) = 0.79
  - Predictive probability of success @ 100 = 0.086

54

54

Predictive Probabilities

- Observe 12 / 20 (60%)
  - Need 47 / 80 successes; 59% or better rest of way
  - *p-value* = 0.25, Pr(p>0.5) = 0.81
  - Predictive probability of success @ 100 = 0.54
- Observe 28 / 50 (56%)
  - Need 31/50 successes; 62% or better rest of way
  - *p-value* = 0.24, Pr(p>0.5) = 0.80
  - Predictive probability of success @ 100 = 0.30
- Observe 41 / 75 (54.7%)
  - Need 18/25 successes; 72% or better rest of way
  - *p-value* = 0.24, Pr(p>0.5) = 0.79
  - Predictive probability of success @ 100 = 0.086

55

55

Another trial

- N_G = 100 in Treatment & Control Group
- Testing $p_t > p_c$ using Fisher's Exact Test @ 0.025
- Observe
  - 34/50 in Control Group
  - 41/50 in Treatment Group
- What is predictive probability of success?

56

56

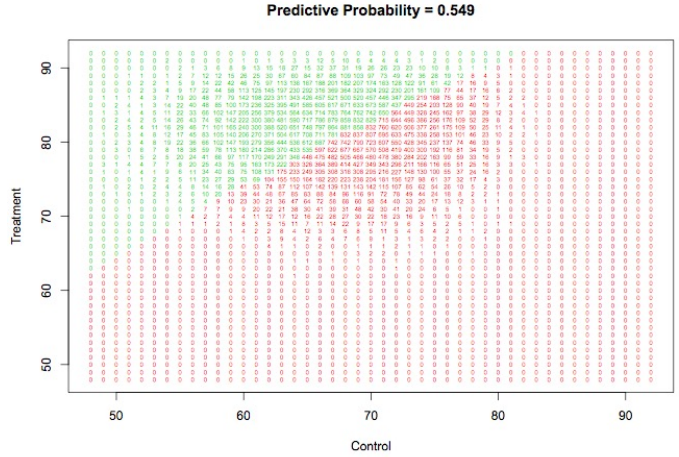
Predictive Probability

```
alpha <- 1; beta <- 1
xc <- 34; nc <- 50
xt <- 41; nt <- 50

pc <- rbeta(100000, alpha+xc, beta+nc-xc)
pt <- rbeta(100000, alpha+xt, beta+nt-xt)

xc.total <- xc + rbinom(100000, 50, pc)
xt.total <- xt + rbinom(100000, 50, pt)

p.values <- rep(NA,100000)
for(i in 1:100000){
  p.values[i] <- fisher.test(
      matrix(c(xc.total[i], 100-xc.total[i],
               xt.total[i], 100-xt.total[i]),nrow=2),
               alternative="less")$p.value
}

> mean(p.value<0.025)
[1] 0.549
```

57

57

GREEN numbers are when it's statistically superior
RED are cases not significant



58

58

But what if we have historical data

- Well known historical data, p_c = 60%±5%
- Expected from pilot studies, p_t = 80% ±15%

59

59

But what if we have historical data

- Well known historical data, p_c = 60%±5%
- Expected from pilot studies, p_t = 80% ±15%
- Beta distribution defined by p~Beta(α,β) has mean & variance

$$E(p)=\frac{\alpha}{\alpha+\beta} \qquad V(p)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- Solve for α & β

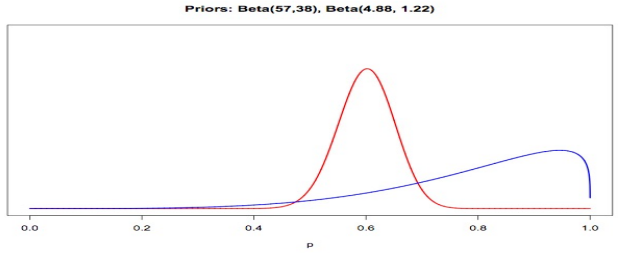
$$\frac{\alpha}{\alpha+\beta}=0.6 \qquad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=0.05^2$$

60

60

But what if we have historical data

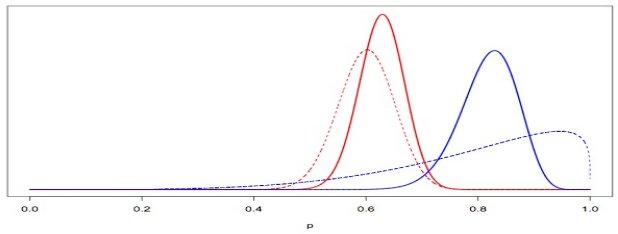
- Well known historical data, $p_c = 60\% \pm 5\%$
  - $\alpha_c = 57$, $\beta_c = 38$
- Expected from pilot studies, $p_t = 80\% \pm 15\%$
  - $\alpha_t = 4.8888$, $\beta_t = 1.2222$



61

61

```
alpha.c <- 57; beta.c <- 38; alpha.t <- 4.888888; beta.t <- 1.222222
xc <- 34; nc <- 50; xt <- 41; nt <- 50
pc <- rbeta(100000, alpha.c+xc, beta.c+nc-xc)
pt <- rbeta(100000, alpha.t+xt, beta.t+nt-xt)
xc.total <- xc + rbinom(100000, 50, pc)
xt.total <- xt + rbinom(100000, 50, pt)
p.values <- rep(NA,100000)
for(i in 1:100000){
  p.values[i] <- fisher.test(matrix(c(xc.total[i], 100-xc.total[i],
                                      xt.total[i], 100-xt.total[i]),nrow=2),
                                      alternative="less")$p.value
                    }
> mean(p.values<0.025)
[1] 0.73422
```

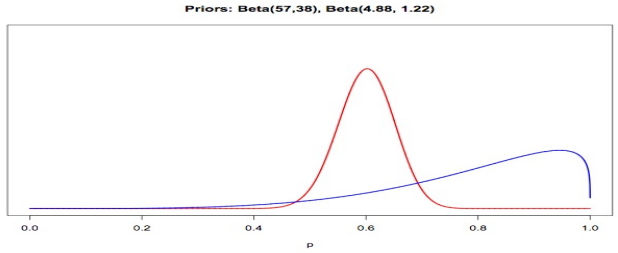


62

62

Downweight Historical Information

- Well known historical data, $p_c = 60\% \pm 5\%$
  - $\alpha_c = 57$, $\beta_c = 38$, 95 patients' worth of info
- Expected from pilot studies, $p_t = 80\% \pm 15\%$
  - $\alpha_t = 4.8888$, $\beta_t = 1.2222$ 6.1 pts' worth of info



63

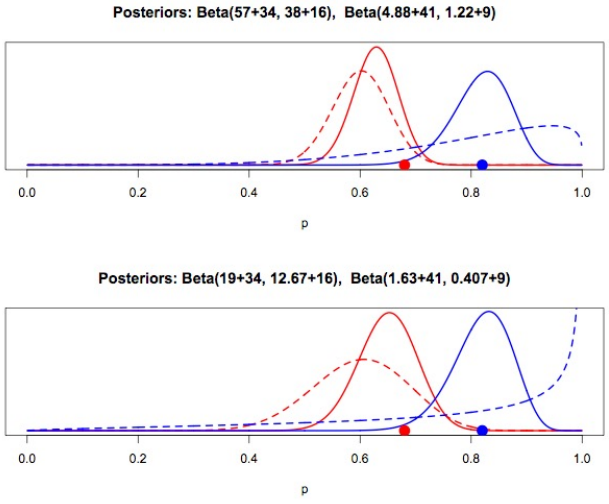
63

Downweight Historical Information

- Well known historical data, $p_c = 60\% \pm 5\%$
  - $\alpha_c = 57$, $\beta_c = 38$, 95 patients' worth of info
  - New data is 50/(50+95) = 34% of information
- Expected from pilot studies, $p_t = 80\% \pm 15\%$
  - $\alpha_t = 4.8888$, $\beta_t = 1.2222$, 6.1 patients' worth of info
  - New data is 50/(50+6) = 89% of information
- Downweight each prior so it includes 1/3 as much information
  - $\alpha_c = 19$, $\beta_c = 12.6667$, 31.67 patients' worth of info
  - $\alpha_t = 1.63$, $\beta_t = 0.407$, 2 patients' worth of info

64

64



65

65

66

Phase 3 Cancer Design

- Binary endpoint, complete response observed at 45 days post treatment
  - Consider CR vs. PFS vs. OS?
  - Primary analysis chi-square test
- Expect 20% improvement vs. control
- Use Bayesian prediction to determine sample size necessary for success in frequentist trial
  - Bayesian 'behind the curtain'

66

67

Statistical Model

- Final analysis: Chi-square test
- Interim analyses with
  - $N = N_c + N_t$ patients enrolled; $n = n_c + n_t$ complete
  - $x_c \sim$ Binomial(n_c, p_c); $p_c \sim$ Beta(1,1)
  - $x_t \sim$ Binomial(n_t, p_t); $p_t \sim$ Beta(1,1)
  - $N = N_c + N_t$ $N_c = n_c + n_c^*$ $N_t = n_t + n_t^*$
  - $x_c^* \sim$ Beta-binomial(n_c^*, $1+x_c$, $1+n_c-x_c$)
  - $x_t^* \sim$ Beta-binomial(n_t^*, $1+x_t$, $1+n_t-x_t$)

$$PP_N = \sum_{x_c^*=0}^{n_c^*} \sum_{x_t^*=0}^{n_t^*} pr\left(x_c^*\right) pr\left(x_t^*\right) I\left\{\chi^2_{p-value}\left(x_c + x_c^*, N_c, x_t + x_t^*, N_t\right) < 0.05\right\}$$

67

68

Design Questions

- What should sample size range be?
  - Most sponsor can do is 300 patients
    - Step 1, calculate power of fixed 300 patient trial

```
> bpower(n1=150, n2=150, p1=0.6, p2=0.8)
  Power
  0.969
```

  - Best case want to go to FDA with ≥150 patients
  - We'll see if 300 is enough, if not we'll go back to the company with evidence they need to up the cap

```
> bpower(n1=150, n2=150, p1=0.6, p2=0.75)
  Power
  0.795
```

Smallest win: 60% (80/150) vs. 72% (108/150) → p=0.03

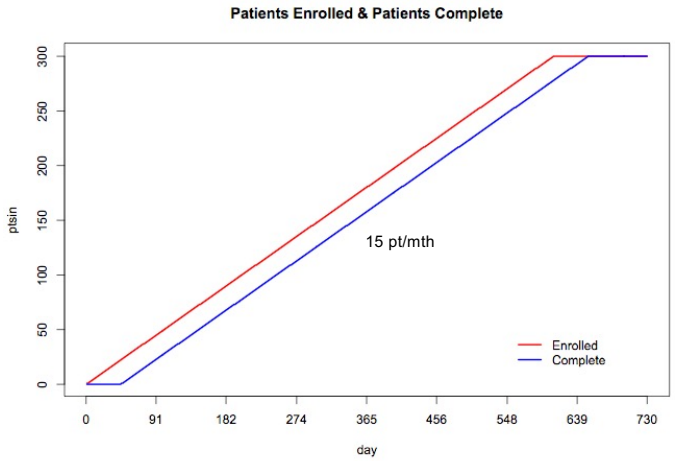
68

Design Questions

- Can we use an adaptive design?
  - Expect 15-25 patients per month
  - "Fast" outcome at 45 days
  - 22-37 outstanding patients at any analysis
  - If we do first look @ 150 patients enrolled
    - 128 with complete data with 15 pt/month accrual
    - 113 with complete data with 25 pt/month accrual
  - Usually accrual ramps up, assume constant here
  - Don't want to interfere with accrual
    - Don't pause accrual at each interim analysis
    - Decide whether to stop accrual while accruing

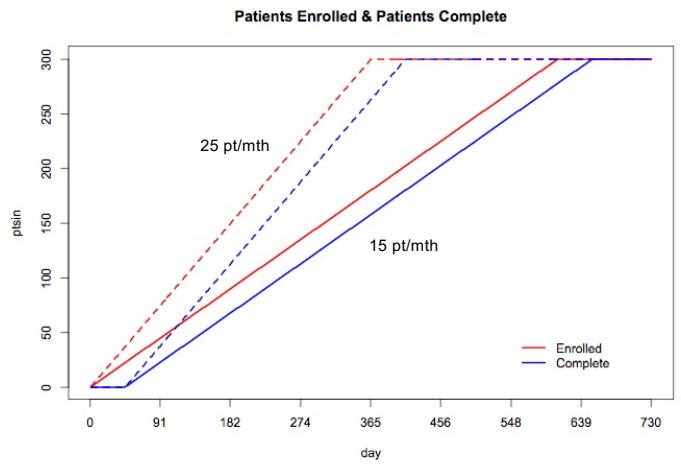
69

69



70

70



71

71

Design Questions

- How often to do interim looks?
  - Every 25 patients is every 1-1⅔ months
  - Manageable, may be CRO fee for every look

72

72

Design Questions

- How to decide when to stop accrual for predicted success?

73

73

Design Questions

- How to decide when to stop accrual for predicted success?
  - Use predictive probabilities
  - At each interim analysis ask

    "If we stop enrolling & wait for all outstanding patients to reach their 45-day outcomes, what is the probability we 'win'?"
  - If high, stop, wait, & analyze

    How high?

    I never want to stop then lose! (and so far haven't)

74

74

Design Questions

- How to decide when to stop accrual for futility (if at all)?

75

75

Design Questions

- How to decide when to stop accrual for futility (if at all)?
  - Use predictive probabilities
  - At each analysis ask

    "If we enrolling to the 300-patient maximum then wait for all patients to reach their 45-day outcomes, what is the probability we 'win'?"
  - If low, stop for futility?

    How low?

    More aggressive, more likely to stop a good trial

76

76

Design Questions

- What priors to use for predictive probabilities Beta dists?
  - Pretty new, let's be conservative with Beta(1,1) for treatment & control
  - Could use historical (or downweighted historical) priors here Incentive to have an 'honest' prior
  - Don't use prior in final analysis, frequentist test
- Stop for predicted success if $PP_N > S_N = 0.90$
- Stop for futility if $PP_{Nmax} < F_N = 0.10$

77

77

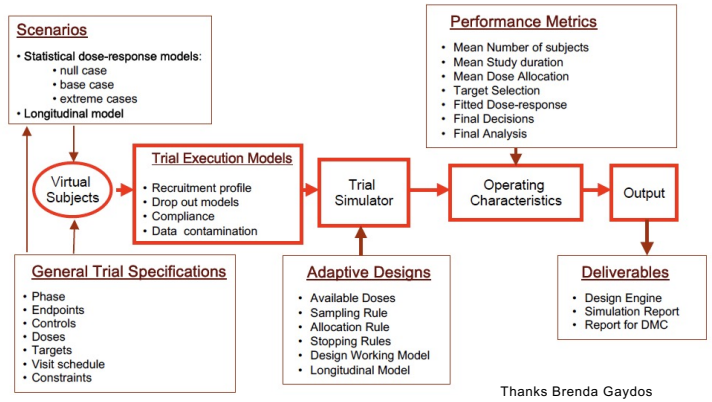
Sketch of my simulation code

- Define when to analyze, priors, cap, accrual rate, alpha level, efficacy
  - Factors I'll change a lot during discussions with sponsor
- Subroutine for patient accrual & randomization
- Subroutine to generate patient response & dropout
- Subroutine for interim analysis
  - Factors in time of analysis, which patients enrolled, which pts have outcomes
  - Outputs predictive probability of success with current *N* and at maximum *Nmax*
- Subroutine for decision
  - Stop for predicted success, stop for cap, stop for futility, keep going
- Final analysis at *n* where trial stopped
- Track trial size, win or lose, reason for stopping, number of looks, trial duration

78

78

Simulation Plan



Thanks Brenda Gaydos

79

79

```
Control Rate=    0.6000
 Exper Rate =    0.8000

Accrual Rate (pts/month):  15.00
        Number of Sims       1000
    Minimum Sample Size       150
    Maximum Sample Size       300
                    CV    0.0250
          Cuts   0.9000   0.1000

                  Mean       SD
    Sample Size 179.60    45.10

                 Lose       Win
   Success     0.008     0.897
       Cap     0.012     0.048
  Futility     0.035     0.000
     Total     0.055     0.945

Look      Lose       Win     Total
 150     0.020     0.565     0.585
 175     0.005     0.118     0.123
 200     0.002     0.091     0.093
 225     0.004     0.069     0.073
 250     0.006     0.028     0.034
 275     0.006     0.026     0.032
 300     0.012     0.048     0.060
 Tot     0.055     0.945     1.000
```

80

80

Control Rate= 0.6000
Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9000 0.1000

	Mean	SD
Sample Size	179.60	45.10

	Lose	Win
Success	0.008	0.897
Cap	0.012	0.048
Futility	0.035	0.000
Total	0.055	0.945

Look	Lose	Win	Total
150	0.020	0.565	0.585
175	0.005	0.118	0.123
200	0.002	0.091	0.093
225	0.004	0.069	0.073
250	0.006	0.028	0.034
275	0.006	0.026	0.032
300	0.012	0.048	0.060
Tot	0.055	0.945	1.000

Fixed trial of 300 provided 96.9% power

This design provides 94.5% power with average sample size just 180 patients

81

81

Control Rate= 0.6000
Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9000 0.1000

	Mean	SD
Sample Size	179.60	45.10

	Lose	Win
Success	0.008	0.897
Cap	0.012	0.048
Futility	0.035	0.000
Total	0.055	0.945

Look	Lose	Win	Total
150	0.020	0.565	0.585
175	0.005	0.118	0.123
200	0.002	0.091	0.093
225	0.004	0.069	0.073
250	0.006	0.028	0.034
275	0.006	0.026	0.032
300	0.012	0.048	0.060
Tot	0.055	0.945	1.000

Control Rate= 0.6000
Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9000 0.0000

	Mean	SD
Sample Size	182.65	49.86

	Lose	Win
Success	0.013	0.894
Cap	0.026	0.067
Futility	0.000	0.000
Total	0.039	0.961

Look	Lose	Win	Total
150	0.011	0.586	0.597
175	0.000	0.097	0.097
200	0.001	0.082	0.083
225	0.000	0.071	0.071
250	0.001	0.022	0.023
275	0.000	0.036	0.036
300	0.026	0.067	0.093
Tot	0.039	0.961	1.000

82

82

Stopping Boundaries, S_n, F_n

- Need not be constant
- We stopped for predicted success but lost at the first interim analysis in 1.1% of trials
  - I never want this to happen if I can avoid it!
- Let S_n be the success stopping bound
- Let F_n be the futility stopping bound
- Current $S_n = 0.9$ & $F_n = 0.1$ for all n
- Could choose $S_n = 0.99$ for small n
  & $S_n = 0.9$ for higher n

83

83

Control Rate= 0.6000
Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9000 0.0000

	Mean	SD
Sample Size	182.65	49.86

	Lose	Win
Success	0.013	0.894
Cap	0.026	0.067
Futility	0.000	0.000
Total	0.039	0.961

Look	Lose	Win	Total
150	0.011	0.586	0.597
175	0.000	0.097	0.097
200	0.001	0.082	0.083
225	0.000	0.071	0.071
250	0.001	0.022	0.023
275	0.000	0.036	0.036
300	0.026	0.067	0.093
Tot	0.039	0.961	1.000

Control Rate= 0.6000
Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9500 0.0000

	Mean	SD
Sample Size	186.47	53.61

	Lose	Win
Success	0.001	0.905
Cap	0.032	0.062
Futility	0.000	0.000
Total	0.033	0.967

Look	Lose	Win	Total
150	0.000	0.520	0.520
175	0.001	0.135	0.136
200	0.000	0.110	0.110
225	0.000	0.054	0.054
250	0.000	0.053	0.053
275	0.000	0.033	0.033
300	0.032	0.062	0.094
Tot	0.033	0.967	1.000

84

84

Control Rate= 0.6000 Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9500 0.0000

	Mean	SD
Sample Size	186.47	53.61

	Lose	Win
Success	0.001	0.905
Cap	0.032	0.062
Futility	0.000	0.000
Total	0.033	0.967

Look	Lose	Win	Total
150	0.000	0.520	0.520
175	0.001	0.135	0.136
200	0.000	0.110	0.110
225	0.000	0.054	0.054
250	0.000	0.053	0.053
275	0.000	0.033	0.033
300	0.032	0.062	0.094
Tot	0.033	0.967	1.000

85

85

Control Rate= 0.6000 Exper Rate = 0.8000

Accrual Rate (pts/month): 15.00
Number of Sims 1000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9500 0.0500

	Mean	SD
Sample Size	183.82	46.57

	Lose	Win
Success	0.001	0.915
Cap	0.014	0.048
Futility	0.022	0.000
Total	0.037	0.963

Look	Lose	Win	Total
150	0.012	0.513	0.525
175	0.003	0.139	0.142
200	0.004	0.108	0.112
225	0.001	0.061	0.062
250	0.000	0.056	0.056
275	0.003	0.038	0.042
300	0.014	0.048	0.063
Tot	0.037	0.963	1.000

86

86

Control Rate= 0.6000 Exper Rate = 0.7500

Accrual Rate (pts/month): 15.00
Number of Sims 5000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9500 0.0500

	Mean	SD
Sample Size	217.45	59.78

	Lose	Win
Success	0.009	0.639
Cap	0.083	0.152
Futility	0.116	0.000
Total	0.209	0.791

Look	Lose	Win	Total
150	0.044	0.260	0.304
175	0.017	0.100	0.117
200	0.012	0.086	0.098
225	0.016	0.068	0.084
250	0.018	0.067	0.085
275	0.019	0.057	0.076
300	0.083	0.152	0.235
Tot	0.209	0.791	1.000

87

87

Control Rate= 0.6000 Exper Rate = 0.6000

Accrual Rate (pts/month): 15.00
Number of Sims 5000
Minimum Sample Size 150
Maximum Sample Size 300
CV 0.0250
Cuts 0.9500 0.0500

	Mean	SD
Sample Size	187.32	49.97

	Lose	Win
Success	0.002	0.020
Cap	0.066	0.012
Futility	0.900	0.000
Total	0.968	0.032

Look	Lose	Win	Total
150	0.519	0.008	0.527
175	0.117	0.002	0.119
200	0.079	0.002	0.081
225	0.079	0.003	0.082
250	0.062	0.002	0.064
275	0.046	0.002	0.048
300	0.066	0.012	0.078
Tot	0.968	0.032	1.000

88

88

Enough!

- Settle on
  - Success Bound = 0.95
  - Futility Bound = 0.10
- Type I error was 0.028 -- too high
  - Pivotal trial, we need this to be ≤ 0.025
  - Hard to calculate analytically
  - Need to simulate over many scenarios
  - Then convince ourselves & FDA we've explored the whole null space

89

89

Intuition Check

- Use critical value = 0.025
- Simulate with 4 accrual rates, 10k sims/scenario
- Will the Type I error rates change with accrual rate? If so how?
- How will sample sizes change?

Accrual (pts/mth)	Mean N	Type I error
5		HIGHER OR LOWER
15*	177	0.030
25		
50		

*Slightly different than previous slide because 10,000 sims each

90

90

Intuition Check

- Use critical value = 0.025
- Simulate with 4 accrual rates, 10k sims/scenario
- Will the Type I error rates change with accrual rate? If so how?
- How will sample sizes change?

Accrual (pts/mth)	Mean N	Type I error
5	172	0.039
15	177	0.030
25	182	0.028
50	195	0.027

91

91

Find Critical Value for $\alpha = 0.025$

- Assume accrual won't be slower than 15/month
- Explore range of true p_c & p_t
- Find right critical value by trial & error
  - 10,000 sims each using 0.6 vs. 0.6
  - Sqrt(0.025*0.975/10000) = 0.0016

Critv	0.40	0.50	0.60	0.70	0.80
0.025			0.030		

92

92

Find Critical Value for α = 0.025

- Assume accrual won't be slower than 15/month
- Explore range of true p_c & p_t
- Find right critical value by trial & error
  - 10,000 sims each using 0.4 vs. 0.4 to 0.8 vs. 0.8
  - Sqrt(0.025*0.975/10000) = 0.0016

Critv	0.40	0.50	0.60	0.70	0.80
0.025			0.030		
0.020	0.024	0.026	0.026	0.024	0.025

93

93

Find Critical Value for α = 0.025

- Assume accrual won't be slower than 15/month
- Explore range of true p_c & p_t
- Find right critical value by trial & error
  - 10,000 sims each using 0.4 vs. 0.4 to 0.8 vs. 0.8
  - Sqrt(0.025*0.975/10000) = 0.0016

Critv	0.40	0.50	0.60	0.70	0.80
0.025			0.030		
0.020	0.024	0.026	0.026	0.024	0.025
0.018	0.024	0.021	0.023	0.023	0.020

94

94

Find Critical Value for α = 0.025

- Assume accrual won't be slower than 15/month
- Explore range of true p_c & p_t
- Find right critical value by trial & error
  - 10,000 sims each using 0.4 vs. 0.4 to 0.8 vs. 0.8
  - Sqrt(0.025*0.975/10000) = 0.0016

Critv	0.40	0.50	0.60	0.70	0.80
0.025			0.030		
0.020	0.024	0.026	0.026	0.024	0.025
0.018	0.024	0.021	0.024	0.023	0.020
0.019	0.022	0.026	0.024	0.024	0.024

Let's go with 0.018

If a real trial I'd run 100,000 or 1M sims and try to get as much power as possible

95

95

Example Trial #1

```
Simulation #  14        Analysis #  150
 Group          N     Obs     Suc
 Control       75      68      35      51%
 Treatment     75      68      49      72%
 P_N    =  0.9360 >  0.950 ? No,    P_Nmax =  0.9180 <  0.100 ? No
 Continue to enroll
```

96

96

Example Trial #1

```
Simulation #  14       Analysis #  150
 Group          N    Obs    Suc
 Control       75     68     35     51%
 Treatment     75     68     49     72%
 P_N   =  0.9360 >  0.950 ? No,   P_Nmax =  0.9180 <  0.100 ? No
 Continue to enroll

Simulation #  14       Analysis #  175
 Group          N    Obs    Suc
 Control       88     73     39     53%
 Treatment     87     72     53     74%
 P_N   =  0.9370 >  0.950 ? No,   P_Nmax =  0.9360 <  0.100 ? No
 Continue to enroll
```

97

97

Example Trial #1

```
Simulation #  14       Analysis #  150
 Group          N    Obs    Suc
 Control       75     68     35     51%
 Treatment     75     68     49     72%
 P_N   =  0.9360 >  0.950 ? No,   P_Nmax =  0.9180 <  0.100 ? No
 Continue to enroll

Simulation #  14       Analysis #  175
 Group          N    Obs    Suc
 Control       88     73     39     53%
 Treatment     87     72     53     74%
 P_N   =  0.9370 >  0.950 ? No,   P_Nmax =  0.9360 <  0.100 ? No
 Continue to enroll

Simulation #  14       Analysis #  200
 Group          N    Obs    Suc
 Control      100     91     48     53%
 Treatment    100     90     68     76%
 P_N   =  >.9999 >  0.950 ? YES,  P_Nmax =  0.9900 <  0.100 ? No
 Stop for predicted success
```

98

98

Example Trial #1

```
Simulation #  14       Analysis #  150
 Group          N    Obs    Suc
 Control       75     68     35     51%
 Treatment     75     68     49     72%
 P_N   =  0.9360 >  0.950 ? No,   P_Nmax =  0.9180 <  0.100 ? No
 Continue to enroll

Simulation #  14       Analysis #  175
 Group          N    Obs    Suc
 Control       88     73     39     53%
 Treatment     87     72     53     74%
 P_N   =  0.9370 >  0.950 ? No,   P_Nmax =  0.9360 <  0.100 ? No
 Continue to enroll

Simulation #  14       Analysis #  200
 Group          N    Obs    Suc
 Control      100     91     48     53%
 Treatment    100     90     68     76%
 P_N   =  >.9999 >  0.950 ? YES,  P_Nmax =  0.9900 <  0.100 ? No
 Stop for predicted success

Simulation #  14   Final Analysis  200
     Group      N    Obs    Suc
 Control      100    100     52     52%
 Treatment    100    100     76     76%
 Successful trial,   p-value = 0.001 < 0.0180
```

99

99

Example Trial #2

```
Simulation #  10       Analysis #  150
 Group          N    Obs    Suc
 Control       75     66     40     61%  (need to see +20
 Treatment     75     65     44     68%   successes for win @
   150)
 P_n   =  0.0000 >  0.950 ? No,   P_Nmax =  0.2590 <  0.100 ? No
 Continue to enroll

Simulation #  10       Analysis #  175
 Group          N    Obs    Suc
 Control       88     80     47     59%
 Treatment     87     79     51     65%
 P_n   =  0.0000 >  0.950 ? No,   P_Nmax =  0.1020 <  0.100 ? No
 Continue to enroll

Simulation #  10       Analysis #  200
 Group          N    Obs    Suc
 Control      100     90     55     61%  (need to see +18
 Treatment    100     89     57     64%   successes for win @
   300)
 P_n   =  0.0000 >  0.950 ? No,   P_Nmax =  0.0360 <  0.100 ? YES
 Stop for futility
 Unsuccessful trial
```

100

100

Final Operating Characteristics
$S_n = 0.95, F_n = 0.10$

p_c	p_t	Mean N	Futility	Max & Win	PredSuc & Win	Power
0.60	0.60	175	0.937	0.046 0.009	0.016 0.015	0.024
0.60	0.65	199	0.775	0.145 0.041	0.081 0.075	0.117
0.60	0.70	220	0.478	0.247 0.114	0.275 0.267	0.381
0.60	0.75	216	0.195	0.216 0.143	0.590 0.580	0.723
0.60	0.80	189	0.039	0.088 0.073	0.873 0.868	0.942

101

101

Final Operating Characteristics
$S_n = 0.95, F_n = 0.05$

p_c	p_t	Mean N	Futility	Max & Win	PredSuc & Win	Power
0.60	0.60	185	0.913	0.071 0.009	0.017 0.015	0.025
0.60	0.65	212	0.716	0.200 0.053	0.084 0.079	0.132
0.60	0.70	231	0.407	0.314 0.131	0.280 0.271	0.401
0.60	0.75	221	0.143	0.256 0.155	0.601 0.591	0.746
0.60	0.80	190	0.025	0.095 0.074	0.880 0.876	0.950

102

102

Final Operating Characteristics vs. Fixed Frequentist Trials

p_c	p_t	B-A Mean N	B-A Power	F-Power 300	F-Power BA Mean
0.60	0.60	175 185	0.024 0.025	0.025	0.025
0.60	0.65	199 212	0.12 0.13	0.14	0.11
0.60	0.70	220 231	0.38 0.40	0.44	0.34
0.60	0.75	216 221	0.72 0.75	0.79	0.66
0.60	0.80	189 190	0.94 0.95	0.969	0.86

103

103

Digression

104

104

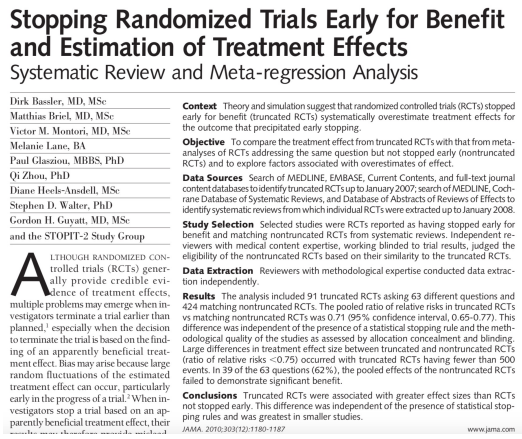
Imagine

- Imagine investigators do a case-control study
- Identify cases
  - Patients with hypertension
- Identify controls
  - People without hypertension with the same demographics (age, gender, marital status)
- See statistically significant increase in blood pressure between cases & controls
- Would JAMA publish this paper?

105

105

- They did: Bassler et al, March 23/31, 2010, V303, No12, 1180-1187.



Stopping Randomized Trials Early for Benefit and Estimation of Treatment Effects

Systematic Review and Meta-regression Analysis

Dirk Bassler, MD, MSc
Matthias Briel, MD, MSc
Victor M. Montori, MD, MSc
Melanie Lane, BA
Paul Glasziou, MBBS, PhD
Qi Zhou, PhD
Diane Heels-Ansdell, MSc
Stephen D. Walter, PhD
Gordon H. Guyatt, MD, MSc
and the STOPIT-2 Study Group

ALTHOUGH RANDOMIZED CONtrolled trials (RCTs) generally provide credible evidence of treatment effects, multiple problems may emerge when investigators terminate a trial earlier than planned,[1] especially when the decision to terminate the trial is based on the finding of an apparently beneficial treatment effect. Bias may arise because large random fluctuations of the estimated treatment effect can occur, particularly early in the progress of a trial.[2] When investigators stop a trial based on an apparently beneficial treatment effect, their

Context Theory and simulation suggest that randomized controlled trials (RCTs) stopped early for benefit (truncated RCTs) systematically overestimate treatment effects for the outcome that precipitated early stopping.

Objective To compare the treatment effect from truncated RCTs with that from meta-analyses of RCTs addressing the same question but not stopped early (nontruncated RCTs) and to explore factors associated with overestimates of effect.

Data Sources Search of MEDLINE, EMBASE, Current Contents, and full-text journal content databases to identify truncated RCTs up to January 2007; search of MEDLINE, Cochrane Database of Systematic Reviews, and Database of Abstracts of Reviews of Effects to identify systematic reviews from which individual RCTs were extracted up to January 2008.

Study Selection Selected studies were RCTs reported as having stopped early for benefit and matching nontruncated RCTs from systematic reviews. Independent reviewers with medical content expertise, working blinded to trial results, judged the eligibility of the nontruncated RCTs based on their similarity to the truncated RCTs.

Data Extraction Reviewers with methodological expertise conducted data extraction independently.

Results The analysis included 91 truncated RCTs asking 63 different questions and 424 matching nontruncated RCTs. The pooled ratio of relative risks in truncated RCTs vs matching nontruncated RCTs was 0.71 (95% confidence interval, 0.65-0.77). This difference was independent of the presence of a statistical stopping rule and the methodological quality of the studies as assessed by allocation concealment and blinding. Large differences in treatment effect size between truncated and nontruncated RCTs (ratio of relative risks <0.75) occurred with truncated RCTs having fewer than 500 events. In 39 of the 63 questions (62%), the pooled effects of the nontruncated RCTs failed to demonstrate significant benefit.

Conclusions Truncated RCTs were associated with greater effect sizes than RCTs not stopped early. This difference was independent of the presence of statistical stopping rules and was greatest in smaller studies.

JAMA. 2010;303(12):1180-1187 www.jama.com

106

106

From Abstract Study Selection

"Selected studies were RCTs reported as having stopped early for benefit and matching nontruncated RCTs from systematic reviews. Independent reviewers with medical content expertise, working blinded to trial results, judged the eligibility of the nontruncated RCTs based on their similarity to the truncated RCTs."

107

107

From Abstract Results

- Large differences in treatment effect size between truncated and nontruncated RCTs occurred ….
- In 39 of the 63 questions (62%), the pooled effects of the nontruncated RCTs failed to demonstrate significant benefit.

108

108

Control Rate=	0.6000	
Exper Rate =	0.7500	
Accrual Rate (pts/month):	15.00	
Number of Sims	5000	
Minimum Sample Size	150	
Maximum Sample Size	300	
CV	0.0250	
Cuts	0.9500	0.0500
	Mean	SD
Sample Size	217.45	59.78

	Lose	Win
Success	0.009	0.639
Cap	0.083	0.152
Futility	0.116	0.000
Total	0.209	0.791

Look	Lose	Win	Total
150	0.044	0.260	0.304
175	0.017	0.100	0.117
200	0.012	0.086	0.098
225	0.016	0.068	0.084
250	0.018	0.067	0.085
275	0.019	0.057	0.076
300	0.083	0.152	0.235
Tot	0.209	0.791	1.000

- Example 1 Revisited
- Truth is 15% benefit
- But 23.5% of time trial goes to maximum … and 8.3% it fails to be stat sig.
- The reason it goes to max is because data is ambiguous
- So of course the ones that go to max have small effects

109

109

S. Berry, Carlin, Connor

"To illustrate the issue, consider a clinical trial in which analysis is as follows: participants found to be performing better are retrospectively placed in the experimental group and participants found not to be performing well are retrospectively placed in the control group; a statistically significant difference in outcome is found when the groups are compared. It is clear that post-treatment selection of participants, based on their outcomes, would be responsible for any observed difference."

"This is logically equivalent to the analysis reported by Bassler et al."

110

110

Goodman, D. Berry, Wittes

"So comparing the truncated trials to the nontruncated trials is similar to comparing completed trials with large effects with those with lower effects. The difference the authors observed was both predictable and uninformative."

"Bias is a property of study procedures; it is not logically applicable to a subset of results."

Goodman SN. Systematic reviews are not biased by results from trials stopped early for benefit. *J Clin Epidemiol.* 2008;61(1):95-96.

111

111

$p_t = 0.8$ vs. $p_c = 0.6$

n=180 → 80% Power

- What is average effect size in the statistically significant trials?

112

112

$p_t = 0.8$ vs. $p_c = 0.6$
n=180 → 80% Power

- What is average effect size in the statistically significant trials?
- What is the average effect size in 100,000 simulated trials?

113

113

```
pvalue <- NULL; effect <- NULL

for(i in 1:100000){
x.c <- rbinom(1, 90, 0.6)
x.t <- rbinom(1, 90, 0.8)
mat <- rbind(c(x.c, 90-x.c), c(x.t, 90-x.t))
test <- chisq.test(mat)
pvalue[i] <- test$p.value
effect[i] <- x.t/90 - x.c/90
print(i)
}

hist(effect, breaks=seq(-.2, .6, by=0.025))
par(new=T)
hist(effect[pvalue<0.05], breaks=seq(-.2, .6, by=0.025), col=2, main=" ", xlab="
", ylab= " " )

> mean(pvalue < 0.05)      ### CHECK power = 80%
[1] 0.80313
> mean(effect)             ### CHECK mean effect = 0.20
[1] 0.2003593
>
> mean(effect[pvalue < 0.05])
[1] 0.2233924
> mean(effect[pvalue >= 0.05])
[1] 0.1063962
>
> 0.80 * .2233924 + 0.20 * 0.1063962
[1] 0.1999932
```

114

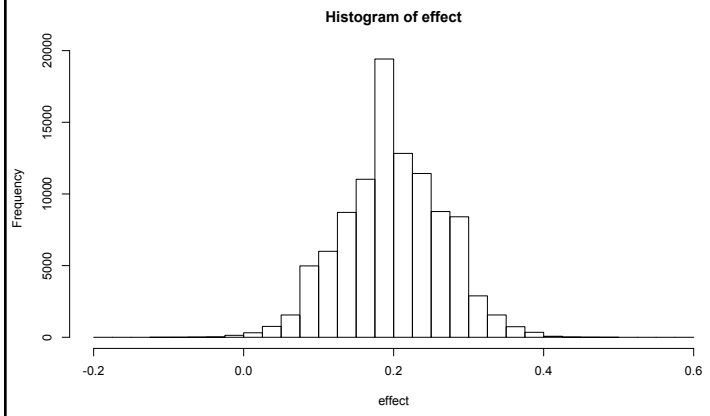
114

```
count <- 0
outcome <- matrix(nrow=8281, ncol=5)
for(xc in 0:90){
  for(xt in 0:90){
  count <- count + 1
  prob.of.pair <- dbinom(xc, 90, 0.6) * dbinom(xt, 90, 0.8)

  mat <- rbind(c(xc, 90-xc), c(xt, 90-xt))
  test <- chisq.test(mat)
  effect <- xt/90 - xc/90

  outcome[count, ] <- c(xc, xt, prob.of.pair, test$p.value, effect)
  print(c(xc, xt))
}}

outcome <- data.frame(outcome)
names(outcome) <- c("xc","xt","pr","pvalue","effect")

> sum(outcome$pr[outcome$pvalue < 0.05])
[1] 0.80168

> sum((outcome$effect * outcome$pr) [outcome$pvalue < 0.05]) /
sum(outcome$pr[outcome$pvalue<0.05])
[1] 0.2231661

> sum((outcome$effect * outcome$pr) [outcome$pvalue > 0.05]) /
sum(outcome$pr[outcome$pvalue>0.05])
[1] 0.1063544
```

115

115

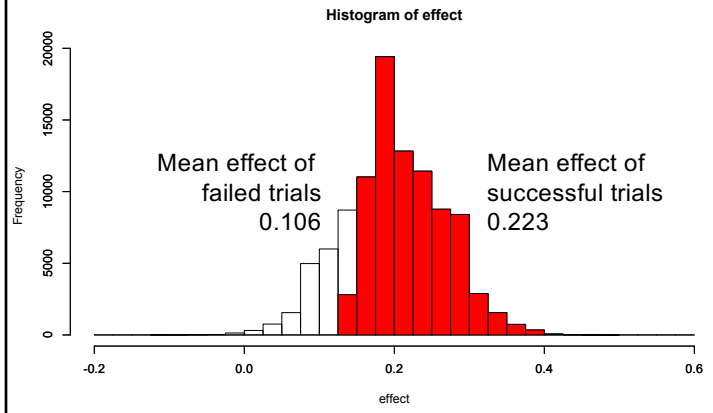
100k sims 0.8 vs. 0.6, n=180



116

116

100k sims 0.8 vs. 0.6, n=180



117

117

Revisit Example #1

- Binary outcome
- Adaptive trial from 150 to 300 patients
- Expected difference 60% vs. 80%

118

118

```
Control Rate=     0.6000
 Exper Rate =     0.8000

Accrual Rate (pts/month):  15.00
        Number of Sims      1000
    Minimum Sample Size      150
    Maximum Sample Size      300
                    CV   0.0250
          Cuts   0.9500   0.1000

                   Mean      SD
    Sample Size  183.20   48.53

               Lose      Win
   Success    0.001    0.892
       Cap    0.015    0.065
  Futility    0.027    0.000
     Total    0.043    0.957

Look     Lose      Win    Total
 150    0.017    0.546    0.564
 175    0.006    0.118    0.124
 200    0.001    0.093    0.094
 225    0.000    0.054    0.054
 250    0.002    0.049    0.051
 275    0.002    0.032    0.034
 300    0.015    0.065    0.080
 Tot    0.043    0.957    1.000
```

119

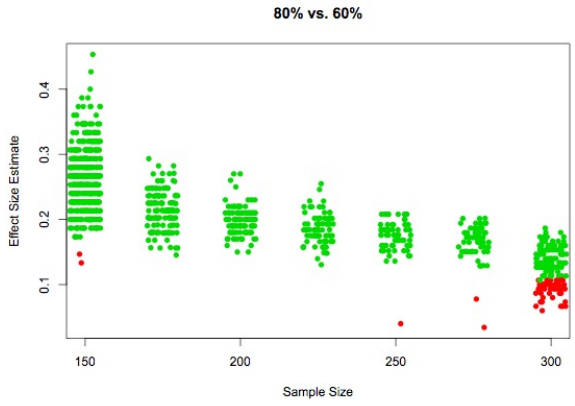
119

```
Control Rate=     0.6000
 Exper Rate =     0.6000

Accrual Rate (pts/month):  15.00
        Number of Sims      1000
    Minimum Sample Size      150
    Maximum Sample Size      300
                    CV   0.0250
          Cuts   0.9500   0.1000

                   Mean      SD
    Sample Size  176.31   44.02

               Lose      Win
   Success    0.002    0.019
       Cap    0.041    0.009
  Futility    0.929    0.000
     Total    0.972    0.028

Look     Lose      Win    Total
 150    0.634    0.006    0.640
 175    0.103    0.004    0.107
 200    0.073    0.003    0.076
 225    0.047    0.003    0.050
 250    0.042    0.002    0.044
 275    0.033    0.001    0.034
 300    0.041    0.009    0.050
 Tot    0.972    0.028    1.000
```

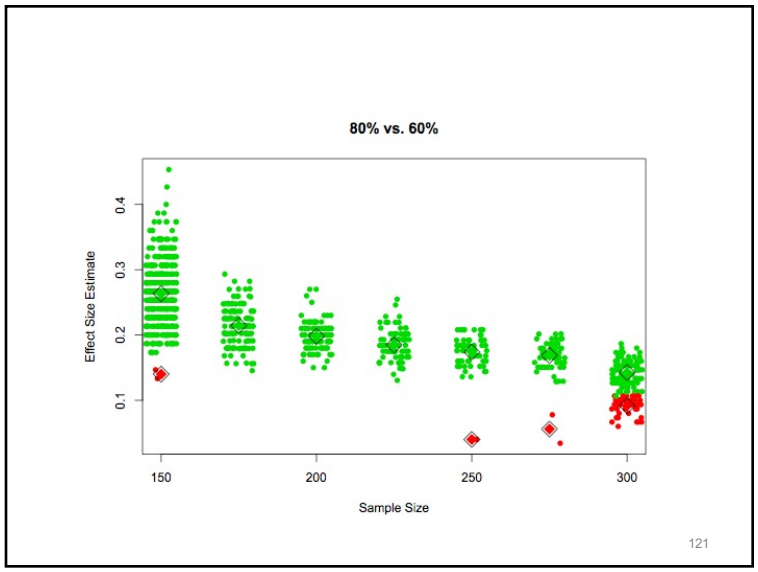
120

120

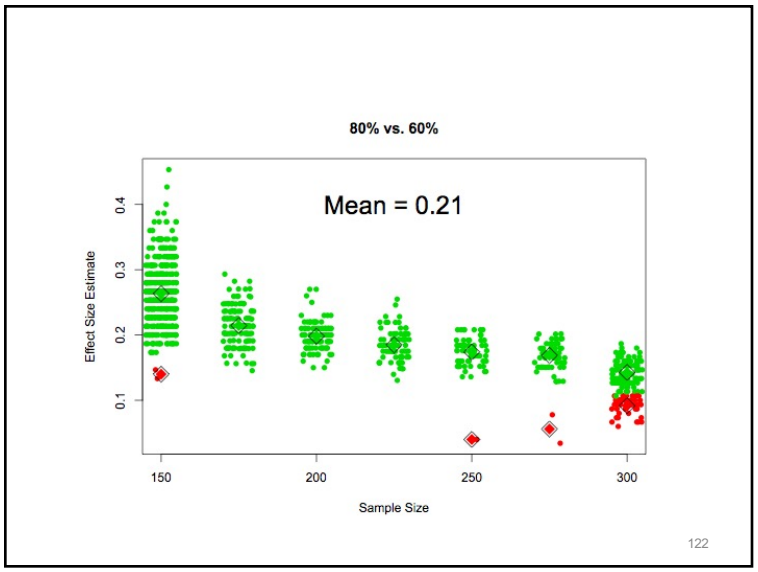
Example 1



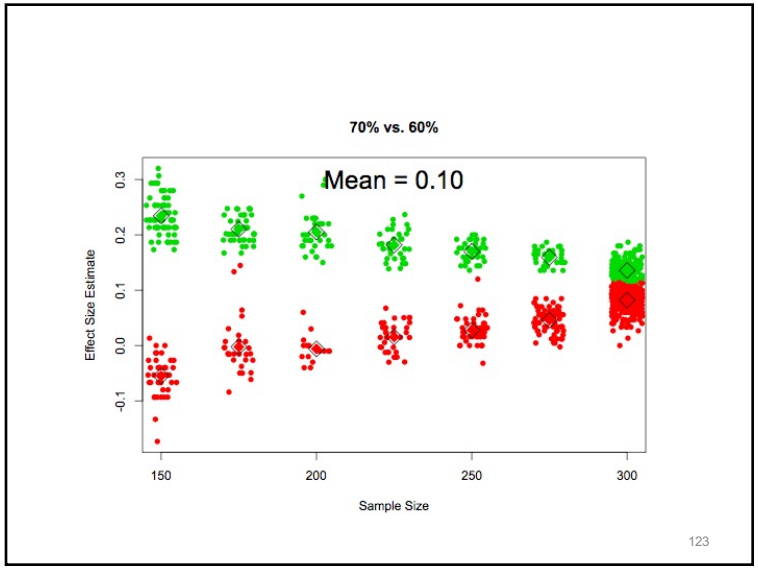
120



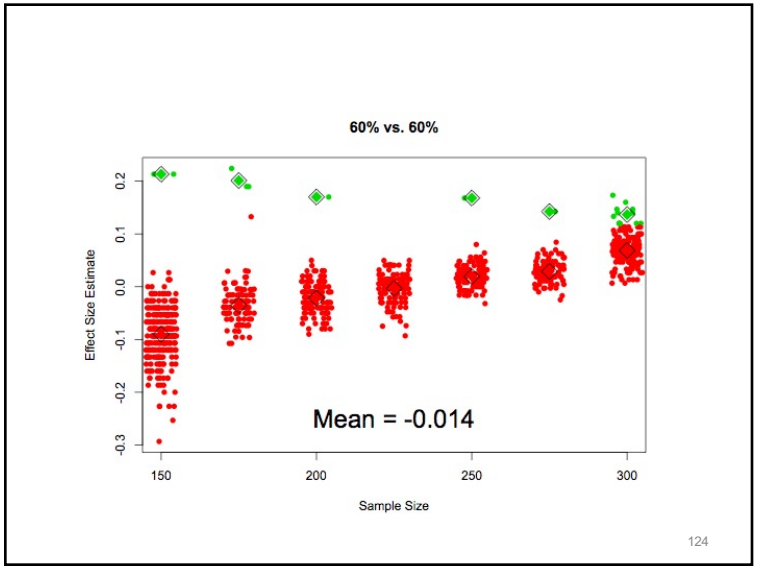
121



122



123



124

Summary

- A process is biased
- Individual trials are not biased
- Individual trials do vary about their true mean
- Larger trials have narrower CIs
- They stopped early because it was a random observation in the right or left tail
- Tradeoff – is it worth deciding earlier and offering benefit to those outside the trial?
- Many adaptive trials are larger so tighter CIs

125

125

Example #2: Goldilocks Trial with 2 Endpoints & Informative Prior on Longitudinal Model

126

126

Background

- Medical device to treat atrial fibrillation (AF)
- Used during open cardiac surgery
  - Only used when surgery being done for other reason
  - e.g., CABG, Valve replacement
- Label was to 'ablate cardiac tissue' not 'treat AF'
- Trial needed to produce evidence of safety and efficacy for treatment of AF
- Controlled trial not possible due to extensive use

127

127

Background

- Early safety study with matched controls failed to enroll
  - Matched control having same cardiac surgery without AF treatment component
  - Stopped @ 32 months when 39 cases & just 5 controls enrolled
- FDA suggested to company to explore Bayesian adaptive trial with safety & efficacy OPCs

128

128

Objective Performance Criteria

- Efficacy OPC (6m)
  - AF free & off AF drugs at 6 months
  - Goal: 70%, $\delta_E = 10\%$
  - Based upon published rates of this procedure
    - 10 papers had 60.1% efficacy
- Safety OPC (1m)
  - Free of significant adverse event
  - Goal: 13.95%, $\delta_S = 5\%$
  - Based upon published SAE rates in Cut & Sew MAZE

129

129

Statistical Endpoints

- Show $\Pr(p_E > 0.60) > 0.975$
  - $70\% - \delta_E = 70\% - 10\% = 60\%$
- Show $\Pr(p_S < 0.1895) > 0.95$
  - $13.95\% + \delta_S = 13.95\% + 5\% = 18.95\%$
- Achievable in 100 patients if
  - observed efficacy ≥ 70%
  - observed safety ≤ 12%
  - basically point estimates have to match or beat OPCs
- $p_E, p_S \sim$ Beta(1,1) priors for both endpoints

130

130

Goldilocks Design

- Enroll 50 - 100 patients
  - Must have 20 patients at 6 months or skip analysis
- Interim analyses every 5 patients
- Final sample size based upon predictive probabilities
- Expect to enroll 5 patients per month
  - ~30 patients enrolled without complete 6m data

131

131

Stopping Decisions

- P_n = Pr(Meet Efficacy & Safety Goals with current sample size n | Current Data)
  - If $P_n \geq S_n$ then stop accrual for predicted success
  - $S_n = 0.90$ for n=50-65
  - $S_n = 0.85$ for n=70-80
  - $S_n = 0.80$ for n=85-95
- P_{max} = Pr(Meet Efficacy @ Safety Goals with 100 patients | Current Data)
  - If $P_n \leq F_n$ then stop trial for futility
  - $F_n = 0.05$ for n=50-70
  - $F_n = 0.10$ for n=75-95

132

132

Longitudinal Model

- Efficacy outcome is AF-free and off AADs at 6m
- Interim outcome at 3-months is whether patients are AF-free already
- Predict 6m outcomes using Beta-Binomial

Group	α	β	Prior Mean
No 3m data	5	1	83%
In AF	4.2	1.8	70%
AF-free	5.4	0.6	90%

133

133

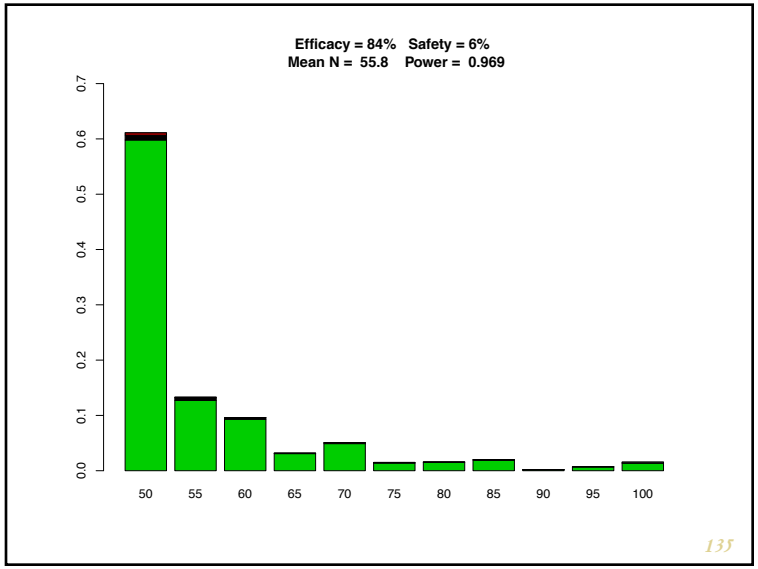
Operating Characteristics for Trial with $p_T = 0.84, p_S = 0.08$

Sample Size	Proportion Of Trials	Stop for Futility	Stop Early For Success	& Lose	& Win
50	0.440	0.008	0.432	0.011	0.421
55	0.150	0.003	0.147	0.007	0.140
60	0.109	0.006	0.102	0.005	0.097
65	0.033	0.004	0.029	0.002	0.027
70	0.063	0.002	0.061	0.002	0.058
75	0.034	0.006	0.027	0.002	0.025
80	0.031	0.011	0.020	0.000	0.020
85	0.042	0.002	0.040	0.000	0.040
90	0.009	0.006	0.003	0.000	0.003
95	0.019	0.003	0.016	0.000	0.016
100	0.070	---	0.070	0.011	0.058
Total	1.000	0.053	0.947	0.042	0.906

Mean Sample Size = 61.6, SD = 15.6

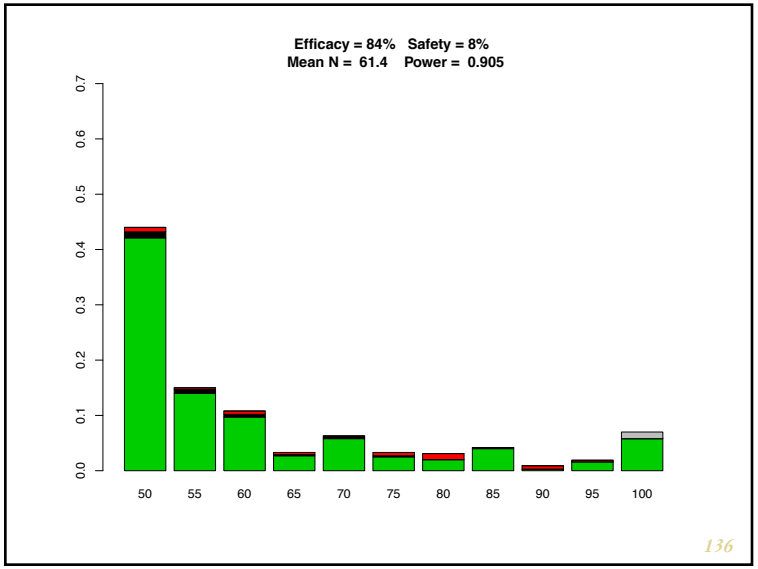
134

134



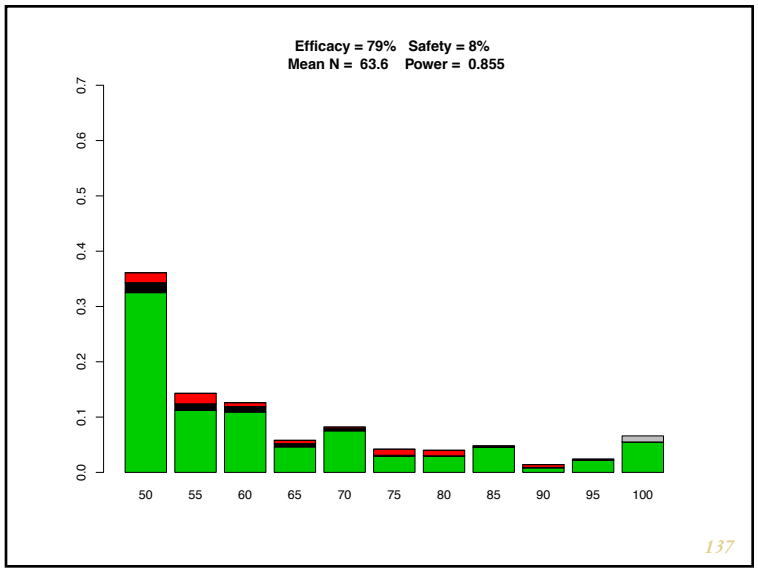
135

135

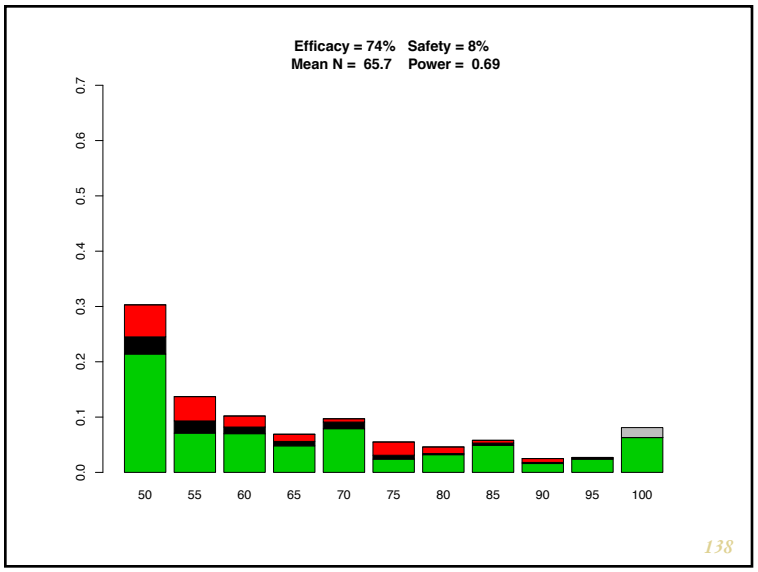


136

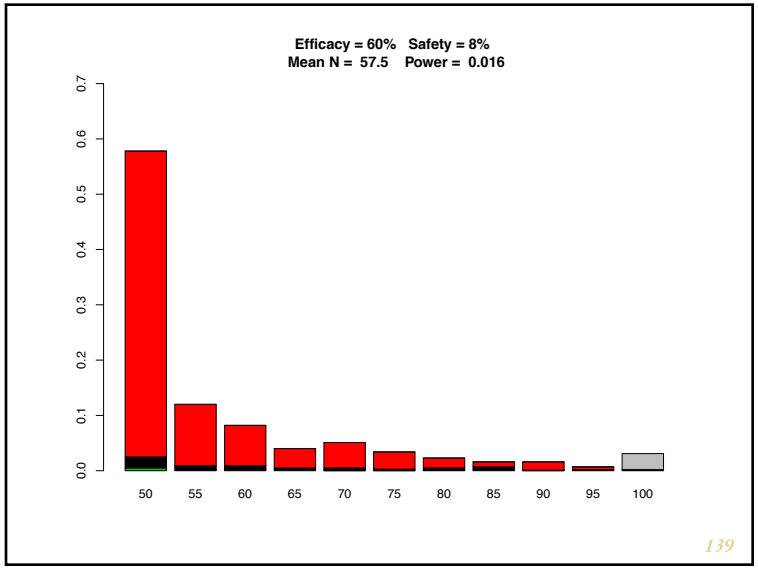
136



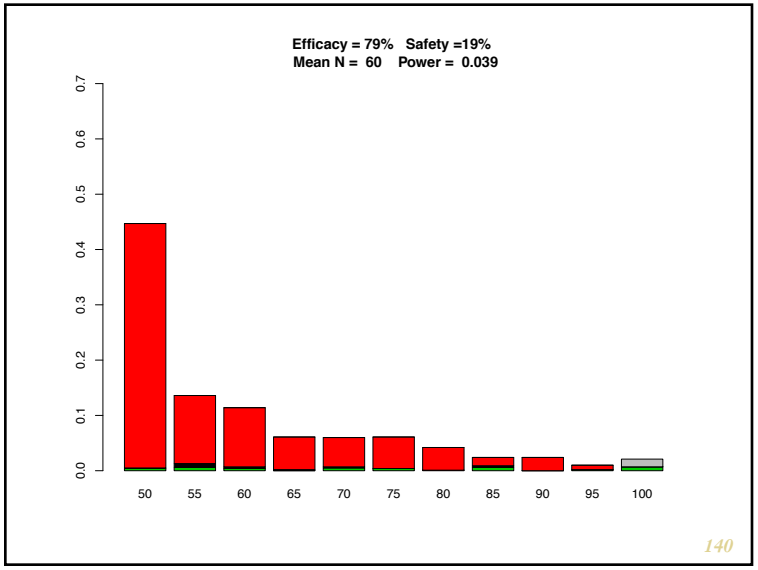
137



138



139



140

Interim Analysis

- No look at 50 patients
- At 55-patients August 24, 2009
  - All patients through 30-day safety, 5/55 had SAEs
  - 24/29 efficacy successes at 6-months
  - 21 subjects remain under surveillance
  - 37/50 successes would show
    $\Pr(p_t > 0.60 \mid 37 \text{ of } 50) = 0.978 > 0.975$
  - Total number of efficacy successes
    $X = 24 + x_o + x_+ + x_-$

141

141

Interim Analysis

- x_o = 5 enrolled with < 3mo follow-up
  - $x_o \sim$ Beta-Bin(n_0 = 5, α=5+24, β=1+5)
- x- = 3 enrolled not AF-free at 3mo
  - $x_- \sim$ Beta-Bin(n_- = 3, α=4.2+3, β=1.8+1)
- x+ = 13 enrolled AF-free at 3mo
  - x+ ~ Beta-Bin(n+ = 13, □ =5.4+17, □ =0.6+3)
- $\Pr(24+x_o+x_-+x_+ \geq 37) = 0.988$

142

142

Interim Analysis

- x_o = 5 enrolled with < 3mo follow-up
  - $x_o \sim$ Beta-Bin(n_0 = 5, α=5+24, β=1+5)
- x- = 3 enrolled not AF-free at 3mo
  - $x_- \sim$ Beta-Bin(n_- = 3, α=4.2+3, β=1.8+1)
- x+ = 13 enrolled AF-free at 3mo
  - x+ ~ Beta-Bin(n+ = 13, □ =5.4+17, □ =0.6+3)
- $\Pr(24+x_o+x_-+x_+ \geq 37) = 0.988$

Longitudinal Priors were right on

5/6 = .83
24/29 = .83

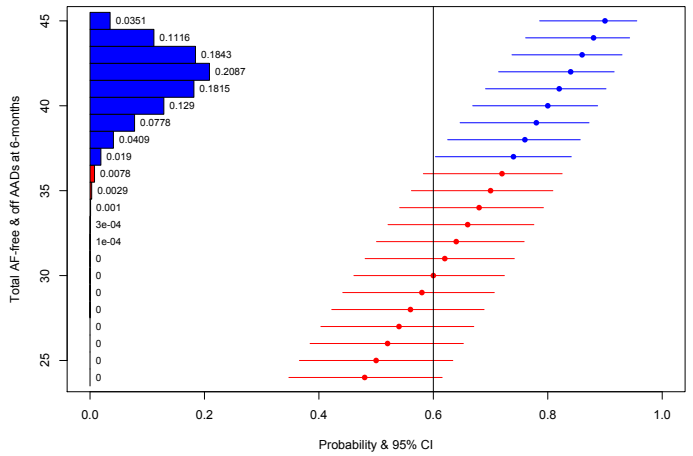
4.2/6 = .70
3/4 = .75

5.4/6 = .90
17/20 = .85

143

143

Prediction of 21 remaining pts based on 29 observed pts



144

144

Sample Size Analysis at 55 pts

```
Current Patients Enrolled: 55
Current patients not contributing to efficacy: 5
Current Safety Events: 5 of 55 patients
Current Efficacy Success: 24 of 29 patients
Current Efficacy Successes: 24 of 29 patients
Current Efficacy Successes: 3 of 4 Efficacy Failures at 3 months
Current Efficacy Successes: 17 of 20 Efficacy Successes at 3 months

0 enrolled patients to predict for 1mo safety outcomes
45 future patients to predict for 1mo safety outcomes
5 enrolled patients with <3mo to predict for efficacy outcomes
3 enrolled patients with AF at 3mo to predict for 6mo efficacy outcomes
13 enrolled patients without AF at 3mo to predict for 6mo efficacy outcomes
45 future patients to predict for 6mo efficacy outcomes

Predicted Safety Events with Current Accrual: 5 ( 5 - 5 ) of 55 patients
       5  or fewer needed for safety success
Predicted Safety Events with Maximum Accrual: 9.7 ( 6 - 16 ) of 100 patients
       12  or fewer needed for safety success
Predicted Efficacy Successes with Current Accrual: 41.5 ( 37 - 45 ) of 50 patients
       37  or more needed for efficacy success
Predicted Efficacy Successes with Maximum Accrual: 78.8 ( 69 - 86 ) of 95 patients
       67  or more needed for efficacy success

Decision Rule: Stop Enrolling Due to Predicted Success

      Prob Win Efficacy Prob Win Safety Prob Win Both
Now             0.988           1.000          0.988
Max N           0.992           0.846          0.838
```

.988 > .90
Stop for
predicted success

145

145

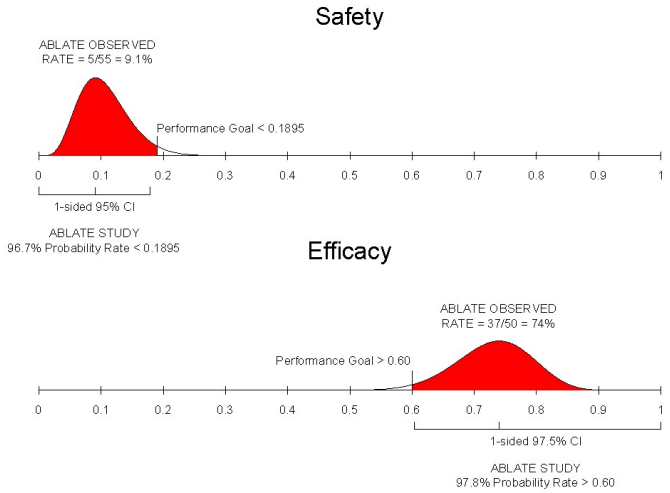
Stopped Accrual for Predicted Success

- Accrual stopped with 55 patients in
- Continue to follow 21 enrolled patients
- Perform final analysis on complete data

- Final Data
  - 5/55 SAEs
  - 37/50 AF-free and off AADs

146

146



147

147

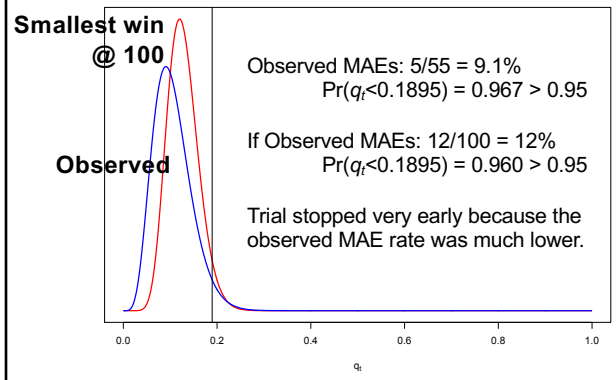
Post Trial Discussion with FDA

- Efficacy barely won
  - One less success would have failed to meet primary endpoint
  - Not a robust win, in part due to post-hoc changes related to inc/excl criteria
- Some concern with n=55
  - But this was possible based upon design
  - Safety OPC = 0.1395, observed 0.091
  - Efficacy OPC = 0.70, observed 0.74

148

148

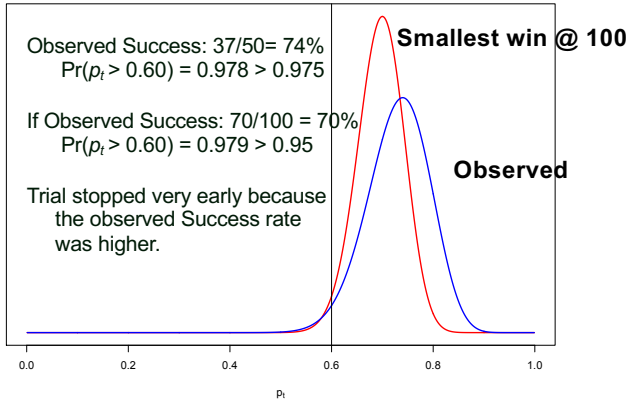
Safety: Compare Stopping at n=55 to Maximum Trial Size n=100



149

149

Efficacy: Compare Stopping at n=50 to Maximum Trial Size n=100



150

150

FDA Advisory Panel Vote Oct 2011

- Is there reasonable assurance that the AtriCure Synergy Ablation System is effective …?
  - 9 for, 0 against
- Is there reasonable assurance that the AtriCure Synergy Ablation System is safe…?
  - 5 for, 4 against, 1 abstain (chair broke 4-4 tie)
  - Largely due to patients needing pacemakers
- Do the benefits … outweigh the risks …?
  - 5 for, 3 against, 1 abstain

151

151

heartwire

ARRHYTHMIA/EP

AtriCure AF ablation system gets cautious thumbs-up from FDA advisors

OCTOBER 26, 2011 Michael O'Riordan

Recommend 0 Tweet 4 +1 0 Share

Comments Read later Print Font size A A A Cite

Rockville, MD (updated) – The **AtriCure Synergy Ablation System** squeaked by today in a meeting of the **Food and Drug Administration**'s Circulatory System Devices advisory panel when panelists gave a cautious nod of approval for the device.

Five panel members believed the benefits of the ablation system outweighed the risks when used in the treatment of atrial-fibrillation (AF) patients undergoing open concomitant coronary artery bypass graft (CABG) surgery and/or valve replacement or repair. Three panelists expressed doubts about the system and cautioned against device approval, voting that they did not believe the benefits outweighed the risks.

One panelist abstained from voting on the benefit/risk trade-off question.

In a vote on efficacy alone, all panelists believed the ablation system is effective in restoring sinus rhythm, but they were split for the vote on safety. Chair of the advisory panel, **Dr John Hirschfeld** (University of Pennsylvania, Philadelphia), cast the deciding vote on safety, saying he believes there is reasonable assurance the device is safe for use in patients who meet the indication criteria. Overall, the panel voted 9 to 0 on efficacy and 5 to 4 on safety (with one abstention).

Panel member **Dr David Slotwiner** (Long Island Jewish Medical Center, New Hyde Park, NY) voted in favor of the ablation system, saying that he believes the benefits outweigh the risks.

"I think it's effective at creating these ablation lesions, and I think it's effective in many people for maintaining sinus rhythm, although what that means [clinically] for many patients remains unanswered," he said. "But I hope, mostly, that if an approval is granted, it will allow us to get more information and to educate more surgeons so that the procedure becomes more widely available and we understand better who will benefit the most."

152

FDA Approved Dec 14, 2011

- **Study Design (from device label)**
- ABLATE was a multi-center, prospective, non-randomized study based on a Bayesian adaptive design that provides high probability of demonstrating safety and effectiveness of the AtriCure Synergy Ablation System for the treatment of permanent atrial fibrillation. The safety and effectiveness of the device was compared to performance goals derived from historical information. The Bayesian adaptive clinical design incorporated interim analyses of the data to determine the point of completion of trial enrollment. Enrollment was targeted to be between 50 and 100 subjects at 20 sites. The study was designed to have an initial assessment of results at the point that 50 subjects were enrolled with a minimum of 20 subjects completing their six-month follow-up visit. Nine investigational sites enrolled 55 subjects.

153

153

Lessons

- Ensure minimum sample size will suffice
  - Not just statistical, but impactful
  - Company did a continue access protocol to get more patients during review, leading to panel
- Ensure data isn't coded optimistically
- Ensure inclusion / exclusion criteria rigorously followed
- Goldilocks gets the size 'just right' but that means you can be close to 'just wrong' if some data changes post hoc

154

154

Example 3: SHINE TRIAL

with Karen Johnston, Valerie Durkalski
Kristine Broglio, & Will Meurer

- Trial for SOC vs. tight glycemic control after stroke
- Designed as Group Sequential
- Run with "Shadow" Bayesian Trial
- Design papers online
  - Compares GSD to Goldilocks Trial

155

155

Platform Trials

& Master Protocols

156

156

Woodcock & Lavange, NEJM 2017

- High-quality evidence is what we use to guide medical practice. The standard approach to generating this evidence — a series of clinical trials, each investigating one or two interventions in a single disease — has become ever more expensive and challenging to execute. As a result, important clinical questions go unanswered.
- A methodologic innovation responsive to this need involves coordinated efforts to evaluate more than one or two treatments in more than one patient type or disease within the same overall trial structure. Such efforts are referred to as master protocols, defined as one overarching protocol designed to answer multiple questions.

157

157

Woodcock & Lavange, NEJM 2017

- Master protocols may involve one or more interventions in multiple diseases or a single disease, as defined by current disease classification, with multiple interventions, each targeting a particular biomarker-defined population or disease subtype. Included under this broad definition of a master protocol are three distinct entities: umbrella, basket, and platform trials (Table 1 and Figs. 1 and 2). All constitute a collection of trials or substudies that share key design components and operational aspects to achieve better coordination than can be achieved in single trials designed and conducted independently.

158

158

Master Protocols

- May or may not compare treatment across groups
  - One structure, but each TX vs. common control
  - Reported as multiple trials (e.g. 1 per intervention)
  - Sites have one set of rules, execute like 1 trial
- Intensive pretrial discussion among sponsors
  - data use, publication rights, and the timing of regulatory submission
- Matchmaker
  - Therapies to targeted subpopulations

159

159

Master Protocols

Table 1. Types of Master Protocols.

Type of Trial	Objective
Umbrella	To study multiple targeted therapies in the context of a single disease
Basket	To study a single targeted therapy in the context of multiple diseases or disease subtypes
Platform	To study multiple targeted therapies in the context of a single disease in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a decision algorithm

NEJM 377, 1, p63, Table 1

160

160

Master Protocols

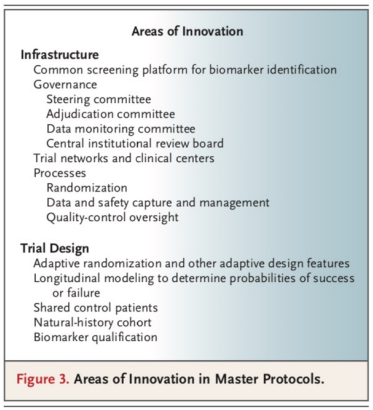


Figure 3. Areas of Innovation in Master Protocols.

- Master protocols come in different sizes and shapes but share many commonalities.
- Increased planning efforts and coordination to satisfy the objectives of different stakeholders.
- Maximum information is obtained from the research effort
- Infrastructure required for implementation increases data quality and trial efficiencies, as compared with those in stand-alone trials.
- Can last many years, even decades, with innovations from the laboratory translating quickly to clinical evaluation.

NEJM 377, 1, p63, Figure 3

161

161

Platform Trials

162

162

Asking the Right Question

- Current Clinical Trials

  Is this drug effective and safe?

  More precisely

  What is the probability of the observed data assuming the treatment is no good?

163

163

Asking the Right Question

- Current Clinical Trials
  - Is this drug effective and safe compared to a placebo?
  - Is this drug effective & safe compared to the SOC
- Correction Question
  - What is the best treatment for this Patient?
  - What is the best treatment for this type of patient?

164

164

Traditional Trial Design

- Single treatment vs. Control
- Homogenous patient population
- 1 or 2 questions per 1 trial
- Start with assuming a particular control group effect and a particular (usually optimistic) treatment group effect
- Assume 'average' effect relevant to all patients
- Calculate a sample size as if we know the true effect

165

Platform Trial

- An experimental infrastructure to evaluate multiple treatments, often for a group of diseases, and intended to function continually and be productive beyond the evaluation of any individual treatment
  - Designed around a group of related diseases rather than a single treatment
  - Disease focused not treatment focused
  - Dynamic list of available treatments, assigned with response-adaptive randomization
  - Preferred treatments may depend on health system, patient, or disease-level characteristics

166

VIEWPOINT

The Platform Trial
An Efficient Strategy for Evaluating Multiple Treatments

Scott M. Berry, PhD
Berry Consultants LLC, Austin, Texas; and Department of Biostatistics, University of Kansas Medical Center, Kansas City.

Jason T. Connor, PhD
Berry Consultants LLC, Austin, Texas; and University of Central Florida College of Medicine, Orlando.

Roger J. Lewis, MD, PhD
Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California; and Berry Consultants LLC, Austin, Texas.

The drug development enterprise is struggling. The development of new therapies is limited by high costs, slow progress, and a high failure rate, even in the late stages of development. Clinical trials are most commonly based on a "one population, one drug, one disease" strategy, in which the clinical trial infrastructure is created to test a single treatment in a homogeneous population.

This approach has been largely unsuccessful for multiple diseases, including sepsis, dementia, and stroke. Despite promising preclinical and early human trials, there have been numerous negative phase 3 trials of treatments for Alzheimer disease[1] and more than 40 negative phase 3 trials of neuroprotectants for stroke.[2] Effective treatments for such diseases will likely require combining treatments to affect multiple targets in complex cellular pathways and, perhaps, tailoring treatments to subgroups defined by genetic, proteomic, metabolomic, or other markers.[3]

There has been increasing interest in efficient trial strategies designed to evaluate multiple treatments and combinations of treatments, in heterogeneous patient benefits when evaluating potentially synergistic combination treatments (eg, treatment A, treatment B, treatment C, and all combinations) if the starting point is the testing of each treatment in isolation.

What Is a Platform Trial?

A platform trial is defined by the broad goal of finding the best treatment for a disease by simultaneously investigating multiple treatments, using specialized statistical tools for allocating patients and analyzing results. The focus is on the disease rather than any particular experimental therapy. A platform trial is often intended to continue beyond the evaluation of the initial treatments and to investigate treatment combinations, to quantify differences in treatment effects in subgroups, and to treat patients as effectively as possible within the trial. Although some of the statistical tools used in platform trials are frequently used in other settings and some less so, it is the integrated application of multiple tools that allows a platform trial to address its multiple goals. The **Table** summarizes the general differences between a traditional clinical trial and a platform trial

JAMA. Published online March 23, 2015. doi:10.1001/jama.2015.2316

167

From: **The Platform Trial: An Efficient Strategy for Evaluating Multiple Treatments**

JAMA. Published online March 23, 2015. doi:10.1001/jama.2015.2316

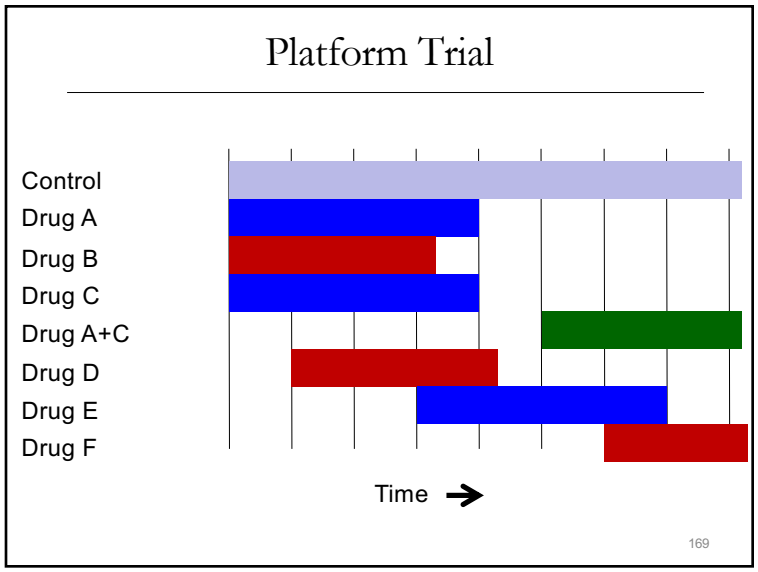
Table. General Characteristics of Traditional and Platform Trials[a]

Characteristic	Traditional Trial	Platform Trial
Scope	Efficacy of a single agent in a homogeneous population	Evaluating efficacy of multiple agents in a heterogeneous population; explicitly assumes treatment effects may be heterogeneous
Duration	Finite, based on time required to answer the single primary question	Potentially long-term, as long as there are suitable treatments requiring evaluation
No. of treatment groups	Prespecified and generally limited	Multiple treatment groups; the number of treatment groups and the specific treatments may change over time
Stopping rules	The entire trial may be stopped early for success or futility or harm, based on the apparent efficacy of the single experimental treatment	Individual treatment groups may be removed from the trial, based on demonstrated efficacy or futility or harm, but the trial continues, perhaps with the addition of new experimental treatment(s)
Allocation strategy	Fixed randomization	Response-adaptive randomization
Sponsor support	Supported by a single federal or industrial sponsor	The trial infrastructure may be supported by multiple federal or industrial sponsors or a combination

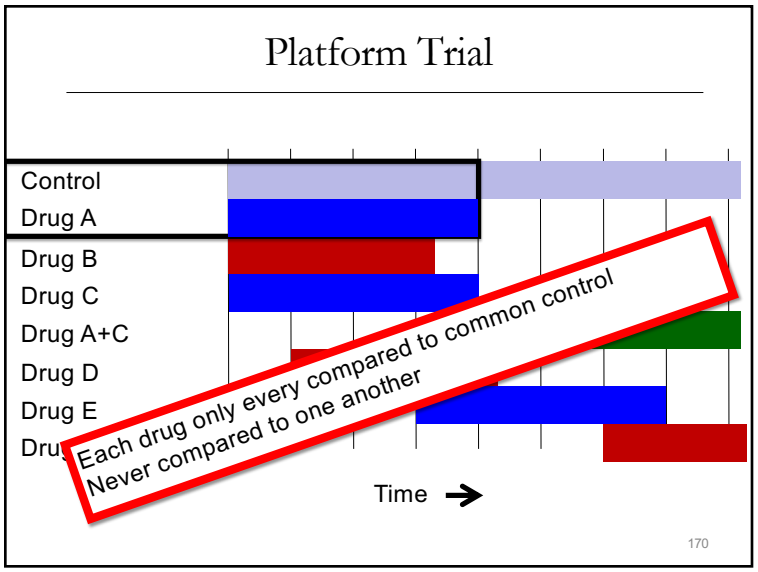
[a] Platform trials and similar trials may also be called basket, bucket, umbrella, or standing trials.

Table Title:
General Characteristics of Traditional and Platform Trials[a]

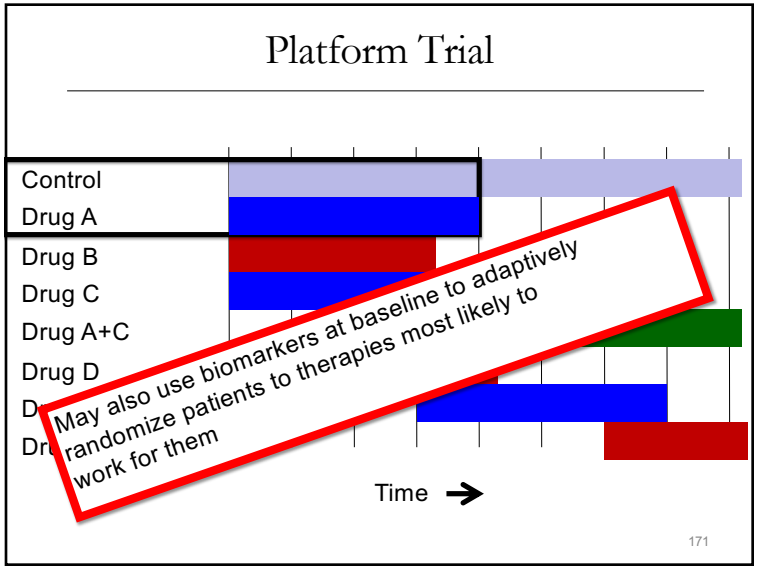
168



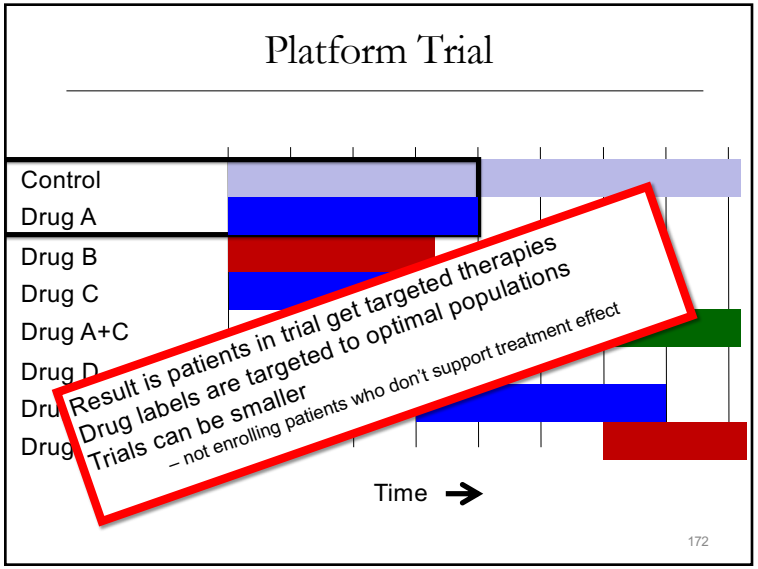
169



170

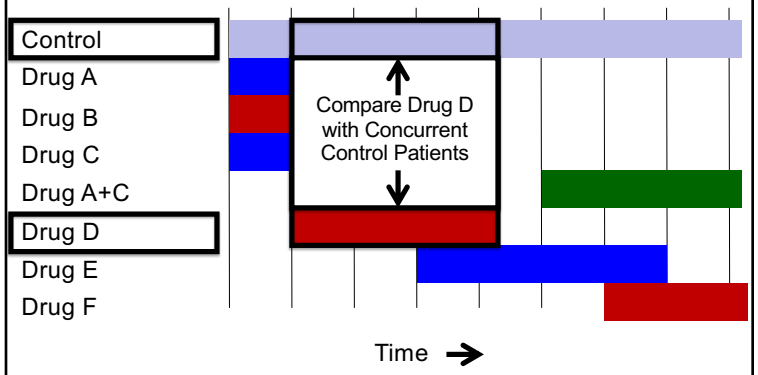


171



172

Platform Trial

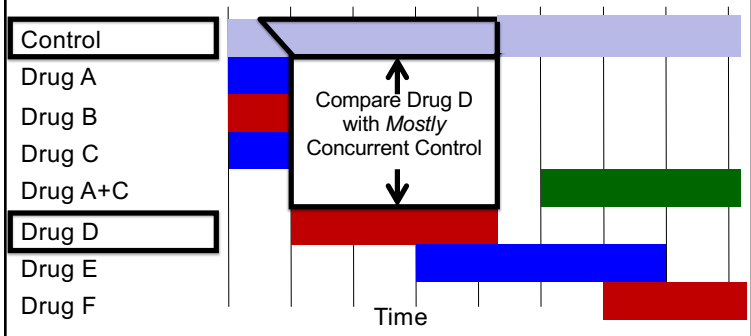


173

173

Time Machine

Model how controls change over time, if similar, then use some controls outside concurrent window

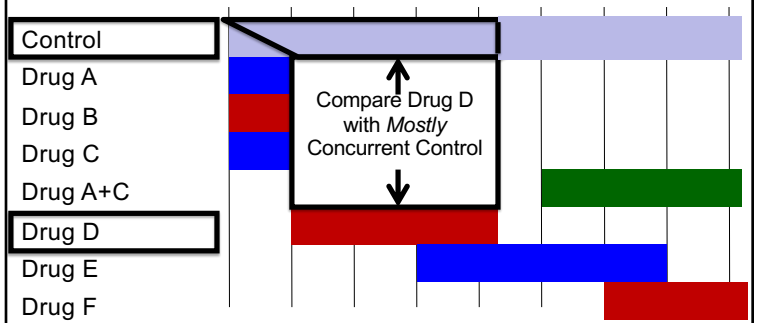


174

174

Time Machine

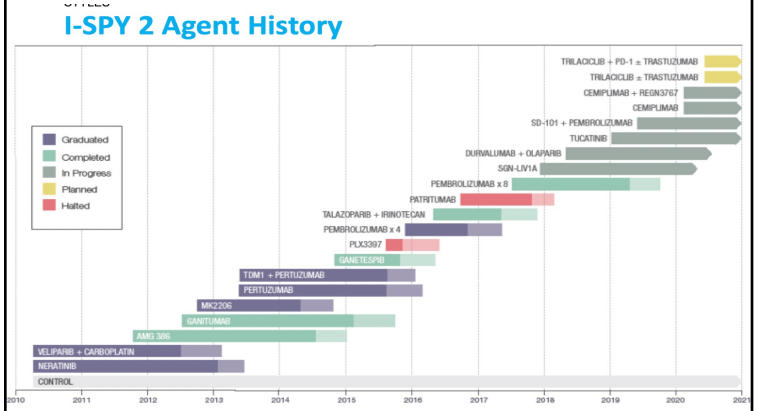
If controls change little over time, then use more weight from non-concurrent controls, increases power & efficiency



175

175

I-SPY2



Used with permission from Dr. Laura Esserman, I-SPY2 PI

176

176

Platform Trials are Happening

- Cancer
  - I-SPY2 in Breast Cancer
  - GBM AGILE in Glioblastoma multiforme
  - LUNG-MAP in Lung Cancer
  - PANCAN in Pancreatic Cancer
- Alzheimer's
  - EPAD: European Prevention of Alzheimer's Dementia
  - DIAN: Dominantly Inherited Alzheimer's Network
- ALS
  - Healey ALS Platform Trial, Phase 2/3 with 5 drugs

177

177

Platform Trials are Happening

- Infection diseases
  - Gates Foundation sponsored Ebola design
  - NIH Ebola design
  - PREPARE: European Consortium for Disease Preparedness
    - Pandemic flu, Butler at al Lancet, Jan 2020
    - REMAP CAP (Community Acquired Pneumonia) ongoing, REMAPCAP.org
- COVID-19
  - RECOVERY
  - ACTT by NIAID -- the Remdesivir trial
  - SOLIDARITY by WHO, 4 arms
  - REMAP-COVID by International consortium critical care trial
  - PRINCIPLE in UK, pre-hospital trial
  - ISPY-COVID: UCSF & WISDOM Network, Phase 2
  - ACTIV by NIH

178

178

Characteristics of Modern Platform Trials	Cancer Trials							
	I-SPY 2	MICAT	BATTLE	LUNG-MAP	UK MATRIX	GBM-AGILE	Alzheimer's	Ebola
Screen markers for all pts	✔	✔	✔	✔	✔	✔	—	—
Master protocol	✔	✔	✔	✔	✔	✔	✔	✔
Many regimens	✔	✔	✔	✔	✔	✔	✔	✔
Combination therapies	✔	✔	✔			✔	✔	✔
Sequential therapies		✔				✔		
Assembly line	✔	✔			✔	✔	✔	✔
Learn off-target effects	✔	✔	✔			✔	—	—
Pair regimens/biomarkers	✔	✔	✔			✔	—	—
Common control arm	✔	✔				✔	✔	✔
Adaptive randomization	✔	✔	✔			✔	✔	✔
Adaptive sample size	✔	✔				✔	✔	✔
Early "curable" disease	✔					✔	✔	✔
Registration endpoint	✔					✔	✔	✔
Seamless phases						✔		
Longitudinal modeling	✔	✔				✔	✔	
Bayesian	✔	✔	✔			✔	✔	✔

From Don Berry

179

179

Article

CLINICAL TRIALS

Clinical Trials
1–9
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1740774515626362
ctj.sagepub.com
SAGE

Efficiencies of platform clinical trials: A vision of the future

Benjamin R Saville[1,2] and Scott M Berry[1,3]

Abstract

Background: A "platform trial" is a clinical trial with a single master protocol in which multiple treatments are evaluated simultaneously. Adaptive platform designs offer flexible features such as dropping treatments for futility, declaring one or more treatments superior, or adding new treatments to be tested during the course of a trial.

Methods: A simulation study explores the efficiencies of various platform trial designs relative to a traditional two-arm strategy.

Results: Platform trials can find beneficial treatments with fewer patients, fewer patient failures, less time, and with greater probability of success than a traditional two-arm strategy.

Conclusion: In an era of personalized medicine, platform trials provide the innovation needed to efficiently evaluate modern treatments.

Keywords

Platform trial, master protocol, multi-arm, adaptive, Bayesian, clinical trial design

180

Design

CLINICAL TRIALS

Clinical Trials
1–9

DOI: 10.1177/1740774515621721
ctj.sagepub.com
SAGE

A response adaptive randomization platform trial for efficient evaluation of Ebola virus treatments: A model for pandemic response

Scott M Berry[1,2], Elizabeth A Petzold[3], Peter Dull[4], Nathan M Thielman[5], Coleen K Cunningham[6], G Ralph Corey[5], Micah T McClain[6], David L Hoover[7], James Russell[8], J McLeod Griffiss[7] and Christopher W Woods[3,5,6]

Abstract

The outbreak of Ebola virus disease in West Africa is the largest ever recorded. Numerous treatment alternatives for Ebola have been considered, including widely available repurposed drugs, but initiation of enrollment into clinical trials has been limited. The proposed trial is an adaptive platform design. Multiple agents and combinations will be investigated simultaneously. Additionally, new agents may enter the trial as they become available, and failing agents may be removed. In order to accommodate the many possible agents and combinations, a critical feature of this design is the use of response adaptive randomization to assign treatment regimens. As the trial progresses, the randomization ratio evolves to favor the arms that are performing better, making the design also suitable for all-cause pandemic preparedness planning. The study was approved by US and Sierra Leone ethics committees, and reviewed by the US Food and Drug Administration. Additionally, data management, drug supply lines, and local sites were prepared. However, in response to the declining epidemic seen in February 2015, the trial was not initiated. Sierra Leone remains ready to rapidly activate the protocol as an emergency response trial in the event of a resurgence of Ebola. (ClinicalTrials.gov Identifier: NCT02380625.) In summary, we have designed a single controlled trial capable of efficiently identifying highly effective or failing regimens among a rapidly evolving list of proposed therapeutic alternatives for Ebola virus disease and to treat the patients within the trial effectively based on accruing data. Provision of these regimens, if found safe and effective, would have a major impact on future epidemics by providing effective treatment options.

181

EBOLA

Thanks to: Scott Berry,
Elizabeth Petzold,
Chris Woods, David Hoover

Berry Consultants
Statistical Innovation

Duke Clinical Research Institute

ClinicalRM
20th Anniversary 1994-2014

182

182

The Problem: Ebola Treatment Trial

- Acknowledge universe of possible treatments
  - Will evolve over time
  - Recognition that combinations may play an important role
- Uncertainty over role of standard of care
- Our Goal: To determine best treatment for treating ebola
  - Not a trial to determine if a single drug X works

183

183

EV-003 Adaptive Platform Design

- Reviewed and approved by:
  - Duke University IRB
  - University of Sierra Leone ethics committee
- *Master Protocol* dictates trial behavior, each treatment included as an appendix
- Multiple Agents
  - Primary & Secondary agents
  - Combination + Single agents
- Response Adaptive Randomization (RAR)
  - Run by a single algorithm
  - Assigns treatment regimens that are performing better using collection of primary endpoint data
- Protocol is built so trial arms evolve (part of the protocol!), trial is **perpetual**
- Endpoint is 14-day mortality

184

184

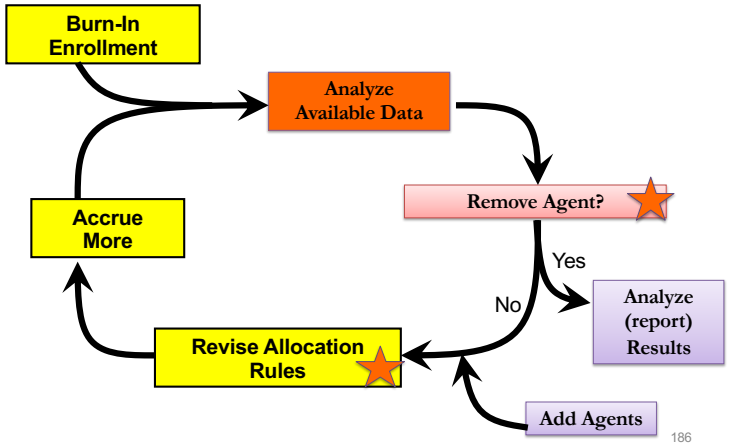
Primary/Secondary Agents

- All arms receive optimized standard of care (SOC)
- Primary and Secondary agents
  - Primary: Expected capability to work as single agent (e.g. anti-viral efficacy)
  - Secondary: Expected to work with other agents (not given alone)

Regimens		Treatments					
		P1	P2	P3	P4	S1	S2
Treatments	P1						
	P2						
	P3						
	P4						

185

Adaptive Platform Design



186

Design Details

Burn-In Enrollment

- Endpoint: Death (Dichotomous, events are bad)
- Start with burn-in period to all 10 regimens
  - Equal randomization to 10 arms
  - 30 subjects / 3 per arm
- After burn-in
  - Response adaptive randomization
  - Proportional to probability regimen is optimal
    - Adjusted for information
  - Continue perpetually (committee can change vote)

187

187

Decision Criteria (In/Out)

Analyze (report) Results

- If there is a less than 0.01 probability an agent is part of the optimal regimen
  - Candidate for futility
- If the probability an agent is in the optimal regimen is greater than 0.95
  - Report to the steering committee for public dissemination
- If a regimen has at least a 0.95 probability of being superior to SOC Alone then SOC Alone is reported for removal

188

188

Allocation Rules

Revise Allocation Rules

- If a SOC it gets minimum of 20%...
- Randomize to regimens with probability proportional to:

$$r_{ij} \sim \frac{\Pr\left(\pi_{ij} = \max\left(\pi\right)\right)}{n_{ij}+1}$$

189

189

Statistical Model

Analyze Available Data

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_{X=1}^{M}[X] + \sum_{X=1}^{M}\sum_{Y=X+1}^{M}[X,Y] + \lambda_{TIME}$$

- Priors:

$$[X] \sim N\left(0,1^2\right) \qquad [X,Y] \sim N\left(0,0.2^2\right)$$

190

190

Statistical Model

Analyze Available Data

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_{X=1}^{M}[X] + \sum_{X=1}^{M}\sum_{Y=X+1}^{M}[X,Y] + \lambda_{TIME}$$

- Priors:

$$[X] \sim N\left(0,1^2\right) \qquad [X,Y] \sim N\left(0,0.2^2\right)$$

N(0,1) has 95% CI from about 1/7 to 7.

191

191

Statistical Model

Analyze Available Data

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_{X=1}^{M}[X] + \sum_{X=1}^{M}\sum_{Y=X+1}^{M}[X,Y] + \lambda_{TIME}$$

- Priors:

$$[X] \sim N\left(0,1^2\right) \qquad [X,Y] \sim N\left(0,0.2^2\right)$$

N(0,0.2²) has 95% CI from about 2/3 to 3/2.

192

192

Statistical Model

Analyze Available Data

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_{X=1}^{M}[X] + \sum_{X=1}^{M}\sum_{Y=X+1}^{M}[X,Y] + \lambda_{TIME}$$

- Priors:

$$[X] \sim N(0,1^2) \qquad [X,Y] \sim N(0,0.2^2)$$

- Time:
  - Incorporate time "buckets" to model time trend or 'drift'

$$[\lambda] \sim NDLM(0,\tau^2)$$

193

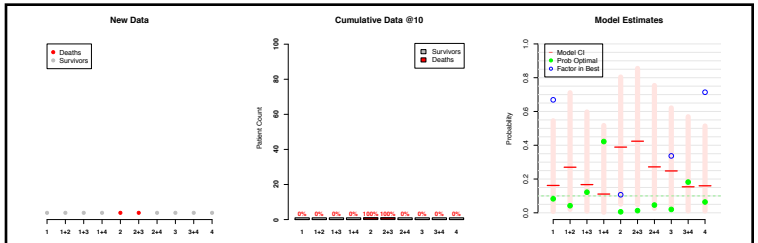
193

Example Trial

Regimens		Agents			
		1	2	3	4
Agents	1				
	2				
	3				
	4				

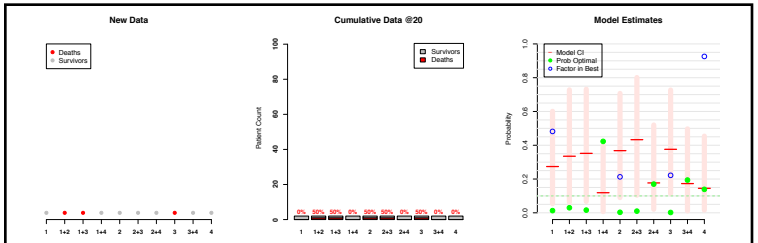
194

194



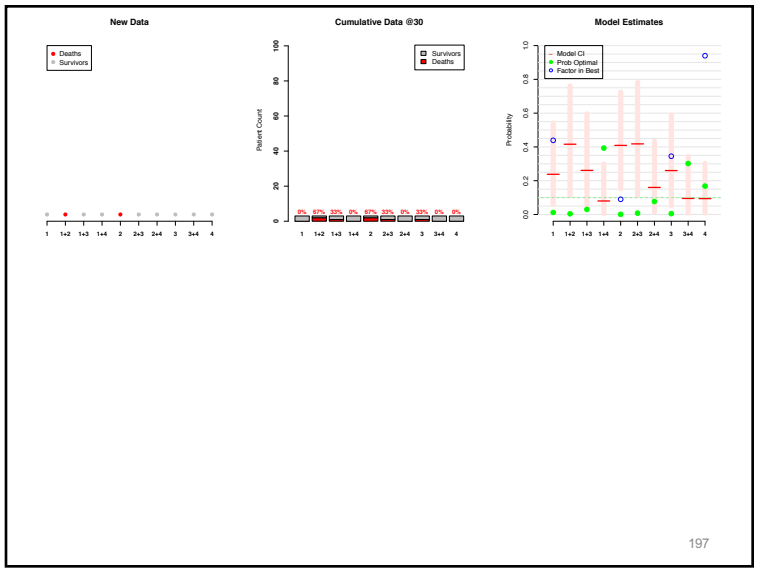
195

195

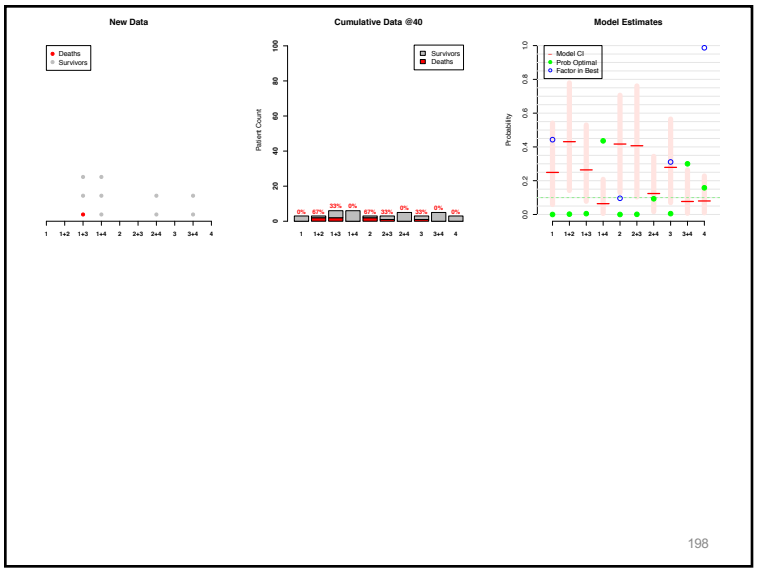


196

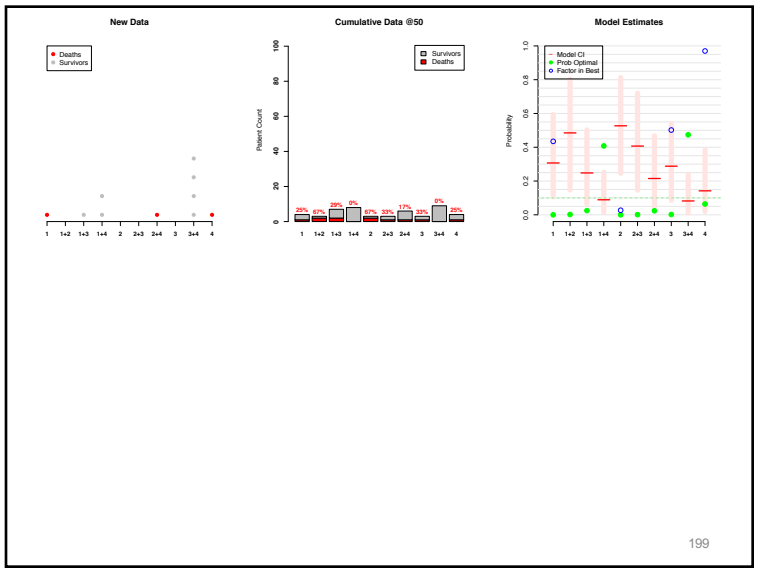
196



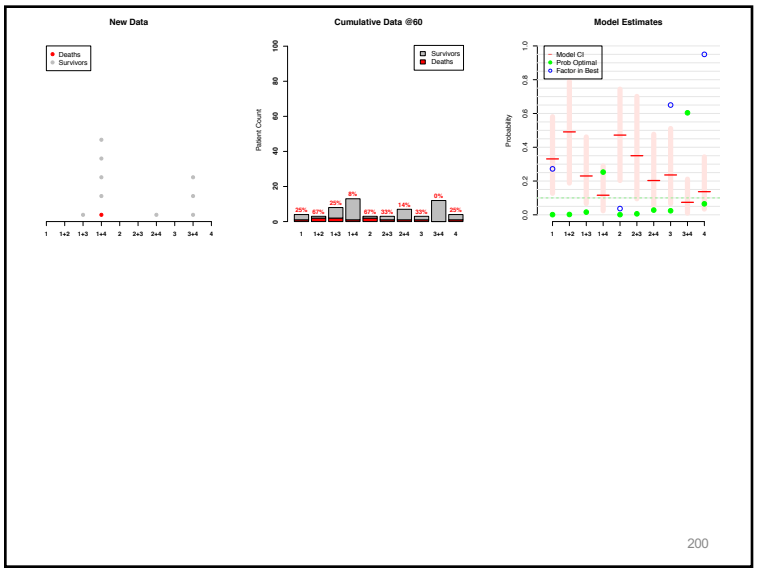
197



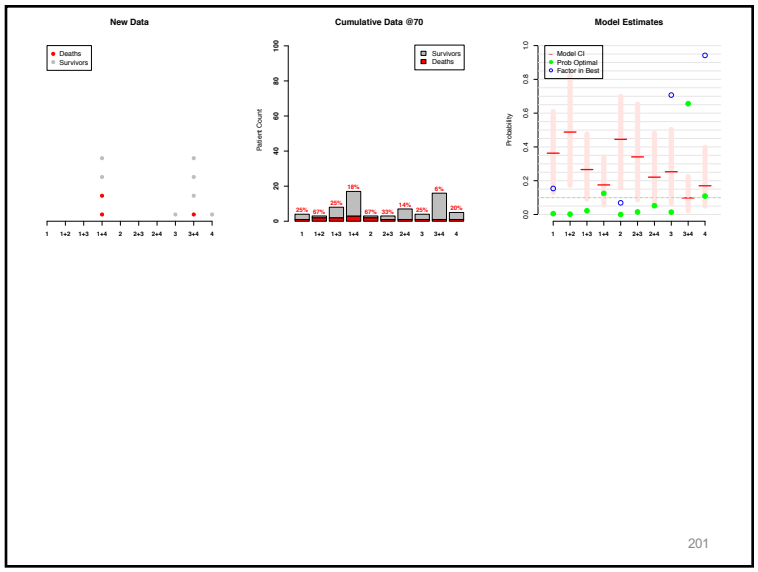
198



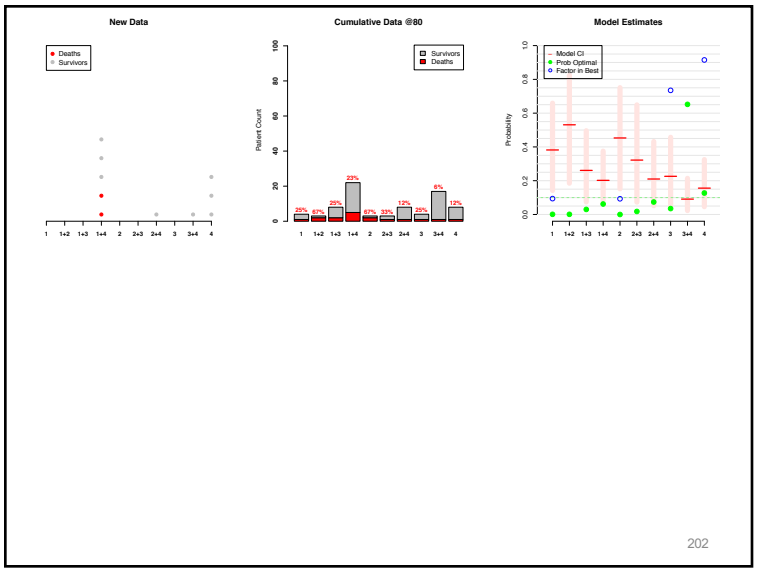
199



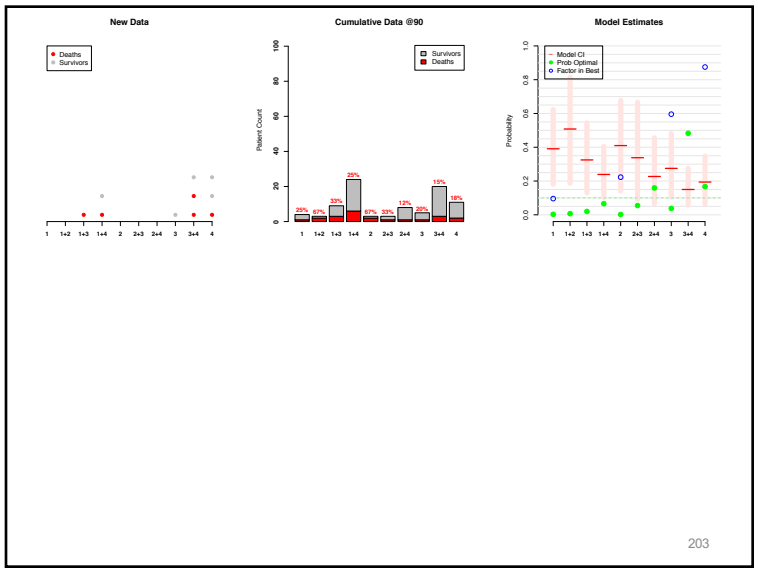
200



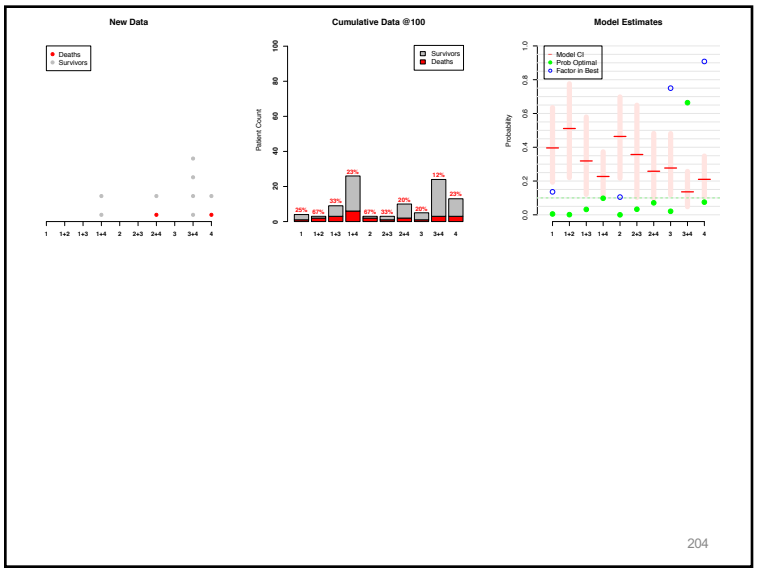
201



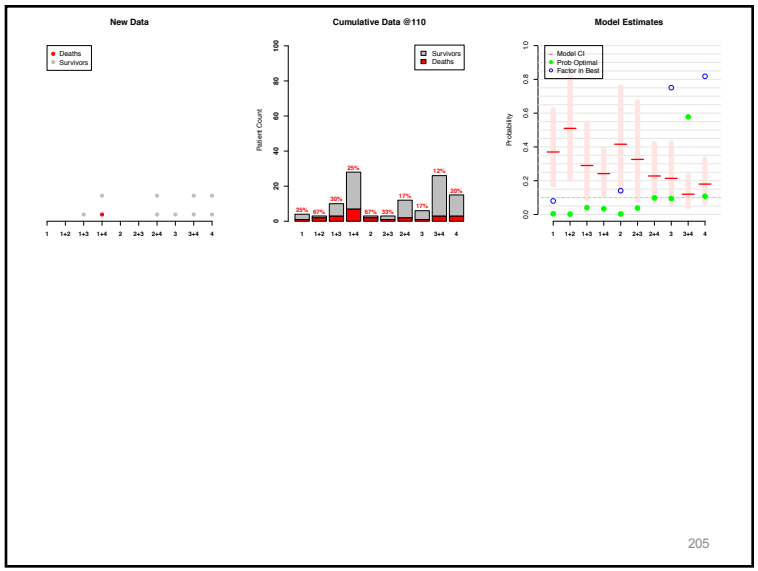
202



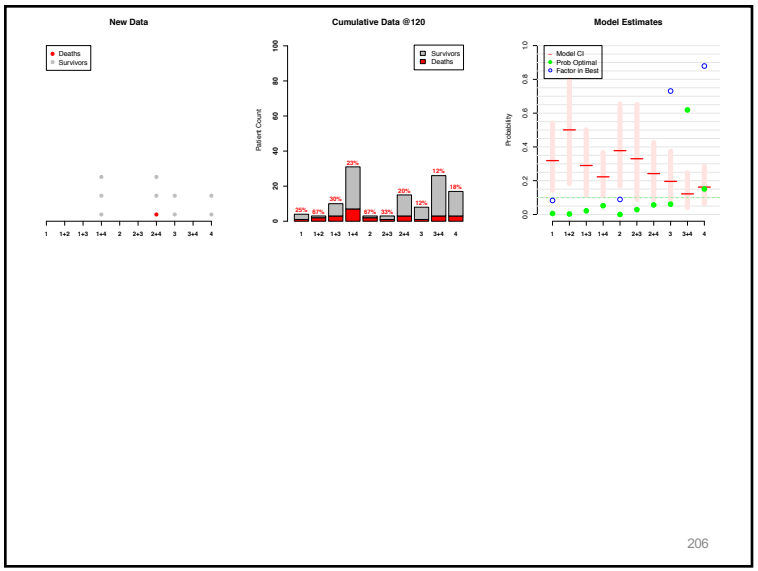
203



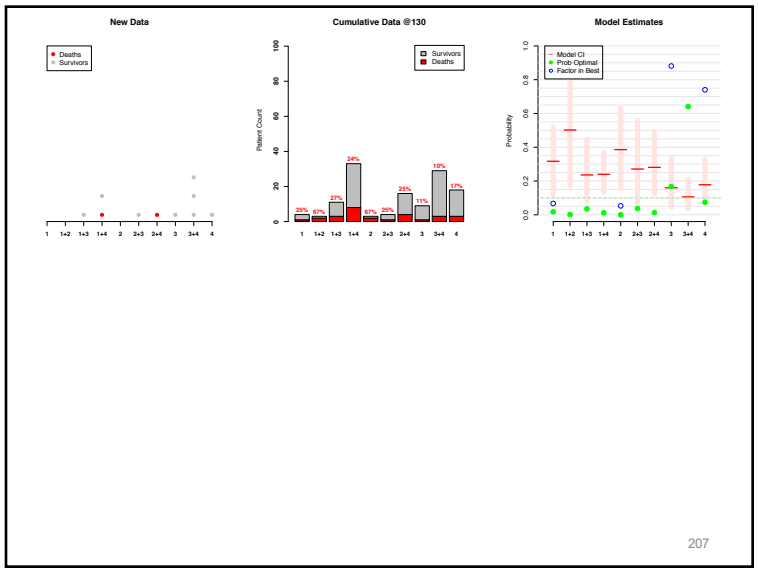
204



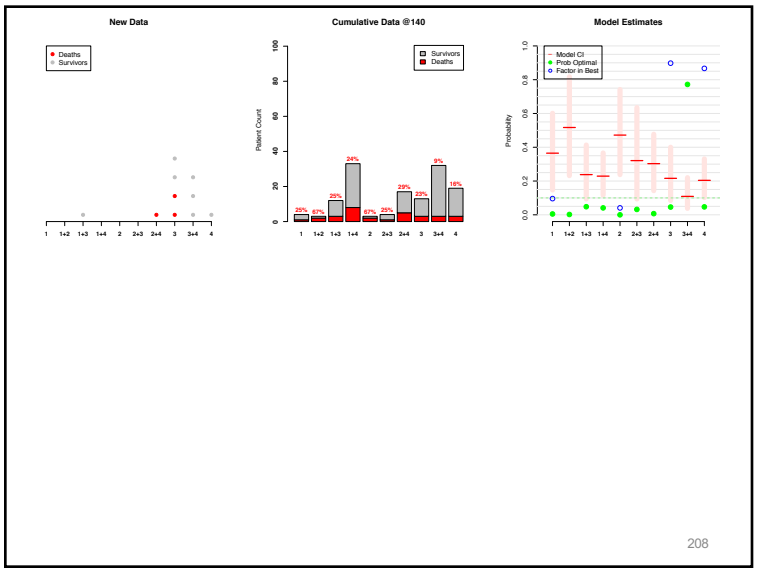
205



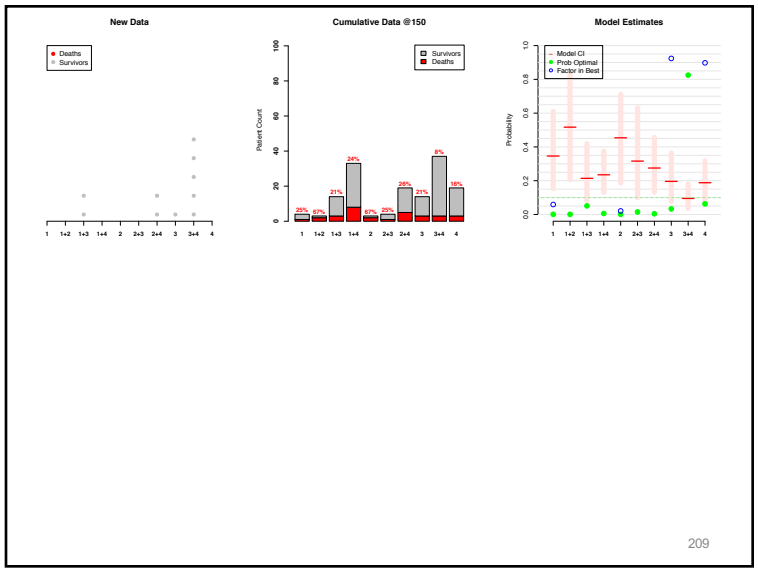
206



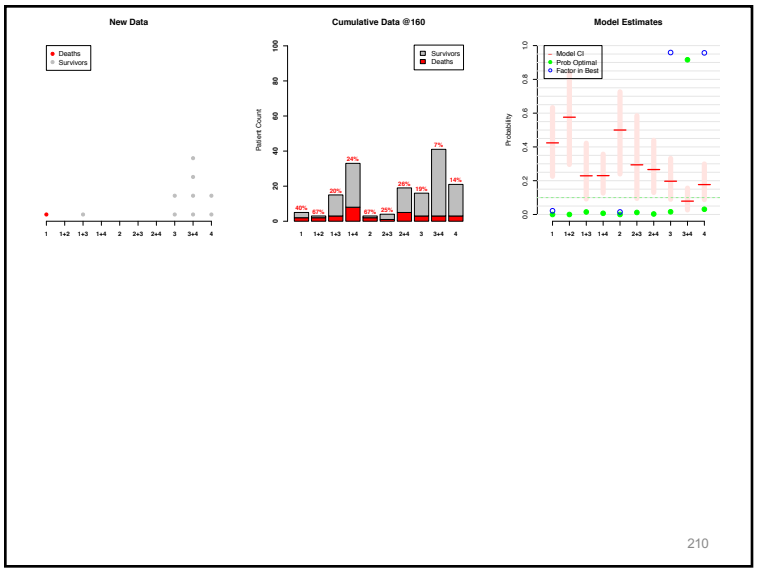
207



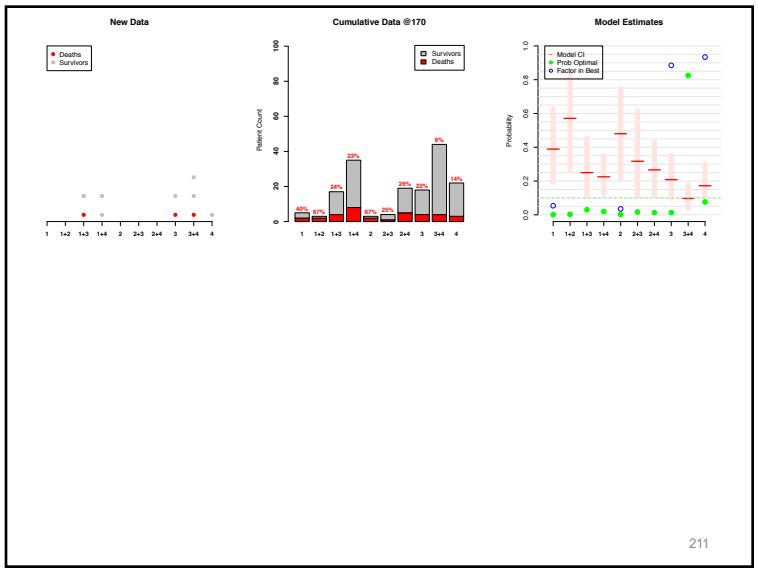
208



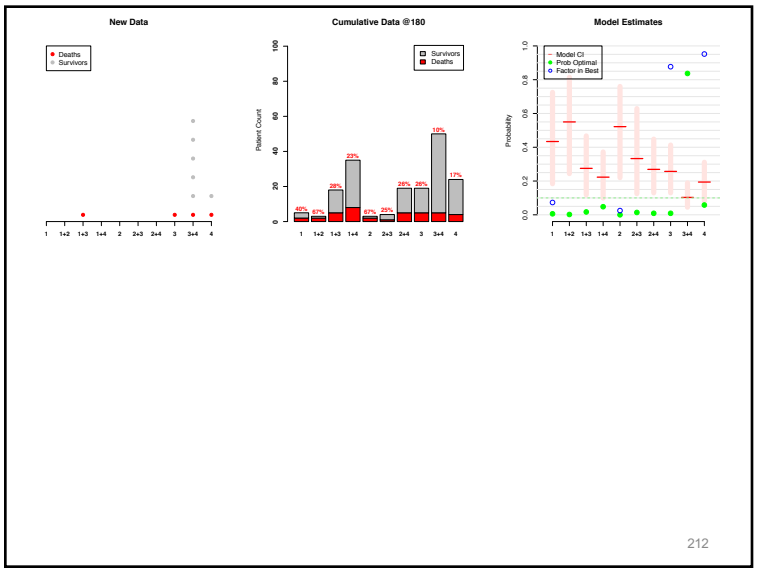
209



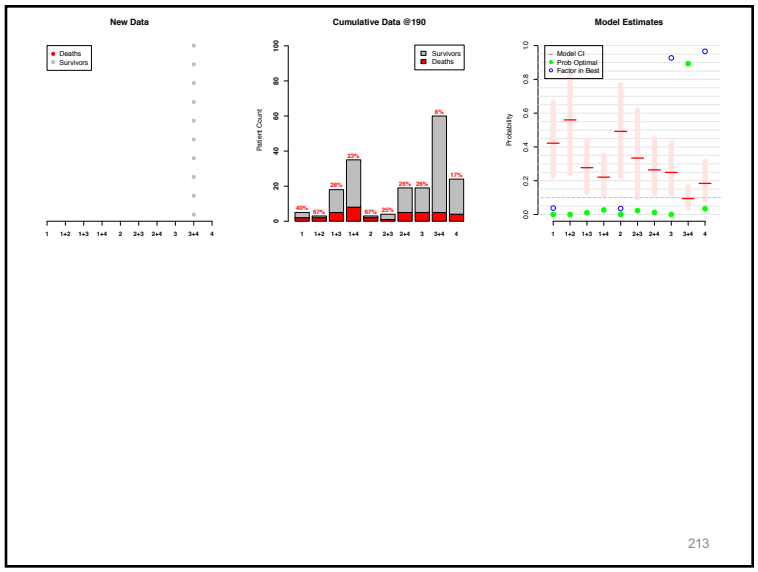
210



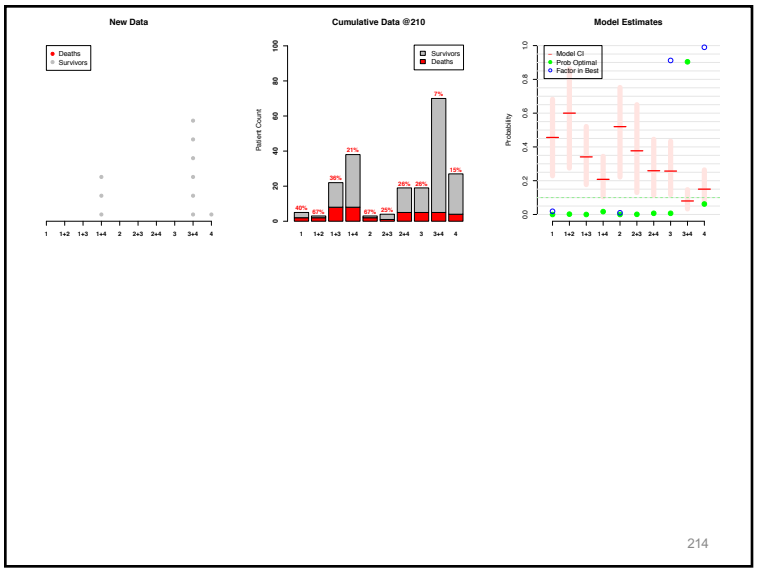
211



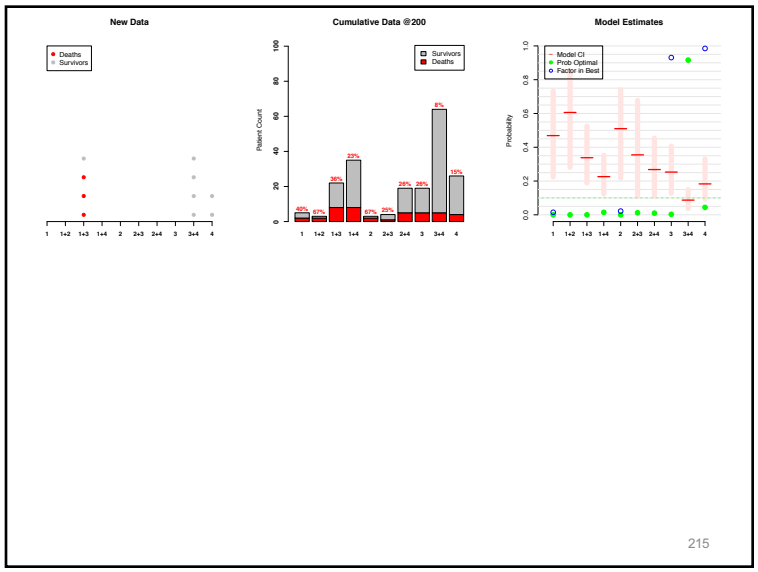
212



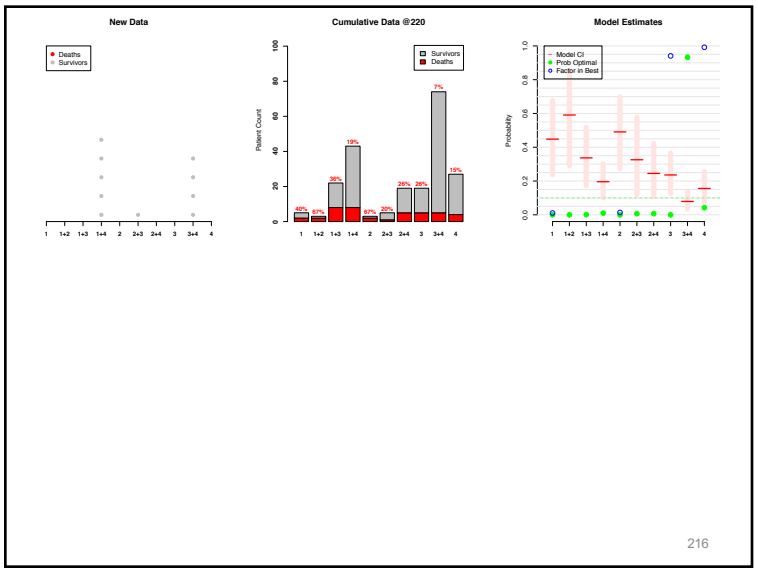
213



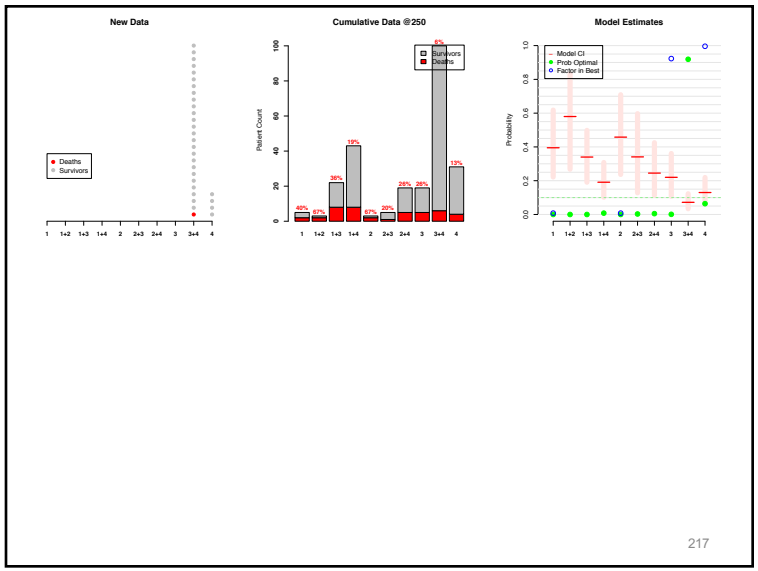
214



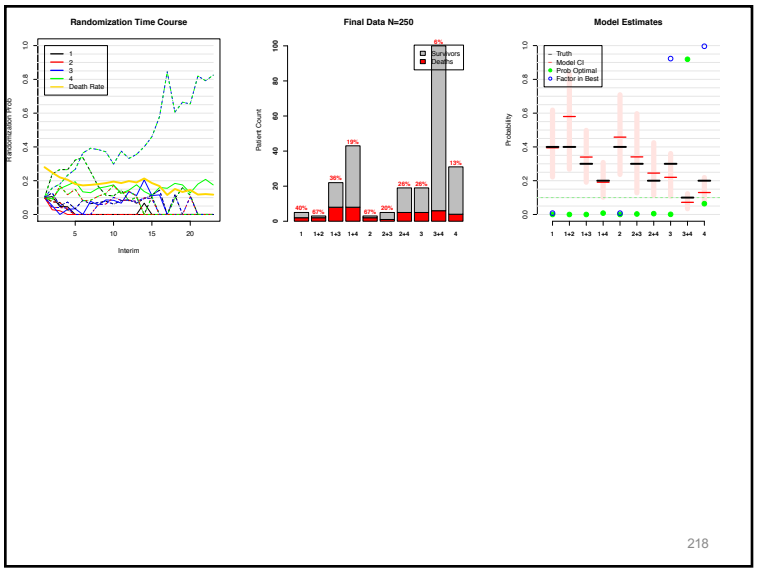
215



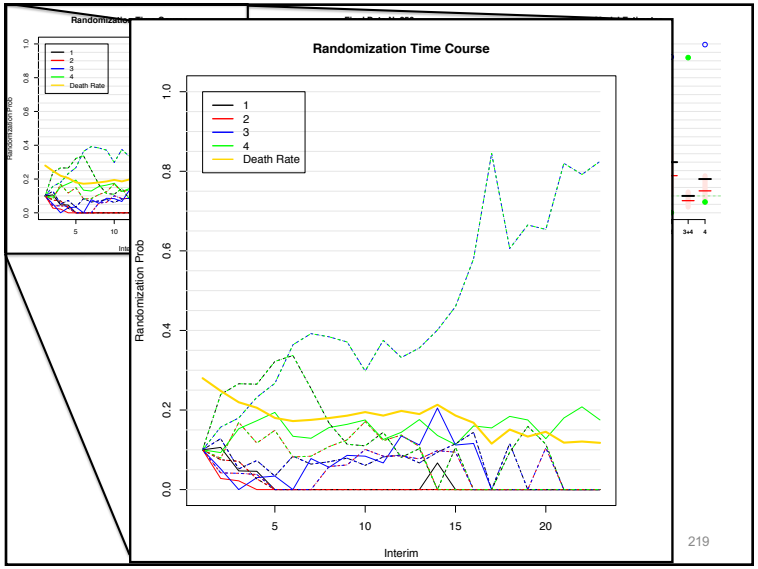
216



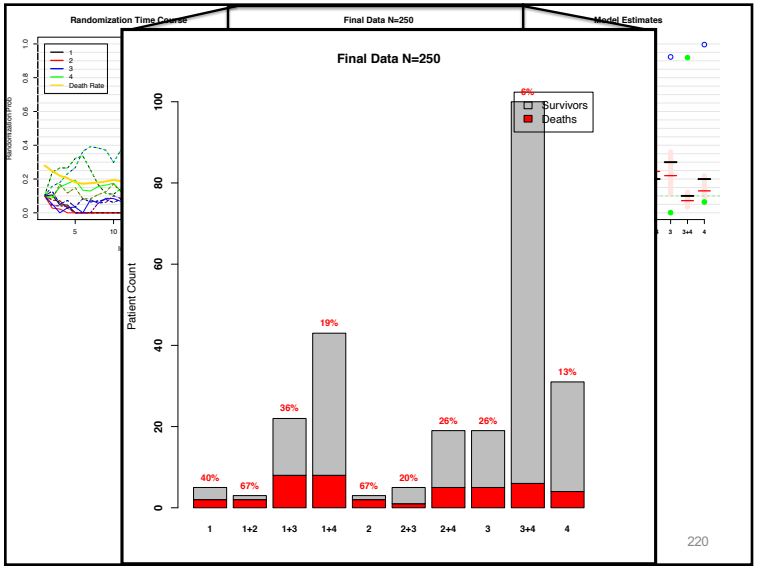
217



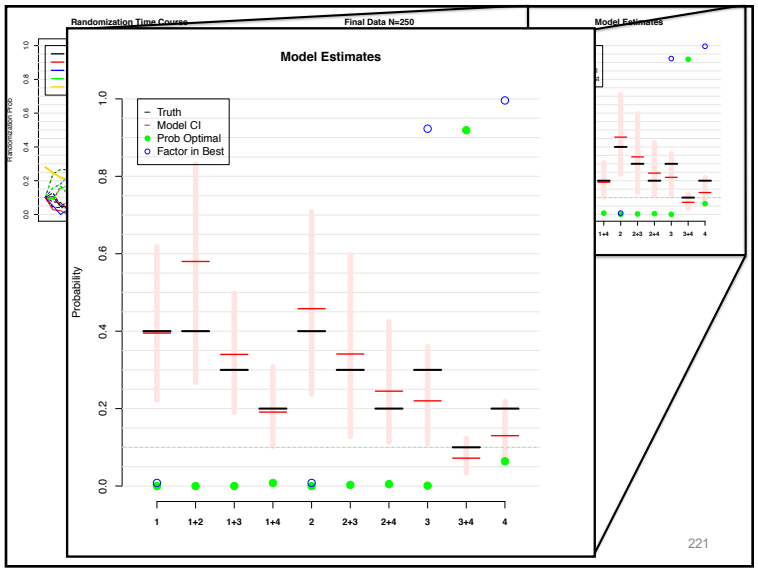
218



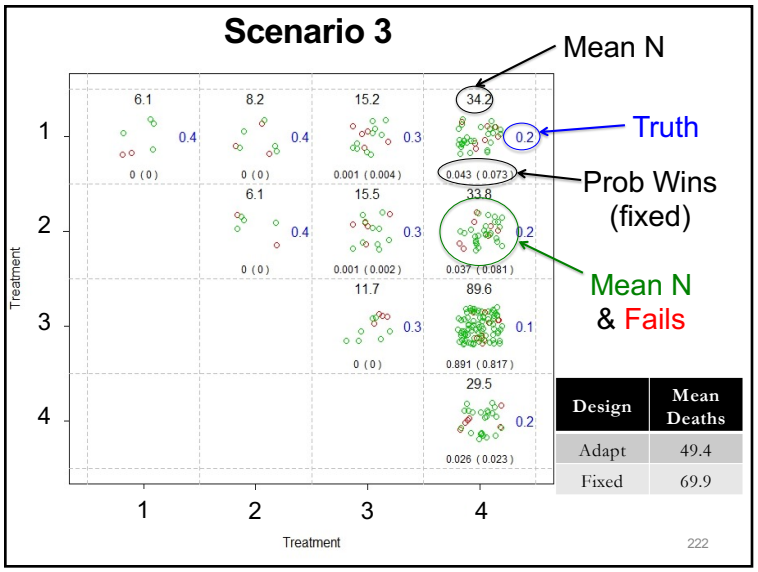
219



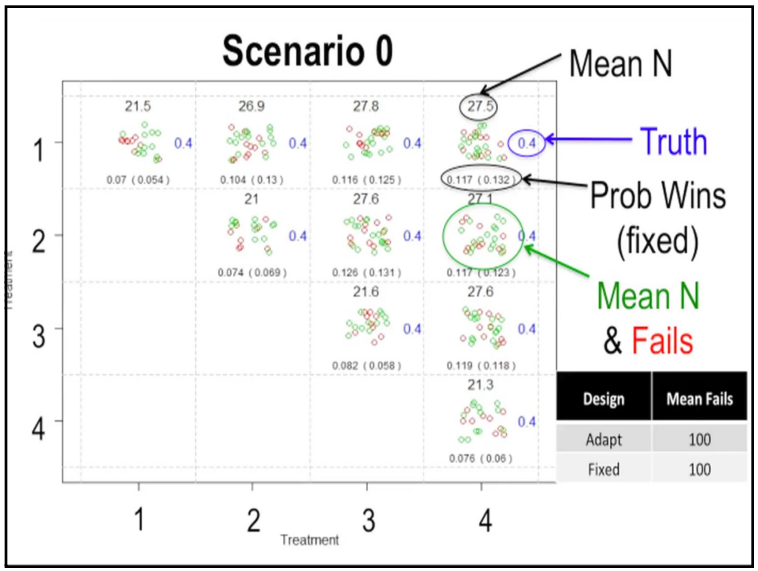
220



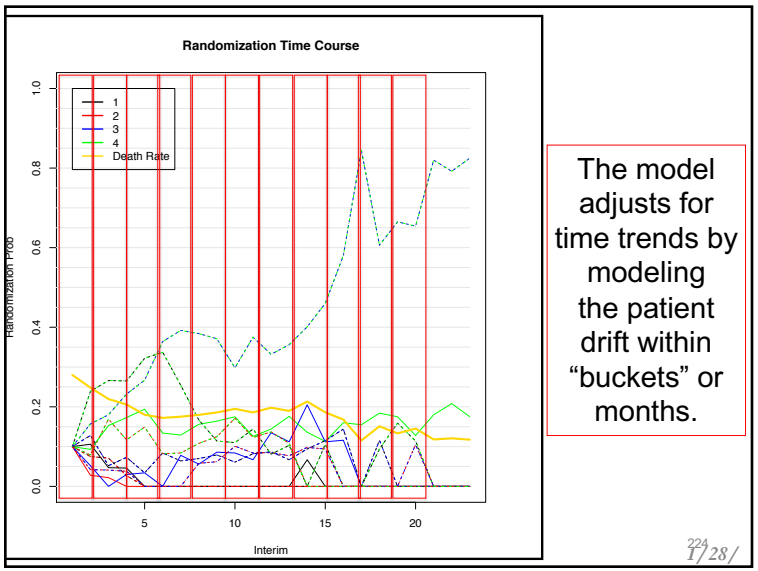
221



222



223



The model adjusts for time trends by modeling the patient drift within "buckets" or months.

224

Summary

- Incredibly powerful design for finding effective therapies and combinations in the universe of treatments
  - Type III Error (the question never asked!)
- Allows the arms to evolve internally and externally to changing science
- Improved Embedded Care: Efficiently and quickly identifies best agents, *while treating patients more effectively*
- Have design ready—on the shelf for next pandemic
  - A number of parameters can be optimized quickly
  - Protocol ready (add appendices)
  - Models + simulations ready
- Need independent committee to decide which drugs to plug in

225

225

Platform Example 2

The Role of Biomarkers in Treatments & Trials

226

226

Testing a New Treatment

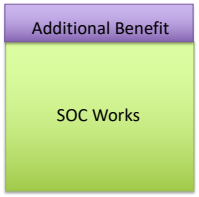
- Standard of Care works in 40%

227

227

10% of Patients Benefit

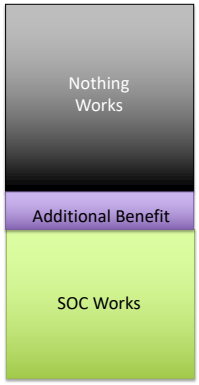
- Standard of Care works in 40%
- New therapy works in 50%



228

228

50% still untreatable

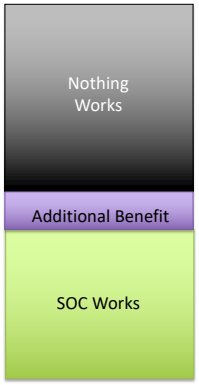


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%

229

229

50% still untreatable

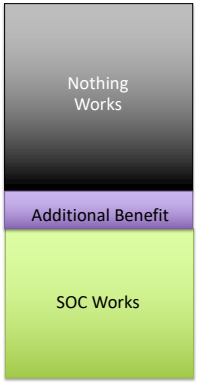


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%
- How many patients do we need to have 90% chance to see a 'statistically significant' difference?

230

230

Need 1036 patients for 90% Power

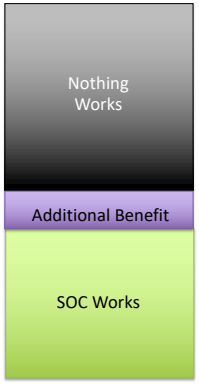


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%

231

231

Need 1036 patients for 90% Power

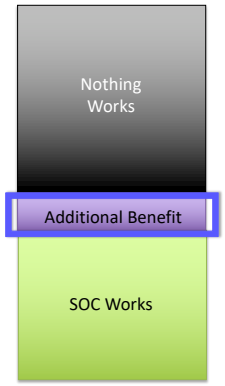


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%
- 90% of patients you enroll tell you nothing

232

232

Need 1036 patients for 90% Power



- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%

- 90% of patients you enroll tell you nothing
- What if you knew which 10% of patients benefited?

233

233

What if you KNEW which 10% Benefit

- If you just enrolled the purple patients how many patients do you need for 90% power?

234

234

What if you KNEW which 10% Benefit

- If you just enrolled the purple patients you need 8 patients for 100% power
- If you could perfectly predict
  - 0/4 on standard of care
  - 4/4 on new treatment
  - Fisher's exact test p-value = 0.029

235

235

What if you^KNEW which 10% Benefit

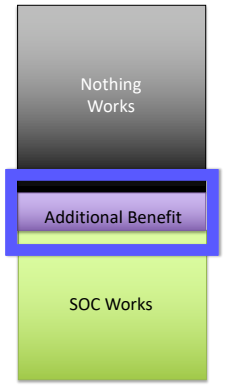
sorta

- Enroll 20% to capture the 10%
- 25% cured by SOC
- 25% still not cured
- 50% of enrolled patients benefit

236

236

What if you sorta^KNEW which 10% Benefit

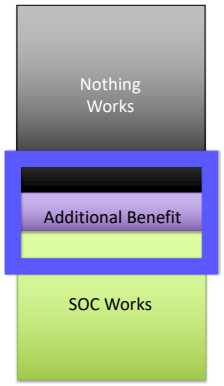


- Enroll 20% to capture the 10%
- 25% cured by SOC
- 25% still not cured
- 50% of enrolled patients benefit
- Need 36 patients for 90% power

237

237

What if you kinda sorta^KNEW which 10% Benefit



- Enroll 30% to capture the 10%
- 33% cured by SOC
- 33% not cured
- 33% of enrolled patients benefit

238

238

What if you kinda sorta^KNEW which 10% Benefit

- Enroll 30% to capture the 10%
- 33% cured by SOC
- 33% not cured
- 33% of enrolled patients benefit
- Need 90 patients for 90% power

239

239

Platform Example 2

240

240

GBM AGILE

Adaptive Global Innovative Learning Environment

Trial Design V1

EXAMPLE TRIAL ONLY
TRIAL HAS CHANGED DRAMATICALLY SINCE THIS

Thanks to Todd Graves & Don Berry

241

Statistical Model

- Primary outcome: Overall Survival
- Time-to-event model including
  - Age
  - Tumor Size
  - Performance Status
  - Site (to be defined)
  - Drug
  - Drug × Biomarker
  - Drug × Biomarker × Biomarker
- Flexible to add drugs & biomarkers on the fly

242

242

Biomarkers → Signatures

Newly diagnosed MGMT-unmethylated	Newly diagnosed MGMT-methylated
Recurrent GBM MGMT-unmethylated	Recurrent GBM MGMT-methylated

243

243

2 × 2 Biomarkers → 4 Signatures

Newly diagnosed MGMT-unmethylated	Newly diagnosed MGMT-methylated
Recurrent GBM MGMT-unmethylated	Recurrent GBM MGMT-methylated

244

244

2 × 2 Biomarkers → 3 Signatures

Newly diagnosed MGMT-unmethylated	Newly diagnosed MGMT-methylated
Recurrent GBM MGMT-unmethylated	Recurrent GBM MGMT-methylated

245

245

2 × 2 Biomarkers → 1 Signature

Newly diagnosed MGMT-unmethylated	Newly diagnosed MGMT-methylated
Recurrent GBM MGMT-unmethylated	Recurrent GBM MGMT-methylated

246

246

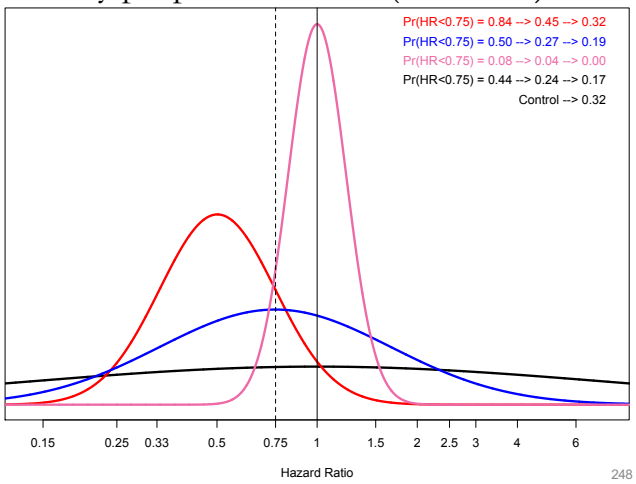
Response-adaptive randomization

- Randomize separately within signature
- Randomization probability proportional to Pr(HR < 0.75)
- If randomization probability < 5%, round to 0
- If N < 50, min rand prob = 1/ # of drugs
- Probability randomize to control = Probability randomize to best drug
- Update monthly

247

247

Why proportional to Pr(HR<0.75)?



248

248

Graduation

A drug graduates if, *within any signature*,

- Pr(HR < 1) > 99%
- Min 75 patients on that drug overall
- Min 300 pt-months exposure on that signature

When a drug graduates

- Drug out of trial
- Data for all subtypes delivered to sponsor

249

249

Futility

A drug is removed from the trial for futility if

- Pr(HR < 0.75) < 5% for all signatures
- At least 50 patients

Or

- Been enrolling for 3 years

Stop at Max N=150 over all signatures

250

250

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	
Unmethlylated		

251

251

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	But not here
Unmethlylated	Or here	Or here

Identify it works in red lasso:
We made the right choice

252

252

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	But not here
Unmethlylated	Or here	Or here

Identify it works in red lasso:
Have we made the right choice?
Is this a Type 1 error?
Call this a SUPERSET error

253

253

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	And here
Unmethlylated	But not here	Or here

Identify it works in red lasso:
We made the right choice

254

254

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	And here
Unmethlylated	But not here	Or here

Identify it works in red lasso:
Did we made the right choice?
We made a "Type 2 error"
Call this a SUBSET error

255

255

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	And here
Unmethlylated	But not here	Or here

Identify it works in red lasso:
Did we made the right choice?
We got one right but made a "Type 1 Error" & "Type 2 error"!
Call this a "MIXED TYPE ERROR"

256

256

Factors We Can Tune

- Max N per drug
- Signatures (Biomarker-drug interactions)
- Randomization algorithm
- Futility rule
  - Pr(HR<0.75)
  - Min N
  - Max time allowed to accrue
- Graduation rule
  - Pr(HR<1)
  - Min N, Min Exposure

257

257

Learn & Confirm Using Biomarkers

- Make confirmatory trials dramatically smaller
  - Or learn & confirm within a trial
- Lead us toward personalized medicine
  - What works best in whom?
- May require larger platforms trials, data sharing & adaptive randomization to efficiently identify
- Different drugs work in different types of patients
  - Not one trial, one patient type
  - Learn, confirm, perpetually

258

258

Challenges in Platform Trials

- Complexity in trial implementation and planning
- Collaborations across sponsors - who initiates the planning?
- Timely communication between participating sites and data coordinating units
- Sponsors sacrifice autonomy in running the trial
- Determining shared costs
- Identifying what to report when
  - iSpy2 has rules for 'graduating'
  - When to report subgroup results broadly?

259

259

Platform Trial Efficiencies

- Useful for evaluating combinations of treatments and for direct comparisons between competing treatments
  - Decide a prior whether each vs. control or vs. each other
- Do not require a new trial infrastructure for every treatment under investigation
- Easier for regulators given evidence comes in common form
- Always new drugs on horizon
  - Even if lull, get SOC in the process
- Can build in Phase 2/3 design
  - Learn & confirm
- Need to prospectively define control group
  - Concurrent controls
  - 'Time machine'
  - What if control group changes

260

260

Conclusions

- Adaptive trial designs can be used to create a seamless process in which new evidence about effectiveness is immediately used to improve patient care
- A platform trial can extend this process beyond a single treatment or few treatments
- Current work is focused on embedding this approach into the health care infrastructure
- Patients will benefit if we merge clinical trials and decision support into a single, continuous process

261

261

Thank you!

- Thank you for a great class.

- Please complete evaluations

To access evaluations, log in to https://si.biostat.washington.edu/user/login, click "My Account" in the upper right, the evaluations will appear on your dashboard.

After you have completed your evaluations, a link to download the certificate of completion will appear within 24 hours.

262

262