Bayesian Adaptive Clinical Trial Design

Jason Connor
ConfluenceStat

Jason@ConfluenceStat.com
412-860-3113

Day 2

1

NOTE TO SELF:
START THE RECORDING

2

ESSET Code

3

Definitions, Trial Parameters

```
rm(list=ls())
## All times in months
library(VGAM)
v = list(
  ### Event, success probabilities for IV, IV+2nd therapy, Oral, Oral + 2nd therapy
  S3 = c(## There are success rates for the three groups
     0.50,      # fPHT
     0.50,      # LVT
     0.50       # VPA
  ),
  ## Maximum sample size & max sample size for Stage 1
  MaxN = 795,
  # Priors
  a = rep(1, 3),
  b = rep(1, 3),
  # First look and look every
  firstlook = 300,
  firststop = 400
  lookevery = 100,
  # Min to randomized
  minpr = 0.05,
  # simulations
  nsims = 1000,
  badlim = 0.25,
  # critv to (a) for 'best'
  #          (b) for 'worst'
  #          (c) to stop for futility (i.e Pred prob a winner or loser id'd)
  #          (d) for worse than 25%
  critv = c(.975, .975, 0.05, 0.05)
)
```

Response Rates

MaxN

Priors

Sample Size &
Timing of Looks

Critical values for stopping

4

```
simtrials <- function(v){

  co <- ppcutoffs(v$critv[3])

  #out.mat
  # (1) N
  # (2-4) N per group
  # (5-7) Rank as 1, 2, 3 (according to prob best)
  # (8)   Sig best (1 2 or 3 or 0 if none)
  # (9)   Sig worst (1 2 or 3 or 0 if none)
  # (10) Final conclusion
  #           1 = overall futility stop,
  #           2 = stop early for winner
  #           3 = stop early for winner & loser
  #           4 = stop early for loser and futility (not possible in ours)
  #           5 = max overall futility
  #           6 = max and loser
  #           7 = max and winner
  #           8 = max & winner & loser
  #  (11-13) Final Pr(best)
  #  (14-16) Final Pr(2nd)
  #  (17-19) Final Pr(worst)
  #  (20-22) Successes per group
  #  (23-25) Ever drop arm? (rand goes to 0 at any pt)
```

Creates a big matrix to store simulation results

5

```
out.mat <- matrix(NA, nrow=v$nsims, ncol=25)
  for(s in 1:v$nsims){
    ad <- c(1,1,1)
    ## Rand assignment for first FirstLook pts & generate outcome
    group <- rep(NA, v$MaxN)
    group[1:v$firstlook] <- rand.new(v$firstlook, c(1,1,1))
    y <- rep(NA, v$MaxN)
    y[1:v$firstlook] <- sim.endpoint(group[1:v$firstlook], v$S3)
    look1 <- interim(v$firstlook, y, group, v, co)
#    print(round(look1,3))
    # Track if arm every dropped
    ad <- ad * as.numeric(look1[12:14]>0)
    n.now <- v$firstlook
    print(c(s,n.now))
 ## Now loop through Stage 1
    while(look1[1]==1){
      new <- min(v$MaxN-n.now, v$lookevery)
      group[(n.now+1):(n.now+new)] <- rand.new(new, look1[12:14])
      y[(n.now+1):(n.now+new)] <- sim.endpoint(group[(n.now+1):(n.now+new)], v$S3)
      look1 <- interim(n.now+new, y, group, v, co)
#      print(round(look1,3))
      ad <- ad * as.numeric(look1[12:14]>0)
      n.now <- n.now+new
      print(c(s,n.now))
    }
```

Simulate group assignment & response to tx

First interim look

Simulate group assignment & response to tx

Do interim looks

6

```
    mx <- look1[3:5];    mn <- look1[6:8]
    winner <- ifelse(max(mx) > v$critv[1], (1:3)[mx==max(mx)], 0)
    loser <-  ifelse(max(mn) > v$critv[2], (1:3)[mn==max(mn)], 0)
   if(look1[2]==1){
      whystop <- 1    ## futility
     }else if(look1[2]==3){
       if(loser>0){
         whystop <- 3
       }else{
         whystop <- 2
       }
     }else if(look1[2]==2){
       if(winner==0 & loser==0)   { whystop <- 5}
       else if(winner>0 & loser>0){ whystop <- 8}
       else if(winner>0)          { whystop <- 7}
       else if(loser>0)           { whystop <- 6}
       else{print("error why stop at max?")}
       else{print("error, why did trial stop?")}

 out.mat[s,1:25] <- c(n.now, look1[18:20], order(mx), winner, loser,
                 whystop,look1[c(3,4,5,9,10,11,6,7,8,15,16,17)],1-ad)
}
    out.mat <- data.frame(out.mat)
    names(out.mat) <- c("N","N1","N2","N3",…
     return(out.mat)
```

See if best or worst identified

See if stopping rules met

Print out simulation results

7

```
rand.new <- function(N, p, minp){
  ### Returns randomization codes (1:3) for N patients
  ### requires prob vector, p, of lenght 3.
  ### If
  if(prod(p ==c(1,1,1))==1){
     out <- rep(sample(1:3,3), ceiling(N/3))
     out <- out[1:N]
   }else{
     out <- rep(sample(1:3, N, prob=p, replace=T))
   }
    return(out)
}


## Simulates data for new patients using inputs group assignement and success rate
(length 3)
sim.endpoint <- function(group, successrate){
  out <- rbinom(length(group), 1, successrate[group])
}
```

Takes how many patients to simulate, N
Rand prob , p
And Min rand prob minp

Simulate a success or failure for each based given their group assignment
group is vector of 1,2 or 3
successrate is length 3

8

Predictive Probability Cutoffs Lookup Matrix

```
### Creates a lookup matrix to make the predictive probability stopping algorithm
go faster.
### Creates a 99.9% confidence interval, then basically sees if its' highly
likely that the stop rate is less than the cutoff
ppcutoffs <- function(critv){
    whenstop <- cbind(rep(0,1000),rep(0,1000))
    for(i in 50:1000){
    x <- ceiling(critv*i)
    while(as.numeric(binom.test(x,i,conf.level=0.999)$conf.int[1])<critv){
    x <- x+1
    }
    whenstop[i,1] <- x
    }
    whenstop[1:49,1] <- whenstop[50,1]

    for(i in 50:1000){
    x <- ceiling(critv*i)
    while(as.numeric(binom.test(x,i,conf.level=0.999)$conf.int[1])>=critv){
    x <- x-1
    }
    whenstop[i,2] <- x
    }
    return(whenstop)
}
```

9

```
interim <- function(N, y, group, v, co){
  ## Runs trial returns:
  #  (1) go (0=stop, 1=keep going)
  #  (2) why stop (1=3-way fut, 2=max n, 3=1 winner)
  #  (3-5) Pr each is best
  #  (6-8) Pr each is worst
  #  (9-11) x/N for each group
  #  (12-14) rand probs
  ns <- table(factor(group[1:N], levels=1:3))
  tab <- table(factor(group[1:N],levels=1:3), factor(y[1:N], levels=0:1))
  post1 <- rbeta(10000, v$a[1]+tab[1,2], v$b[1]+tab[1,1])
  post2 <- rbeta(10000, v$a[2]+tab[2,2], v$b[2]+tab[2,1])
  post3 <- rbeta(10000, v$a[3]+tab[3,2], v$b[3]+tab[3,1])
  vr <- as.numeric(( (v$a+tab[,2])*(v$b+tab[,1])) / ((v$a+v$b+ns)^2 * (v$a+v$b+ns+1)))
  top <- apply(cbind(post1,post2,post3), 1, max)
  bot <- apply(cbind(post1,post2,post3), 1, min)

  best <- c(mean(post1==top), mean(post2==top), mean(post3==top))
  worst <- c(mean(post1==bot), mean(post2==bot), mean(post3==bot))
  middle <- 1-best-worst

  toobad <- 1-c(pbeta(v$badlim, v$a[1]+tab[1,2], v$b[1]+tab[1,1]),
                pbeta(v$badlim, v$a[2]+tab[2,2], v$b[2]+tab[2,1]),
                pbeta(v$badlim, v$a[3]+tab[3,2], v$b[3]+tab[3,1]))

  wt <- sqrt(best * vr / as.numeric(ns));     wt <- wt/sum(wt)
  wt[wt < v$minpr] <- 0;    wt[toobad < v$critv[4]] <- 0
  if(sum(wt) > 0){
  wt <- wt/sum(wt)
  }
```

Does interim analysis
Calc posteriors, new rand probs,
Pred prob of success at max

Calc posteriors

Calc prob each is best & worst

Calc Pr(p<0.25)

Calc new rand prob

10

```
sumtrial <- function(outmat){
  mat <- matrix(nrow=4, ncol=9)
  out <- table(factor(outmat[,10], levels=1:8))
#             Ntotal SDN phat Rank1  Rank2  Rank3  SigBest SigWorst Drop
#      fPHT
#      LVT
#      VPA            --
#      Total
  mat[1:3,1] <- apply(outmat[,2:4], 2, mean)  ### Average Patients per arm
  mat[1:3,2] <- apply(outmat[,2:4], 2, sd)    ### SD patients per arm
  mat[1:3,3] <- c(mean(outmat[,20]/outmat[,2]), mean(outmat[,21]/outmat[,3]),
mean(outmat[,22]/outmat[,4]))                 ### Average successes per arm
  mat[1,4:6] <- table(factor(outmat[,5], levels=3:1))/dim(outmat)[1]  ## Avg Pr Best
  mat[2,4:6] <- table(factor(outmat[,6], levels=3:1))/dim(outmat)[1]  ## Avg Pr middle
  mat[3,4:6] <- table(factor(outmat[,7], levels=3:1))/dim(outmat)[1]  ## Avg Pr Worst
  mat[1:3,7] <- table(factor(outmat[,8], levels=1:3))/dim(outmat)[1]  ## Pr Sig Best
  mat[1:3,8] <- table(factor(outmat[,9], levels=1:3))/dim(outmat)[1]  ## Pr Sig Worst
  mat[1:3,9] <- apply(outmat[,23:25], 2, mean)                        ## Pr Ever Dropped
  mat[4,1] <- mean(outmat[,1])     ### Mean total sample size
  mat[4,2] <- sd(outmat[,1])       ### SD total sample size
  mat[4,3] <- mean(rowSums(outmat[,20:22]) / rowSums(outmat[2:4]))   ### Mean response rate
per arm
  mat[4,4:6] <- NA
  mat[4,7] <- sum(mat[1:3,7])    ### Total prob ID a sig best
  mat[4,8] <- sum(mat[1:3,8])    ### Total prob ID a sig worst
  mat[4,9] <- NA
  mat <- data.frame(mat)
  names(mat) <- c("N","SD","Phat","Best","Mid","Worst","SigBest","SigWorst","Drop")
  dimnames(mat)[[1]] <- c("fPHT","LVT","VPA","Total")
  return(list(out, mat))
}
```

Takes the results of 'simtrials' and Produces prettier output

11

```
#####PRED PROBS; only do if all 3 arms left
  if((N >= v$firststop) & (N < v$MaxN) & (prod(wt>0)> 0)){
    drop <- 0
    left <- v$MaxN - N
    left <- ceiling(rep(left/3, 3))
    ns.total <- ns+left
    winlose <- 0
    counter <- 1

    while((winlose < co[counter,1]) & (winlose >= co[counter,2]) & (counter < 1000)){
      y.end <- tab[,2] + rbetabin.ab(3, left, v$a+tab[,2], v$b+tab[,1])
      post1f <- rbeta(10000, v$a[1]+y.end[1], v$b[1]+ns.total[1]-y.end[1])
      post2f <- rbeta(10000, v$a[2]+y.end[2], v$b[2]+ns.total[2]-y.end[2])
      post3f <- rbeta(10000, v$a[3]+y.end[3], v$b[3]+ns.total[3]-y.end[3])
      topf <- apply(cbind(post1f,post2f,post3f), 1, max)
      botf <- apply(cbind(post1f,post2f,post3f), 1, min)
      bestf <- c(mean(post1f==topf), mean(post2f==topf), mean(post3f==topf))
      worstf <- c(mean(post1f==botf), mean(post2f==botf), mean(post3f==botf))
      winlose <- winlose + ifelse((max(bestf)>v$critv[1]) | (max(worstf)>v$critv[2]),
1, 0)
      counter <- counter + 1
#      print(c(winlose/counter, counter))
    }
    ppwin <- winlose/counter
  }else{
    drop <- 1
    ppwin <- v$critv[3]+1 #  If missing just make bigger than the crit value.
  }
```

Calc pred prob of success At Max N

12

```
  ## Stopping:
  if(N < v$firststop){
    go <- 1
    whystop <- NA
  }else if(N >= v$MaxN){
    go <- 0
    whystop <- 2
  }else if(max(best) > v$critv[1]){
    go <- 0
    whystop <- 3
  }else if(ppwin < v$critv[3]){
    go <- 0
    whystop <- 1
  }else if(wt[1]==0 & wt[2]==0 & wt[3]==0){
    go <- 0
    whystop <- 1
  }else{
    go <- 1
    whystop <- NA
  }

 return(as.numeric(c(go, whystop, best, worst, middle, wt, tab[,2], ns, ppwin, drop)))
}
```

Track IF stop
And WHY stop

13

Predictive Probabilities

14

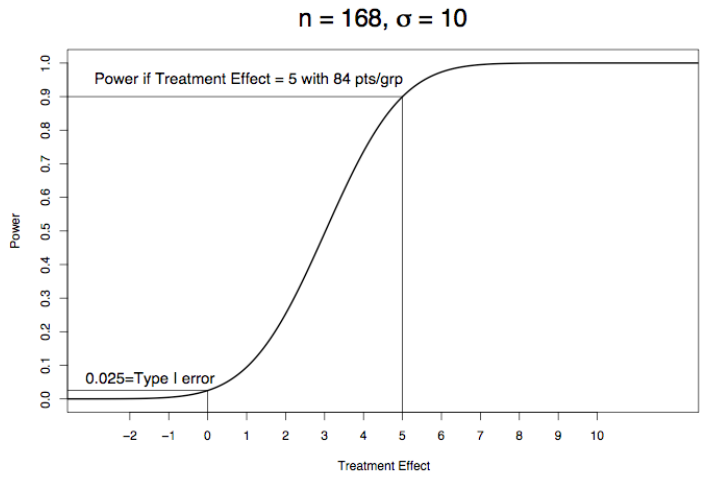
14

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?"

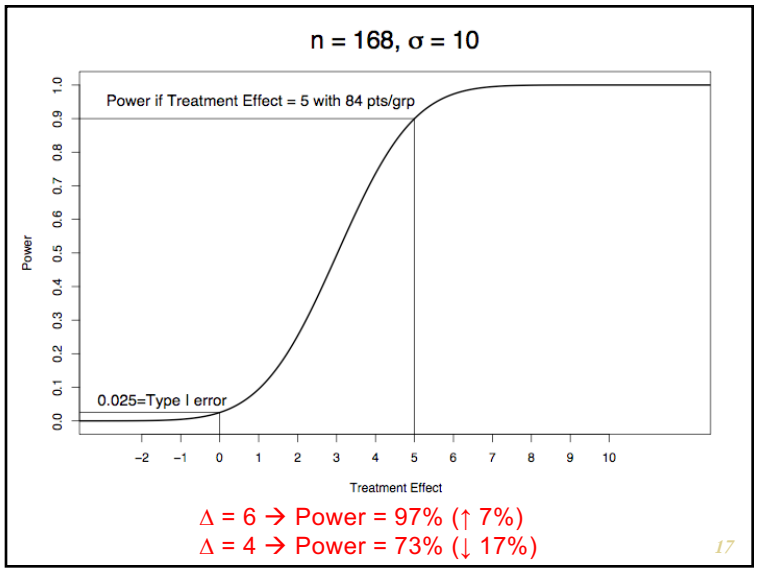
15

15

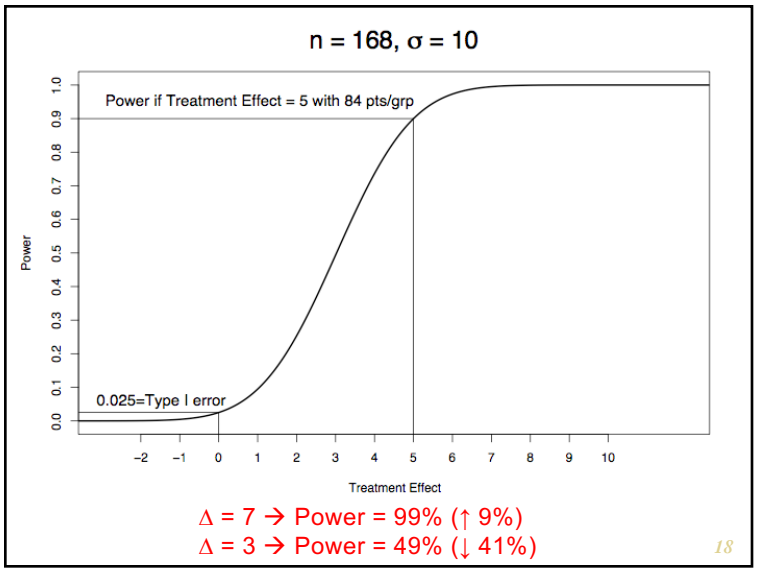


16

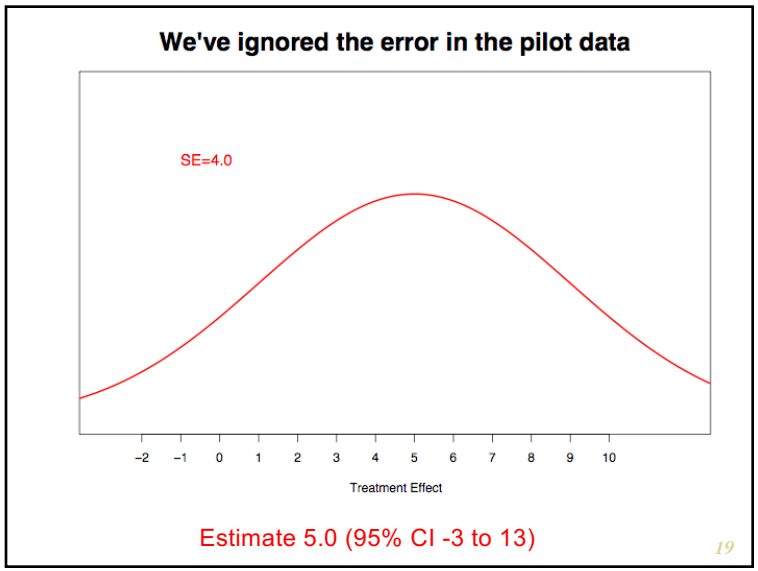
16


n = 168, σ = 10
Power if Treatment Effect = 5 with 84 pts/grp
0.025=Type I error
Power
Treatment Effect
Δ = 6 → Power = 97% (↑ 7%)
Δ = 4 → Power = 73% (↓ 17%)
17

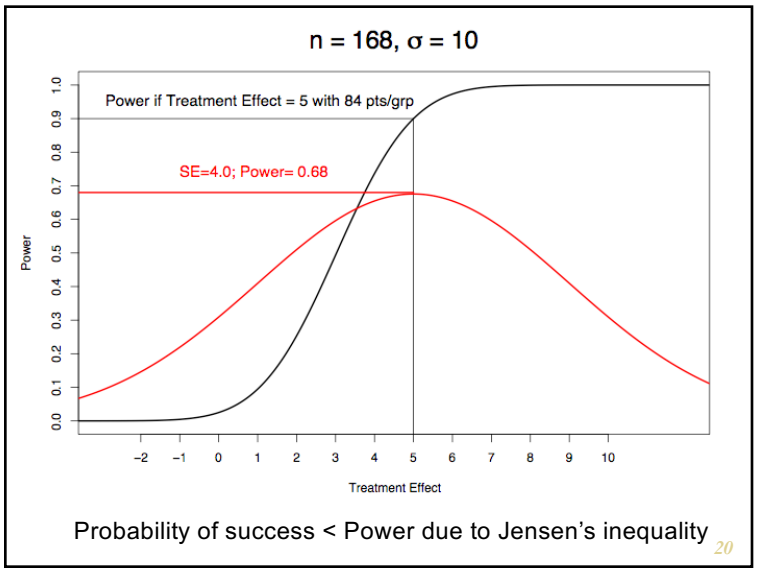
17


n = 168, σ = 10
Power if Treatment Effect = 5 with 84 pts/grp
0.025=Type I error
Power
Treatment Effect
Δ = 7 → Power = 99% (↑ 9%)
Δ = 3 → Power = 49% (↓ 41%)
18

18


We've ignored the error in the pilot data
SE=4.0
Treatment Effect
Estimate 5.0 (95% CI -3 to 13)
19

19


n = 168, σ = 10
Power if Treatment Effect = 5 with 84 pts/grp
SE=4.0; Power= 0.68
Power
Treatment Effect
Probability of success < Power due to Jensen's inequality
20

20

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)

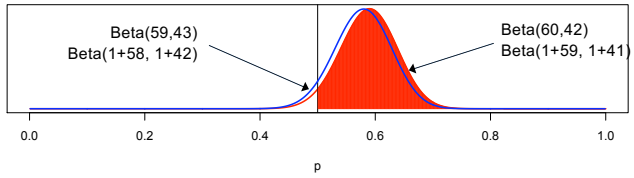
21

21

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



22

22

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

23

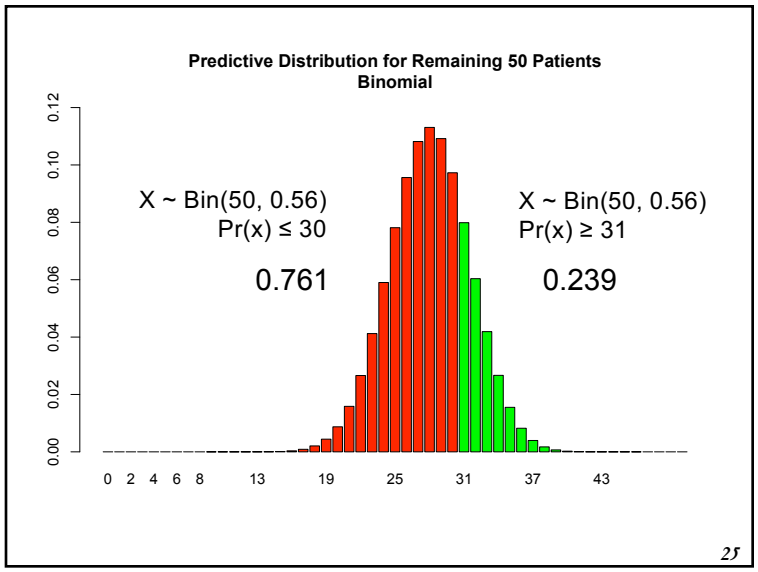
23

Predictive Probabilities

- 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
- Observe data half way through
  - See 28/50 successes
  - Need to see 31/50 to meet threshold
  - What is predictive probability of trial success?

24

24



25

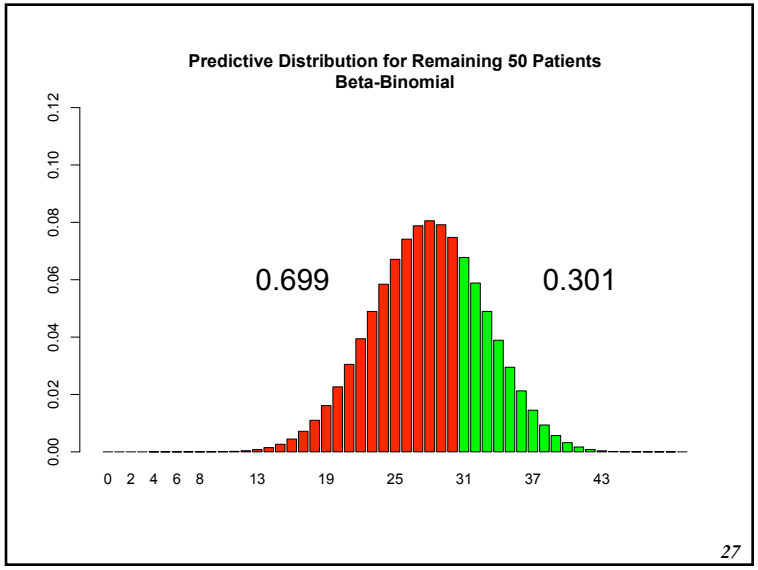
Predictive Probabilities

- 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)

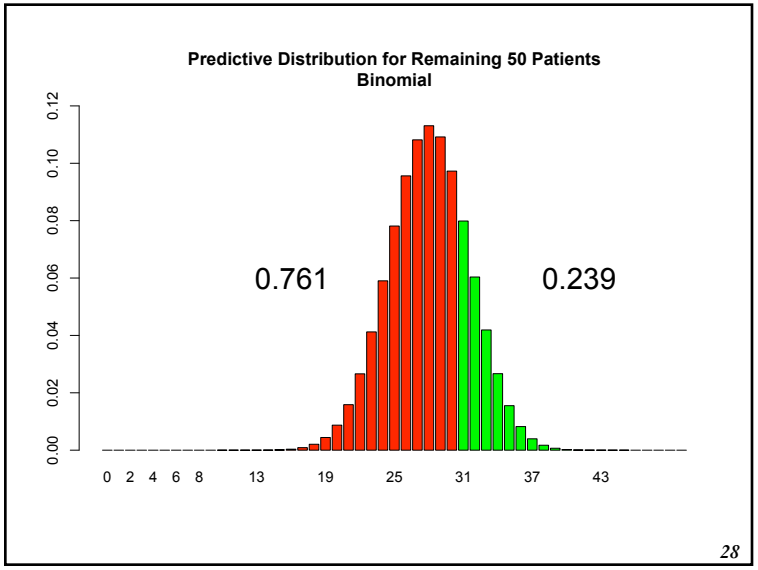
$$\Pr(\text{Win Trial}) = \sum_{x_2=31}^{50} \left\{ \binom{50}{x_2} \frac{B(x_2 + 29, 50 - x_2 + 23)}{B(29,22)} \right\} = 0.301$$

26

26



27



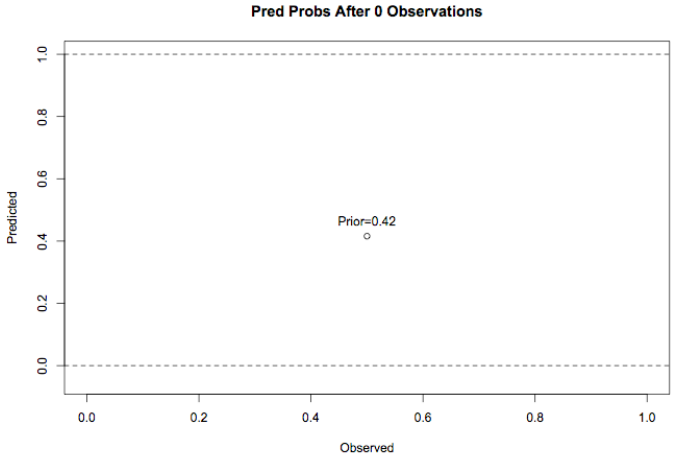
28

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")
```

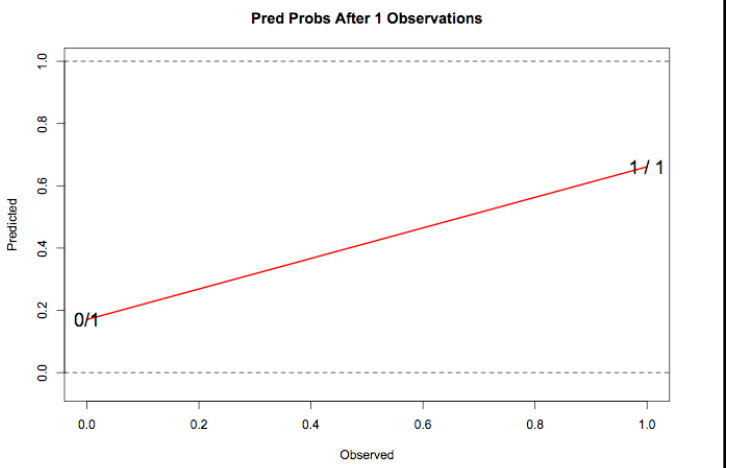
29

29



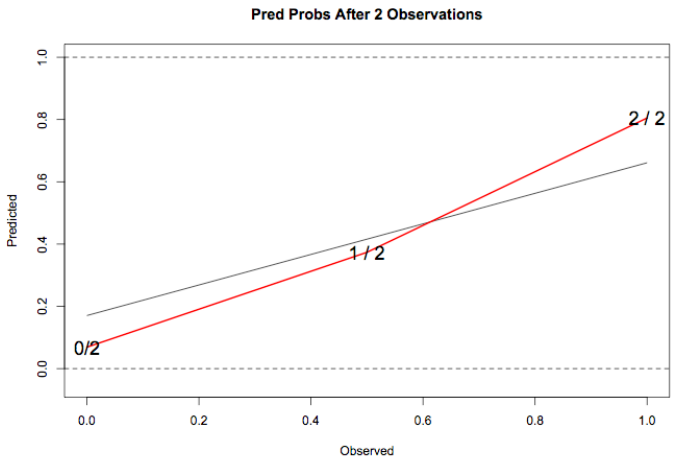
30

30



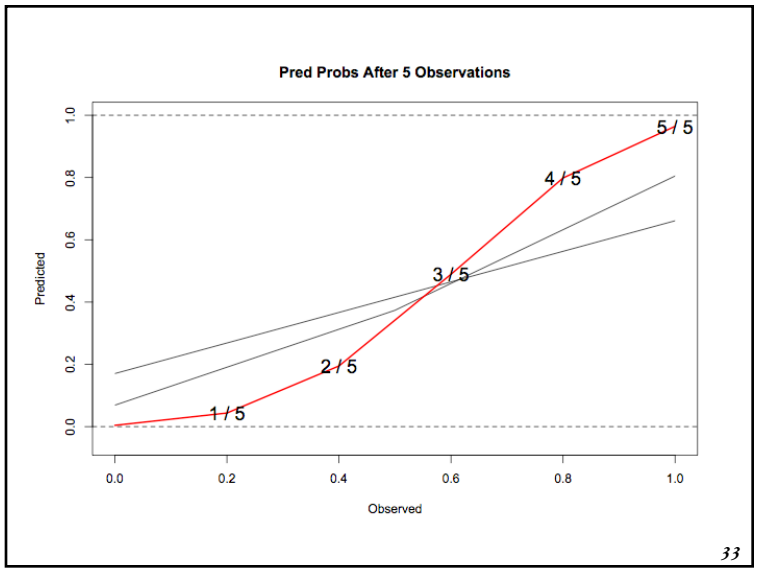
31

31

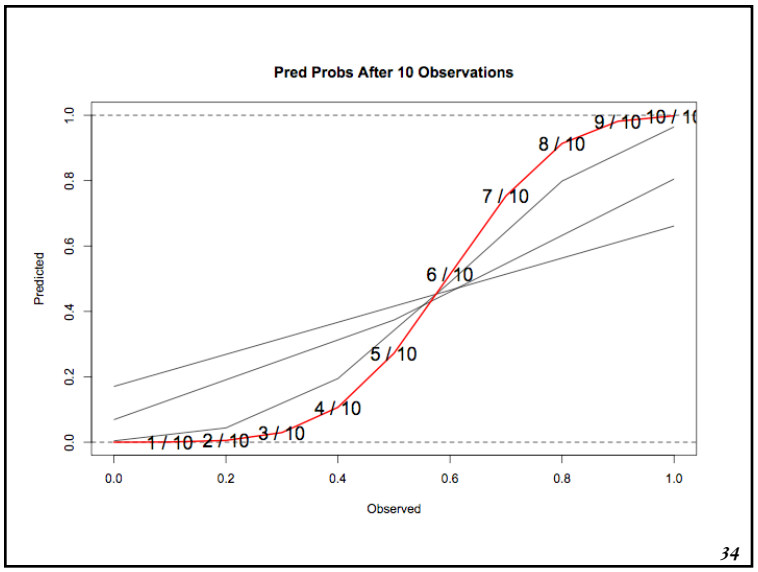


32

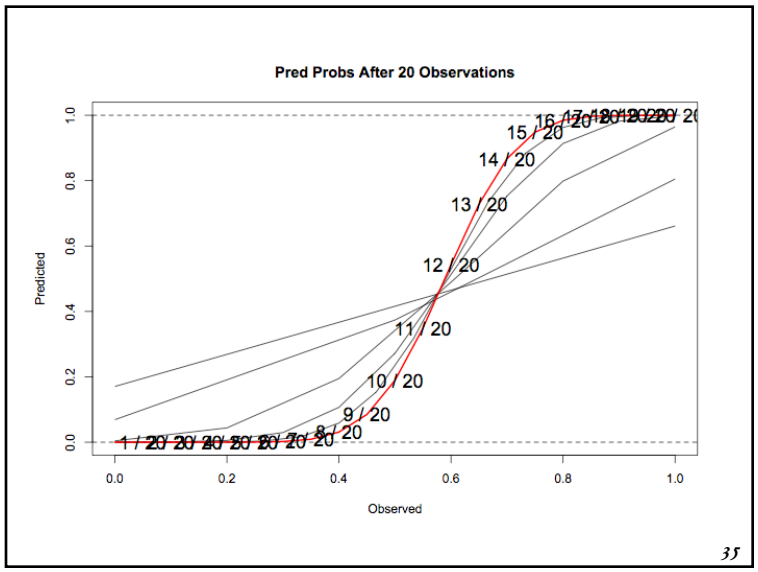
32


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
33

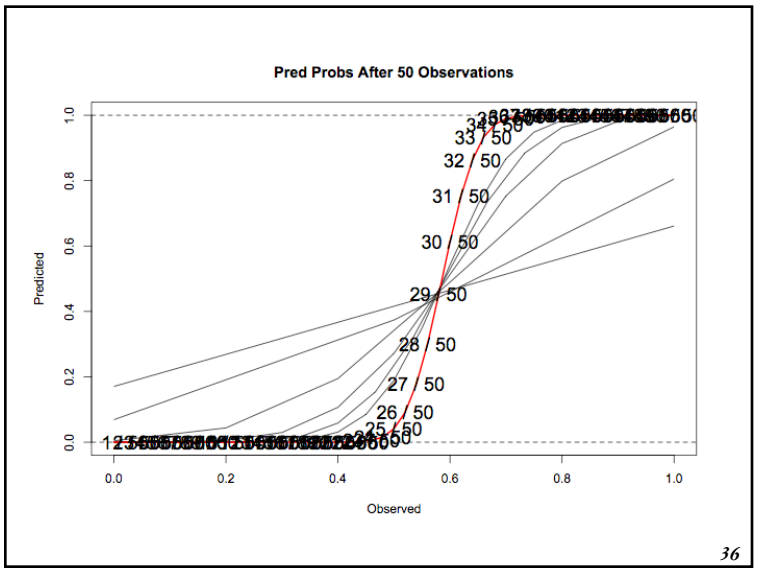
33


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
34

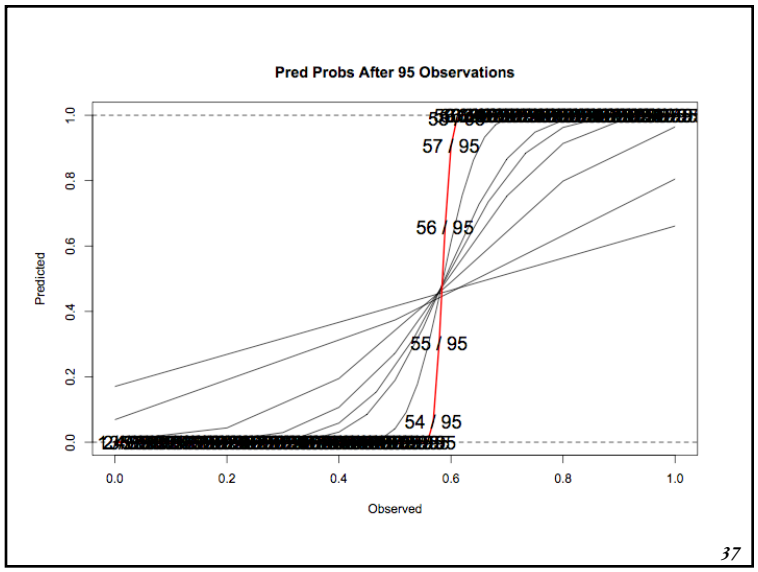
34


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
35

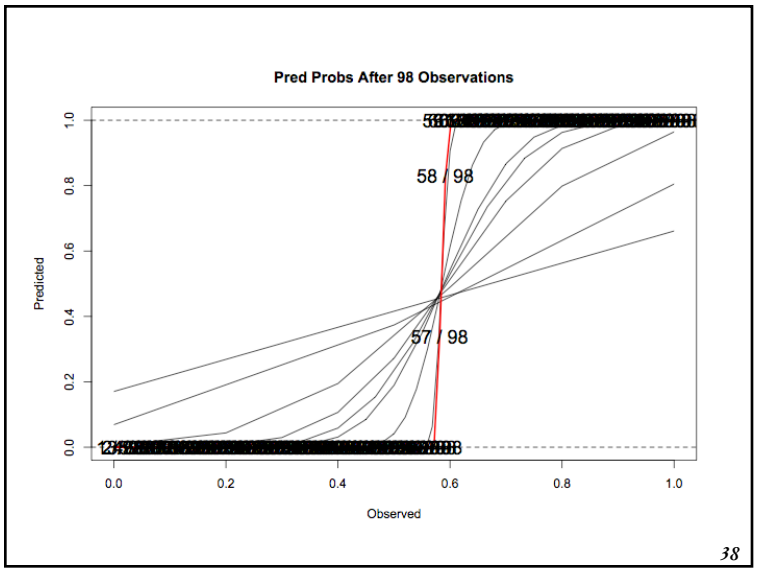
35


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
36

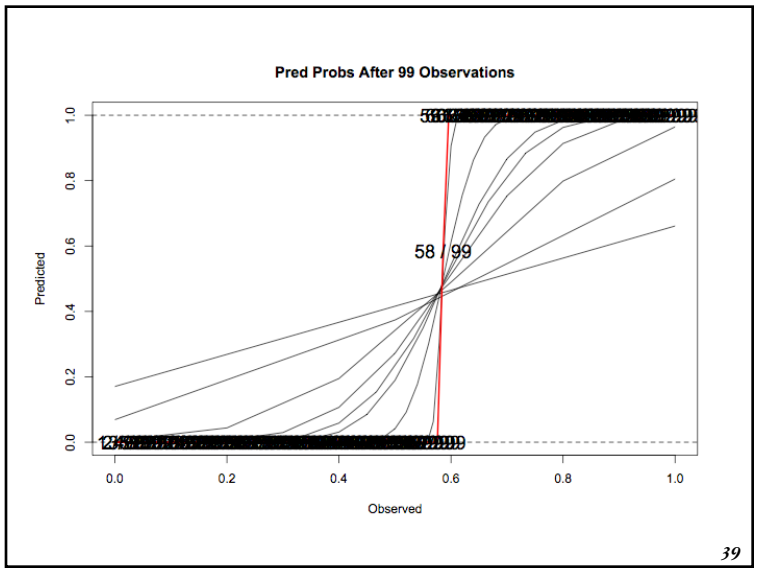
36



37



38



39

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

40

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

41

41

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

42

42

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

43

43

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?

44

44

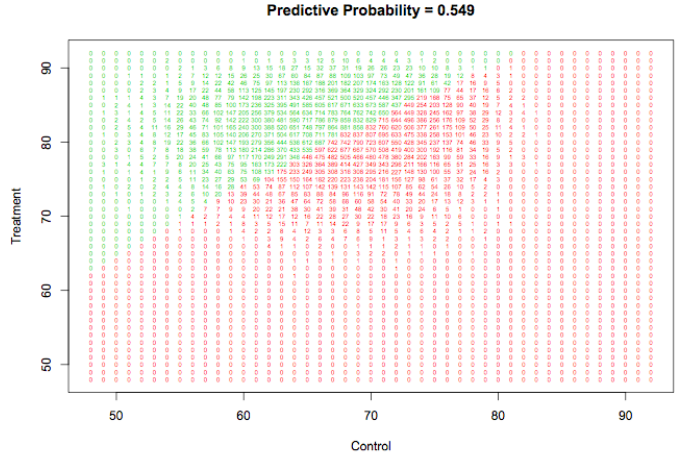
Predictive Probability

```
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
```

45

45

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



46

46

Example: Phase 2 Trials

47

47

Phase 2 Trials

- Early phase results
  - Animal studies showed promise for disease
  - Phase 1 showed non-toxic in healthy humans
- Questions for Phase 2
  - Does the treatment work in humans
  - Which dose is best
  - Which dose(s) to take to Phase 3
  - Is an dose with promising efficacy also safe
  - What is likelihood of Phase 3 success

48

48

Adaptive Randomization Strategies

- Bandits
- Play the Winner
- Randomized Play the Winner
- Randomize ~ Pr(Best Treatment)
- Randomize ~ *f*(Pr(Best Treatment))
- Randomize ~ Dose that gives the most information
- One of these with constraints

49

49

Adaptation

- Multiple trial characteristics may be changed during the course of the trial based on accumulating data
- Must pre-prescribe changes
  - Available Doses
  - Randomization proportions
  - Time of interim analyses
  - Maximum sample size
  - Dose dropping rules
  - Allow doses to re-enter?

50

50

Example In Uterine Cancer

- Phase 2 dose finding trial
- 3-armed RCT
  - Control chemotherapy
  - Control + experimental treatment q2w
  - Control + experimental treatment q1w
- Goals
  - Treat patients effectively & ethically
  - Learn about experimental treatment
  - Explore adaptive designs
    - This company's first attempt at an adaptive design

51

51

Trial Setup

- Primary Outcome
  - Progression Free Survival (PFS)
  - λ_c = Rate of PFS in Control population
  - λ_2 = Rate of PFS in Control + q2w population
  - λ_1 = Rate of PFS in Control + q1w population
- Expectation
  - Control mean PFS = 303 days, median = 210
  - Accrual
    - 1 patient every 3 days for first 45 pts (135 days)
    - 1 patient every 2 days thereafter
- Need to beat control by 10% to be marketable

52

52

Factors to Consider

- Statistical Model
  - Parametric dose-response curve, non-parametric, independent arms
  - Historical vs. vague priors
- How many doses
- Maximum sample size
- Timing of first interim analysis
- Timing of subsequent interim analyses
  - Time based or patient based
- Randomization scheme
- Rules to drop doses
- Rules to allow doses to re-enter
- Rules to stop for futility
- Rules to stop for success
- How long to track patients after last patient enrolled

53

53

Statistical Model

- Assume time-to-progression exponential
- Priors on rates:

  $\lambda_c, \lambda_2, \lambda_1 \sim \Gamma(1, 303 \text{ days})$
- Posteriors

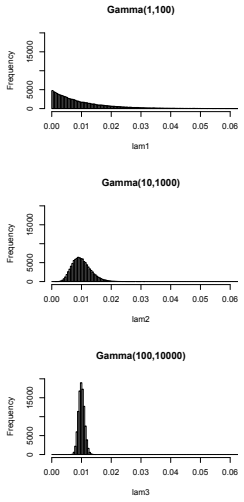
  $\lambda_d \mid \text{Data} \sim \Gamma(1+ \#\text{ Progressors}, 303+\text{Exposure Time})$
- Also calculate probability each dose is best
  - "best" = has lowest PFS rate
  - $p_c = \Pr(\lambda_c<\lambda_2 \,\&\, \lambda_c<\lambda_1)$
  - $p_2 = \Pr(\lambda_2<\lambda_c \,\&\, \lambda_2<\lambda_1)$
  - $p_1 = \Pr(\lambda_1<\lambda_c \,\&\, \lambda_1<\lambda_2)$

54

54

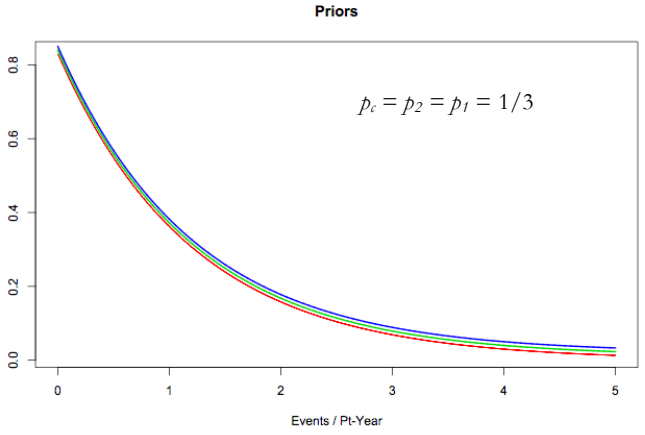
Example

```
> lam1 <- rgamma(100000,   1,   100)
> lam2 <- rgamma(100000,  10,  1000)
> lam3 <- rgamma(100000, 100, 10000)
> par(mfrow=c(3,1))

> mean(lam1 < lam2 & lam1 < lam3)
[1] 0.5738
> mean(lam2 < lam1 & lam2 < lam3)
[1] 0.24854
> mean(lam3 < lam1 & lam2 > lam3)
[1] 0.17766

>hist(lam1,breaks=seq(0,.12, length=250),
          xlim=c(0,.06), ylim=c(0,20000),
          main="Gamma(1,100)")
> hist(lam2,breaks=seq(0,.12, length=250),
          xlim=c(0,.06), ylim=c(0,20000),
          main="Gamma(10,1000)")
> hist(lam3,breaks=seq(0,.12, length=250),
          xlim=c(0,.06), ylim=c(0,20000),
          main="Gamma(100,10000)")
```



55

55



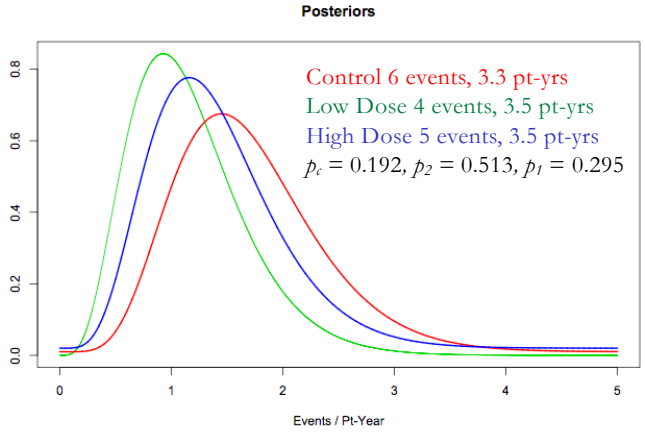
56

56

Statistical Summary

- Primary Outcome: Progression free survival
  - λ_t = PFS rate for Treatment t; t = A, B, C
- Statistical Assumptions and Modeling
  - PFS distributed $y_{i,t} \sim \text{Exp}(\lambda_t)$; t = A, B, C
  - Priors: $\lambda_A, \lambda_B, \lambda_C \sim \Gamma(1, 303)$
    Equals 1 subject with mean 303 days
    median = 210 days
    Median = Mean × log(2) for gamma dist
  - Posteriors:
    $\lambda_t \mid data \sim \Gamma(1 + \#\text{Events}_t, 303 + \text{Exposure}_t)$

57



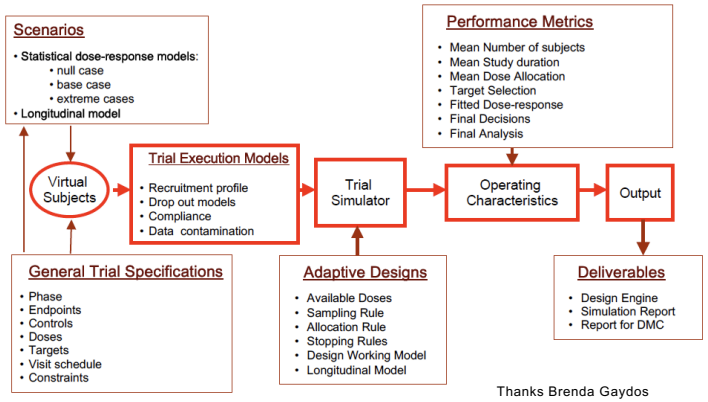
58

Complication I'll ignore

- In fact there were 2 types of patients
  platinum sensitive & platinum refractory
- Expect mean TTP shorter for refractory
  TTP in refractory = 2/7 that of sensitive
- Model event rates as $\gamma\lambda_d$ for refractory
  assume γ same across groups
- Prior on $\log(\gamma)$ ~$N(0,10^2)$
- Means we no longer have conjugate priors
  must use Metropolis-Hastings algorithm

59

Simulation Plan



60

Maximum Sample Size

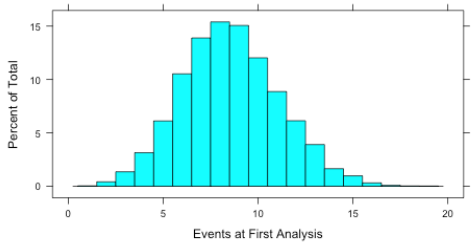
- Oftentimes determined by company resources
- Considered 150 & 195

61

61

Timing of interim analyses

- Expected accrual rate = 3 days per patient
  - 45 patients take 135 days
  - With expected $\lambda_c = 1/303$
  - Expect 8.5 events by 135 days
  - Median 8, IQR 7-10



62

62

Randomization

- Randomize first 45 patients 15:15:15
- Start interim analysis after 45th patient enrolled
- Repeat interim analyses every 15 patients
  - Approximately every 1 month with expected accrual
  - This timing worked logistically
  - Allowed blocks of 15 to ensure patients on each dose
- Open question: How to randomize?

63

63

Randomization Options

- Let r_d = randomization probability to dose d
- Let p_d = probability arm d has lowest (best) progression rate
- Randomization weighting by C

$$r_d = \frac{p_d^C}{p_1^C + p_2^C + p_3^C + \ldots + p_D^C}$$

64

64

Randomization Options

$$r_d = \frac{p_d^C}{p_1^C + p_2^C + p_3^C + \ldots + p_D^C}$$

- $C = 0$, equal randomization ($r_d = 1/\text{Number of Groups}$)
- $C = 1$, proportional to probability best ($r_d = p_d$)
- $C \geq 1$
  - strongly favor 1 arm earlier in the trial, even when treatments are equal
  - more subjects likely assigned to the best treatment
  - C → big means assign all to best treatment, play the leader
- $0 < C < 1$
  - weakly favor better
  - fewer subjects likely assigned to best treatment
  - more even distribution early in trials
  - randomization less affected by early events
- $C = n/N$, trial begins with c = 0 and ends with c = 1

65

65

Rules to Stop

- When to Stop for Success?
  - If $p_2 > 0.95$, stop for success
  - If $p_1 > 0.95$, stop for success
  - Take successful dose to Phase III
- What if experimental doses equally effective?

66

66

Rules to Stop

- When to Stop for Success?
  - If $p_2 > 0.95$, stop for success
  - If $p_1 > 0.95$, stop for success
  - Take successful dose to Phase III
- What if experimental doses equally effective?
- Instead use if $p_C < 0.10$ or 0.05 to success stop?

67

67

Rules to Stop

- When to Stop for Futility?
  - If $p_2 < 0.05$ drop q2w arm
  - If $p_1 < 0.05$ drop q1w arm
  - If both arms dropped, trial ends
  - Allow dropped arms to re-enter?

68

68

Rules to Stop

- When to Stop for Futility?
  - If $p_2 < 0.05$ drop q2w arm
  - If $p_1 < 0.05$ drop q1w arm
  - If $Pr(\lambda_e / \lambda_2 > 1.10 \mid \text{Data}) < 0.05$ drop q2w arm
  - If $Pr(\lambda_e / \lambda_1 > 1.10 \mid \text{Data}) < 0.05$ drop q1w arm
  - If both arms dropped, trial ends
  - Allow dropped arms to re-enter? Yes

69

69

Post Accrual Tracking

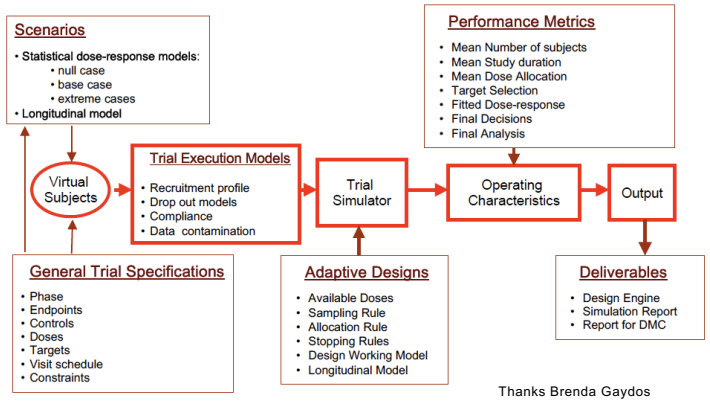
- Choose to track patients for 1-year post accrual
- 70% chance last patient will have event

  $1\text{-}e^{-365/303} = 0.70$
- Under assumed accrual rates & N=195, 83% of patients will have events if $\lambda = 1/303$.

70

70

Simulation Plan



Thanks Brenda Gaydos

71

71

At each interim analysis

1. Calculate:

   Posteriors $\lambda_t \mid data$; $t \in$ A, B, C

   p_t = P(Treatment t is 'Best' treatment | *data*)

   e.g. $p_B = P(\lambda_B \leq \lambda_A \,\&\, \lambda_C \mid \text{data})$

   P(Treatment *t* is ≥10% better than A | *data*)

2. Check superiority and futility stopping/dropping rules
3. Randomize next 15 subjects with probability p_t
4. Repeat steps 1-4 up to 195 subjects

72

Simulation Output

```
 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.168
Stop for Futility    0.004
     Stop for Cap    0.828

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        30.4850   0.214 10.8927   0.003   0.000   0.000
Control + q2w 55.8790   0.392 19.5526   0.492   0.059   0.682
Control + q1w 56.2410   0.394 19.0859   0.505   0.057   0.690

     Total N =     142.605       SD=  20.247
   Pr(Either Beats Placebo) =    0.853
```

73

73

Max N = 150

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.049
Stop for Futility    0.073
     Stop for Cap    0.878

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        47.6570   0.334 18.0027   0.342   0.000   0.000
Control + q2w 47.0040   0.330 19.2468   0.310   0.006   0.051
Control + q1w 47.9440   0.336 19.3273   0.348   0.008   0.052

     Total N =     142.605       SD=  22.106
   Pr(Either Beats Placebo) =    0.081

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.168
Stop for Futility    0.004
     Stop for Cap    0.828

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        30.4850   0.214 10.8927   0.003   0.000   0.000
Control + q2w 55.8790   0.392 19.5526   0.492   0.059   0.682
Control + q1w 56.2410   0.394 19.0859   0.505   0.057   0.690

     Total N =     142.605       SD=  20.247
   Pr(Either Beats Placebo) =    0.853
```

Max N = 195

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.070
Stop for Futility    0.103
     Stop for Cap    0.827

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        60.3100   0.333 25.4370   0.331   0.000   0.000
Control + q2w 60.9040   0.336 28.1304   0.346   0.009   0.063
Control + q1w 59.9710   0.331 27.7830   0.323   0.006   0.061

     Total N =     181.185       SD=  35.625
   Pr(Either Beats Placebo) =    0.102

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.208
Stop for Futility    0.002
     Stop for Cap    0.790

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        35.1840   0.195 13.7992   0.001   0.000   0.000
Control + q2w 72.1780   0.400 27.5021   0.491   0.047   0.757
Control + q1w 72.9830   0.405 27.1835   0.508   0.053   0.766

     Total N =     180.345       SD=  33.923
   Pr(Either Beats Placebo) =    0.907
```

74

74

Max N = 195, Firstlook=45

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.070
Stop for Futility    0.103
     Stop for Cap    0.827

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        60.3100   0.333 25.4370   0.331   0.000   0.000
Control + q2w 60.9040   0.336 28.1304   0.346   0.009   0.063
Control + q1w 59.9710   0.331 27.7830   0.323   0.006   0.061

     Total N =     181.185       SD=  35.625
   Pr(Either Beats Placebo) =    0.102

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.208
Stop for Futility    0.002
     Stop for Cap    0.790

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        35.1840   0.195 13.7992   0.001   0.000   0.000
Control + q2w 72.1780   0.400 27.5021   0.491   0.047   0.757
Control + q1w 72.9830   0.405 27.1835   0.508   0.053   0.766

     Total N =     180.345       SD=  33.923
   Pr(Either Beats Placebo) =    0.907
```

Max N = 195, Firstlook=90

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.057
Stop for Futility    0.089
     Stop for Cap    0.854

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        61.4750   0.332 19.4908   0.348   0.000   0.000
Control + q2w 62.2340   0.336 21.2199   0.322   0.005   0.042
Control + q1w 61.6460   0.333 21.2751   0.330   0.006   0.041

     Total N =     185.355       SD=  27.081
   Pr(Either Beats Placebo) =    0.071

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.199
Stop for Futility    0.000
     Stop for Cap    0.801

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        41.0450   0.224  9.0906   0.001   0.000   0.000
Control + q2w 70.8100   0.387 20.6464   0.499   0.044   0.806
Control + q1w 71.1900   0.389 20.7805   0.500   0.046   0.809

     Total N =     183.045       SD=  28.766
   Pr(Either Beats Placebo) =    0.931
```

75

75

Max N = 195, c = 1

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.070
Stop for Futility    0.103
     Stop for Cap    0.827

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        60.3100   0.333 25.4370   0.331   0.000   0.000
Control + q2w 60.9040   0.336 28.1304   0.346   0.009   0.063
Control + q1w 59.9710   0.331 27.7830   0.323   0.006   0.061

     Total N =     181.185       SD=  35.625
   Pr(Either Beats Placebo) =    0.102

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.208
Stop for Futility    0.002
     Stop for Cap    0.790

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        35.1840   0.195 13.7992   0.001   0.000   0.000
Control + q2w 72.1780   0.400 27.5021   0.491   0.047   0.757
Control + q1w 72.9830   0.405 27.1835   0.508   0.053   0.766

     Total N =     180.345       SD=  33.923
   Pr(Either Beats Placebo) =    0.907
```

Max N = 195, c = 0

```
 Doing Case =            1
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  303.00
Control + q1w    Mean TTP =  303.00


 Stop for Success    0.063
Stop for Futility    0.118
     Stop for Cap    0.819

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        60.0350   0.333 12.3501   0.352   0.000   0.000
Control + q2w 60.0350   0.333 12.3501   0.331   0.009   0.044
Control + q1w 60.0350   0.333 12.3501   0.317   0.008   0.048

     Total N =     180.105       SD=  37.050
   Pr(Either Beats Placebo) =    0.083

 ----------------------------------------------


 Doing Case =            9
Control          Mean TTP =  303.00
Control + q2w    Mean TTP =  606.00
Control + q1w    Mean TTP =  606.00


 Stop for Success    0.195
Stop for Futility    0.004
     Stop for Cap    0.801

          Name  Mean N     % N     SD N    Best     Win Beat P
Control        60.3950   0.333 11.0779   0.003   0.000   0.000
Control + q2w 60.3950   0.333 11.0779   0.488   0.046   0.828
Control + q1w 60.3950   0.333 11.0779   0.509   0.047   0.828

     Total N =     181.185       SD=  33.234
   Pr(Either Beats Placebo) =    0.931
```

76

76

Max N = 195, c = 1

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.070
Stop for Futility 0.103
Stop for Cap 0.827

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	60.3100	0.333	25.4370	0.331	0.000	0.000
Control + q2w	60.9040	0.336	28.1304	0.346	0.009	0.063
Control + q1w	59.9710	0.331	27.7830	0.323	0.006	0.061

Total N = 181.185 SD= 35.625
Pr(Either Beats Placebo) = 0.102

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.208
Stop for Futility 0.002
Stop for Cap 0.790

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	35.1840	0.195	13.7992	0.001	0.000	0.000
Control + q2w	72.1780	0.400	27.5021	0.491	0.047	0.757
Control + q1w	72.9830	0.405	27.1835	0.508	0.053	0.766

Total N = 180.345 SD= 33.923
Pr(Either Beats Placebo) = 0.907

Max N = 195, c = ∞

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.047
Stop for Futility 0.092
Stop for Cap 0.861

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	60.4500	0.330	43.6835	0.347	0.000	0.000
Control + q2w	61.6800	0.336	45.8555	0.339	0.009	0.061
Control + q1w	61.2900	0.334	45.4790	0.314	0.002	0.057

Total N = 183.420 SD= 32.733
Pr(Either Beats Placebo) = 0.092

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.201
Stop for Futility 0.003
Stop for Cap 0.796

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	24.1950	0.134	18.5007	0.004	0.000	0.000
Control + q2w	78.3450	0.435	51.8603	0.498	0.049	0.570
Control + q1w	77.7000	0.431	50.7603	0.498	0.043	0.561

Total N = 180.240 SD= 34.519
Pr(Either Beats Placebo) = 0.772

77

Max N = 195, c = ∞

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.047
Stop for Futility 0.092
Stop for Cap 0.861

Name	Mean N	%N	SD N	Best	Win	Beat P
Control	60.4500	0.330	43.6835	0.347	0.000	0.000
Control + q2w	61.6800	0.336	45.8555	0.339	0.009	0.061
Control + q1w	61.2900	0.334	45.4790	0.314	0.002	0.057

Total N = 183.420 SD= 32.733
Pr(Either Beats Placebo) = 0.092

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.201
Stop for Futility 0.003
Stop for Cap 0.796

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	24.1950	0.134	18.5007	0.004	0.000	0.000
Control + q2w	78.3450	0.435	51.8603	0.498	0.049	0.570
Control + q1w	77.7000	0.431	50.7603	0.498	0.043	0.561

Total N = 180.240 SD= 34.519
Pr(Either Beats Placebo) = 0.772

Max N = 195, c = ∞, every 1

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.099
Stop for Futility 0.120
Stop for Cap 0.781

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	55.6170	0.319	40.6723	0.311	0.000	0.000
Control + q2w	61.1370	0.350	45.0447	0.352	0.006	0.047
Control + q1w	57.8350	0.331	44.5945	0.337	0.006	0.049

Total N = 174.589 SD= 44.094
Pr(Either Beats Placebo) = 0.081

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.263
Stop for Futility 0.004
Stop for Cap 0.733

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	23.5280	0.136	17.2205	0.004	0.000	0.000
Control + q2w	75.4290	0.435	49.9018	0.514	0.043	0.582
Control + q1w	74.5200	0.430	50.4509	0.482	0.046	0.581

Total N = 173.477 SD= 42.012
Pr(Either Beats Placebo) = 0.770

78

Max N = 195, c = 1

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.070
Stop for Futility 0.103
Stop for Cap 0.827

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	60.3100	0.333	25.4370	0.331	0.000	0.000
Control + q2w	60.9040	0.336	28.1304	0.346	0.009	0.063
Control + q1w	59.9710	0.331	27.7830	0.323	0.006	0.061

Total N = 181.185 SD= 35.625
Pr(Either Beats Placebo) = 0.102

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.208
Stop for Futility 0.002
Stop for Cap 0.790

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	35.1840	0.195	13.7992	0.001	0.000	0.000
Control + q2w	72.1780	0.400	27.5021	0.491	0.047	0.757
Control + q1w	72.9830	0.405	27.1835	0.508	0.053	0.766

Total N = 180.345 SD= 33.923
Pr(Either Beats Placebo) = 0.907

Max N = 195, c = n/N

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.070
Stop for Futility 0.106
Stop for Cap 0.824

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	61.3110	0.340	19.6030	0.335	0.000	0.000
Control + q2w	59.4440	0.330	22.8840	0.344	0.006	0.048
Control + q1w	59.6200	0.331	22.5230	0.321	0.007	0.049

Total N = 180.375 SD= 36.095
Pr(Either Beats Placebo) = 0.083

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.212
Stop for Futility 0.001
Stop for Cap 0.787

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	40.8990	0.226	12.3915	0.000	0.000	0.000
Control + q2w	70.4020	0.389	21.1026	0.523	0.055	0.810
Control + q1w	69.4940	0.384	20.5548	0.477	0.063	0.804

Total N = 180.795 SD= 33.749
Pr(Either Beats Placebo) = 0.937

79

Max N = 195, c = 1

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.070
Stop for Futility 0.103
Stop for Cap 0.827

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	60.3100	0.333	25.4370	0.331	0.000	0.000
Control + q2w	60.9040	0.336	28.1304	0.346	0.009	0.063
Control + q1w	59.9710	0.331	27.7830	0.323	0.006	0.061

Total N = 181.185 SD= 35.625
Pr(Either Beats Placebo) = 0.102

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.208
Stop for Futility 0.002
Stop for Cap 0.790

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	35.1840	0.195	13.7992	0.001	0.000	0.000
Control + q2w	72.1780	0.400	27.5021	0.491	0.047	0.757
Control + q1w	72.9830	0.405	27.1835	0.508	0.053	0.766

Total N = 180.345 SD= 33.923
Pr(Either Beats Placebo) = 0.907

Max N = 195, c = n/N

Doing Case = 1

Control Mean TTP = 303.00
Control + q2w Mean TTP = 303.00
Control + q1w Mean TTP = 303.00

Stop for Success 0.070
Stop for Futility 0.106
Stop for Cap 0.824

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	61.3110	0.340	19.6030	0.335	0.000	0.000
Control + q2w	59.4440	0.330	22.8840	0.344	0.006	0.048
Control + q1w	59.6200	0.331	22.5230	0.321	0.007	0.049

Total N = 180.375 SD= 36.095
Pr(Either Beats Placebo) = 0.083

Doing Case = 9

Control Mean TTP = 303.00
Control + q2w Mean TTP = 606.00
Control + q1w Mean TTP = 606.00

Stop for Success 0.212
Stop for Futility 0.001
Stop for Cap 0.787

Fixed was 60

Name	Mean N	% N	SD N	Best	Win	Beat P
Control	40.8990	0.226	12.3915	0.000	0.000	0.000
Control + q2w	70.4020	0.389	21.1026	0.523	0.055	0.810
Control + q1w	69.4940	0.384	20.5548	0.477	0.063	0.804

Total N = 180.795 SD= 33.749
Pr(Either Beats Placebo) = 0.937

80

Design Parameters

- First look @ 45
- Interim analyses every 15 patients
- Maximum = 195 patients
- Success
  - If $P_2 > 0.95$, stop for success
  - If $P_1 > 0.95$, stop for success
  - Take successful dose to Phase III
- Futility
  - If $Pr(\lambda_c / \lambda_2 > 1.10 \mid \text{Data}) < 0.05$ drop q2w arm
  - If $Pr(\lambda_c / \lambda_1 > 1.10 \mid \text{Data}) < 0.05$ drop q1w arm
  - If both arms dropped, trial ends

81

81

Show Individual Trials

- Best way to illustrate the adaptive design is to show example trials to collaborators

82

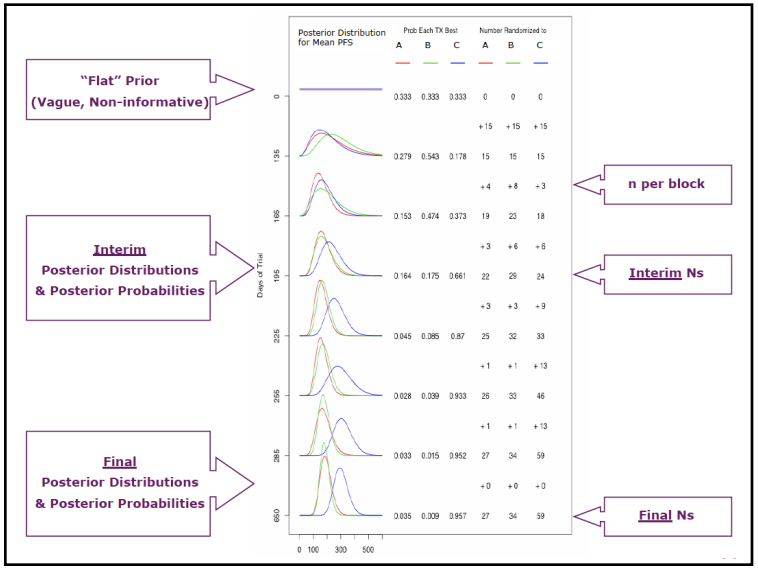
82

Show Individual Trials

- Best way to illustrate the adaptive design is to show example trials to collaborators
- GREAT for debugging!

83

83



84

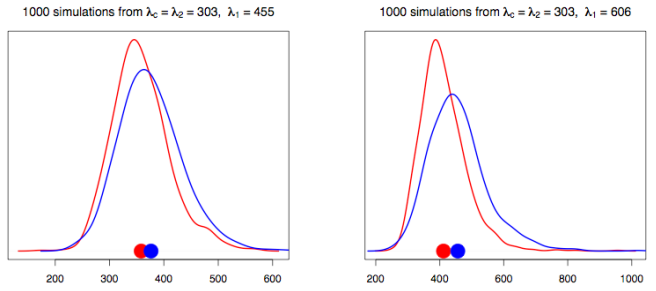
Output I Shared (Make it prettier)

Treatment	Mean PFS	%Δ	Mean N	SD	Pr(Best)	Pr(Win)	Pr(Beat Control)
Control	303		59.7	25.3	0.343	0.000	-----
+q2w	303	No Δ	59.7	28.4	0.322	0.007	0.054
+q1w	303	No Δ	60.0	28.5	0.335	0.008	0.053
Fully Adaptive Trial			179.4	38.7	Pr(Stop for Success) = 0.071 Pr(Stop for Futility) = 0.117 Pr(Stop for Max N) = 0.813 Pr(Either Beats Control) = 0.090		

Treatment	Mean PFS	%Δ	Mean N	SD	Pr(Best)	Pr(Win)	Pr(Beat Control)
Control	303		34.0	14.2	0.001	0.000	-----
+q2w	455	+50%	56.9	27.0	0.099	0.002	0.462
+q1w	606	+100%	79.4	28.6	0.900	0.351	0.881
Fully Adaptive Trial			170.3	43.2	Pr(Stop for Success) = 0.345 Pr(Stop for Futility) = 0.004 Pr(Stop for Max N) = 0.650 Pr(Either Beats Control) = 0.907		

85

85



C=1	C=0		C=1	C=0
141.0	140.4	N	135.3	135.5
40.4	40.3	Pr(q1w Beat Placebo)	75.2	75.2
47.2	33.0	% to Best Dose	51.7	33.0
374	354	Overall Mean PFS	460	404
60.3	39.7	Pr(Avg. Patient Lives Longer)	64.3	35.7

86

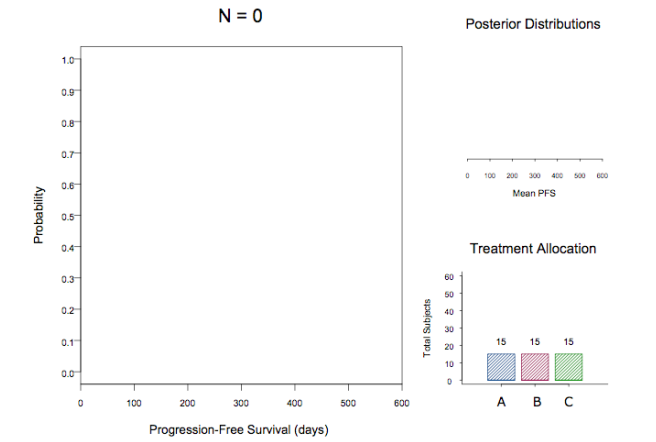
86

The trial is over!

This is how it really went.

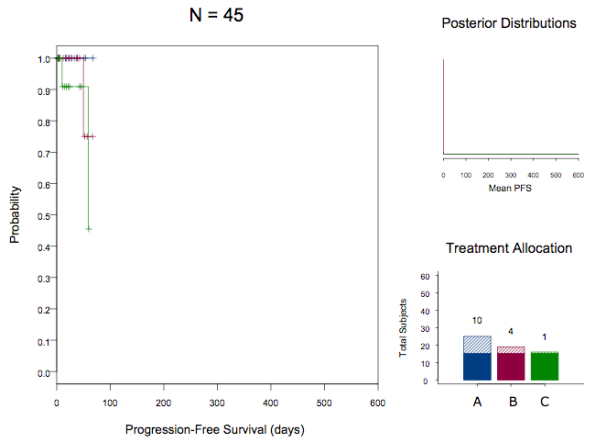
87

87



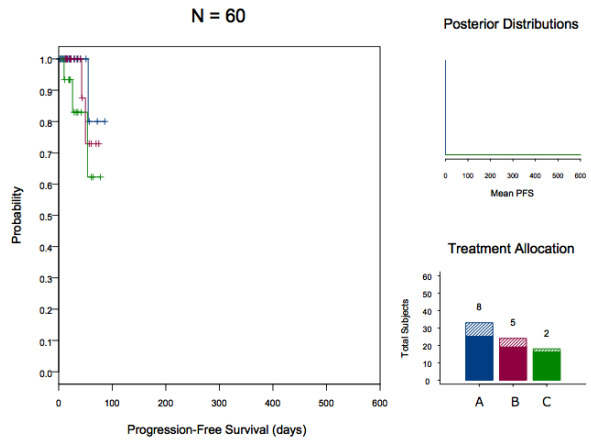
88

Expected @ Day 135; Actual Day 67



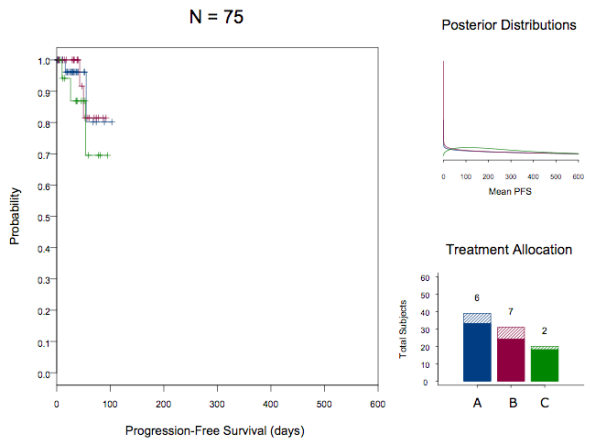
89

Expected @ Day 165; Day 87

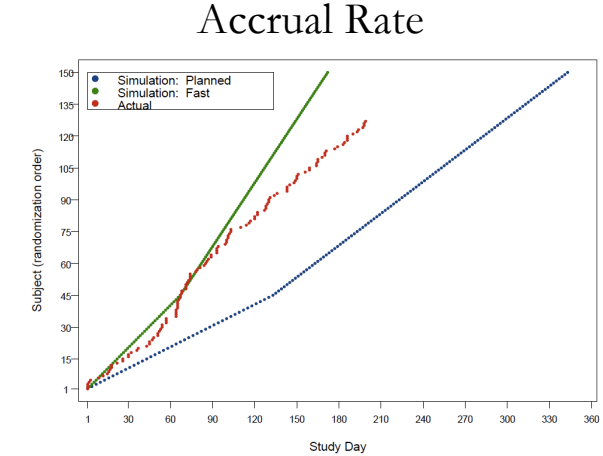


90

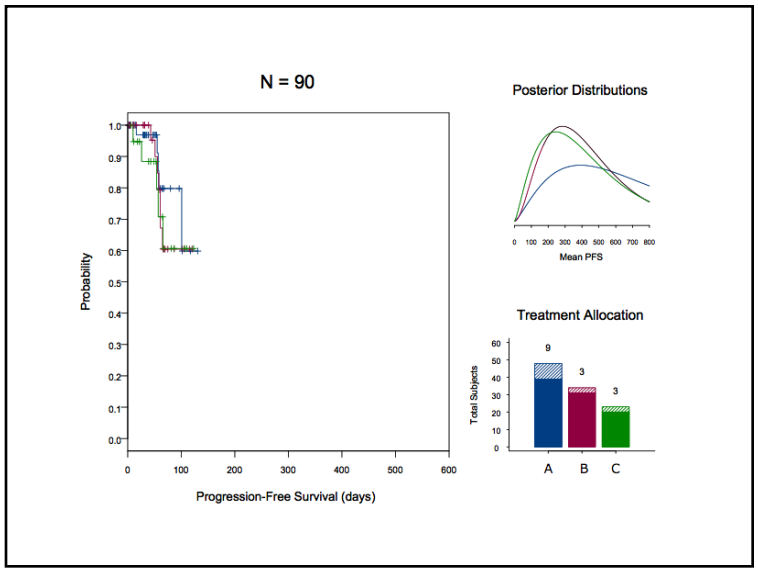
Expected @ Day 195; Day 106



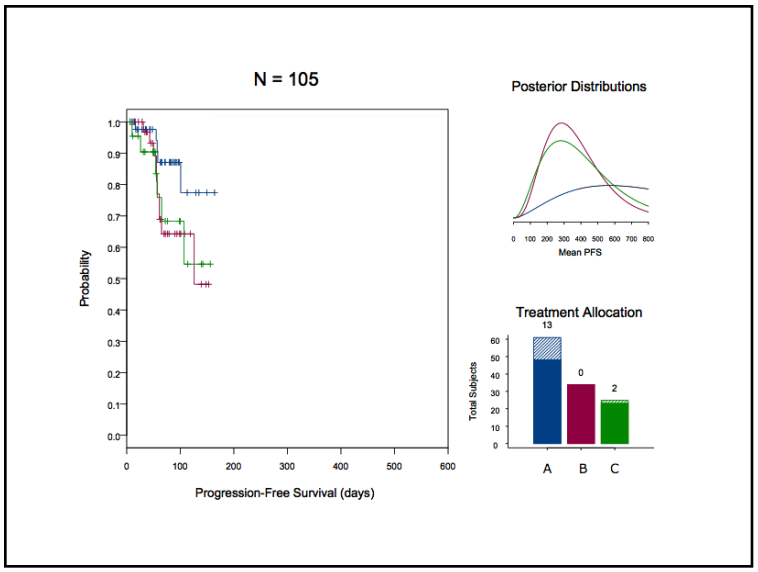
91



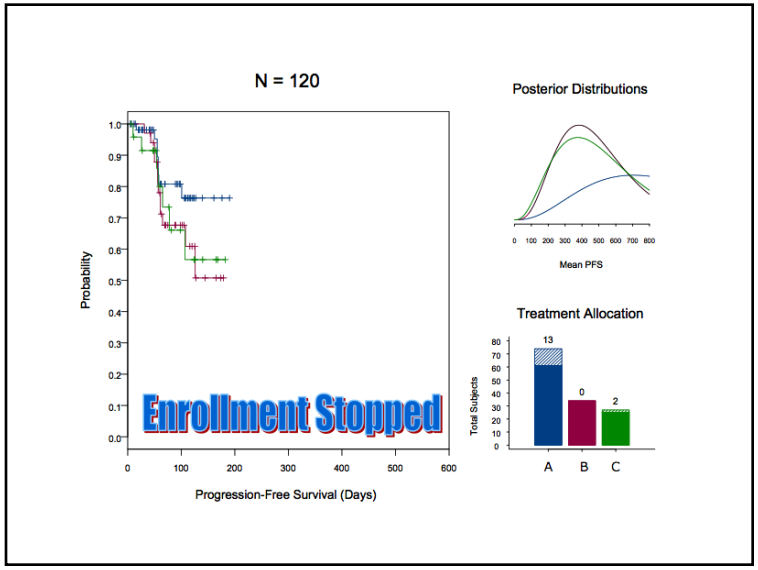
92


N = 90
Posterior Distributions
Probability
Progression-Free Survival (days)
Mean PFS
Treatment Allocation
Total Subjects
9
3
3
A
B
C

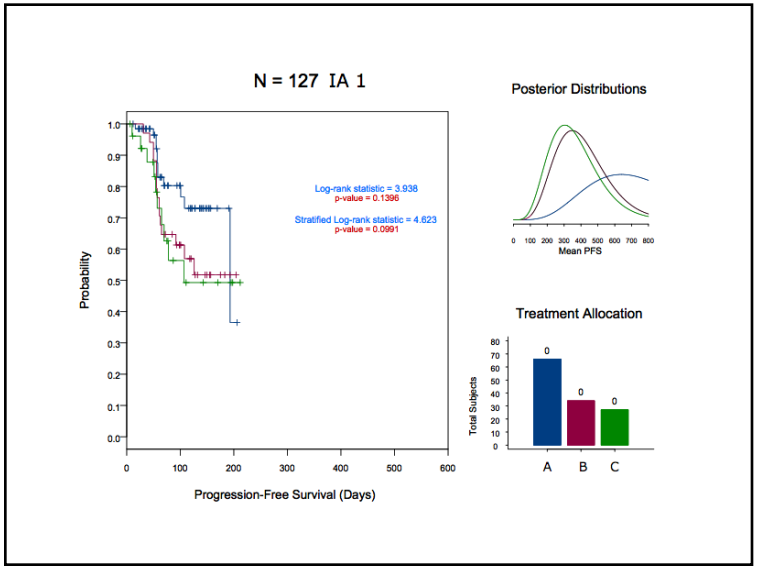
93


N = 105
Posterior Distributions
Probability
Progression-Free Survival (days)
Mean PFS
Treatment Allocation
Total Subjects
13
0
2
A
B
C

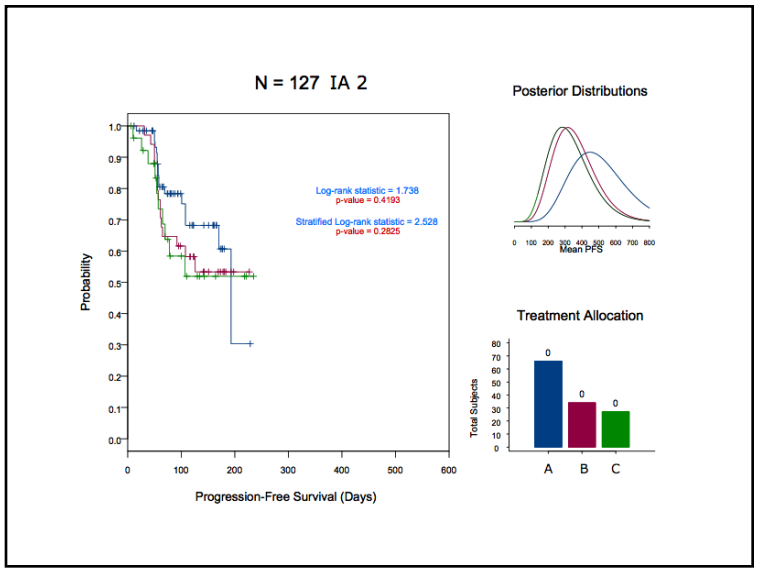
94


N = 120
Posterior Distributions
Probability
Enrollment Stopped
Progression-Free Survival (Days)
Mean PFS
Treatment Allocation
Total Subjects
13
0
2
A
B
C

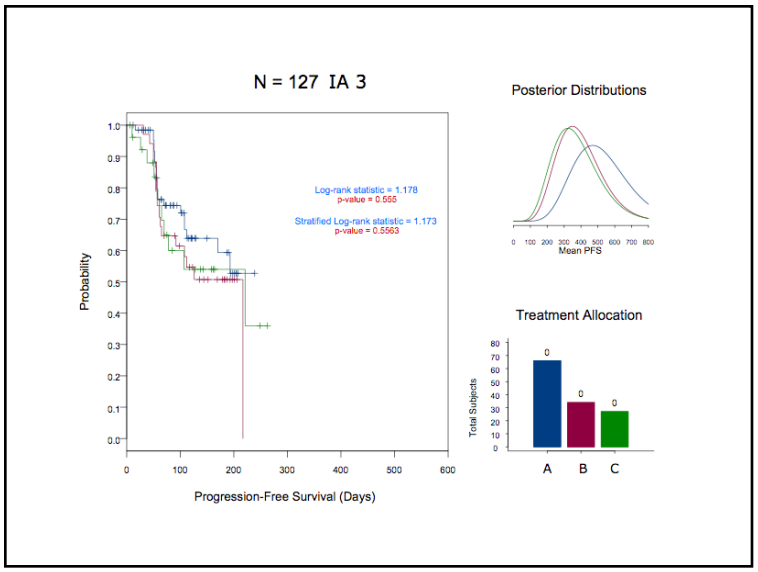
95


N = 127 IA 1
Posterior Distributions
Probability
Log-rank statistic = 3.938
p-value = 0.1396
Stratified Log-rank statistic = 4.623
p-value = 0.0991
Progression-Free Survival (Days)
Mean PFS
Treatment Allocation
Total Subjects
0
0
0
A
B
C

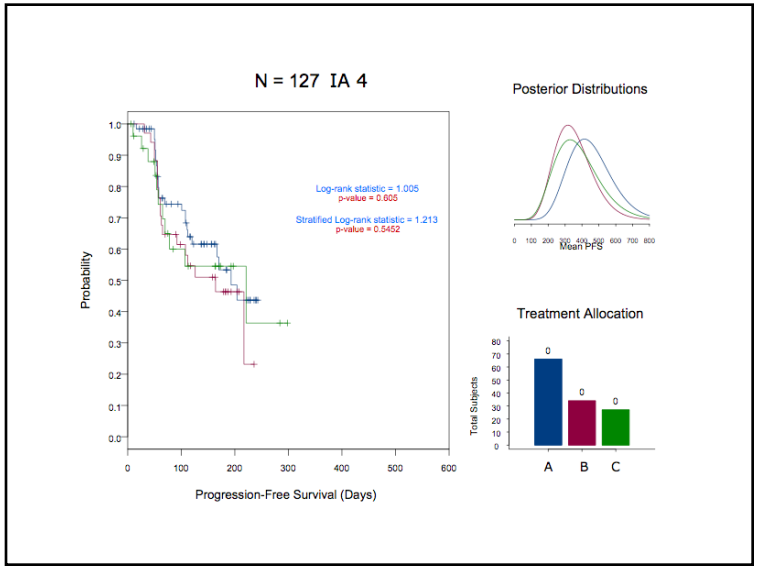
96


N = 127 IA 2
Log-rank statistic = 1.738
p-value = 0.4193
Stratified Log-rank statistic = 2.528
p-value = 0.2825
Probability
Progression-Free Survival (Days)
Posterior Distributions
Mean PFS
Treatment Allocation
Total Subjects
A B C

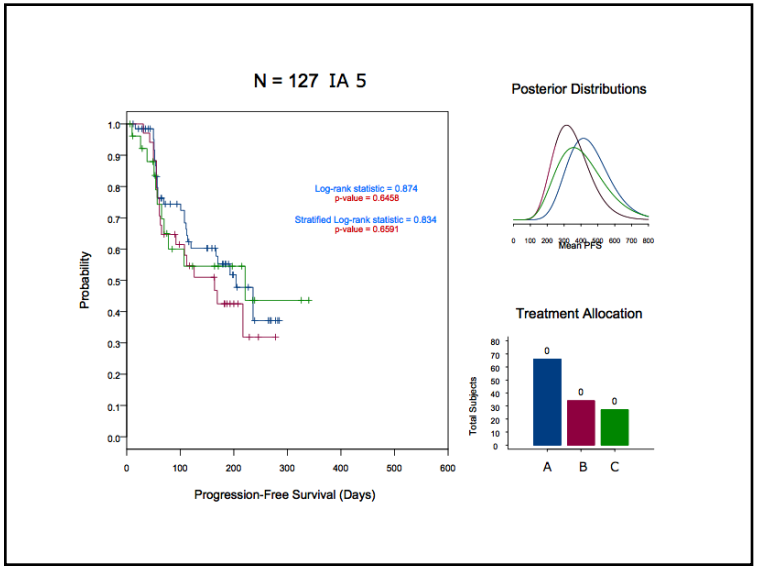
97


N = 127 IA 3
Log-rank statistic = 1.178
p-value = 0.555
Stratified Log-rank statistic = 1.173
p-value = 0.5563
Probability
Progression-Free Survival (Days)
Posterior Distributions
Mean PFS
Treatment Allocation
Total Subjects
A B C

98


N = 127 IA 4
Log-rank statistic = 1.005
p-value = 0.605
Stratified Log-rank statistic = 1.213
p-value = 0.5452
Probability
Progression-Free Survival (Days)
Posterior Distributions
Mean PFS
Treatment Allocation
Total Subjects
A B C

99


N = 127 IA 5
Log-rank statistic = 0.874
p-value = 0.6456
Stratified Log-rank statistic = 0.834
p-value = 0.6591
Probability
Progression-Free Survival (Days)
Posterior Distributions
Mean PFS
Treatment Allocation
Total Subjects
A B C

100

Lessons Learned

$$R_j = \frac{p_j^c}{p_1^c + p_2^c + p_3^c + ... + p_G^c}$$

R_j : randomization probability of treatment j

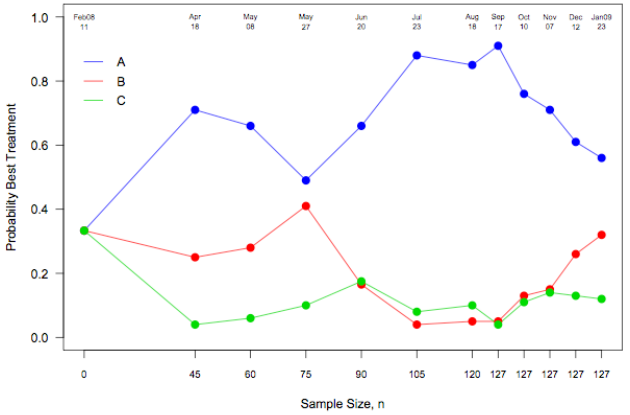
p_j: posterior probability treatment j is the best treatment.

- c = 0, equal randomization ($R_j = 1/G$)
- c = 1, proportional randomization ($R_j = p_j$)
- c ≥ 1
  - more likely to favor 1 arm earlier in the trial, even when treatments are equal
  - more subjects likely assigned to the best treatment
- c < 1
  - randomization less likely to favor one arm earlier in the trial
  - fewer subjects likely assigned to best treatment
- c = n/N, trial begins with c = 0 and ends with c = 1

101

101

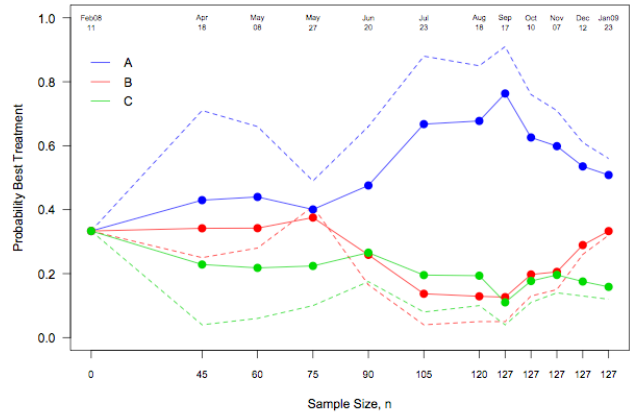
Randomization Assignments



102

102

Randomization using c = n/N



103

103

Summary

104

104

Example: A Prospective Bayesian Adaptive Trial with Hierarchical Borrowing from a Prior Single Arm Study

With Kristine Broglio

10

105

EXCITE Trial Background

- Patients with peripheral artery disease and in-stent restenosis
- Randomized trial of
  - Control: Balloon angioplasty
  - Treatment: Laser ablation
- Primary Efficacy: Freedom from target lesion revascularization at 6 months
- Primary Safety: Freedom from major adverse events at 30 days

10

106

Original Study Design

- Sponsor seeks 510(k) approval
- Maximum of 318 subjects
- Hypotheses:
  - Efficacy superiority (2.5% Type I error)
  - Safety noninferiority 10% margin (5% Type I error)
- OBF interim analysis at 33% information

10

107

Adjunct Analysis

- Randomized trial had slow enrollment
- PATENT: A single arm trial of the laser ablation in Europe completed
- Sponsor asks: can we use the single arm trial to supplement the randomized trial?

10

108

PATENT Trial

- One arm trial
- Efficacy
  - 80 evaluable patients
  - 79% success rate (63/80)
- Safety
  - 90 evaluable patients
  - 4.4% event rate (4/90)

10

109

Hierarchical Borrowing

- Define p_0 as the proportion successes in EXCITE and p_1 as the proportion successes in PATENT
- Model the log-odds of success

$$\gamma_i = \log\left(\frac{p_i}{1-p_i}\right) \; i = 0,1$$

- Assume $\gamma_i \sim N(\mu, \tau^2)$
- Place hyperpriors on μ and τ^2

11

110

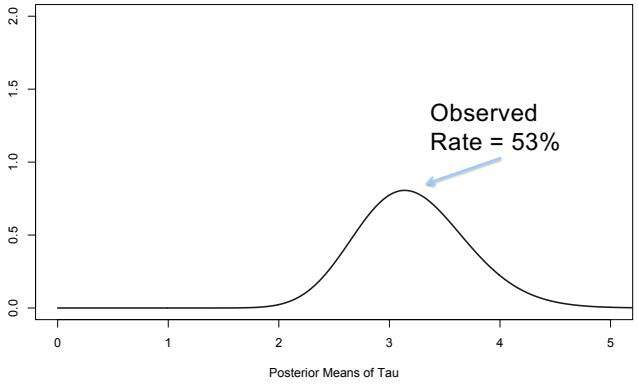
Hierarchical Borrowing

- τ^2 is between study variability
  - $\tau^2 = 0$ corresponds to $\gamma_0 = \gamma_1$ or simple pooling
  - τ^2 = gigantic corresponds to no borrowing
  - τ^2 estimated based on the observed data
  - Estimating τ^2 with 2 studies is hard & means the prior is always informative
  - Allows for a dynamic amount of borrowing
  - $\tau^2 \sim IG(0.025,0.0000025)$.
    - Today I'd use $\tau^2 \sim U(0,5)$ or $\tau^2 \sim U(0,20)$

11

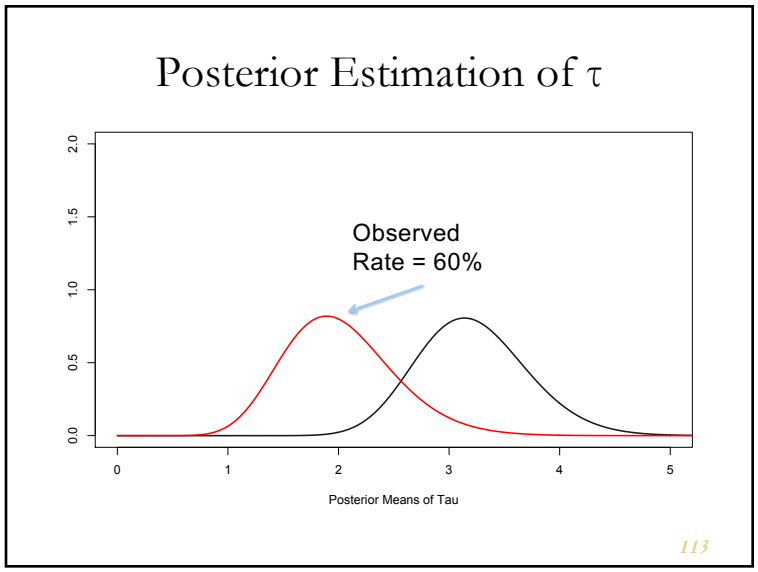
111

Posterior Estimation of τ

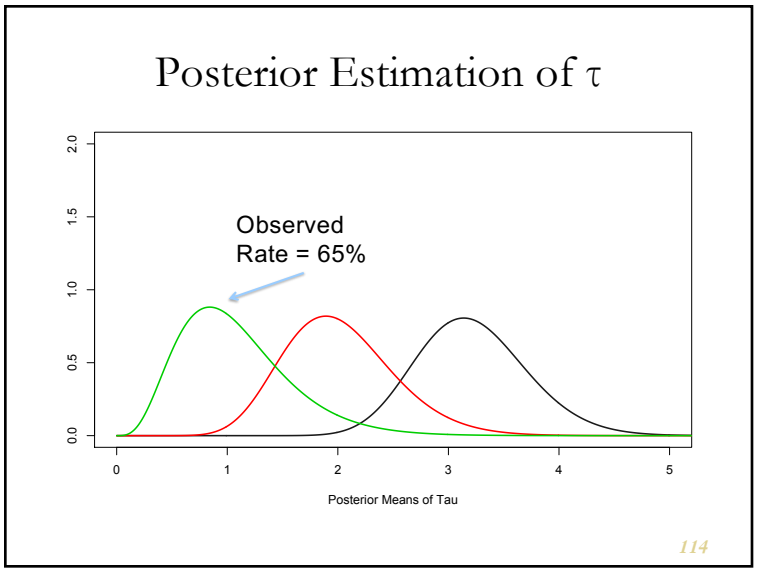


112

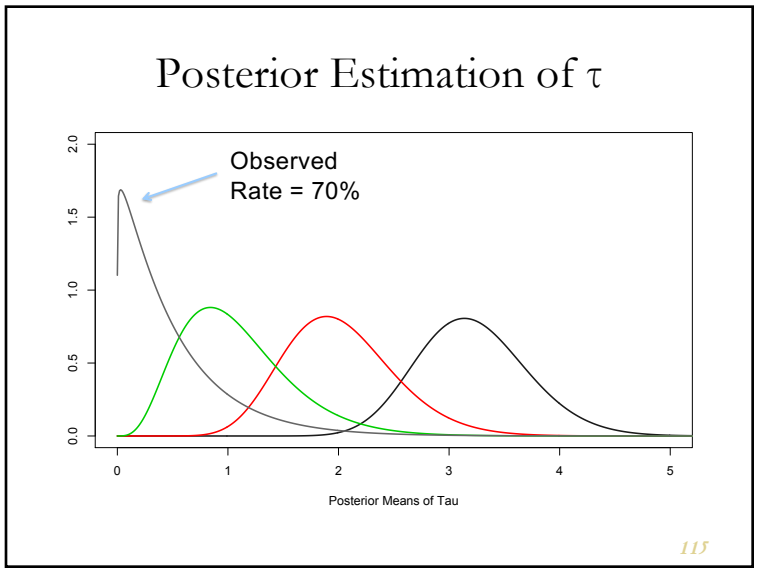
112



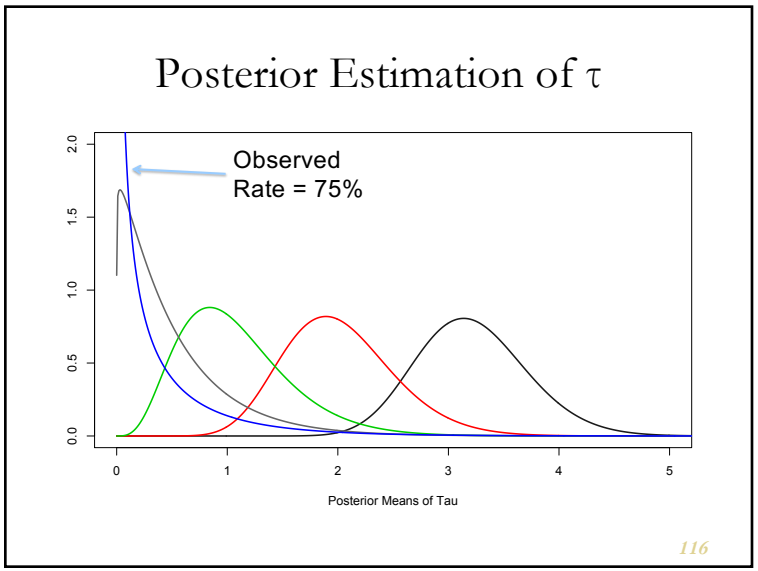
113



114

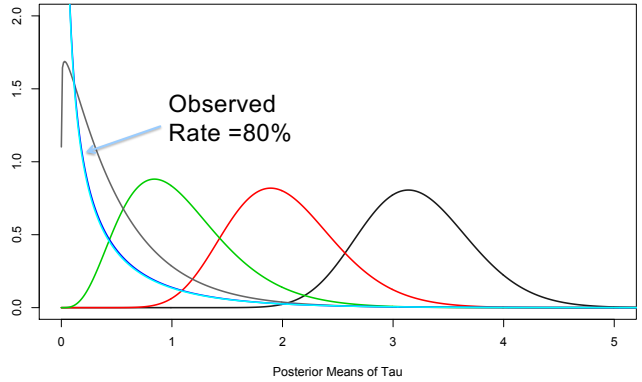


115



116

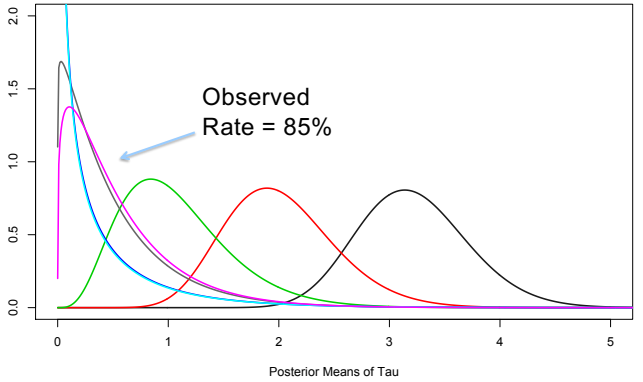
Posterior Estimation of τ



117

117

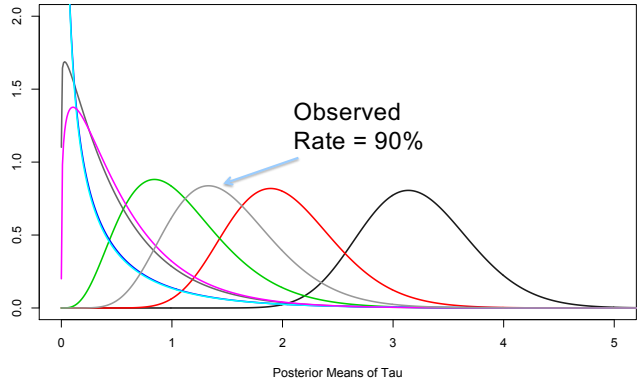
Posterior Estimation of τ



118

118

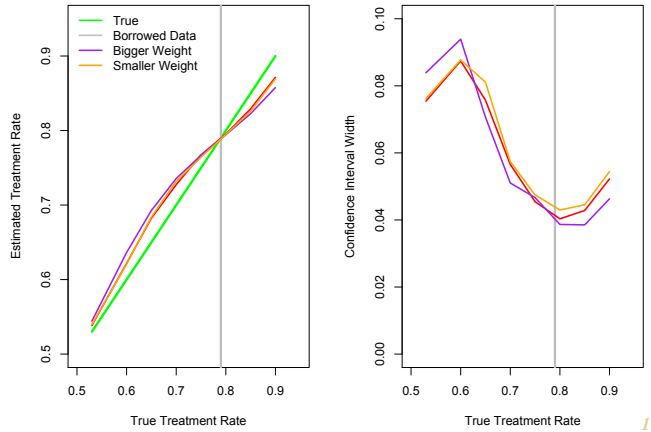
Posterior Estimation of τ



119

119

Estimation



120

120

Adjunct Analysis Plan

- Simulations explored
  - Timing of adjunct analyses
  - Amount of borrowing (Weaker → Stronger)
  - Thresholds for claiming success at each look
  - Accrual rates
- Simulations showed control of overall one-sided Type I error < 5% (both endpoints)

121

121

Adjunct Analysis Plan

- Adjunct analyses based on the number of patients <u>enrolled</u>

			Critical Values for Success	
Analysis	Expected Completers: Laser	Expected Completers: Balloon	**Probability of Superiority for Efficacy**	**Probability of Non-Inferiority for Safety**
200 Patients	89	44	0.998	0.998
250 Patients	119	58	0.9975	0.9975
300 Patients	149	74	0.995	0.995
Final Analysis	190	95	0.979	0.979

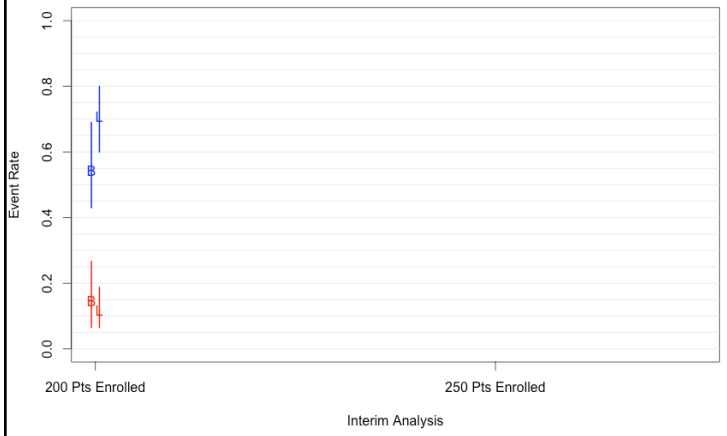
122

122

ACTUAL TRIAL EXECUTION

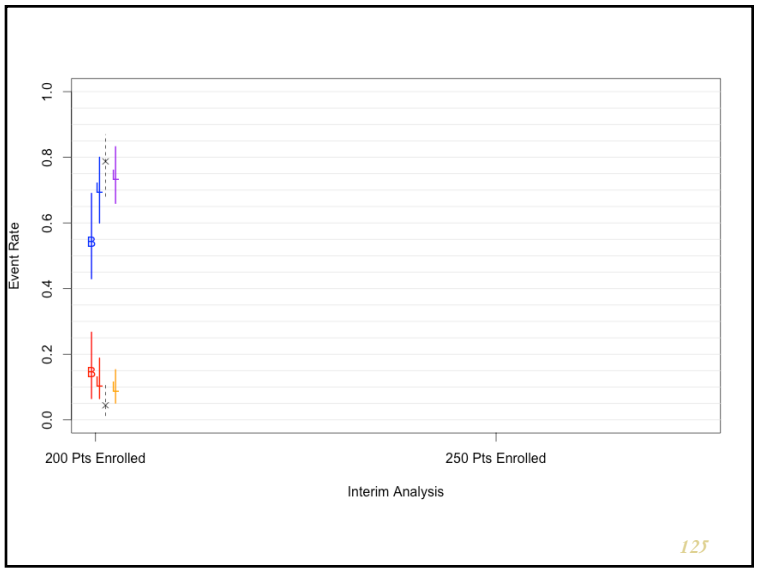
123

123

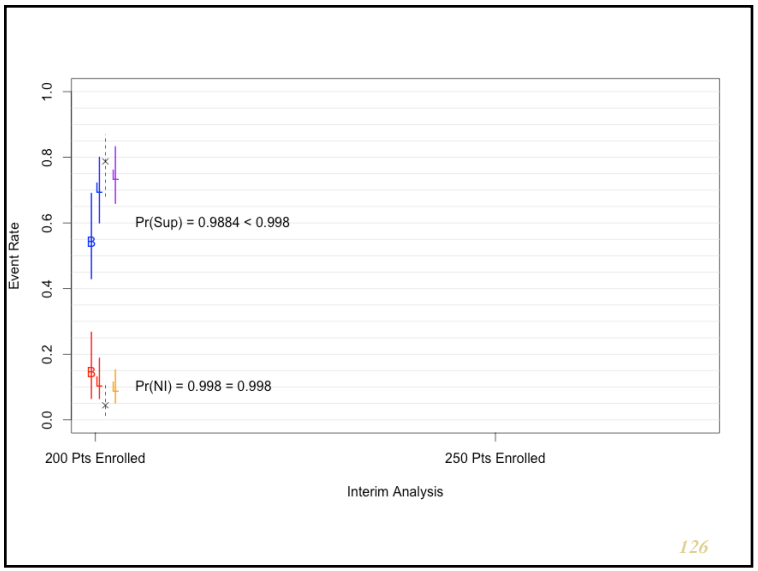


124

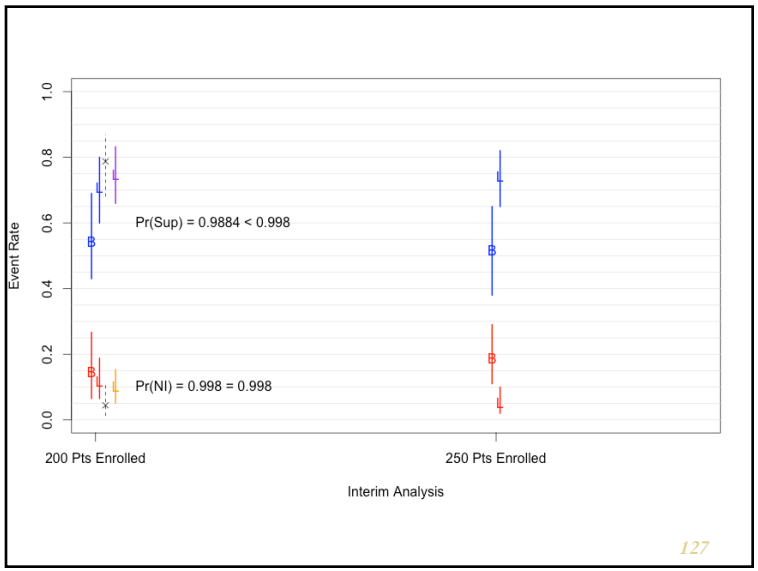
124



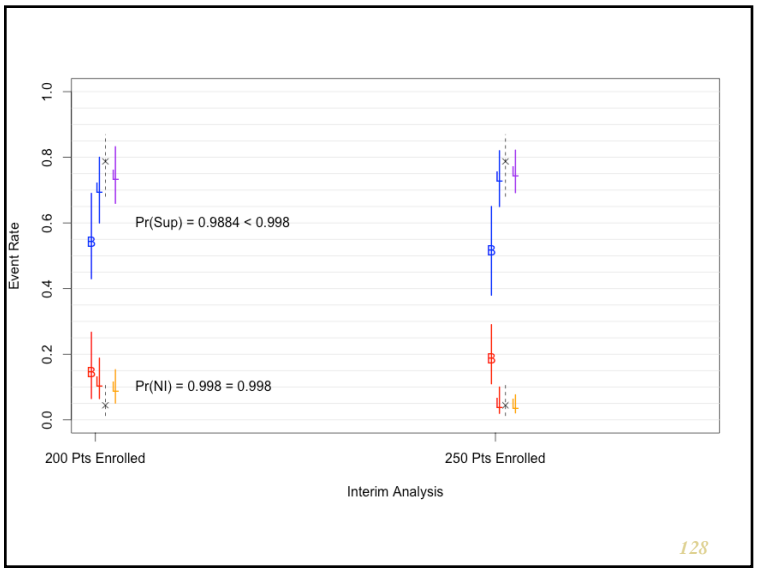
125



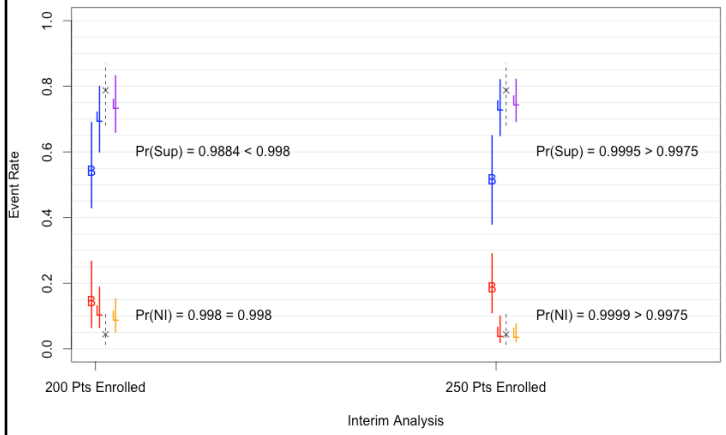
126



127



128



129

129

Compare to the O'Brien Fleming

Interim Analysis	% Information	Nominal P-Value	Observed P-Value
200 Pts	44%	0.0011	0.1005
250 Pts	60%	0.0043	0.006
300 Pts	78%	0.0105	--
Final	100%	0.0208	--

130

130

Conclusions

- Study met the adjunct analysis success criteria in Feb 2014
- 510K approval given in July 2014
- Randomized data showed a benefit in terms of efficacy and safety
- Borrowing from prior data increased precision
- Borrowing via prospectively defined rule
- Borrowing dependent on similarity of new trial with previous trial
- Allowed stopping earlier than an OBF bound

131

131

Platform Trials

& Master Protocols

132

132

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.

133

133

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.

134

134

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

135

135

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

136

136

Master Protocols

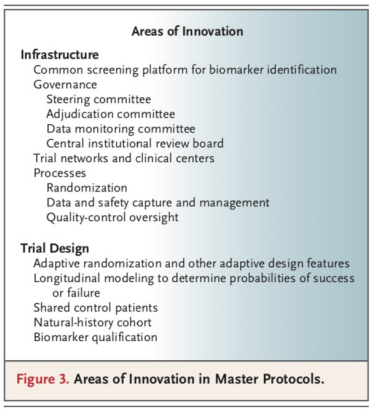


Figure 3. Areas of Innovation in Master Protocols.

- 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

137

137

Platform Trials

138

138

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?

139

139

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?

140

140

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

141

141

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

142

142

Opinion

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

143

143

The JAMA Network

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]

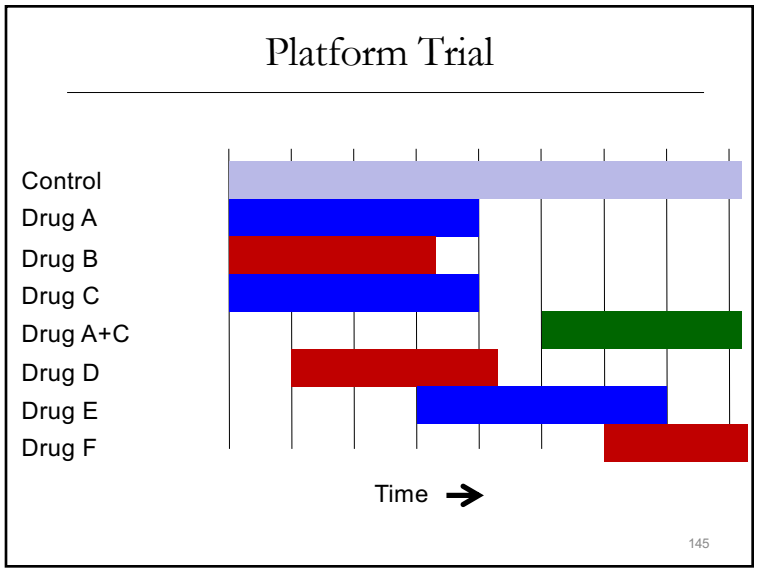
Characteristic	Traditional Trial	Platform Trial
Scope	Efficacy of a single agent in a homogeneous population	Evaluating efficacy of multiple agents in a heterogeneous population; explicitly assumes treatment effects may be heterogeneous
Duration	Finite, based on time required to answer the single primary 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.

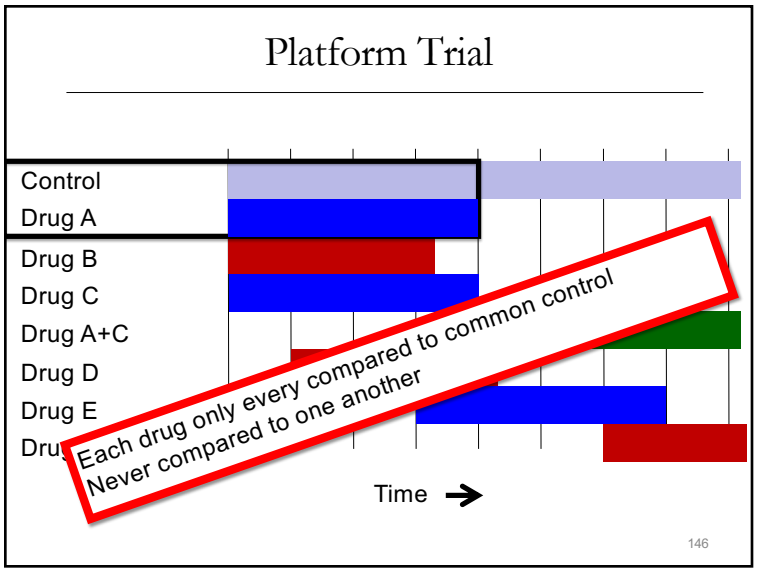
Table Title:
General Characteristics of Traditional and Platform Trials[a]

144

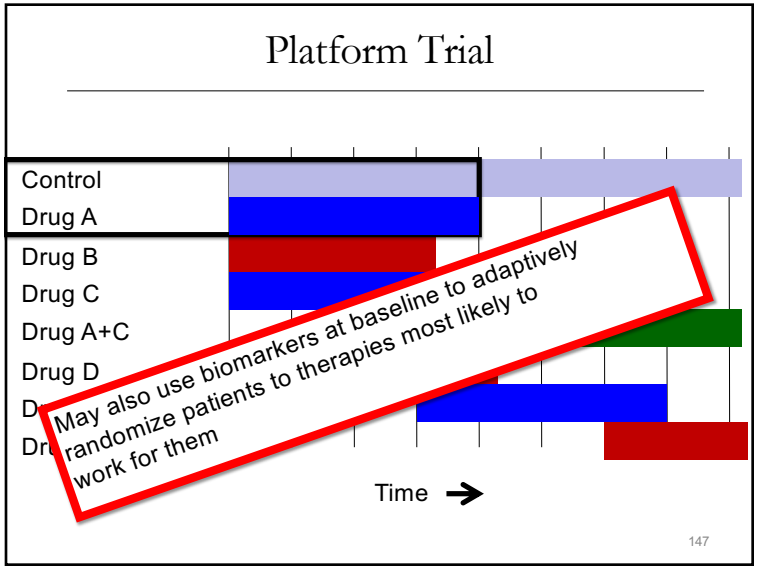
144



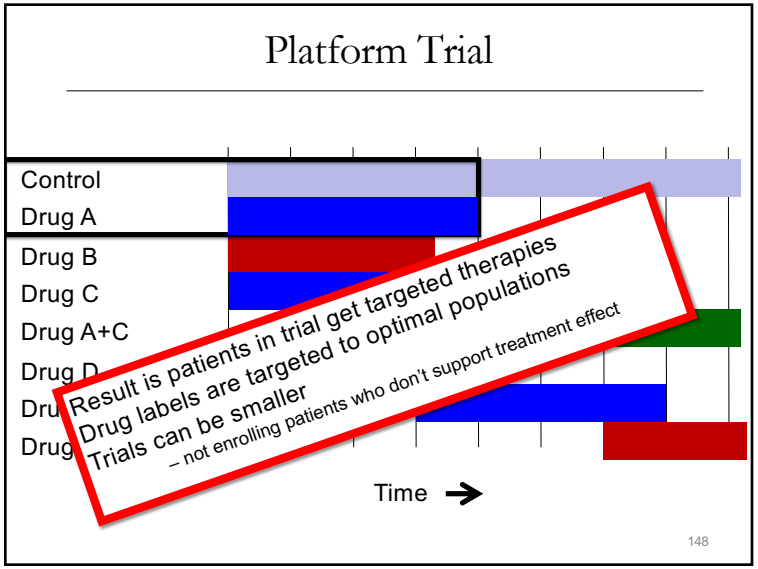
145



146

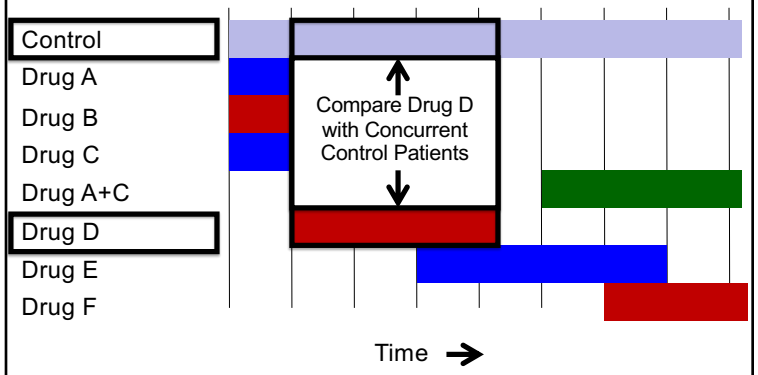


147



148

Platform Trial

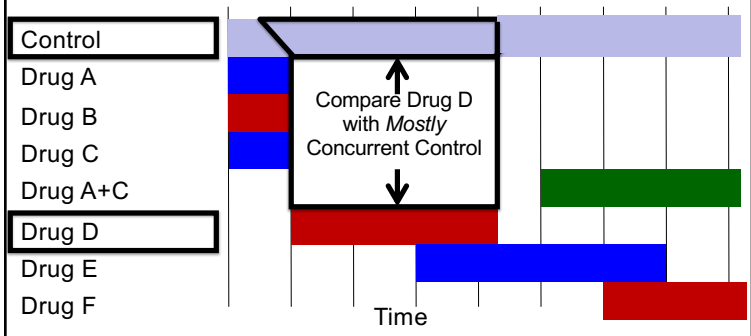


149

149

Time Machine

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

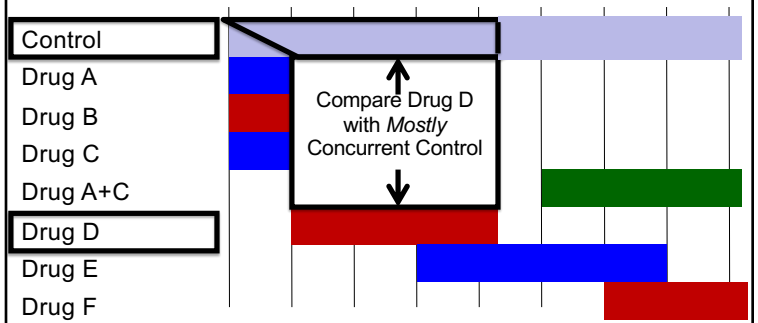


150

150

Time Machine

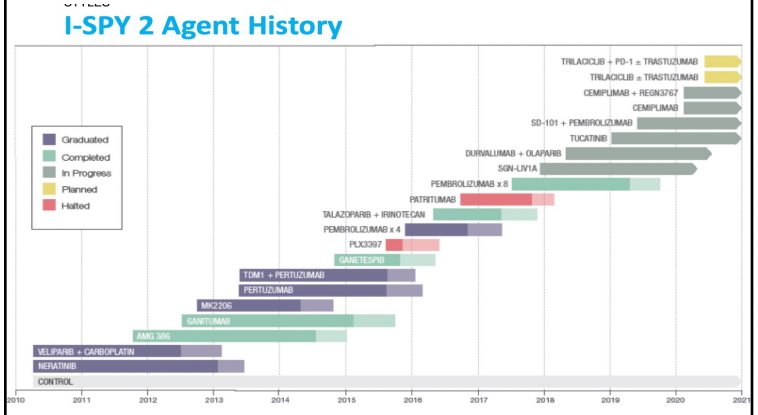
If controls change little over time, then use more weight from non-concurrent controls, increases power & efficiency



151

151

I-SPY2



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

152

152

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

153

153

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

154

154

Characteristics of Modern Platform Trials	Cancer Trials						Alzheimer's	Ebola
	I-SPY 2	MICAT	BATTLE	LUNG-MAP	UK MATRIX	GBM-AGILE		
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

155

155

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

156

156

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.

157

EBOLA

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

Berry Consultants
Statistical Innovation

Duke Clinical Research Institute

ClinicalRM
20th Anniversary 1994-2014

158

158

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

159

159

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

160

160

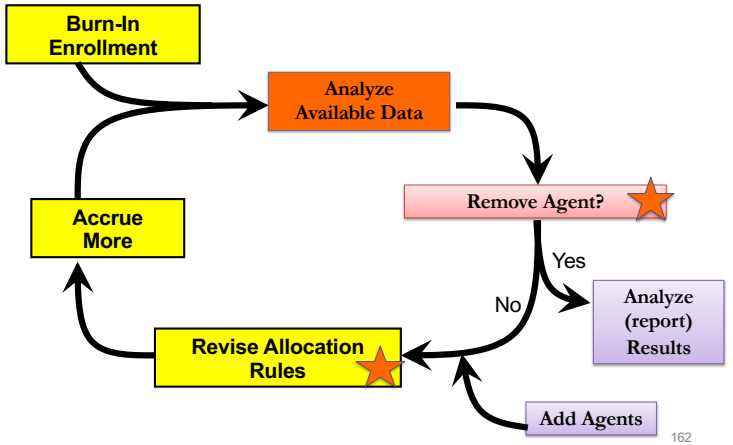
Primary/Secondary Agents

- 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						

161

Adaptive Platform Design



162

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)

163

163

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

164

164

Allocation Rules

Revise Allocation Rules

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

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

165

165

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:

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

166

166

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.

167

167

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.

168

168

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'

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

169

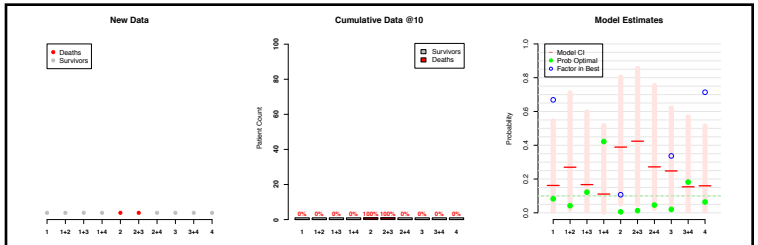
169

Example Trial

Regimens		Agents			
		1	2	3	4
Agents	1				
	2				
	3				
	4				

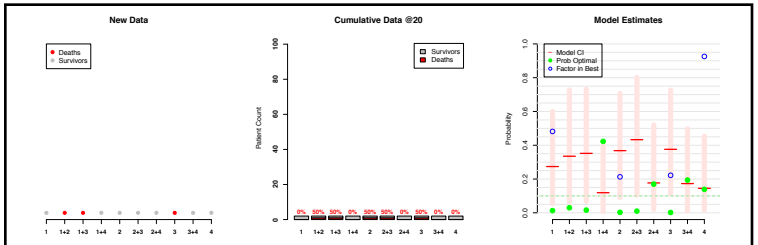
170

170



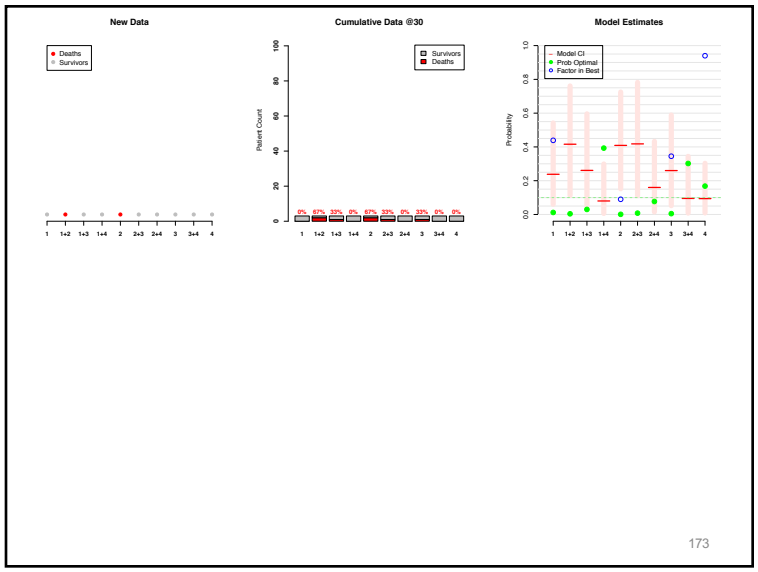
171

171

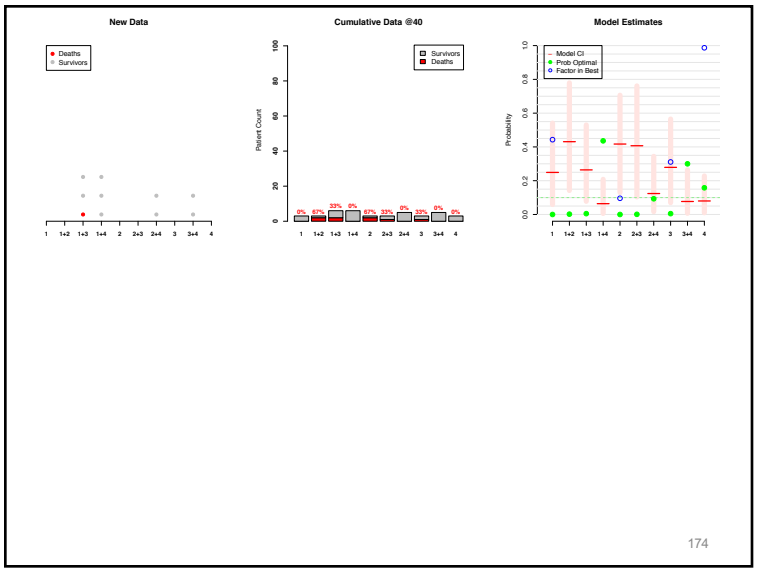


172

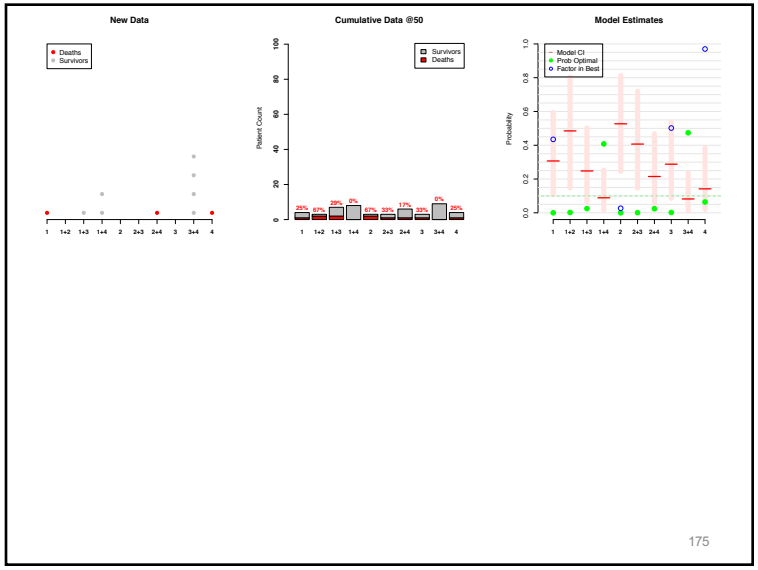
172



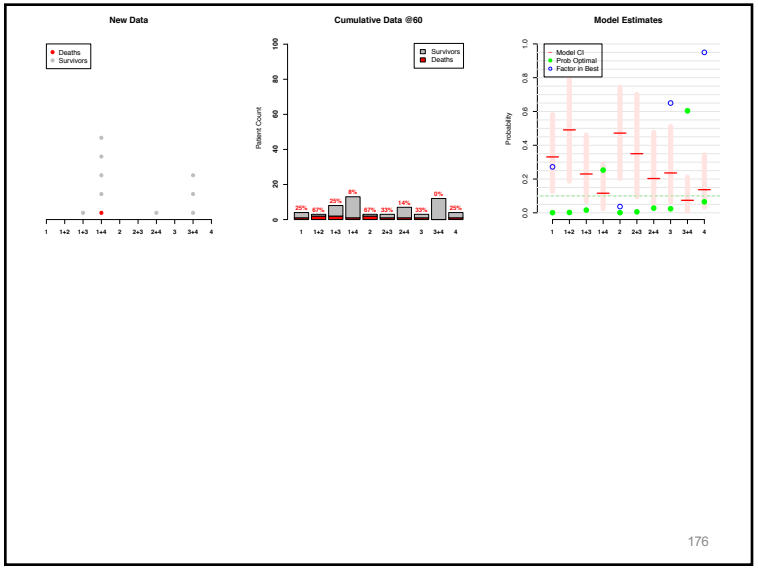
173



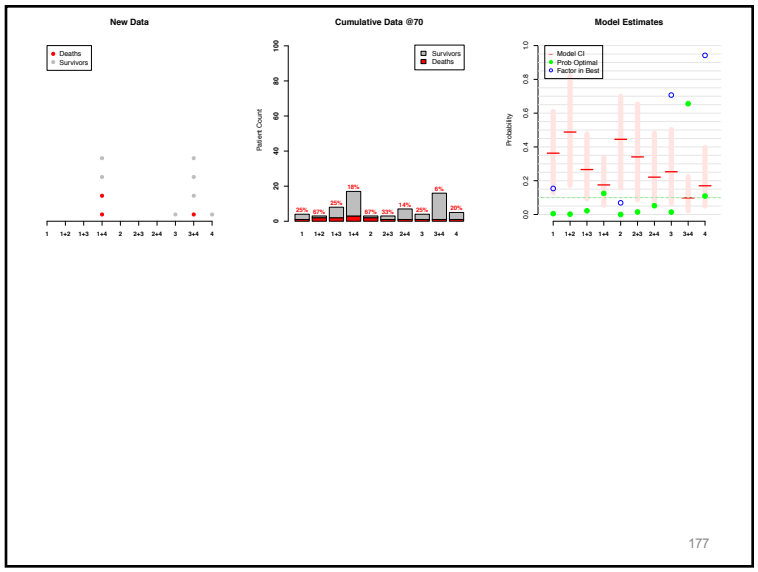
174



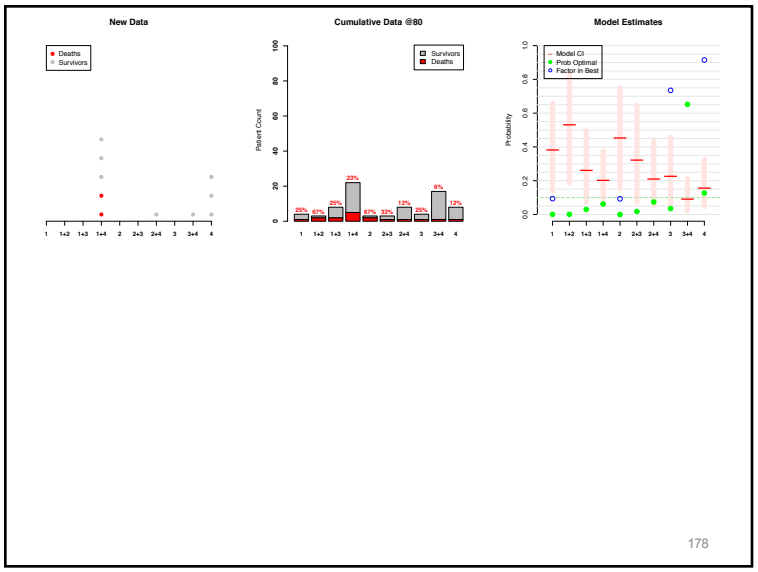
175



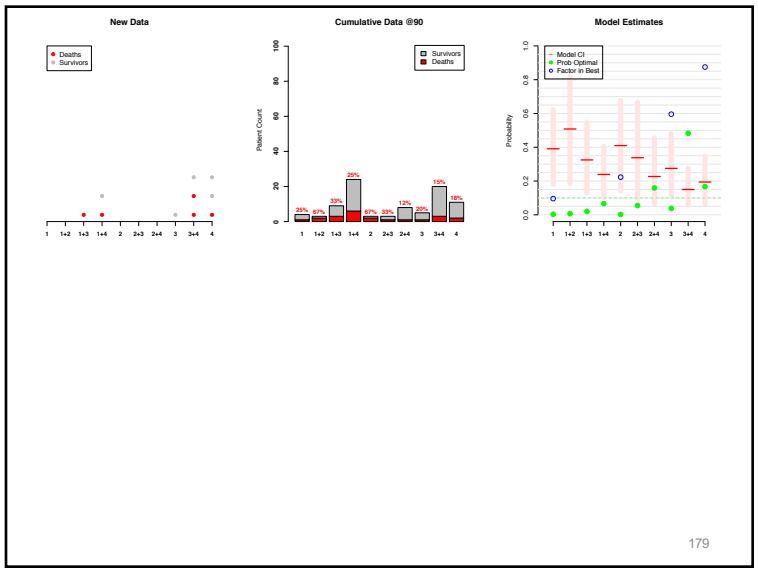
176



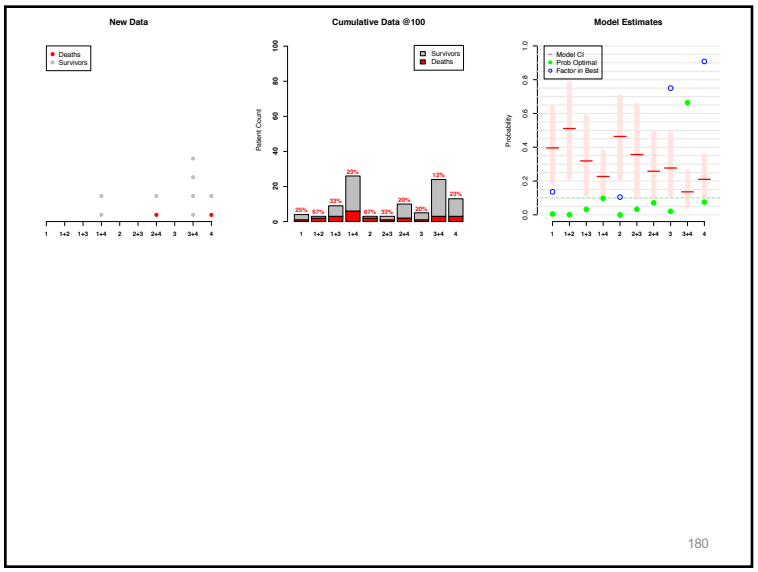
177



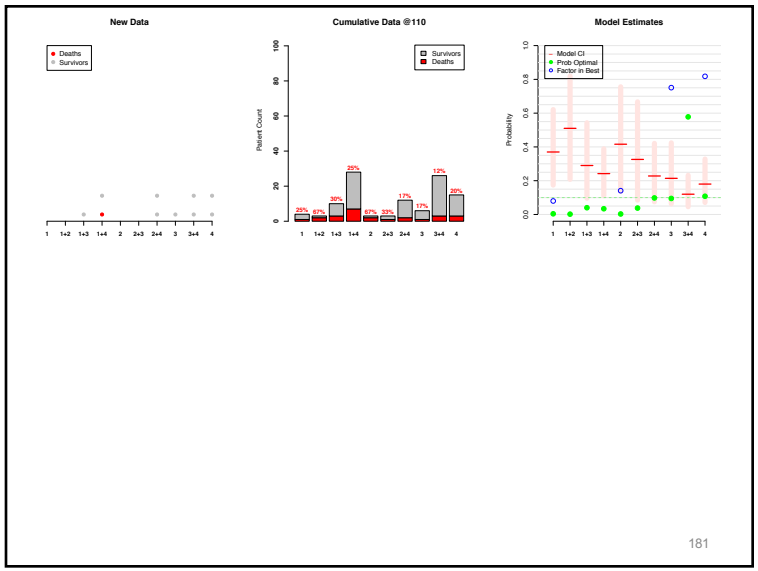
178



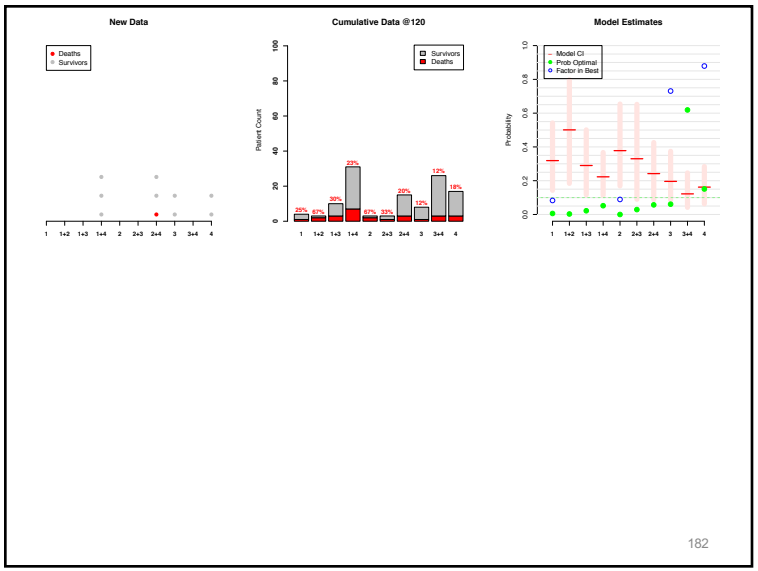
179



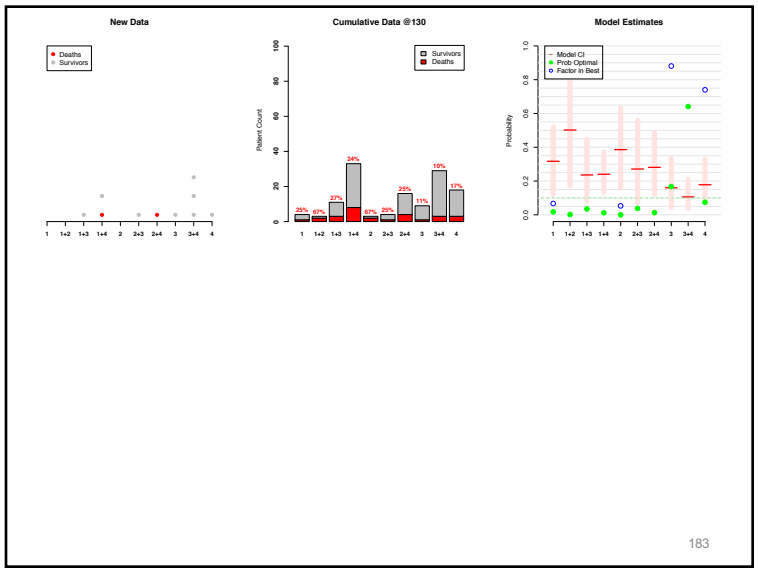
180



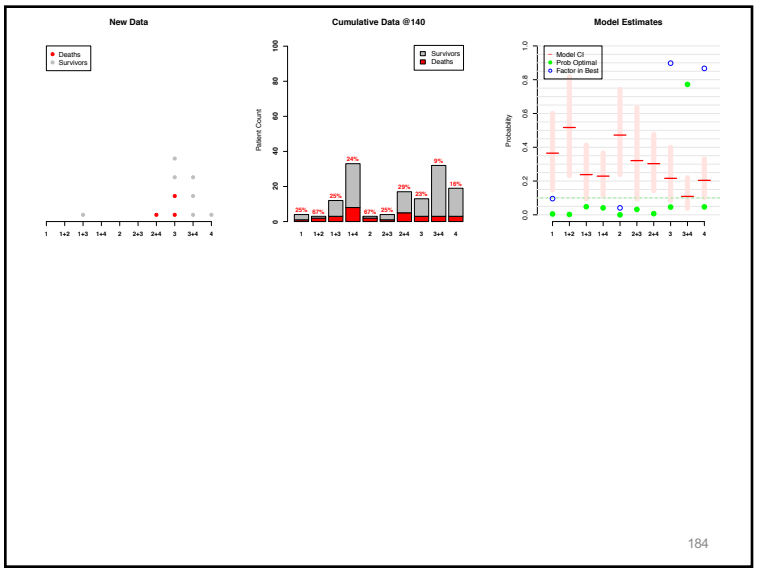
181



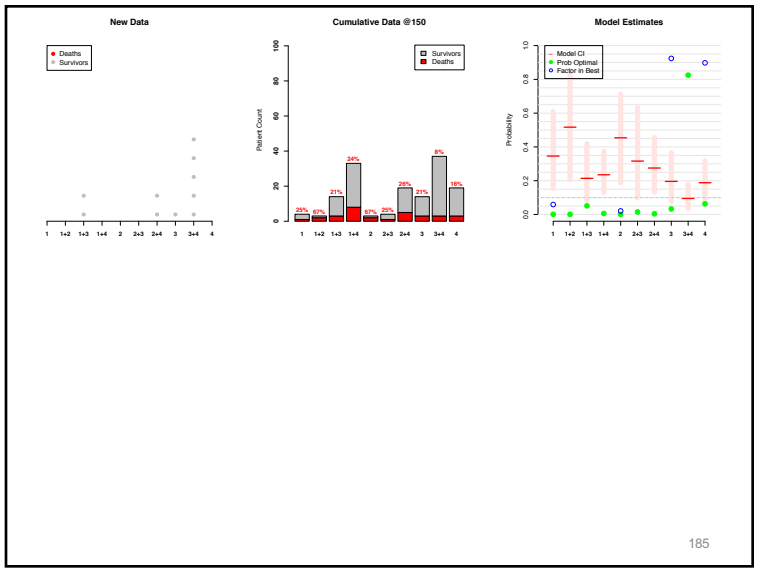
182



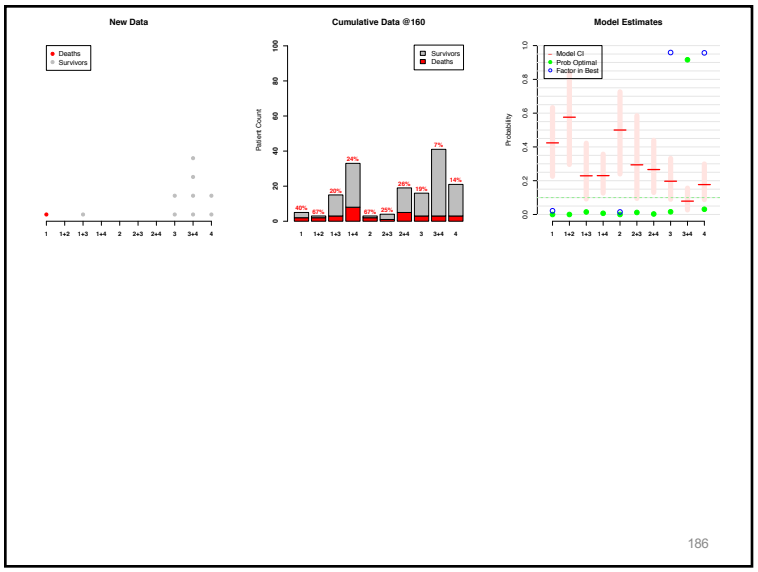
183



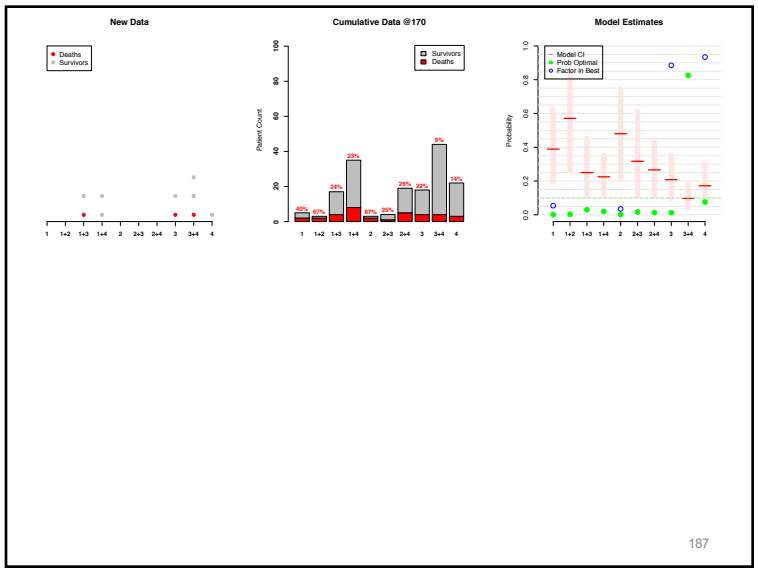
184



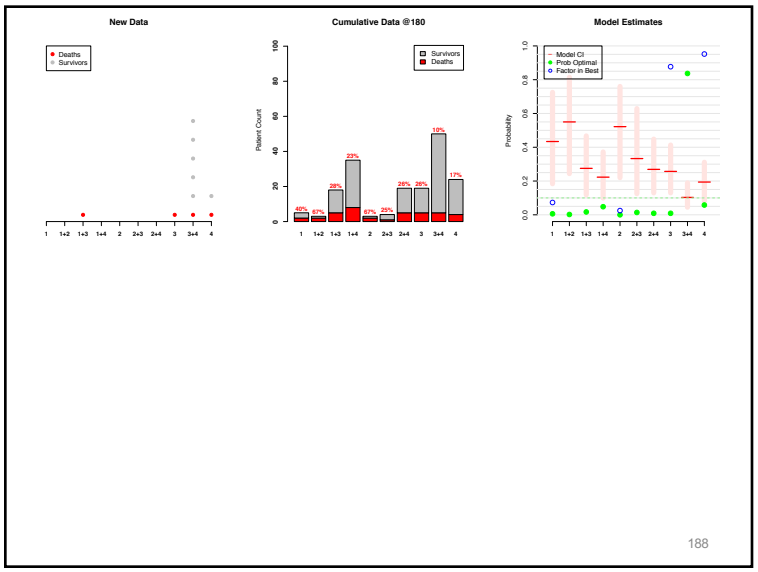
185



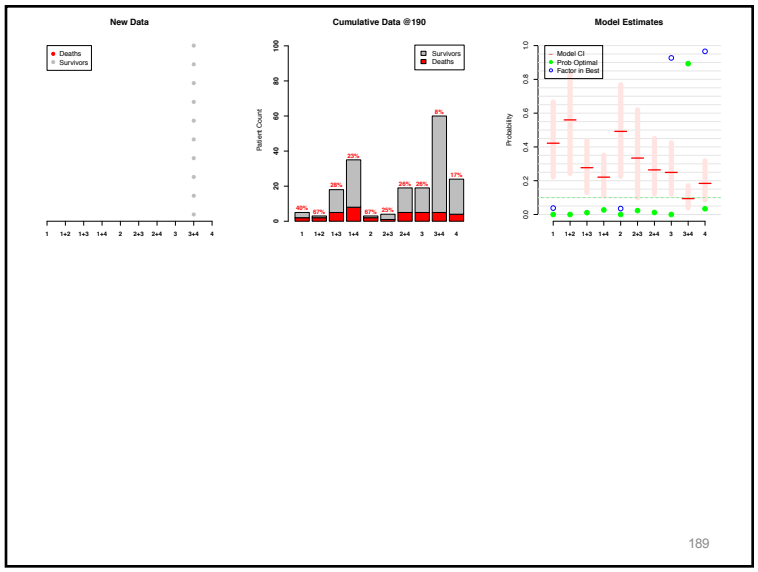
186



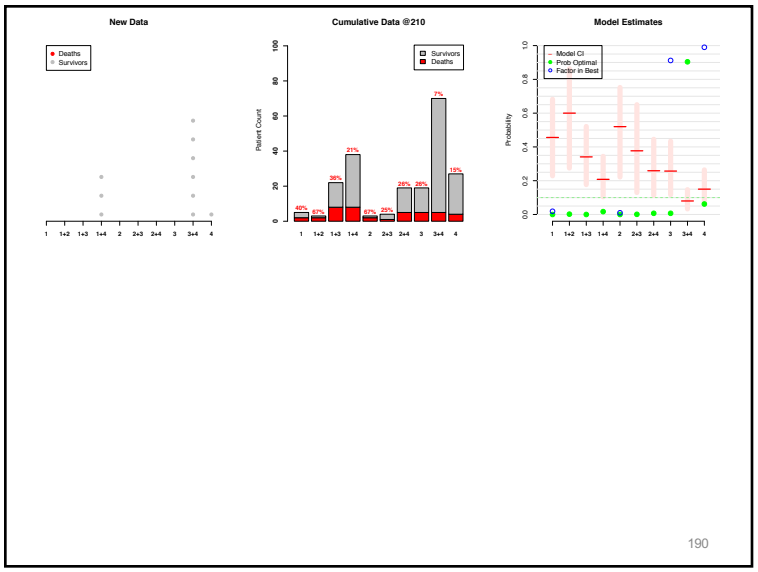
187



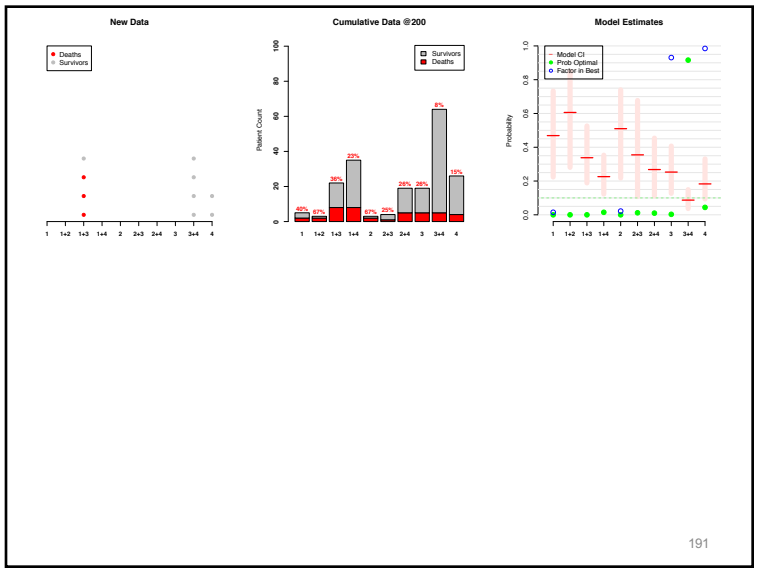
188



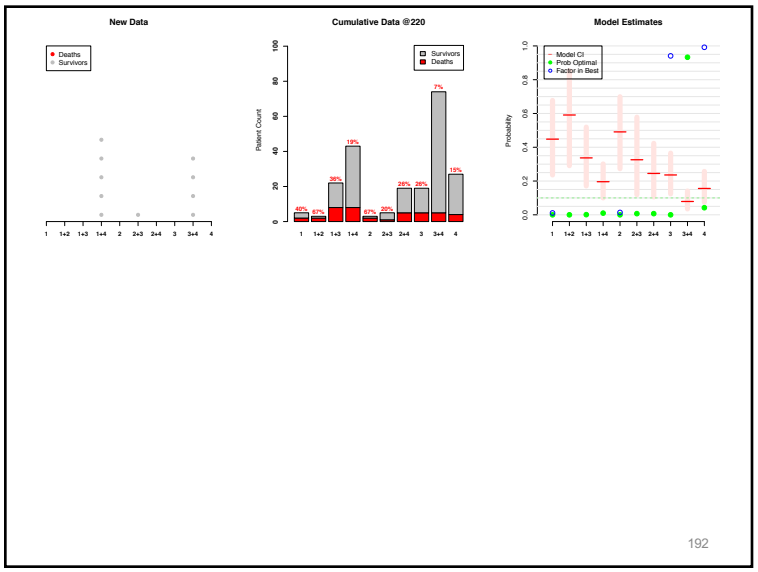
189



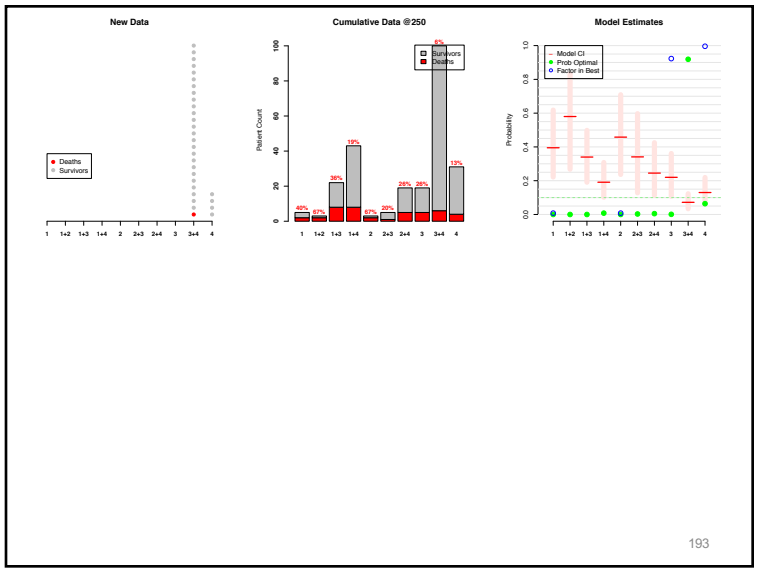
190



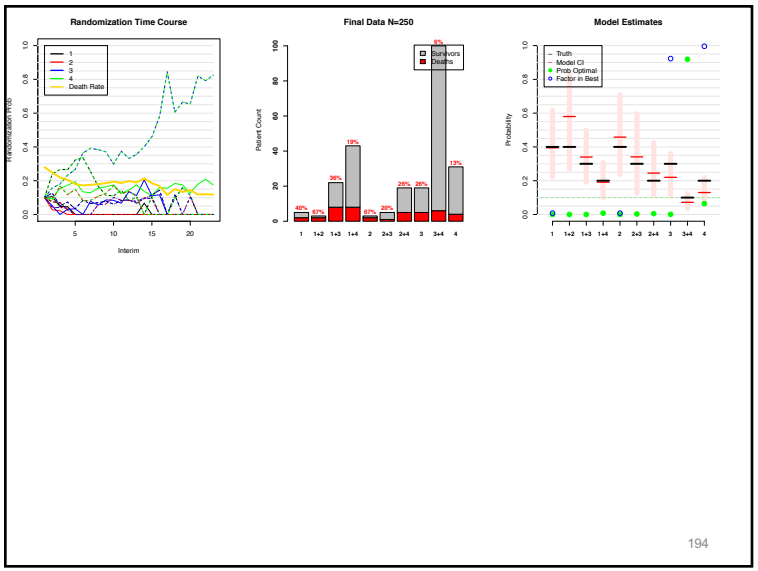
191



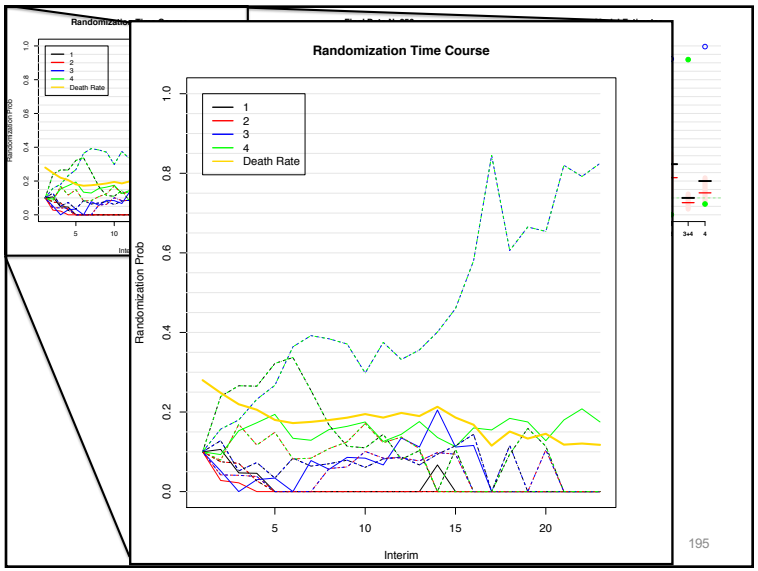
192



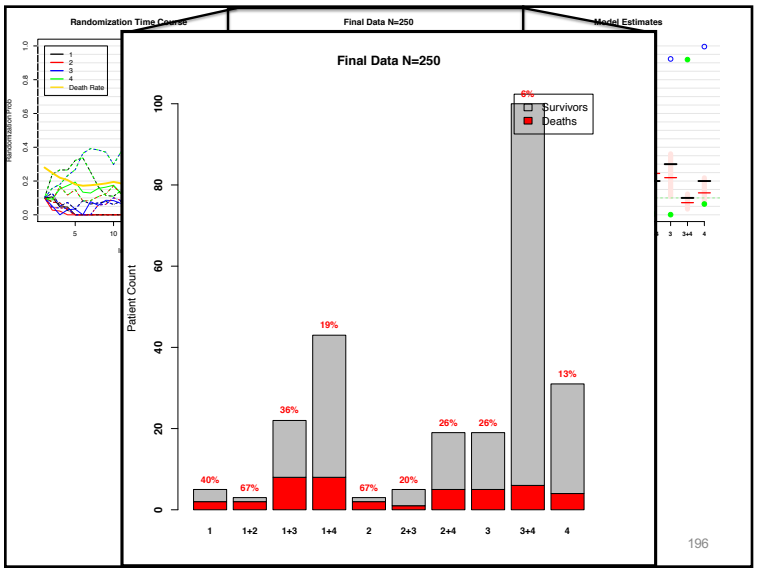
193



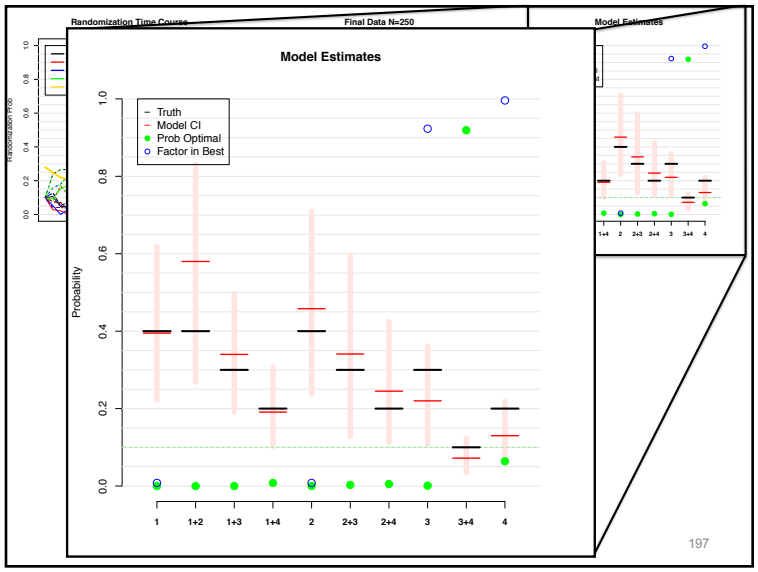
194



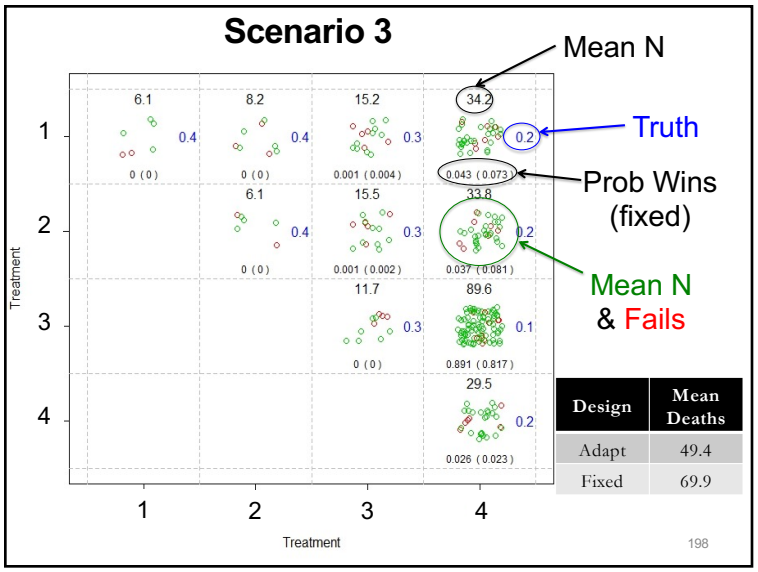
195



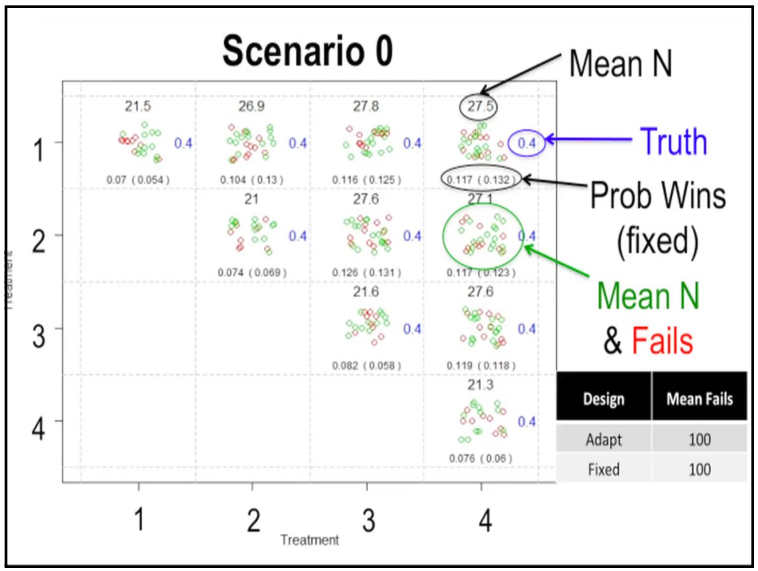
196



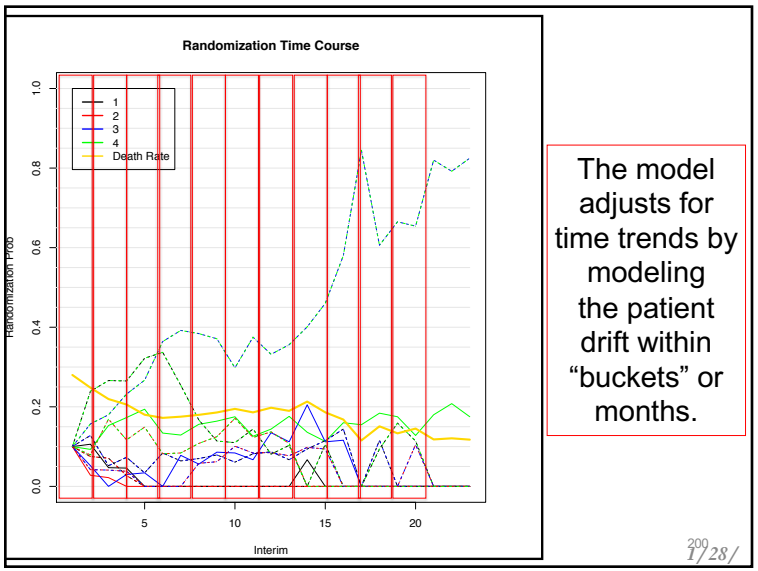
197



198



199



200

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

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

201

201

Platform Example 2

The Role of Biomarkers in Treatments & Trials

202

202

Testing a New Treatment

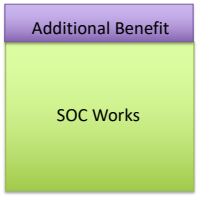
- Standard of Care works in 40%

203

203

10% of Patients Benefit

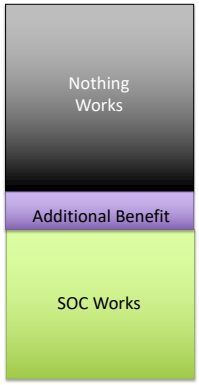
- Standard of Care works in 40%
- New therapy works in 50%



204

204

50% still untreatable

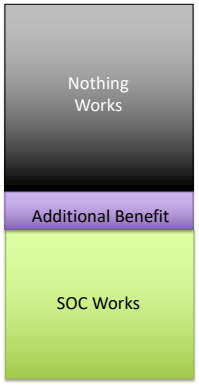


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%

205

205

50% still untreatable

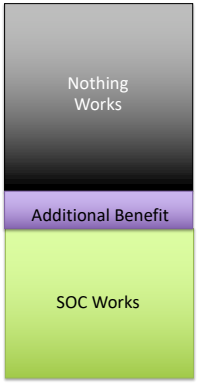


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%
- How many patients do we need to have 90% chance to see a 'statistically significant' difference?

206

206

Need 1036 patients for 90% Power

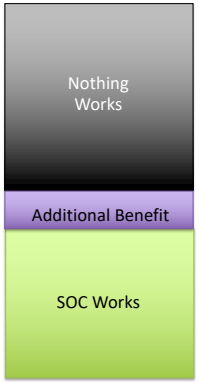


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%

207

207

Need 1036 patients for 90% Power

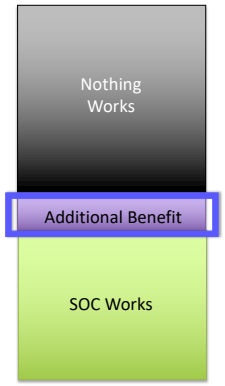


- Standard of Care works in 40%
- New therapy works in 50%
- Nothing works in 50%
- 90% of patients you enroll tell you nothing

208

208

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?

209

209

What if you KNEW which 10% Benefit

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

210

210

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

211

211

What if you^KNEW which 10% Benefit

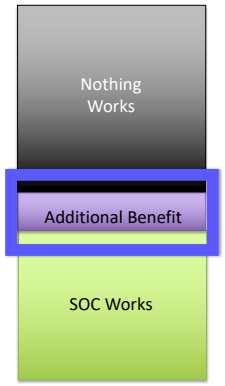
sorta

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

212

212

What if you sorta^KNEW which 10% Benefit

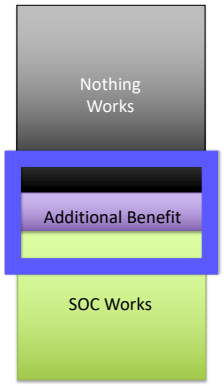


- Enroll 20% to capture the 10%
- 25% cured by SOC
- 25% still not cured
- 50% of enrolled patients benefit
- Need 36 patients for 90% power

213

213

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

214

214

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

215

215

Platform Example 2

216

216

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

217

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

218

218

Biomarkers → Signatures

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

219

219

2 × 2 Biomarkers → 4 Signatures

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

220

220

2 × 2 Biomarkers → 3 Signatures

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

221

221

2 × 2 Biomarkers → 1 Signature

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

222

222

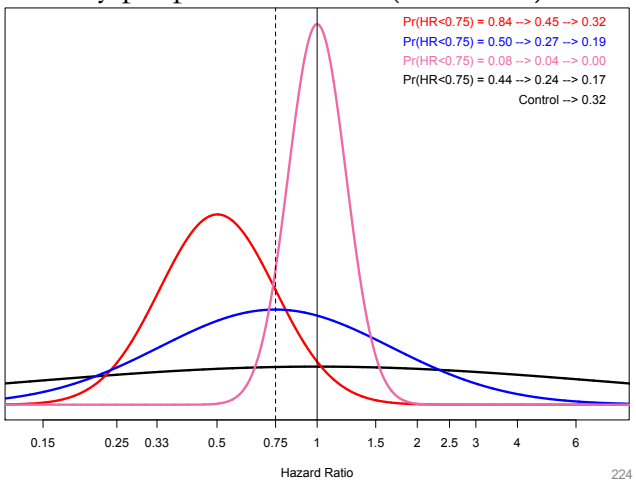
Response-adaptive randomization

- 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

223

223

Why proportional to Pr(HR<0.75)?



224

224

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

225

225

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

226

226

Identifying the Right Target Population

	Newly Diagnosed	Recurrent
Methylated	It works here	
Unmethlylated		

227

227

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

228

228

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

229

229

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

230

230

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

231

231

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"

232

232

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

233

233

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

234

234

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?

235

235

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

236

236

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

237

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.

238

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

239

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

240

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

241

241

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

242

242

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

243

243

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

244

244

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

245

245

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%

246

246

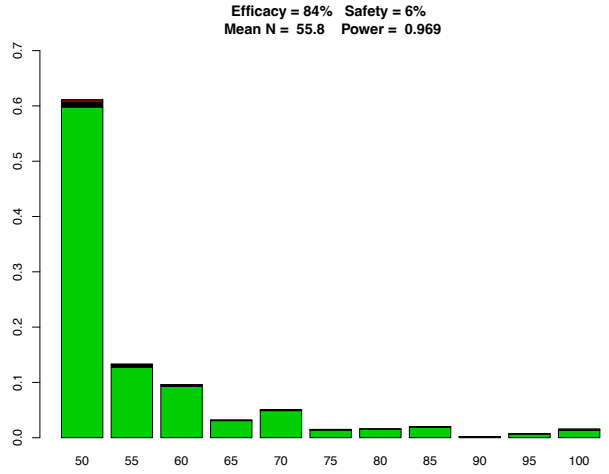
Operating Characteristics for Trial with $p_T = 0.84, p_S = 0.08$

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

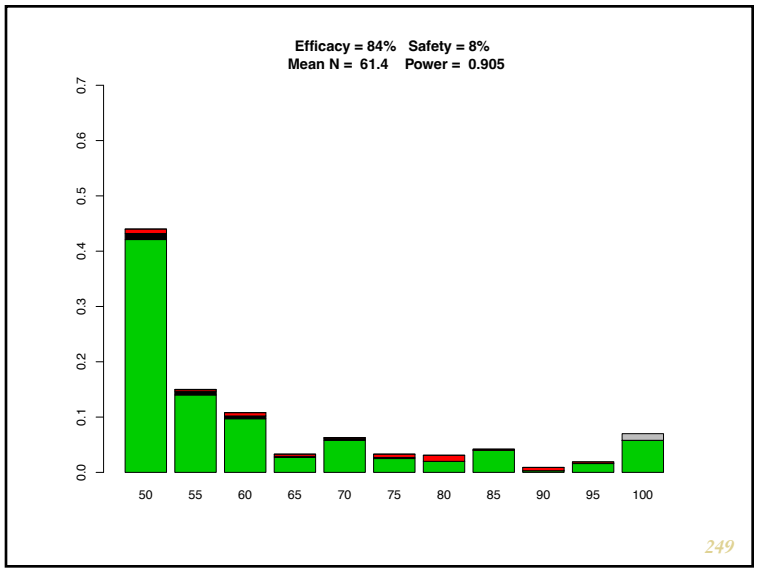
247

247

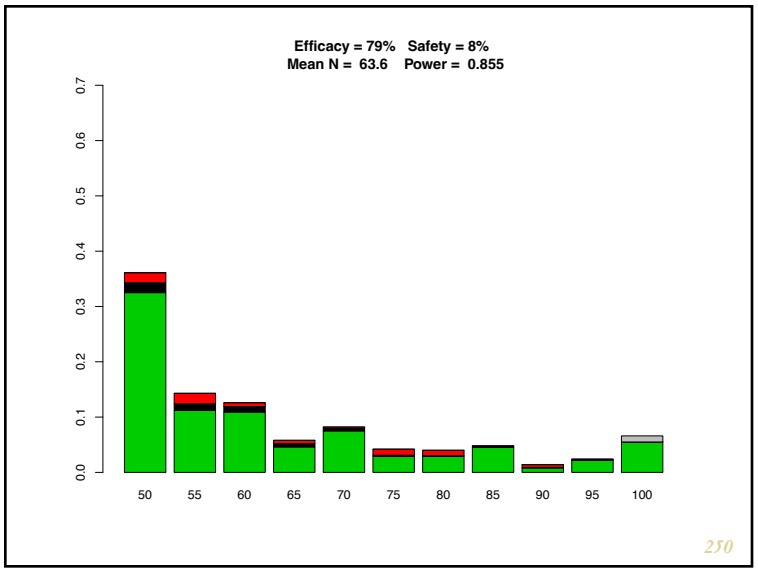


248

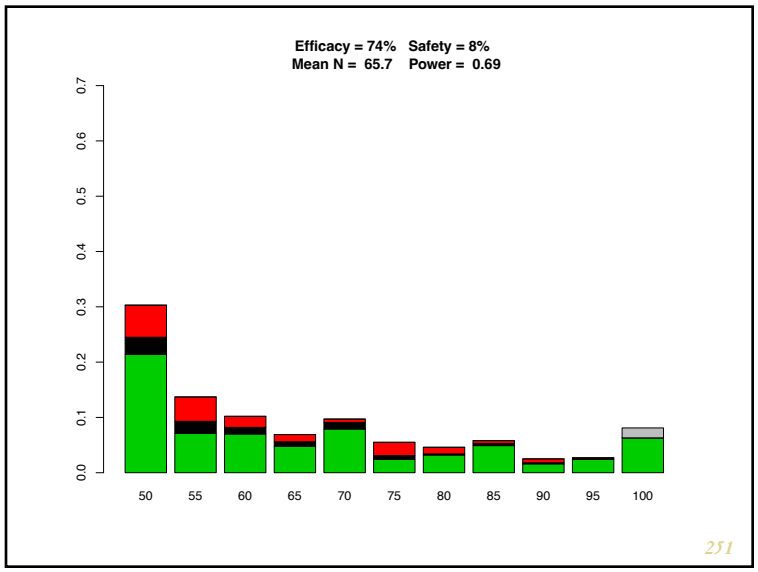
248



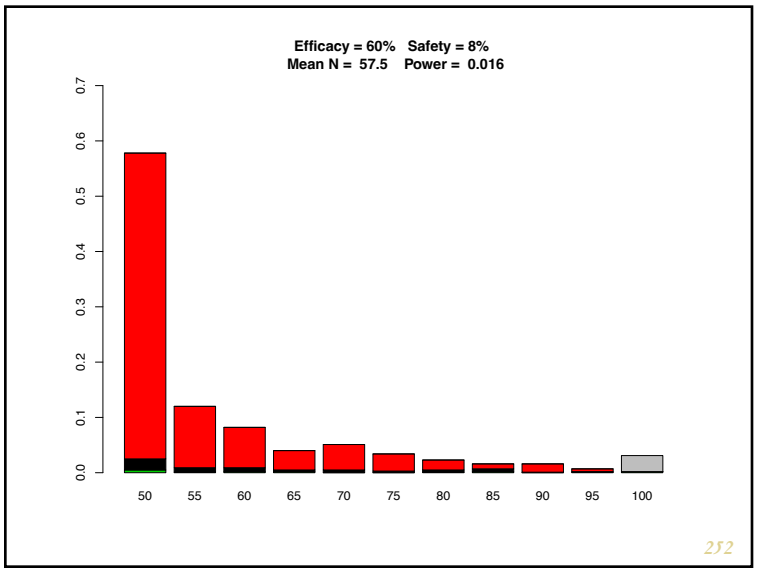
249



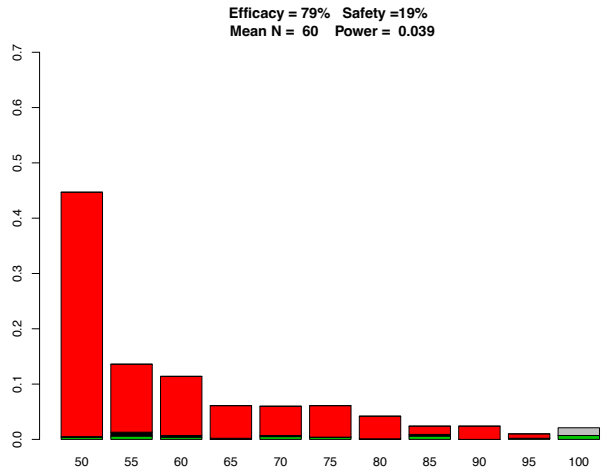
250



251



252



253

253

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
    $X = 24 + x_o + x_+ + x_-$

254

254

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)
- $\Pr(24+x_o+x_-+x_+ \geq 37) = 0.988$

255

255

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

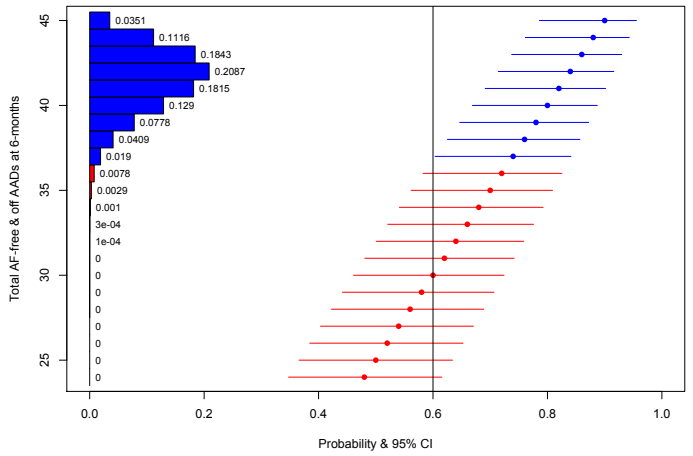
4.2/6 = .70
3/4 = .75

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

256

256

Prediction of 21 remaining pts based on 29 observed pts



257

257

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

258

258

Stopped Accrual for Predicted Success

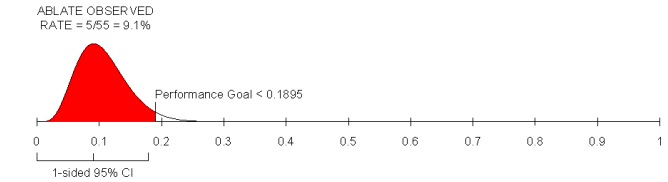
- Accrual stopped with 55 patients in
- Continue to follow 21 enrolled patients
- Perform final analysis on complete data

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

259

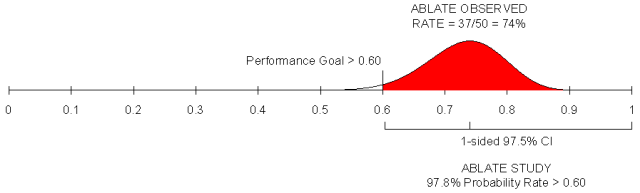
259

Safety



ABLATE STUDY
96.7% Probability Rate < 0.1895

Efficacy



ABLATE STUDY
97.8% Probability Rate > 0.60

260

260

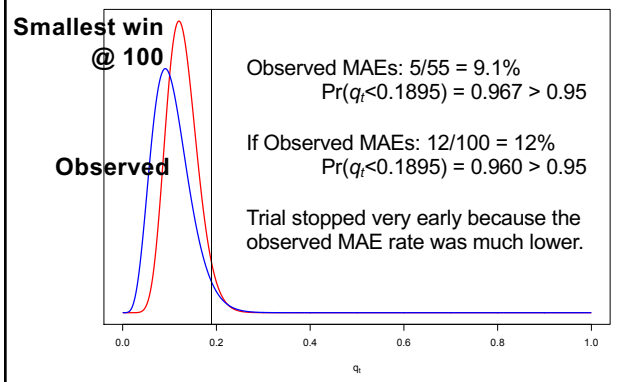
Post Trial Discussion with FDA

- 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

261

261

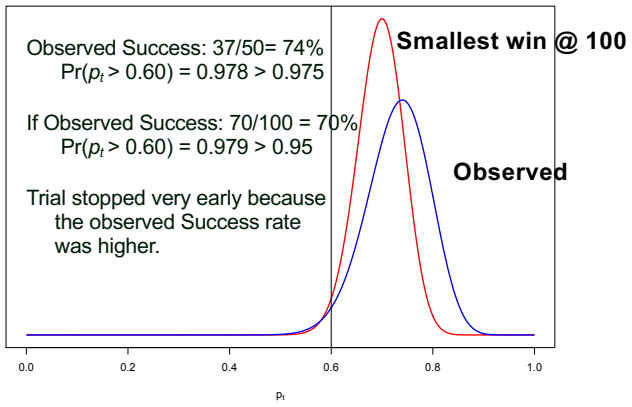
Safety: Compare Stopping at n=55 to Maximum Trial Size n=100



262

262

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



263

263

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

264

264

heartwire

ARRHYTHMIA/EP

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

OCTOBER 26, 2011 Michael O'Riordan

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."

265

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.

266

266

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

267

267