Proof of Concept Concepts

Donald A. Berry
dberry@mdanderson.org

THE UNIVERSITY OF TEXAS
MD ANDERSON
CANCER CENTER

BERRY
CONSULTANTS
STATISTICAL INNOVATION

"Proof of concept"

- "Proof": Impossible
- "Concept": Not completely known
- "of"
- Best can do:
  - Probabilities
  - Utilities

Modeling plays critical role

My perspective Bayesian

THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER

Some Issues

- Sequential learning
- Predictive distributions
- Borrowing strength
  - Historical data
  - Other drugs in class
  - Across patient/disease groups

Design

Design & Analysis

Important Points

- Endpoints*
  - Early
  - Late
  - Modeling both
- Historical information
- Borrowing in class
- Preclinical ⇔ Clinical modeling
- Sequential monitoring*
- Prespecify futility/promotion criteria*
- Utility, marketing, costs*
- Dynamic programming

EXAMPLES

- Predictive probabilities
- Adaptive randomization
  - Drug trial in AML
  - Phase I/II cancer trial
  - Drug screening process
  - Dose finding
  - PoC study in pain
- Modeling early & late endpoints

PREDICTIVE PROBABILITY

- Given results so far in the trial, what is probability of
  - Statistical significance at trial's end?
  - Drug will make money?
  - Futility?
  - Etc?
- Critical role in
  - Trial design
  - Monitoring trials

Example* Using Pred Probs

- Disease: HER2+ neoadj breast cancer
- Endpoint: tumor response
- Balanced randomized: CT±Herceptin
- Sample size planned: n = 164
- DMC; Interim results after n = 34:
  - Not Herc: 4/16 = 25% vs Herc: 12/18 = 67%
  - Consistent with other studies
- Predictive probability of stat sig: 95%
- Update: Adj trials w/12,000 patients!

*Buzdar, et al. ASCO 2004, JCO 2005

ADAPTIVE RANDOMIZATION Giles, et al JCO (2003)

- Troxacitabine (T) in acute myeloid leukemia (AML) combined with cytarabine (A) or idarubicin (I)
- Adaptive randomization to:
  IA vs TA vs TI
- Max n = 75
- End point: Time to CR (< 50 days)

Adaptive Randomization

- Assign 1/3 to IA (standard) throughout (unless only 2 arms)
- Adaptive to TA and TI based on current results: $P(p_E > p_{IA}|data)$
- Results $\rightarrow$

Patient	Prob IA	Prob TA	Prob TI	Arm	CR<50
1	0.33	0.33	0.33	TI	not
2	0.33	0.34	0.32	IA	CR
3	0.33	0.35	0.32	TI	not
4	0.33	0.37	0.30	IA	not
5	0.33	0.38	0.28	IA	not
6	0.33	0.39	0.28	IA	CR
7	0.33	0.39	0.27	IA	not
8	0.33	0.44	0.23	TI	not
9	0.33	0.47	0.20	TI	not
10	0.33	0.43	0.24	TA	CR
11	0.33	0.50	0.17	TA	not
12	0.33	0.50	0.17	TA	not
13	0.33	0.47	0.20	TA	not
14	0.33	0.57	0.10	TI	not
15	0.33	0.57	0.10	TA	CR
16	0.33	0.56	0.11	IA	not
17	0.33	0.56	0.11	TA	CR

Patient	Prob IA	Prob TA	Prob TI	Arm	CR<50
18	0.33	0.55	0.11	TA	not
19	0.33	0.54	0.13	TA	not
20	0.33	0.53	0.14	IA	CR
21	0.33	0.49	0.18	IA	CR
22	0.33	0.46	0.21	IA	CR
23	0.33	0.58	0.09	IA	CR
24	0.33	0.59	0.07	IA	CR
25	0.87	0.13	0	IA	not
26	0.87	0.13	0	TA	not
27	0.96	0.04	0	TA	not
28	0.96	0.04	0	IA	CR
29	0.96	0.04	0	IA	not
30	0.96	0.04	0	IA	CR
31	0.96	0.04	0	IA	not
32	0.96	0.04	0	TA	not
33	0.96	0.04	0	IA	not
34	0.96	0.04	0	IA	CR

Drop TI

Compare n = 75

THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER

Summary of results

CR < 50 days:

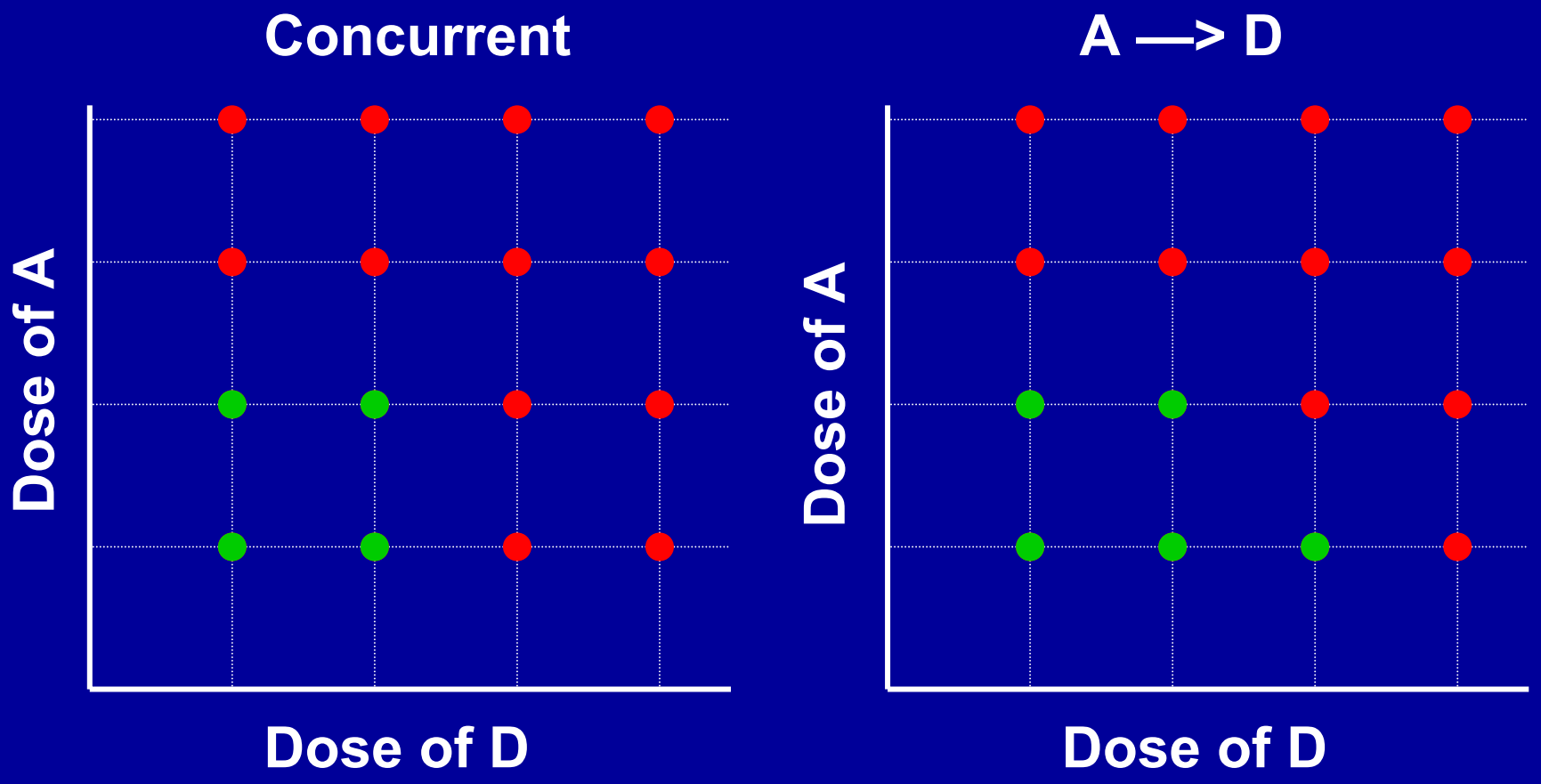
- IA: 10/18 = 56%
- TA: 3/11 = 27%
- TI: 0/5 = 0%

Criticisms . . .

Example: Developing phase I/II cancer trial

- Two drugs, A & D
  - Dosing
  - Sequential or concurrent?
- Phase I and MTD? PI's plans …
- Factorial
- Admissible dose combos
- Adaptive randomization

Dose Admissibility



At any given time

- Expand admissible doses if toxicity allows
- Randomize to admissible doses, adapting to tumor response
- (So might expand dose-range but still focus on lower doses)

PoC trials for many drugs!

- Tumor response
- Goals:
  - Treat effectively
  - Learn quickly
    - Bury losers
    - Promote potential winners

Standard designs

- One drug (or dose) at a time; no drug/dose comparisons
- Typical comparison by null hypothesis: RR = 20%
- Progress hopelessly slow!

Standard 2-stage design

First stage 20 patients:

- Stop if ≤ 4 or ≥ 9 responses
- Else second set of 20

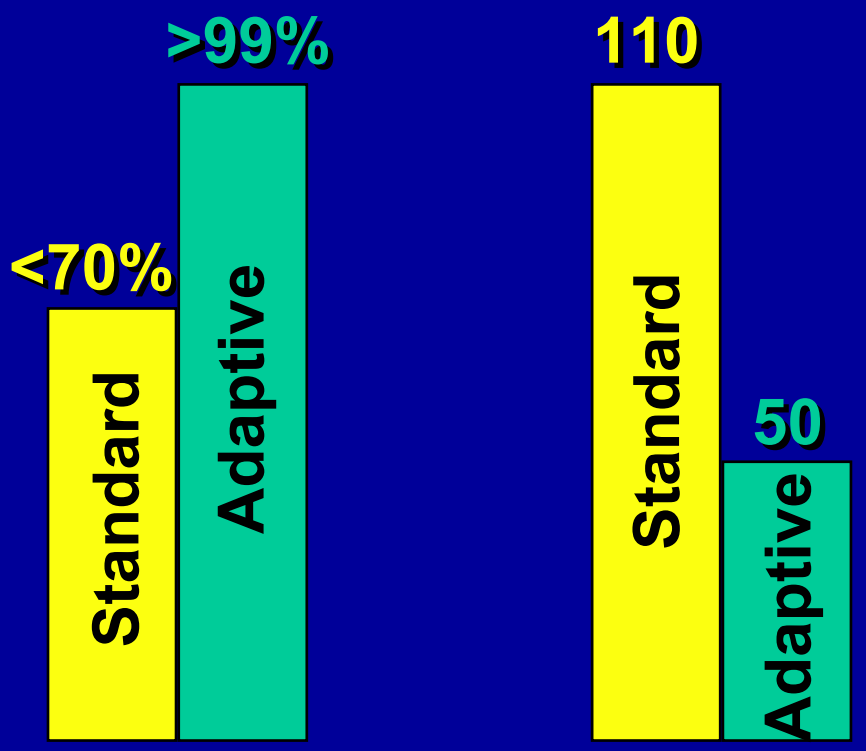
An adaptive allocation

- When assigning next patient, find r = P(rate ≥ 20%|data) for each drug
- Assign drugs in proportion to r
- Add drugs as become available
- Drop drugs that have small r
- Drugs with large r → phase 3

Suppose 10 drugs, 200 patients

9 drugs have mix of RRs 20% & 40%, 1 has 60% ("nugget")

Identify nugget …

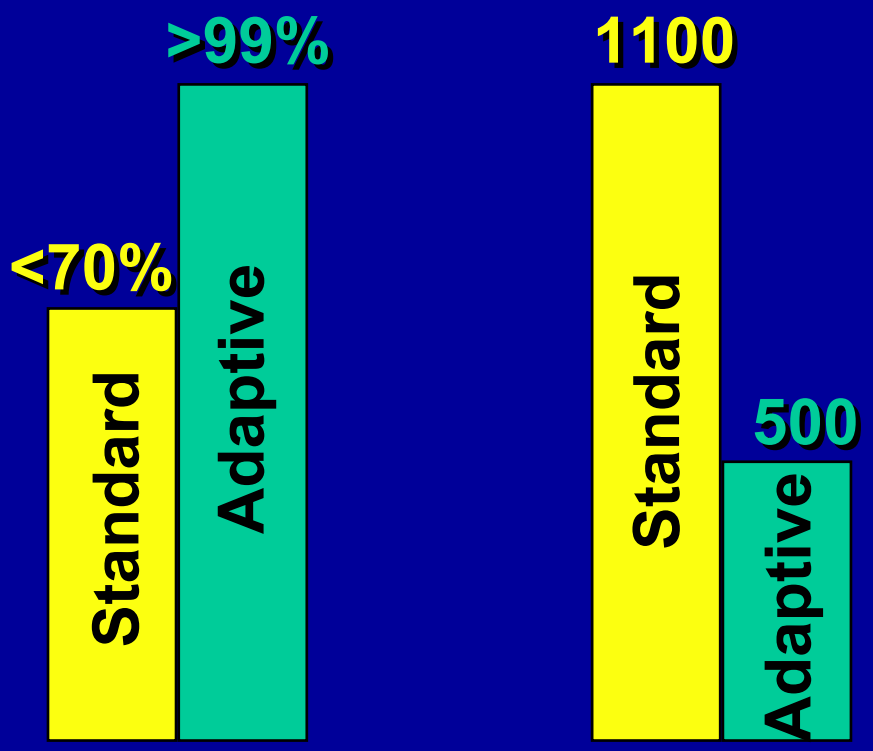


Adaptive also better at finding "40%", sooner

Suppose 100 drugs, 2000 patients

99 drugs have mix of RRs 20% & 40%, 1 has 60% ("nugget")

Identify nugget …



Adaptive also better at finding "40%", sooner

Consequences

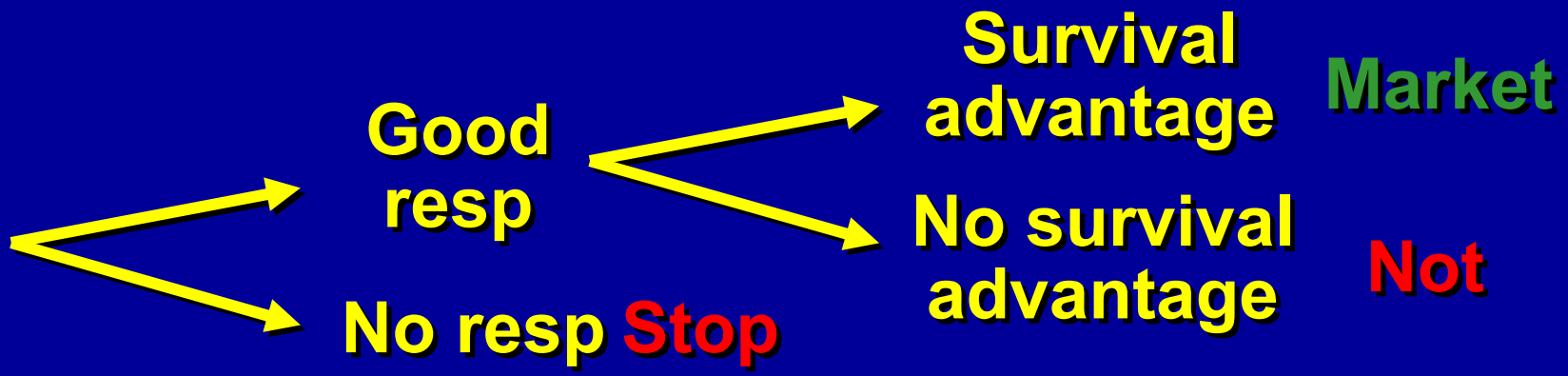
- Treat pts in trial effectively
- Learn quickly
- Attractive to patients, in and out of the trial
- Better drugs identified sooner; move through faster

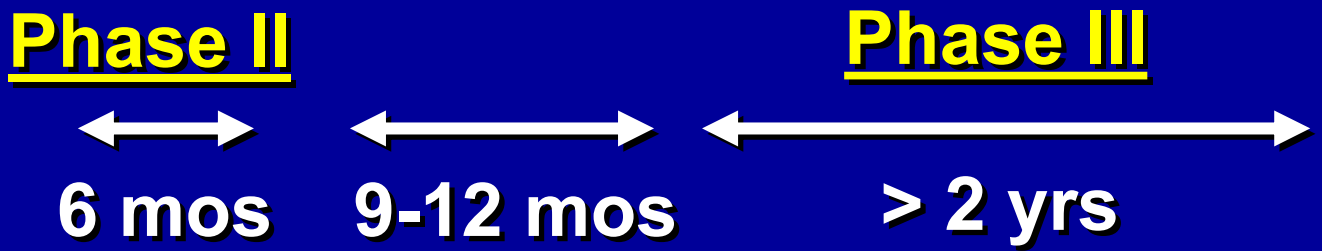
SEAMLESS PHASES II/III*

- Early endpoint (tumor response, biomarker) *may* predict survival?
- May depend on treatment
- Should model the possibilities
- Primary endpoint: survival
- But observe relationships

*Inoue, et al (2002 Biometrics)

Conventional drug development





Seamless phase II/III

Seamless phases

- Phase II: 1 or 2 centers; 10 pts/mo, randomize E vs C
- If pred probs “look good,” expand to Phase III: Many centers; 50 pts/mo (Initial centers continue accrual)
- Max n = 900

[Single trial: survival data combined in final analysis]

Early stopping

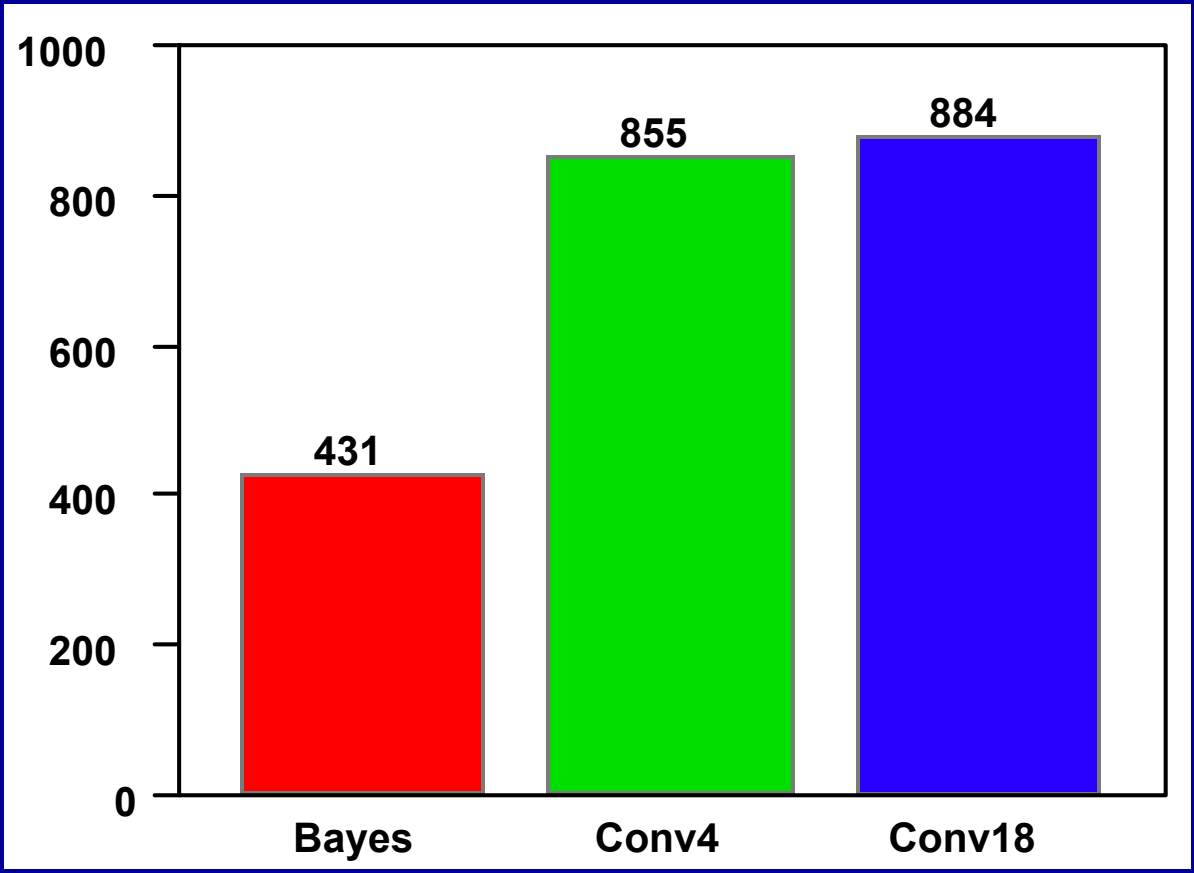
- Use pred probs of stat sig
- Frequent analyses (total of 18) using pred probs to:
  - Switch to Phase III
  - Stop accrual for
    - Futility
    - Efficacy
  - Submit NDA

Comparisons

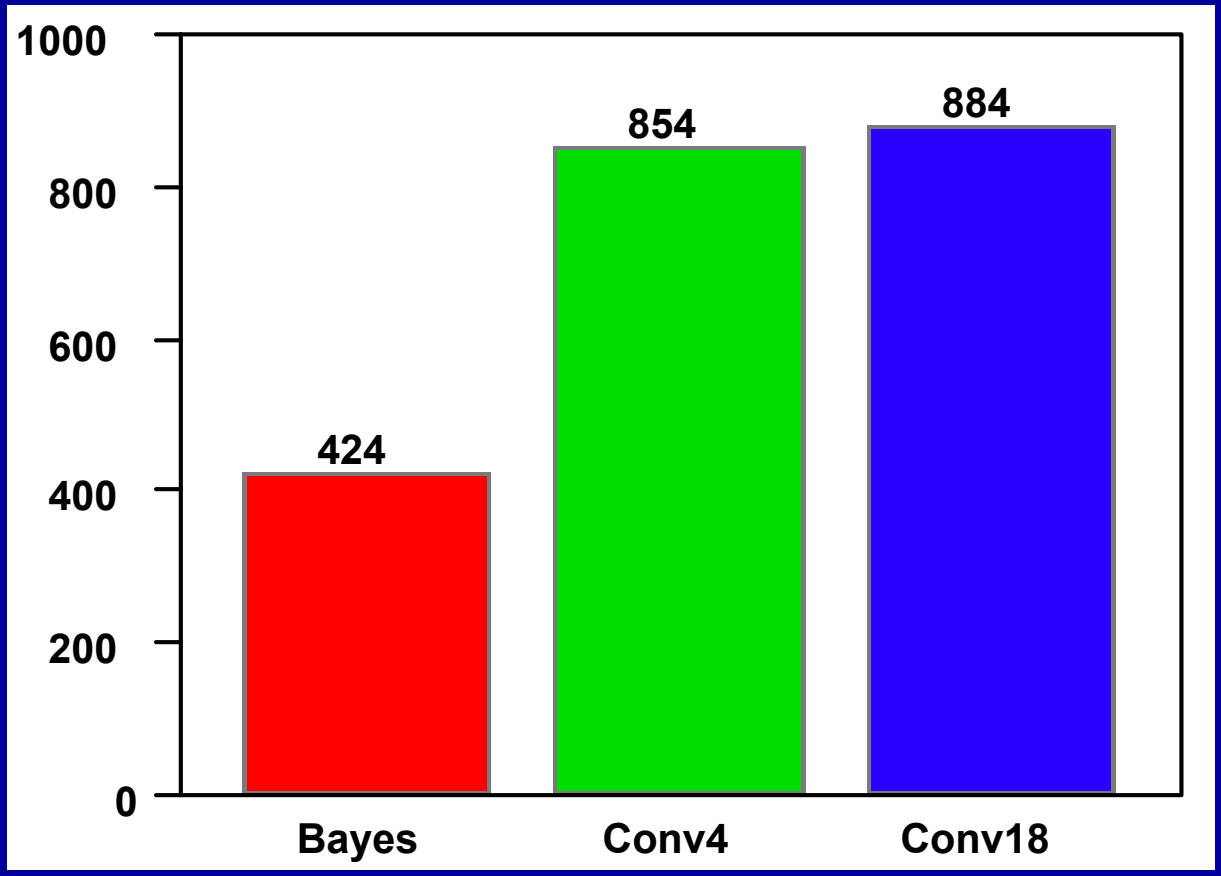
Conventional Phase III designs: Conv4 & Conv18, max N = 900

(O'Brien-Fleming, same power as adaptive design)

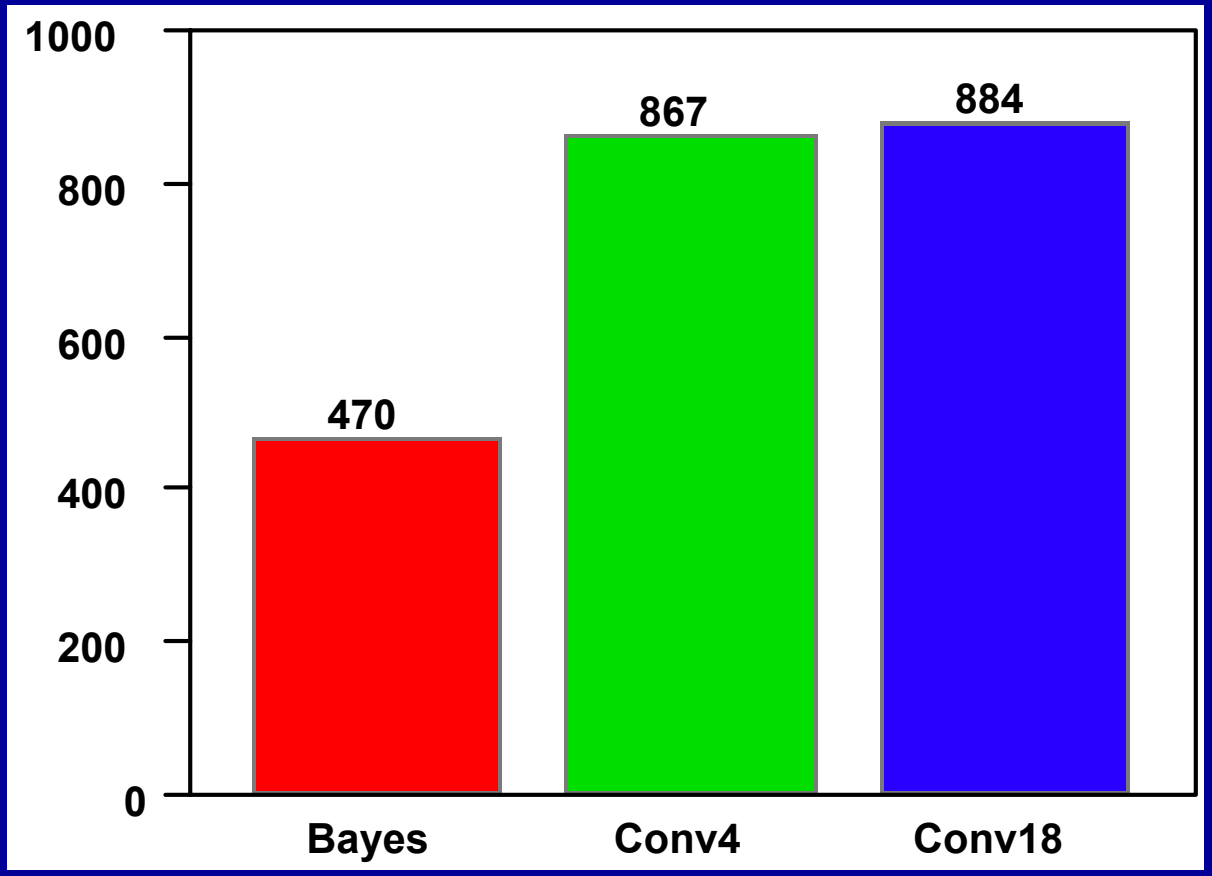
Expected N under H_0



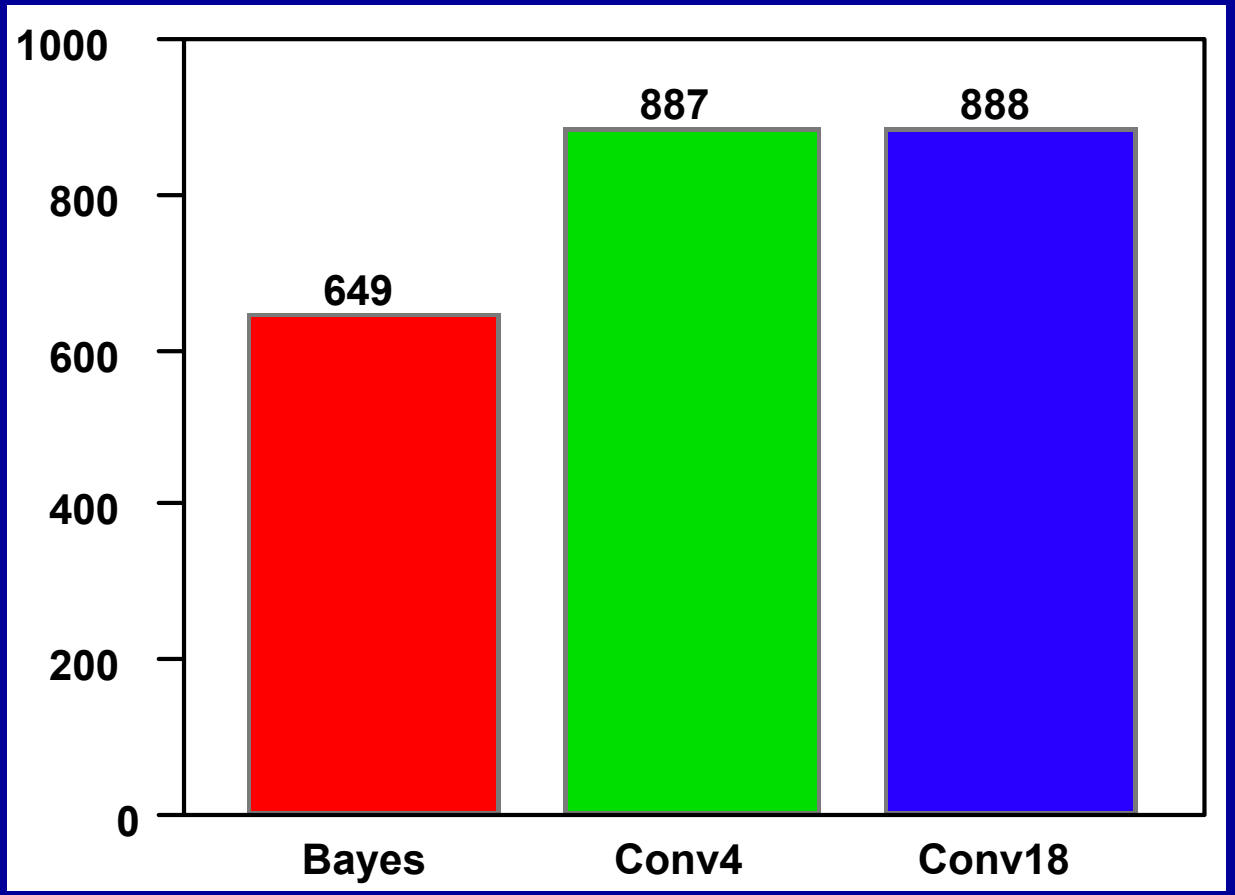
Expected N under H_0*



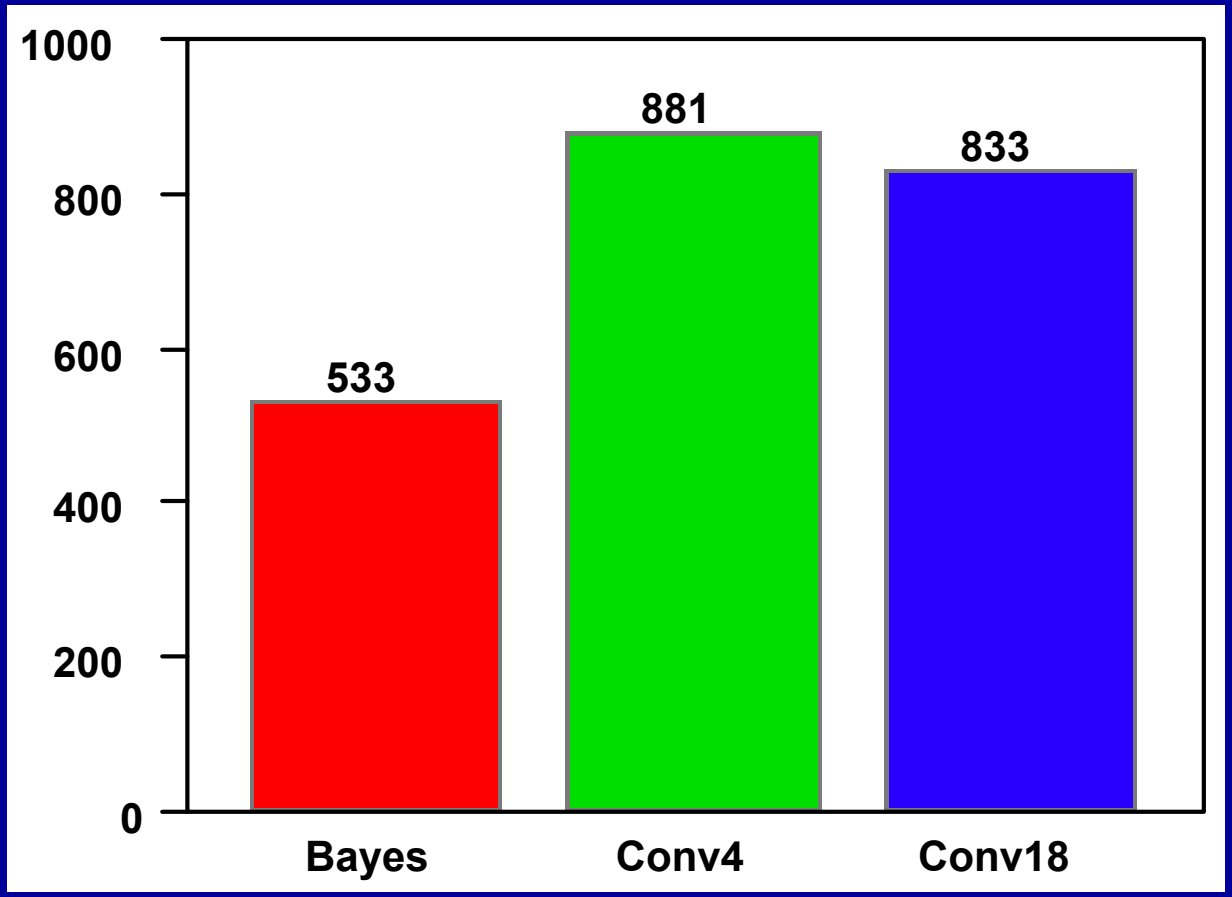
Expected N under H_0**



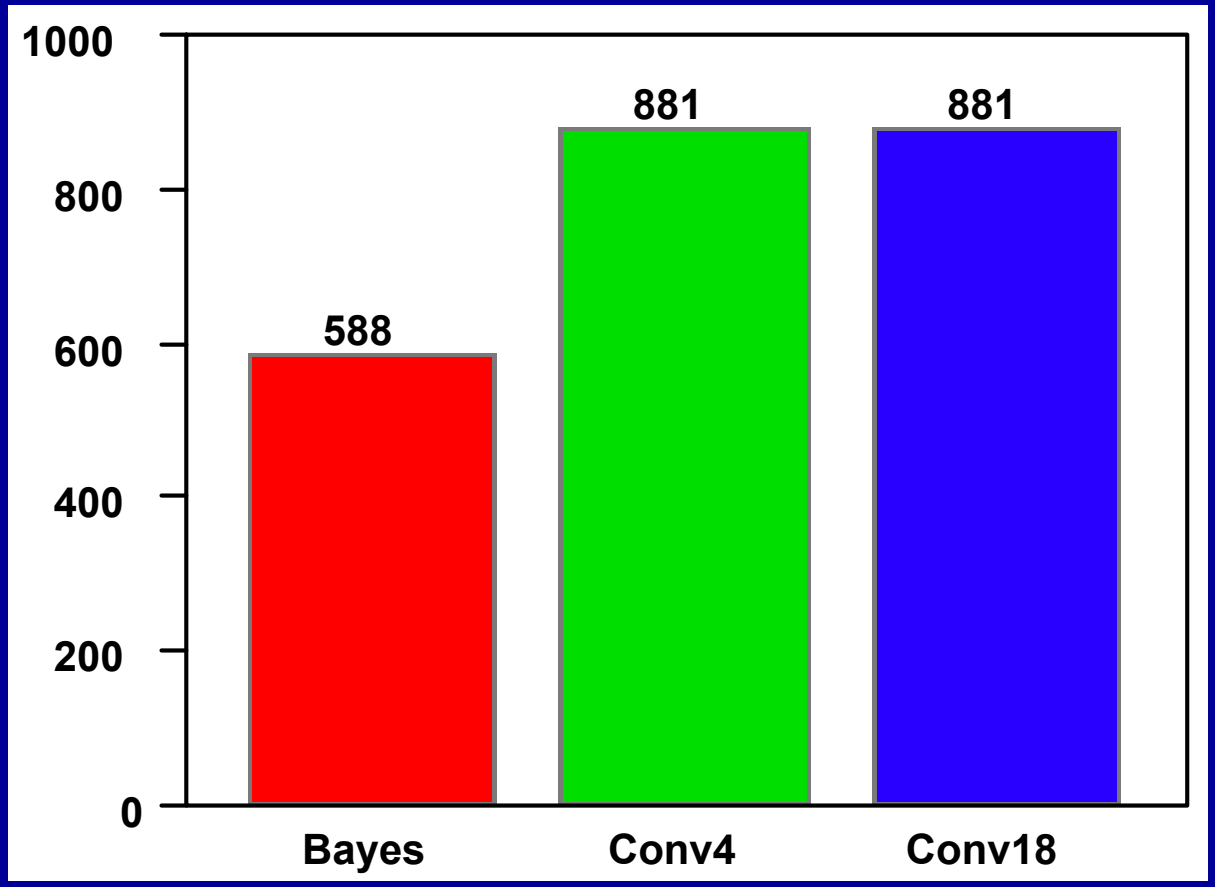
Expected N under H_1



Expected N under H_1*



Expected N under H_1**



Benefits

- Duration of drug development is greatly shortened under adaptive design:
  - Fewer patients in trial
  - No hiatus for setting up phase III
  - All patients used for
    - Phase III endpoint
    - Relation between response & survival

Possibility of large N

- N seldom near 900
- When it is, it's necessary!
- This possibility gives Bayesian design its edge

  [Other reason for edge is modeling response/survival]

Consequences

- Treat pts in trial effectively
- Learn quickly
- Attractive to patients, in and out of the trial
- Better drugs identified sooner; move through faster

Adaptive designs for proof of concept and dose-finding trials

Mike K Smith[1], Mark F. Morris[1], Ieuan Jones[1], Andy P. Grieve[1], Keith Tan[2]

[1] Biostatistics and Reporting, PGRD Sandwich;
[2] Clinical Sciences, PGRD Sandwich

Mike K. Smith at U.S. JSM

- Proof of concept:
  - If drug works well: want dose-response, efficacy & safety
  - If doesn't work: KILL IT!
- Post-herpetic Neuralgia (PHN): Pain after shingles lasting >3 mos

Pain endpoints

- 4-week to 12-week endpoints
- "Average pain score" over *last* 7 days' pain scores
- Change from baseline
- Perhaps not monotonic due to dropouts

Case study: PD-217,014

- Experiment treatment for PHN
- Preclinical tests suggest 2 outcomes:
  - *NO DIFFERENCE* from conventional Rx based on gold standard preclinical test
  - *HUGE* effect based on novel preclinical test
- Normal Dynamic Linear Model (NDLM) as used in stroke dose-response study*

* ref: Berry DA, Müller P, Grieve AP, Smith MK, Parke T, Blazek R, Mitchard N and Krams M. Adaptive Bayesian designs for dose-ranging trials. In *Case Studies in Bayesian Statistics V*, (Ed: Carlin B, Carriquiry A, Gatsonis C,Gelman A, Kass RE, Verdinelli I, West M.) Springer Verlag, (2002) pp99-181.

Design

- Objective: dose that's 1.5 pts better than placebo
- Doses, sample size:
  - Max N = 280
  - 7 active doses + placebo + active comparator
  - N=35/dose; oral dosage form; parallel group
  - Equal allocation to all treatments
  - Interim analyses to drop ineffective doses or stop study
- Continue to max N if any dose ≥ 1.5 pts better than placebo

Operating characteristics: power (average N)

Simulation Scenario	No dose response	Modest improvement	Clinically important improvement
0 interims	0 (280)	0.42 (280)	0.83 (280)
1 interim	0 (156)	0.41 (247)	0.84 (265)
2 interims	0 (118)	0.42 (225)	0.81 (256)
3 interims	0 (97)	0.41 (209)	0.83 (249)

Case study: Pfizer A1461006

- Europe, Canada, Australia
- Started Feb 2004
- First interim Aug 2004
- N=80 subjects evaluable, 110 randomised, 133 recruited

Mike Smith

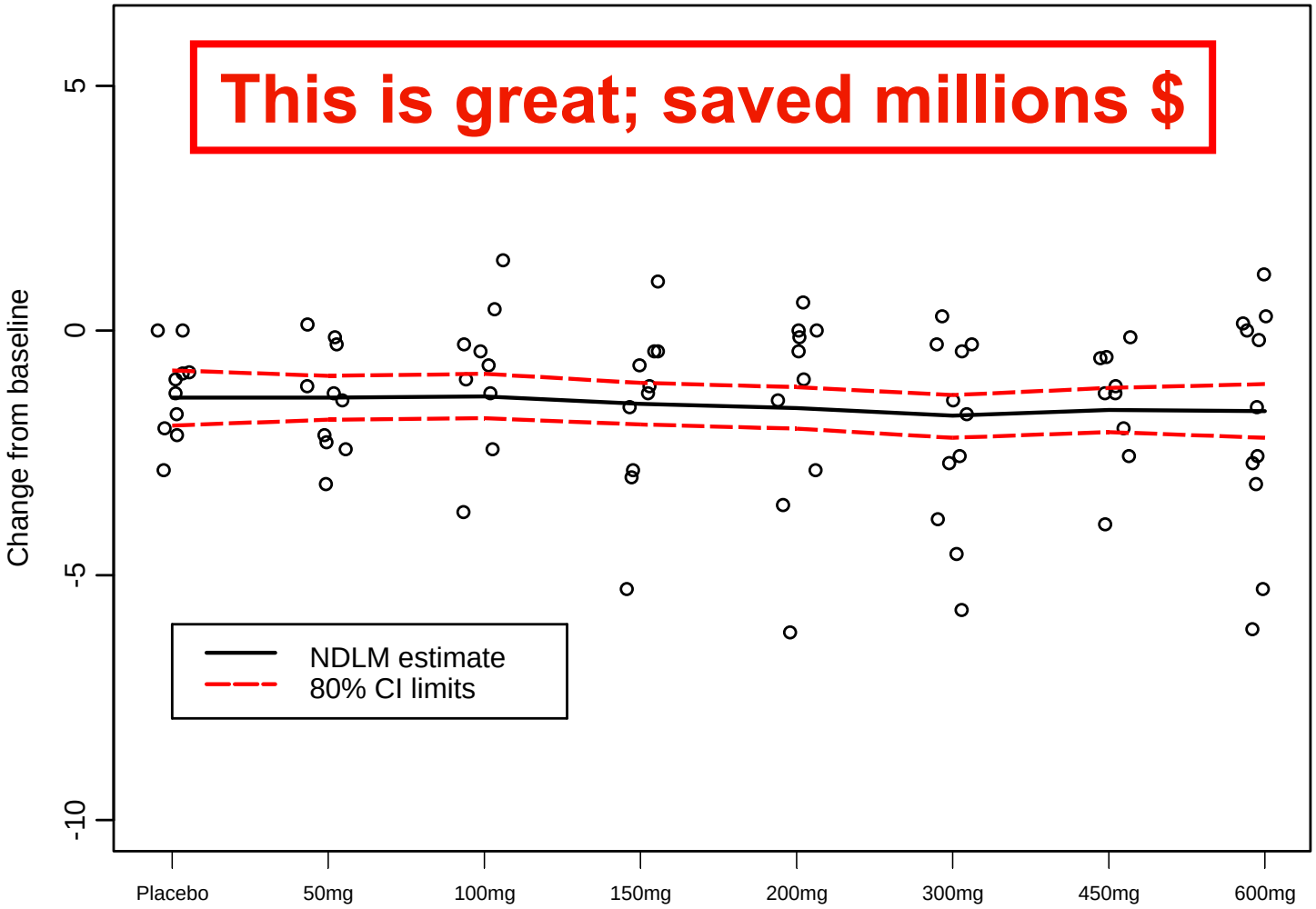
Case study: results @ N=80

			80% Credible Interval		
Dose	NDLM Estimated Dose Effect (Active-Pbo)	SE	Lower	Upper	Posterior prob of futility
50mg	**-0.0**	**0.42**	**-0.86**	**0.90**	**1.00**
100mg	**0.0**	**0.53**	**-0.99**	**1.16**	**1.00**
150mg	**-0.1**	**0.55**	**-1.23**	**0.99**	**1.00**
200mg	**-0.2**	**0.57**	**-1.33**	**0.92**	**1.00**
300mg	**-0.4**	**0.57**	**-1.52**	**0.75**	**1.00**
450mg	**-0.3**	**0.59**	**-1.40**	**0.93**	**1.00**
600mg	**-0.3**	**0.64**	**-1.53**	**1.00**	**1.00**
Number of evaluable subjects in analysis: N = 80					

Mike Smith

Case study: results @ N=80

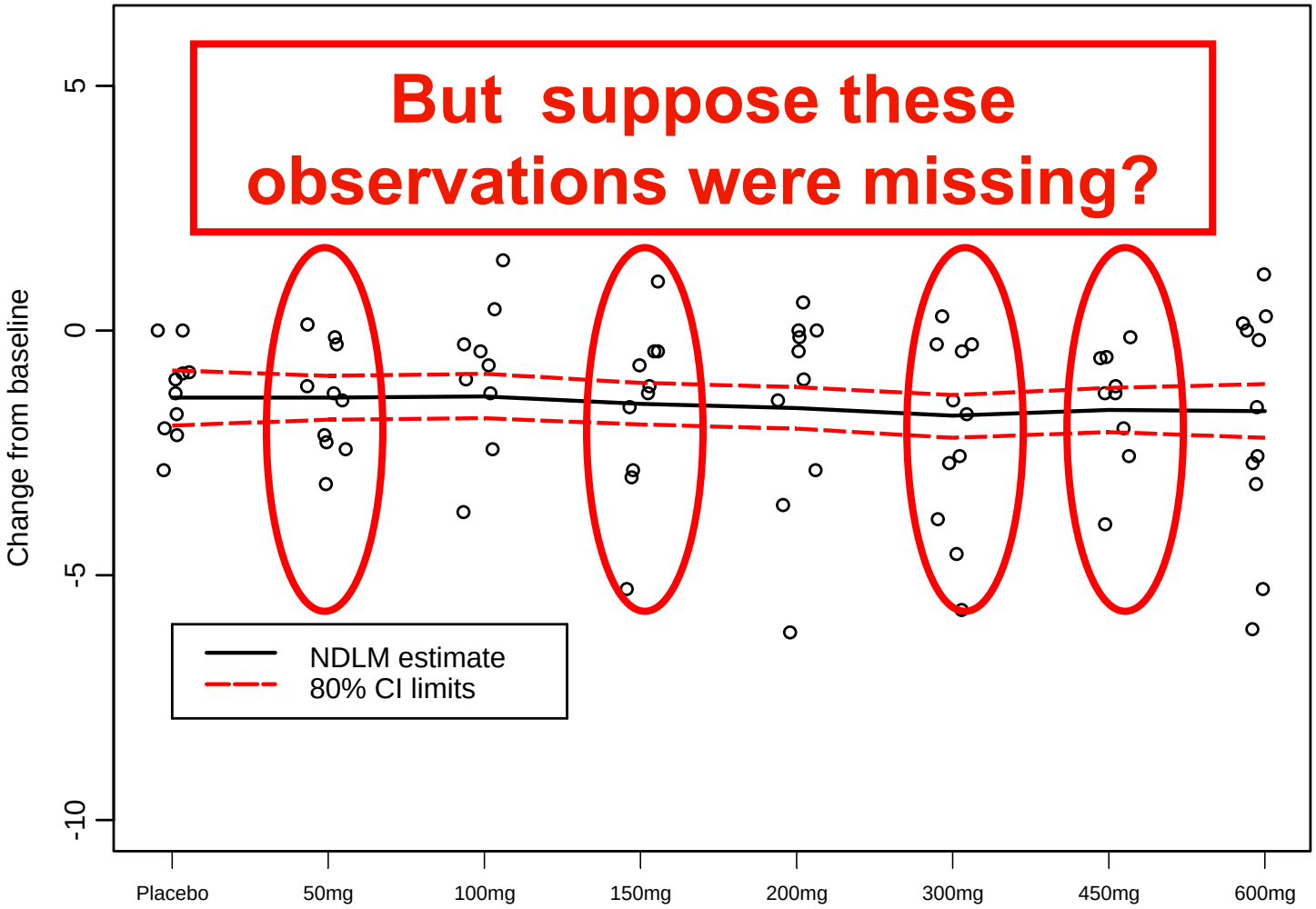
NDLM dose response curve with observed data. Change from baseline in mean pain score



THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER

Case study: results @ N=80

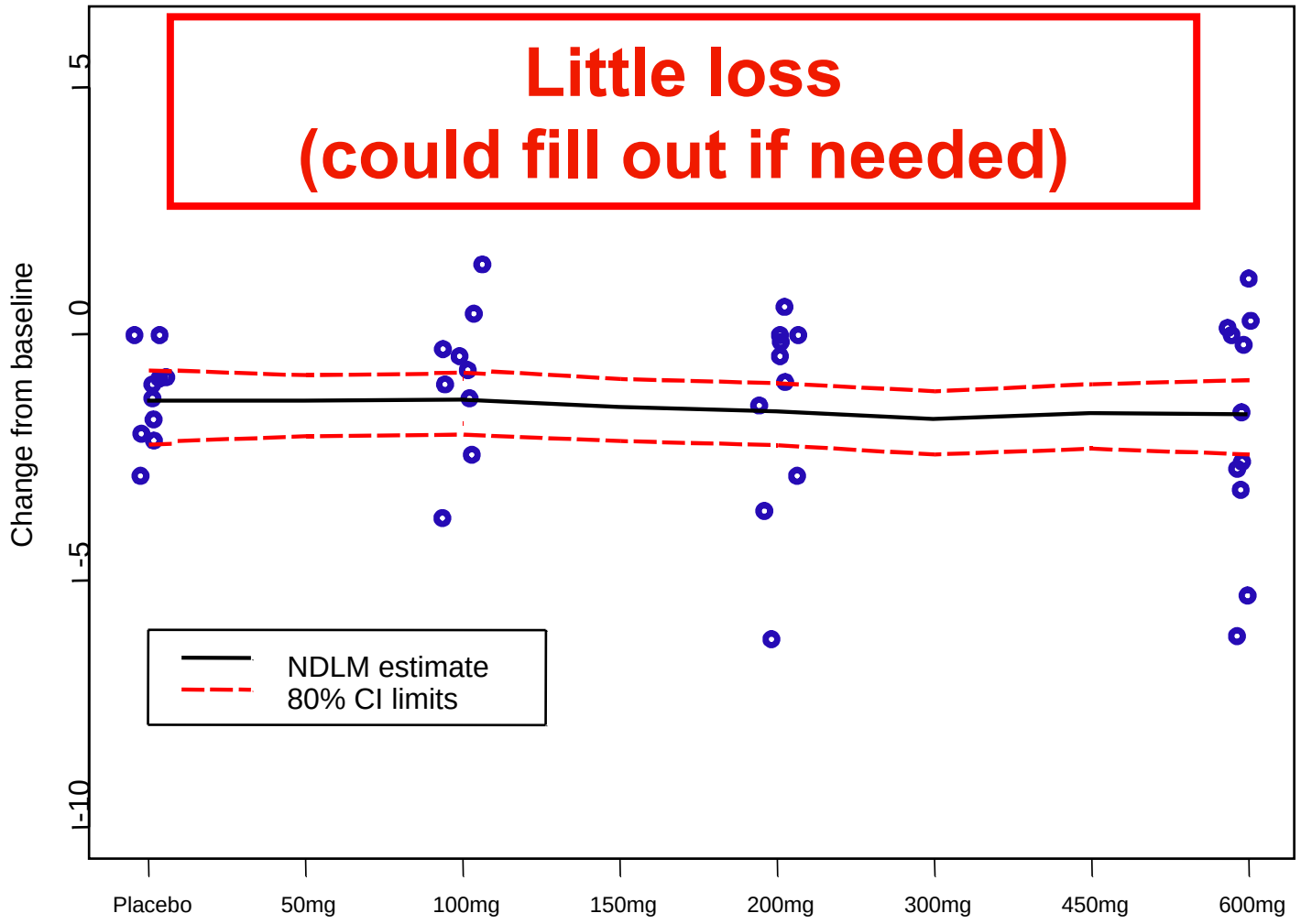
NDLM dose response curve with observed data. Change from baseline in mean pain score



THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER

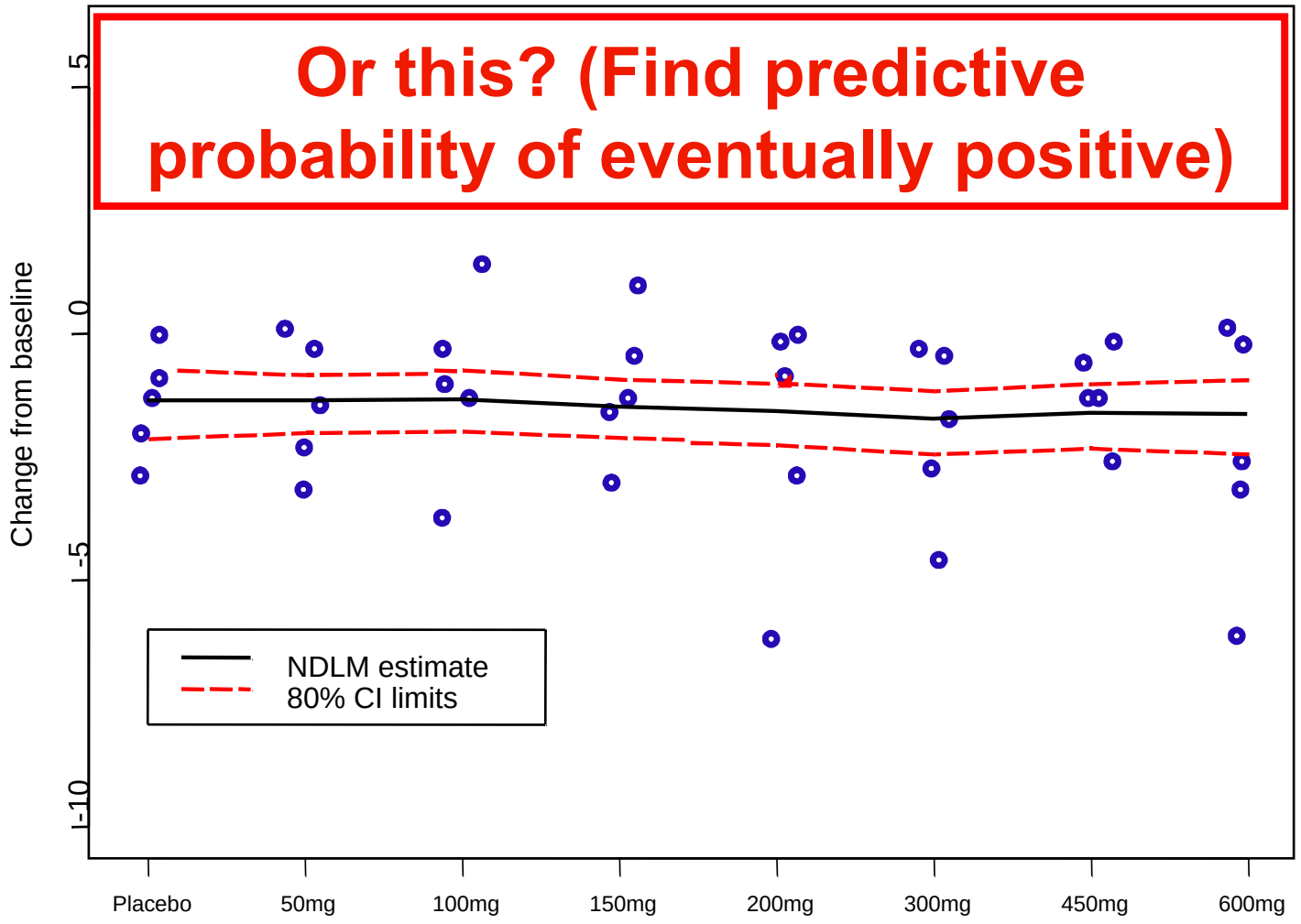
Case study: results @ N=~~80~~ 40

NDLM dose response curve with observed data. Change from baseline in mean pain score



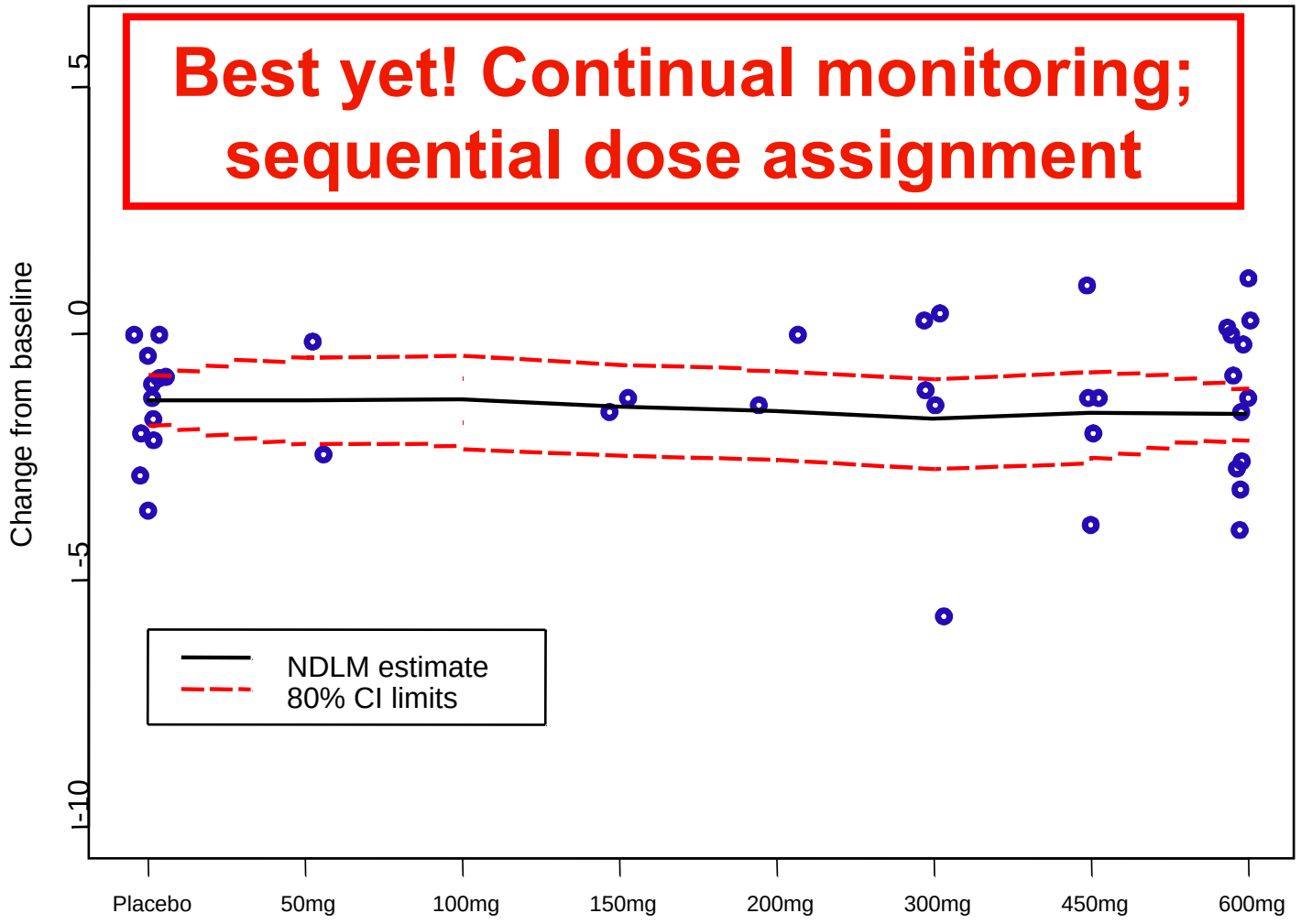
Case study: results @ N=~~80~~ 40

NDLM dose response curve with observed data. Change from baseline in mean pain score



Case study: results @ N=~~80~~ 40

NDLM dose response curve with observed data. Change from baseline in mean pain score



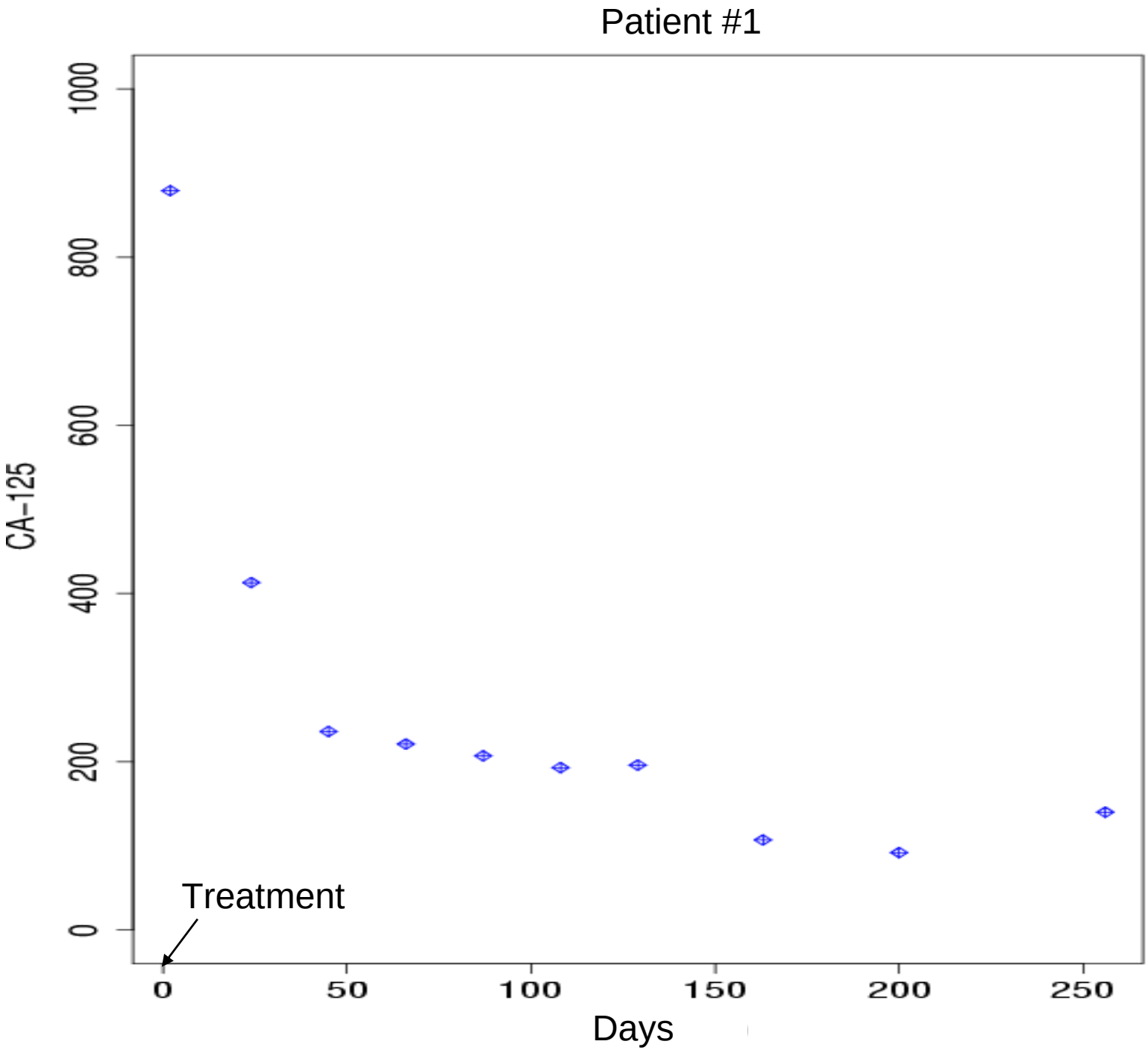
Case study conclusions

- 133 subjects recruited (280 max)
- Little dose effect; little drug effect
- Positive control group gave expected effect
- Pfizer stopped development
- Cost savings > $2M

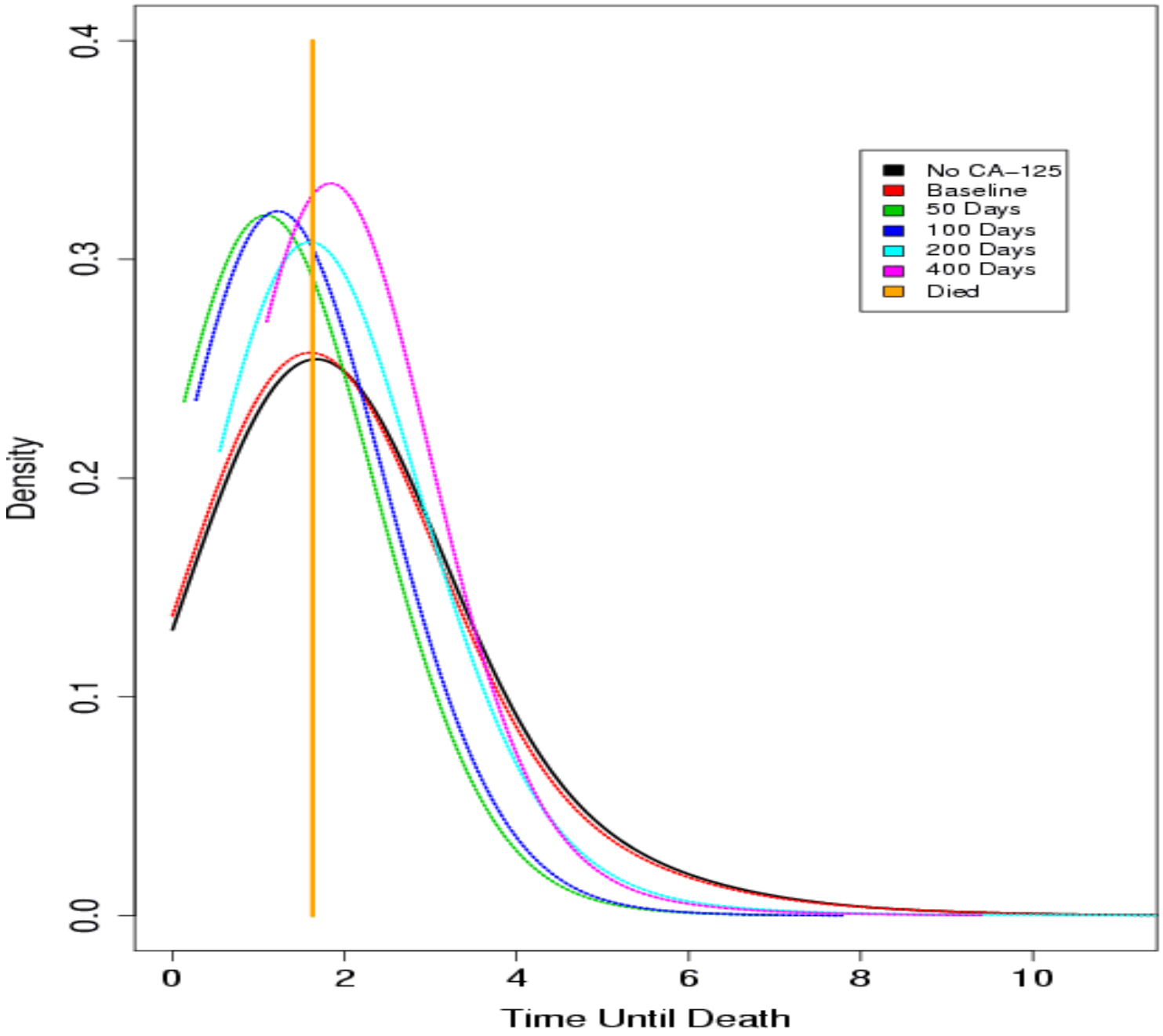
CA125 data & predictive distributions of survival for two of many patients* ——>

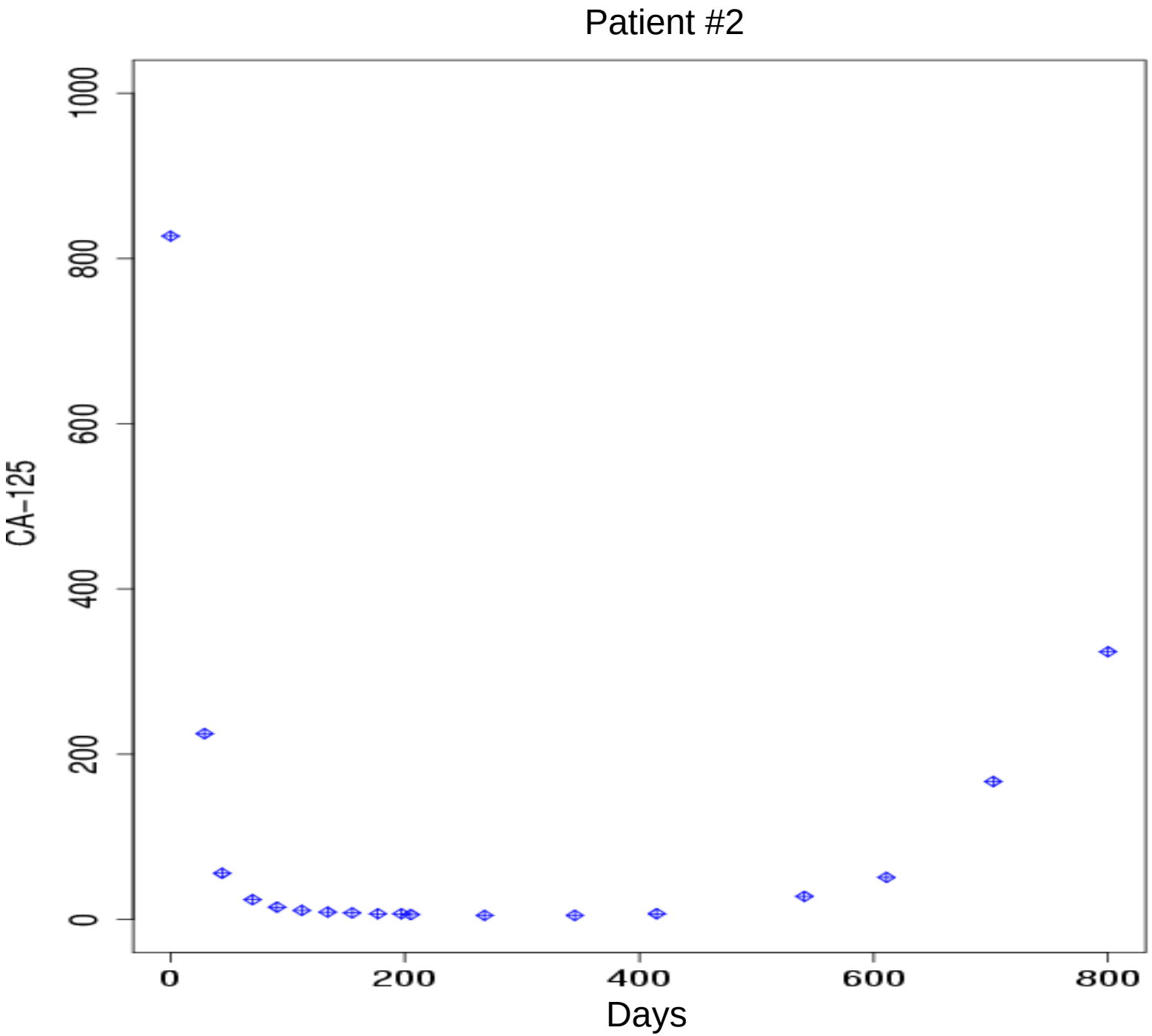
*Modeling due to Scott Berry
<scott@berryconsultants.com>

THE UNIVERSITY OF TEXAS
MD ANDERSON
CANCER CENTER

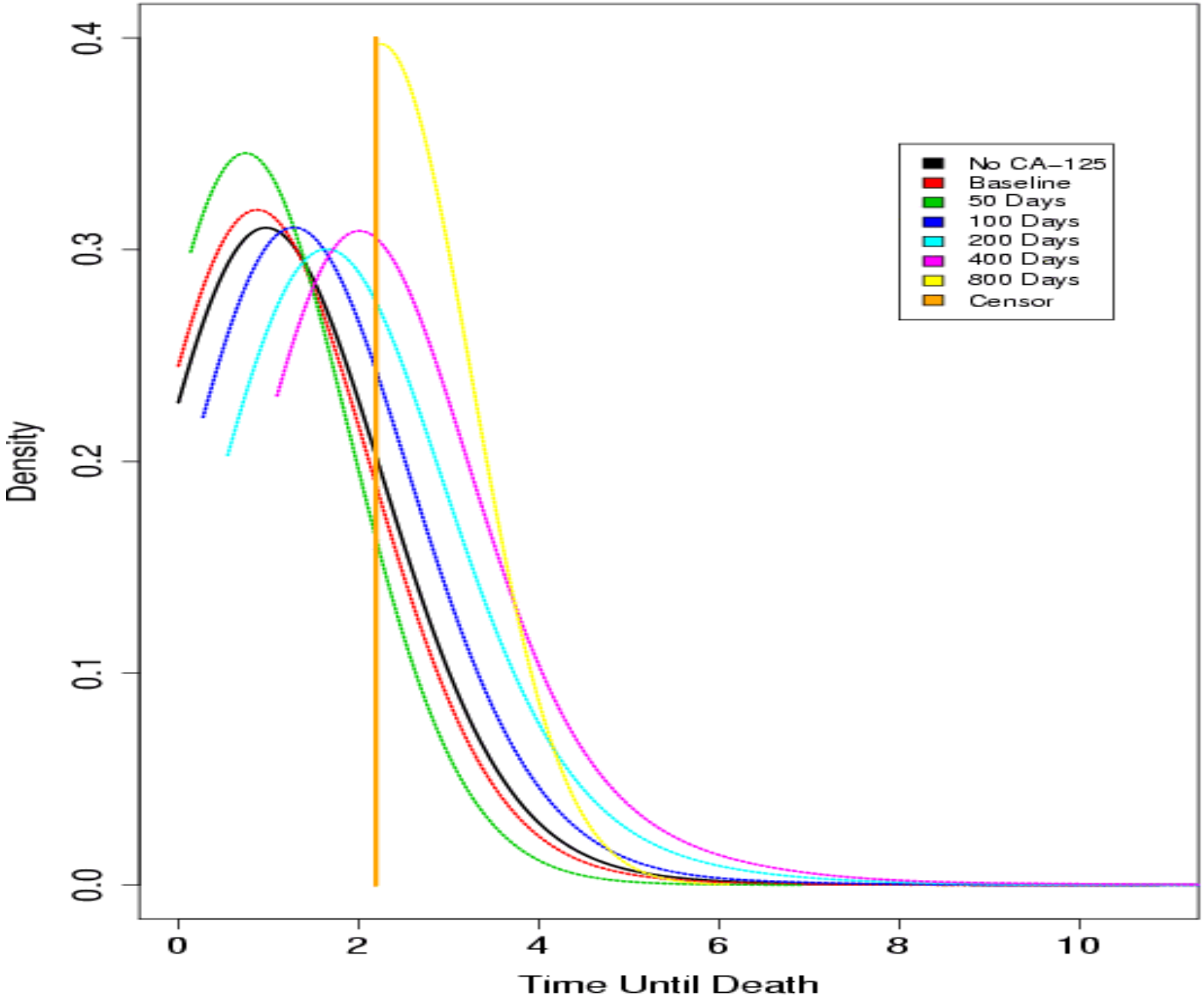

Patient #1
CA-125
1000
800
600
400
200
0
Treatment
0
50
100
150
200
250
Days

Patient #1


Density
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
Time Until Death
No CA–125
Baseline
50 Days
100 Days
200 Days
400 Days
Died


Patient #2
CA-125
0
200
400
600
800
1000
0
200
400
600
800
Days

Patient #2


Density
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
Time Until Death
No CA–125
Baseline
50 Days
100 Days
200 Days
400 Days
800 Days
Censor

Methods

- Analytical
- Multiple imputation

EXAMPLES

- Predictive probabilities
- Adaptive randomization
  - Drug trial in AML
  - Phase I/II cancer trial
  - Drug screening process
  - Dose finding
  - PoC study in pain
- Modeling early & late endpoints