Bayesian Adaptive CNS Trials

Donald A. Berry
dberry@mdanderson.org

THE UNIVERSITY OF TEXAS
MD ANDERSON
CANCER CENTER

Baltimore, July 18, 2006

Janet Woodcock, CDER Director :

"Improved utilization of adaptive and Bayesian methods" could help resolve low success rate of and expense of phase III clinical trials

MSA Partners LLC

2006年6月30日発行 第138号

MSA Partners LLC
516 Fifth Avenue, 9th Fl.
10036
760
pr.com
http://www.msapr.com

The Coming Bayesian Tsunami of Clinical Development

ベイズ統計学パート2

臨床開発にベイズ統計学の「津波」が到来

テキサス大学アンダーソン癌センター生物統計学科主任、ドナルド・A・ベリー博士

30年以上もの間、ドナルド・ベリー博士は臨床試験のデザインと分析におけるベイズ統計学の利用を提唱してきた。この発想は当初、行政機関や製薬業界、および臨床研究における教育を受けておらず認識の乏しい統計学者から無視された。しかし、同氏が統計学の薬剤開発への適用に情熱を傾け続けた結果、数十年後、FDAと製薬会社が同氏に耳を傾け、ベイズ的アプローチの利点を理解するようになった。ベリー博士にこの変化について聞いた。

ドナルド・A・ベリー博士

――薬剤開発のデザインと分析において、統計は乱用されているとお考えですか。

ベリー　統計学者を含め、臨床試験のデザ

OUTLINE

- Bayesian adaptive design?
- Predictive probabilities in design
- Adaptive dose-finding: A stroke trial

Practical Advantages of Bayes

1. On-line learning
2. Predictive probabilities
3. Modeling
   - Hierarchical
   - Longitudinal
4. Decision analysis

Current Use of Bayesian Adaptive Designs

- MDACC (> 300 trials)
- Many device companies (> 20 PMAs, many 510(k)s)
- All top drug companies; many biotechs

Some Bayesian Device Applications

- Orthopedics (esp. spinal implants)
- Diagnostics & screening
- Stents
- Shunts
- Defibrillators
- Bronchial thermoplasty
- Ablation catheters
- PFO closure
- Contraceptives
- Neurostimulation

Some areas of application of Bayesian adaptive drug trials

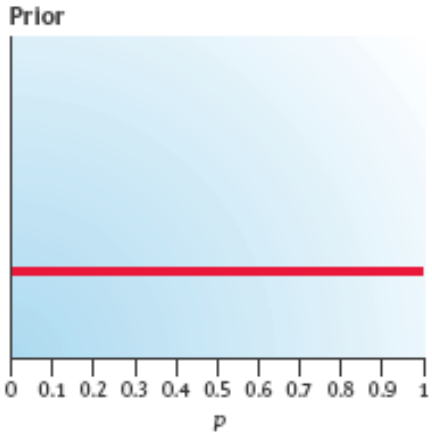
- Oncology
- Migraine
- Rh Arthritis
- Lupus
- Sepsis
- Diabetes
- Obesity
- Stroke
- Spinal Cord Inj
- HIV
- Hep C
- Pre-term labor
- Constipation
- Micturition
- Alzheimer's
- Parkinson's

Bayesian Adaptive Trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Ramping up accrual

Bayesian Updating

- Paired observations, T vs C
- P(S) = P(T wins pair)
- H_0: P(S) = 1/2
- Data: SSFSS FSSSF
  SFSSS SS



Prob: 8/12

Prob: 4/12

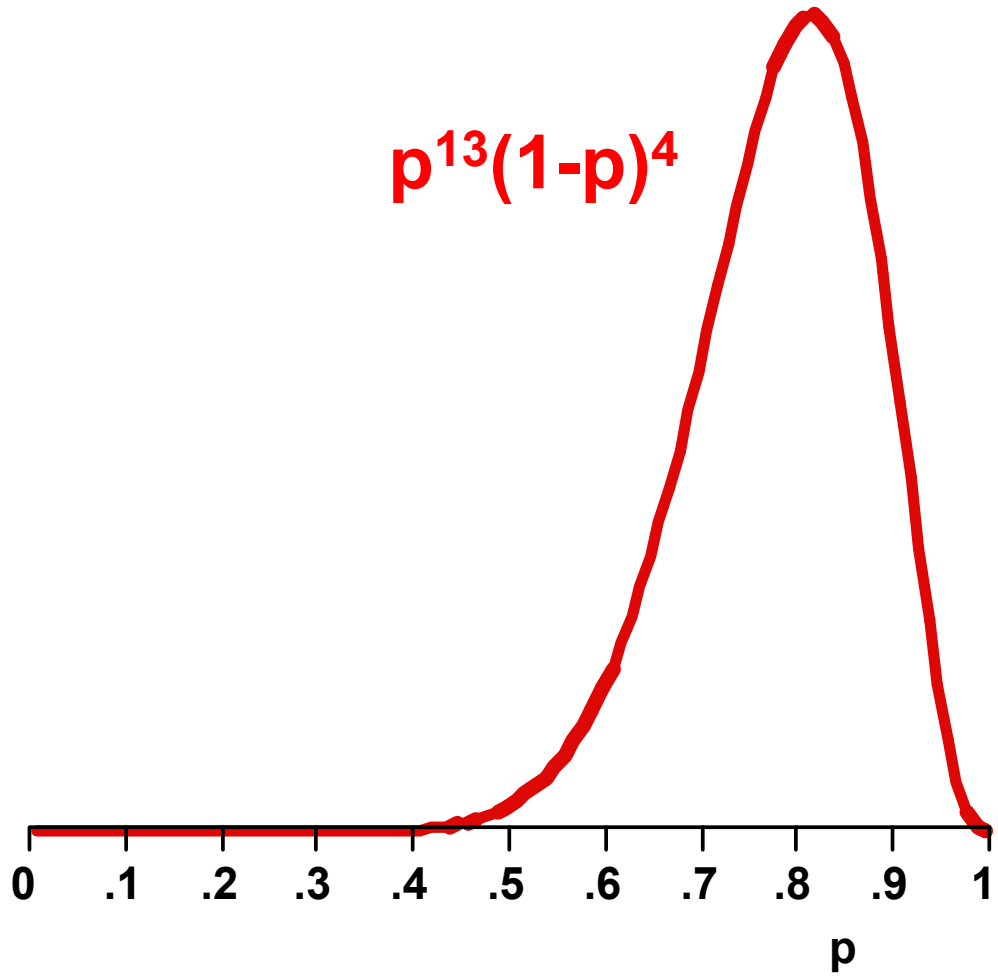
Predictive Probabilities

- Essential for monitoring trials
- Critical component of experimental design

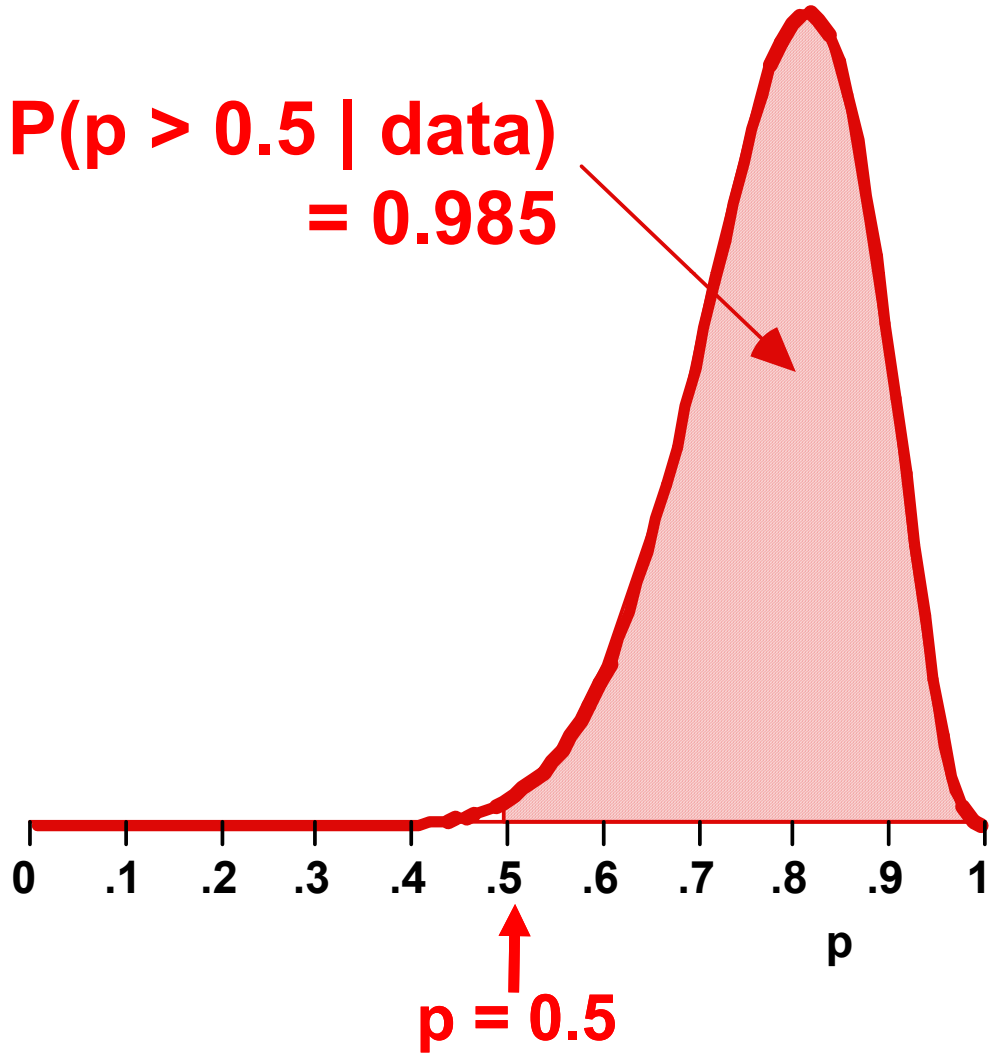
"We must ask where we are and whither we are tending."

— Abraham Lincoln

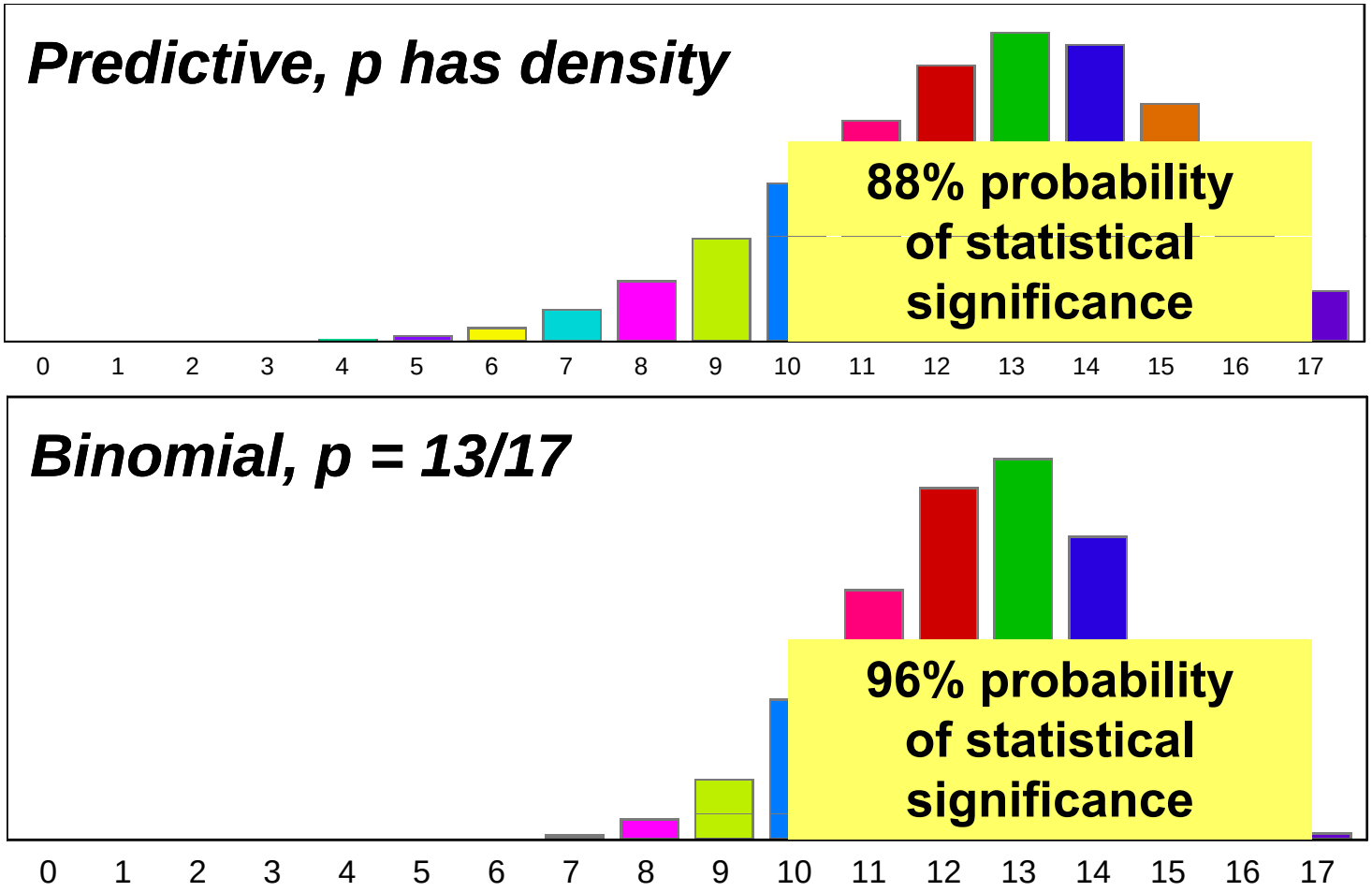
Current (posterior) for p = P(S)



Posterior Probability that p > 0.5 is Area Under Curve



Probabilities for # successes in next 17 observations



Why Bayes?

- Smaller trials (usually!)
- More accurate conclusions
- Objective can include better treatment of patients in trials
- Complicated trials, but control Type I error via simulation

Predicting Trial Results

- Simulate
- Model uncertainty
- Incorporate info (Bayesian-wise) on various outcomes
- Model relationships among early and late endpoints
- Consider alternative designs

Prediction in Adjuvant Breast Cancer Trial

- Patients ≥ 65 yrs old
- Capecitabine (Xeloda®) vs standard therapy (CMF or AC)
- Hypothesis: Noninferiority
- N = 600 to 1800

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 14, 2009 VOL. 360 NO. 20

Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer

Hyman B. Muss, M.D., Donald A. Berry, Ph.D., Constance T. Cirrincione, M.S., Maria Theodoulou, M.D., Ann M. Mauer, M.D., Alice B. Kornblith, Ph.D., Ann H. Partridge, M.D., M.P.H., Lynn G. Dressler, Ph.D., Harvey J. Cohen, M.D., Heather P. Becker, Patricia A. Kartcheske, B.S., Judith D. Wheeler, M.P.H., Edith A. Perez, M.D.,

A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients.

BACKGROUND

Older women with breast cancer are underrepresented in clinical trials, and data on the effects of adjuvant chemotherapy in such patients are scant. We tested for the noninferiority of capecitabine as compared with standard chemotherapy in women with breast cancer who were 65 years of age or older.

METHODS

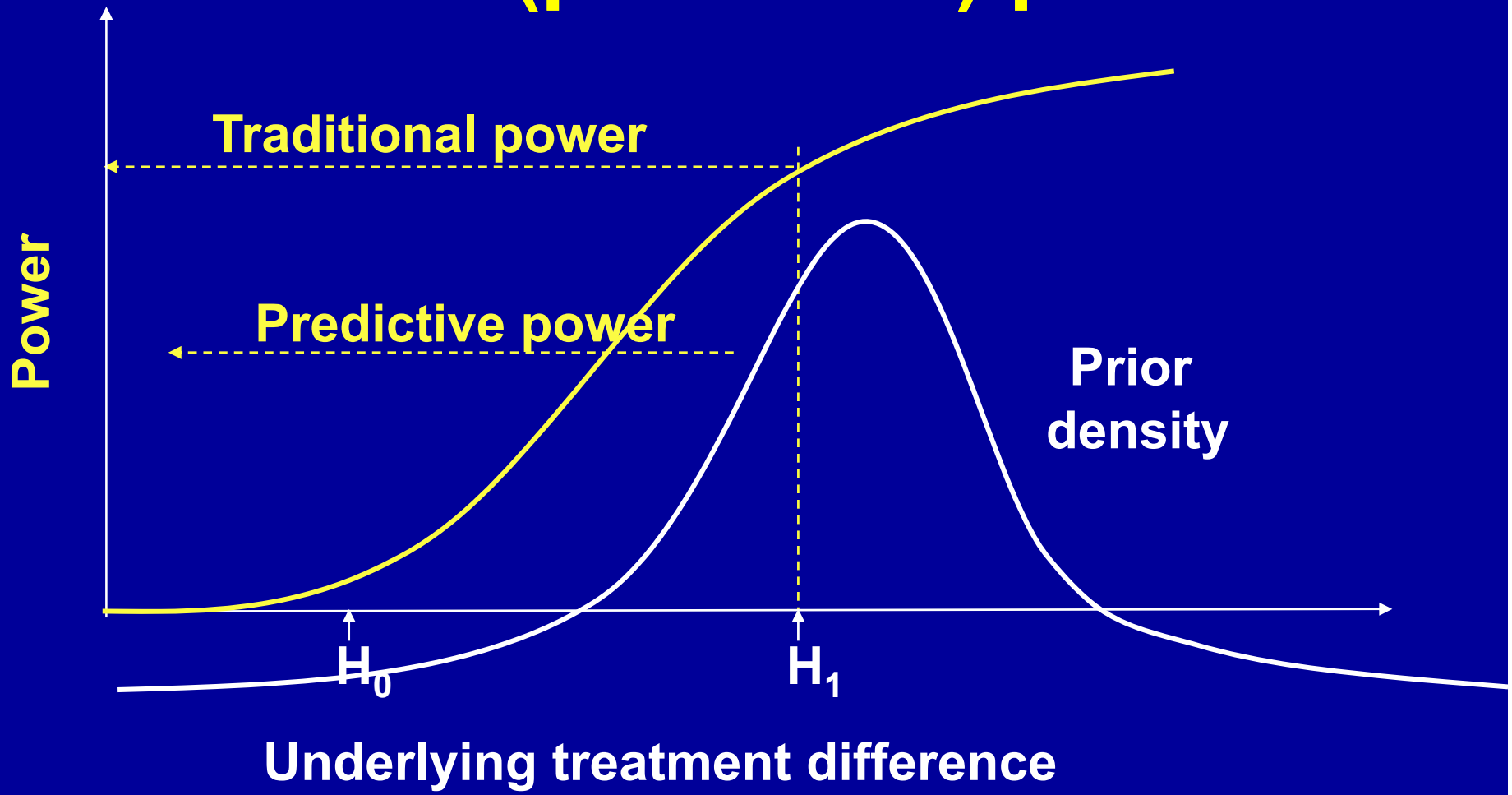
We randomly assigned patients with stage I, II, IIIA, or IIIB breast cancer to standard chemotherapy (either cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide plus doxorubicin) or capecitabine. Endocrine therapy was recommended after chemotherapy in patients with hormone-receptor–positive tumors. A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients. The primary end point was relapse-free survival.

From the University of Vermont, Burlington (H.B.M.); the M.D. Anderson Cancer Center, Houston (D.A.B.); the Cancer and Leukemia Group B (CALGB) Statistical Center, Duke University Medical Center (C.T.C., P.A.K.) and Duke University Medical Center (H.J.C., J.D.W., A.A.M.) — both in Durham, NC; Memorial Sloan-Kettering Cancer Center, New York (M.T., L.N., C.A.H.); CALGB, Chicago (A.M.M., H.P.B.); the Dana–Farber Cancer Institute, Boston (A.B.K., A.H.P., H.J.B., E.P.W.); the University of North Carolina, Chapel Hill (L.G.D.); the North Central Cancer Treat-

A purely statistical reason for sorry performance of drugs in phase III …

Power considerations

True (predictive) power



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

NXY-059 for Acute Ischemic Stroke

Kennedy R. Lees, M.D., Justin A. Zivin, M.D., Tim Ashwood, Ph.D., Antonio Davalos, M.D., Stephen M. Davis, M.D., Hans-Christoph Diener, M.D., James Grotta, M.D., Patrick Lyden, M.D., Ashfaq Shuaib, M.D., Hans-Göran Hårdemark, M.D., and Warren W. Wasiewski, M.D., for the Stroke–Acute Ischemic NXY Treatment (SAINT I) Trial Investigators*

N ENGL J MED 354;6 WWW.NEJM.ORG FEBRUARY 9, 2006

Results of SAINT I:

RESULTS

Among the 1699 subjects included in the efficacy analysis, NXY-059 significantly improved the overall distribution of scores on the modified Rankin scale, as compared with placebo (P=0.038 by the Cochran–Mantel–Haenszel test). The common odds ratio for improvement across all categories of the scale was 1.20 (95 percent confidence interval, 1.01 to 1.42).

Design of SAINT II:

- N = 3200 (up from 1700)
- Power 80% for Odds Ratio 1.20

In SAINT I:

RESULTS

Among the 1699 subjects included in the efficacy analysis, NXY-059 significantly improved the overall distribution of scores on the modified Rankin scale, as compared with placebo (P=0.038 by the Cochran–Mantel–Haenszel test). The common odds ratio for improvement across all categories of the scale was 1.20 (95 percent confidence interval, 1.01 to 1.42).

?

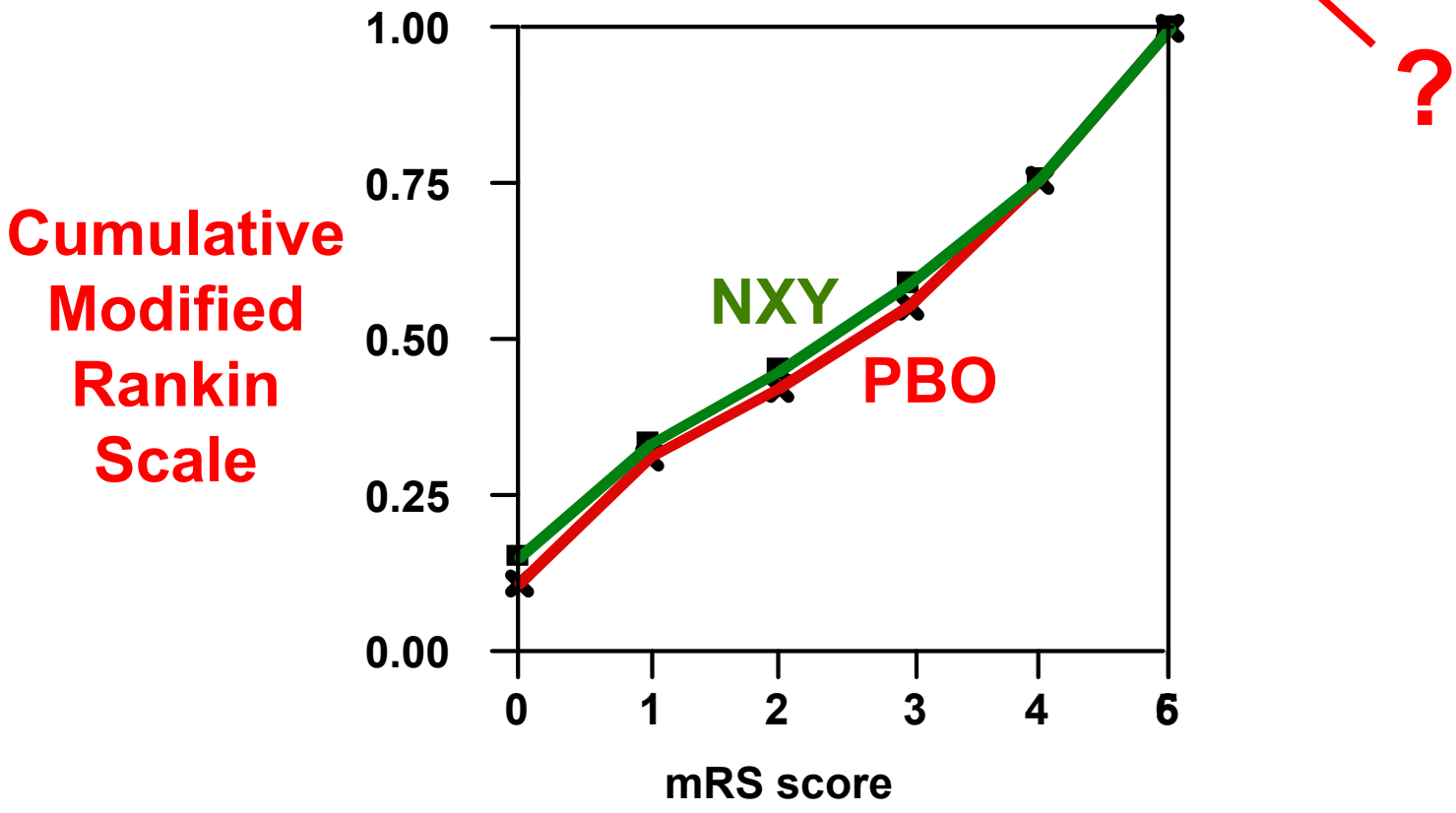


Table 2. Efficacy of the Study Drug at Day 90 or at the Last Rating.*

Outcome Variable	Placebo Group	NXY-059 Group	Difference between NXY-059 and Placebo†	P Value
			% or score (95% CI)	
Modified Rankin scale score (primary end point)				
No. of patients	849	850		
Score — no. (%)				
0	93 (11.0)	131 (15.4)	4	
1	170 (20.0)	153 (18.0)	–2.0	
2	99 (11.7)	97 (11.4)	–0.3	
3	108 (12.7)	121 (14.2)	1.5	
4	175 (20.6)	144 (16.9)	–3.7	
5 (or death)	204 (24.0)	204 (24.0)	0	0.038
Change from baseline in total NIHSS score (coprimary outcome)				
No. of patients	851	851		
Score — LSM ±SE	–1.7±0.5	–1.8±0.5	–0.1 (–1.4 to 1.1)	0.86
Barthel index (dichotomized analysis)				
No. of patients	848	850		
Score, ≥95 — no. (%)	346 (40.8)	368 (43.3)	2.5	0.14
Stroke Impact Scale				
No. of patients	676	669		
Score — LSM ±SE	63.4±1.1	66.2±1.1	2.8 (–0.3 to 5.9)	0.08
EuroQoL EQ-5D (weighted index)				
No. of patients	816	819		
Score — LSM ±SE	0.43±0.013	0.47±0.013	0.04 (–0 to 0.07)	0.06
EuroQoL EQ-5D (VAS)				
No. of patients	671	670		
Score — LSM ±SE	62.0±0.9	64.5±0.9	2.5 (–0.1 to 5.0)	0.05

Not probability of null hypothesis

Advertised power of SAINT II = 80%

Naïve predictive power of SAINT II = 60%

Other info from SAINT I: My pre-probability that SAINT II was going to be positive: 10%

Press release, Oct 26, 2006

"Results from the SAINT II (Stroke Acute Ischemic NXY-059 Treatment) trial: … NXY-059 did not meet its primary outcome of a statistically significant reduction in stroke-related disability, as assessed by the modified Rankin Scale (mRS) (p=0.33, odds ratio 0.94) compared to placebo.

"The company plans no further development of NXY-059 in acute ischemic stroke."

Morals

- P-value is not probability of H_0
- Predictive power < “power”
- Use adaptive design
  - Start trial cautiously
  - Look at the data!!
  - Ask “whither we are tending”
  - Adapt

Types of adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Adaptive randomization
- Identifying responding biomarker profile
- Ramping up accrual

Why adaptive design?

- Smaller trials (usually!)
- More accurate conclusions
- Better treatment of patients in trials, depending on trial goal

REFERENCES

Berry D (1996). *Statistics: A Bayesian Perspective.* Duxbury Press.

Berry D (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery.*

Spiegelhalter D, Abrams K, Myles J (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley.

Websites for JHU/FDA Workshop

www.prous.com/bayesian2004

http://www.cfsan.fda.gov/~frf/bayesdl.html

CDRH Bayesian Guidance

http://www.fda.gov/cdrh/osb/guidance/1601.html

March 12, 2001

BioCentury 100™ Indicators

Week ended 3/9/01

PRICES	VOLUME
2117.52	606.3M shrs
dn 6%	up 14%

The adaptive trial computer simulations described by the authors* start with 16 possible doses or placebo. As data come in each week on patient responses, the next set of doses assigned are allowed to depend on information from patients treated previously.

Bio…

Cov…

Bayes… faster… inform… stances than more familiar frequentist statistics.

Technology Focus

…sian methods among biostatisticians have limited use of the technique. In fact, Bayesian statistics have never been used for regulatory approval of therapeutics.

That is about to

…Center in Houston, and col-…ails an adaptive Bayesian trial …troke drugs. It is to appear in …*es in Bayesian Statistics*".

…ptive trial computer simula-…ribed by the authors start …ssible doses or placebo. As data come in each week on patient responses, the next set of doses assigned are allowed to depend on information from patients treated previously. Thus as data accumulate that a certain dose is more effective, more patients

There's no flexibility in the frequentist straightjacket.

— Donald Berry of M.D. Anderson

*Berry, et al. *Case Studies in Bayesian Statistics* 2001

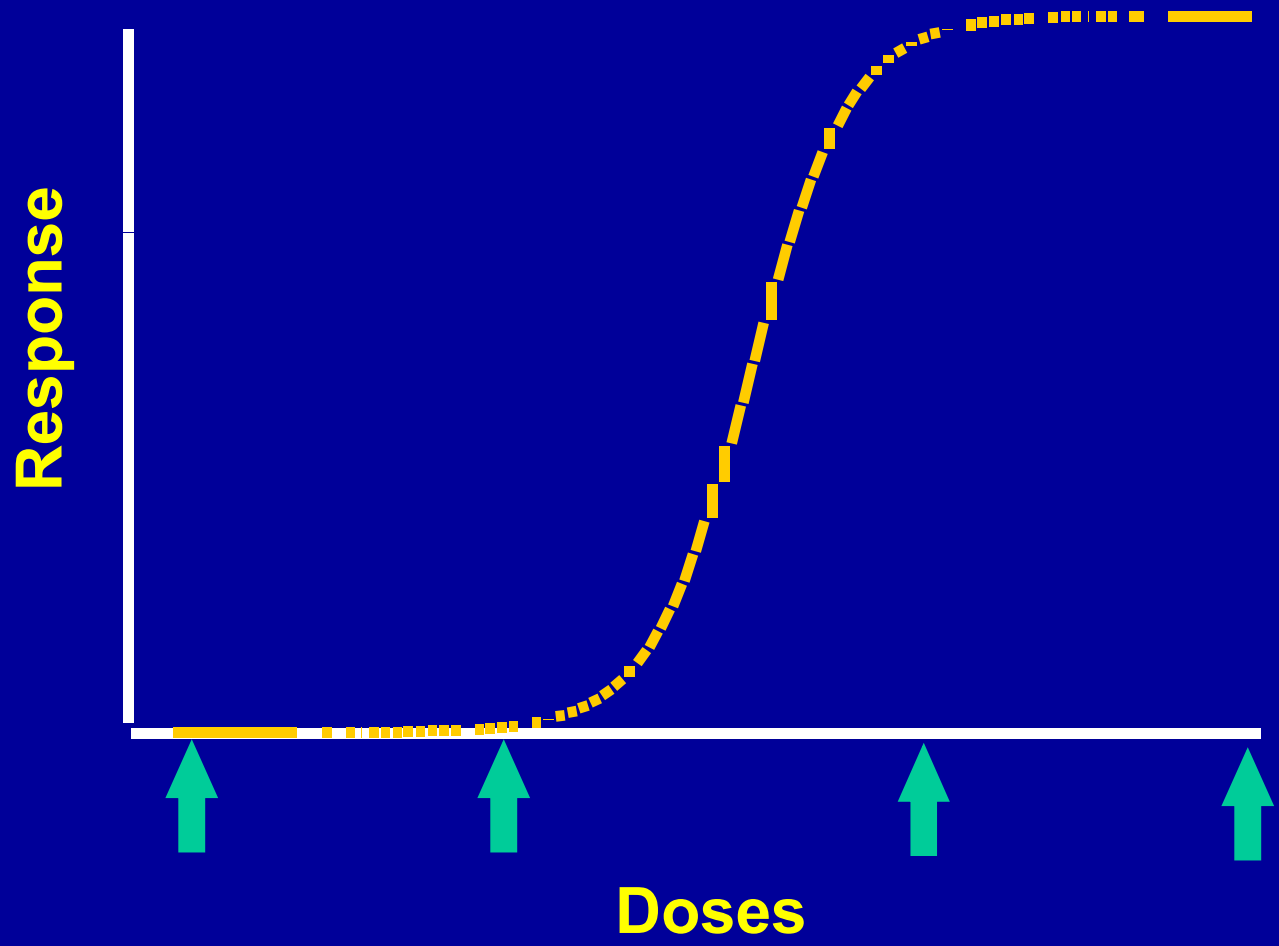
Stroke and adaptive dose-response

- Adaptive doses in Phase II setting: learn efficiently and rapidly about dose-response relationship
- Pfizer trial of a neutrofil inhibitory factor
- Endpoint: stroke scale at week 13
- Early endpoints: weekly stroke scale

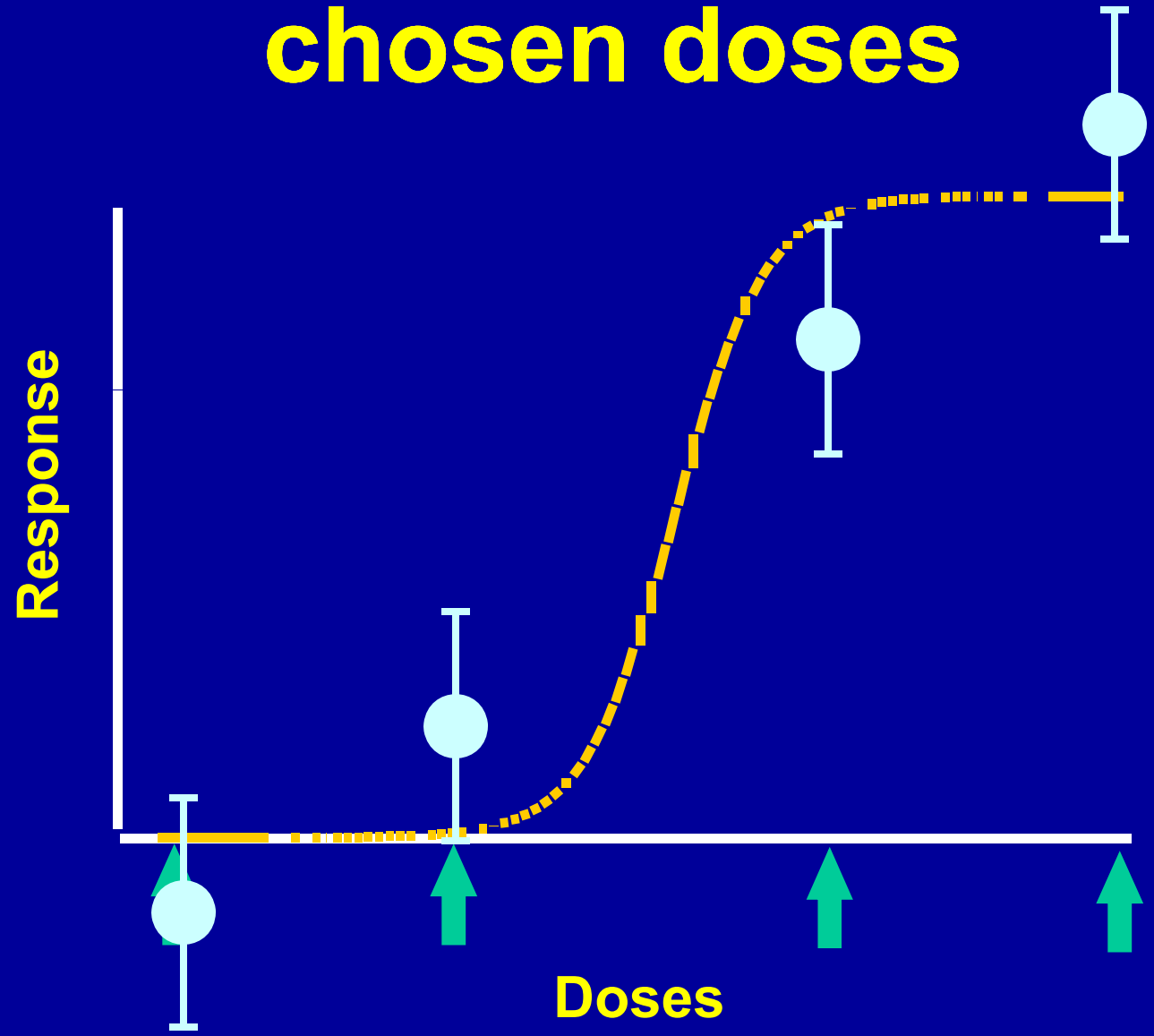
Standard Parallel Group Design

Equal sample sizes at each of k doses.

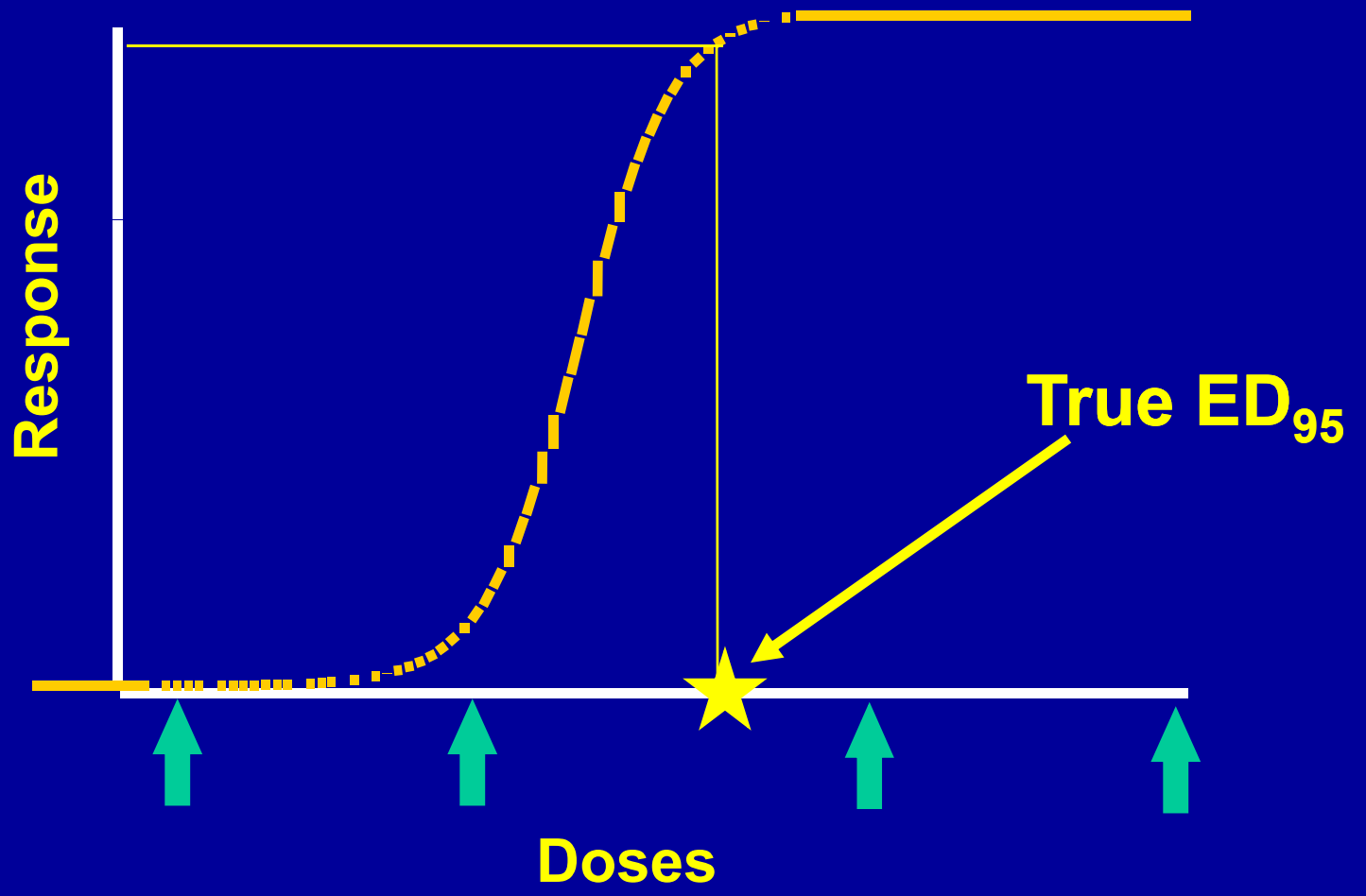
True dose-response curve (unknown)



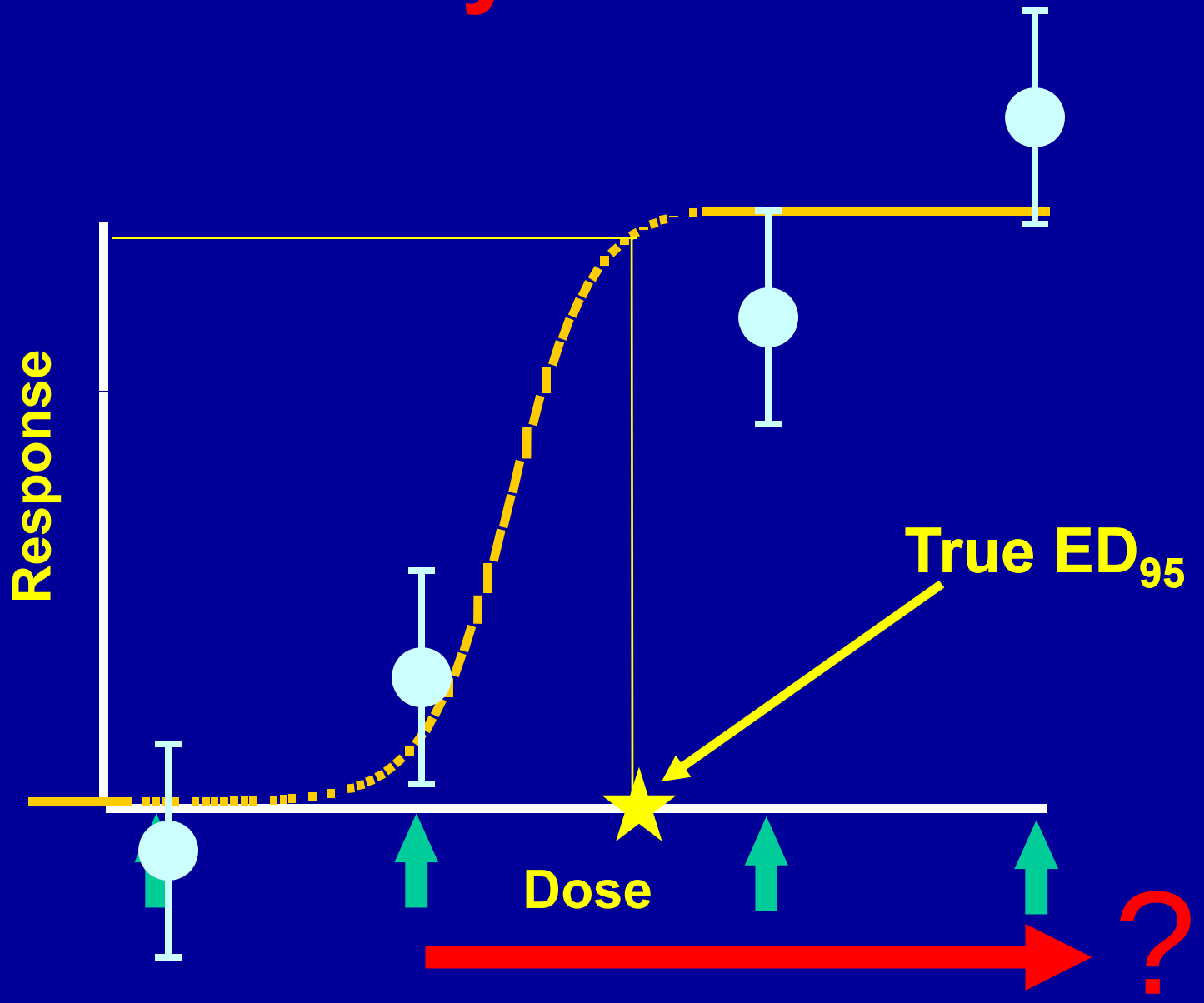
Observe responses (with error) at chosen doses



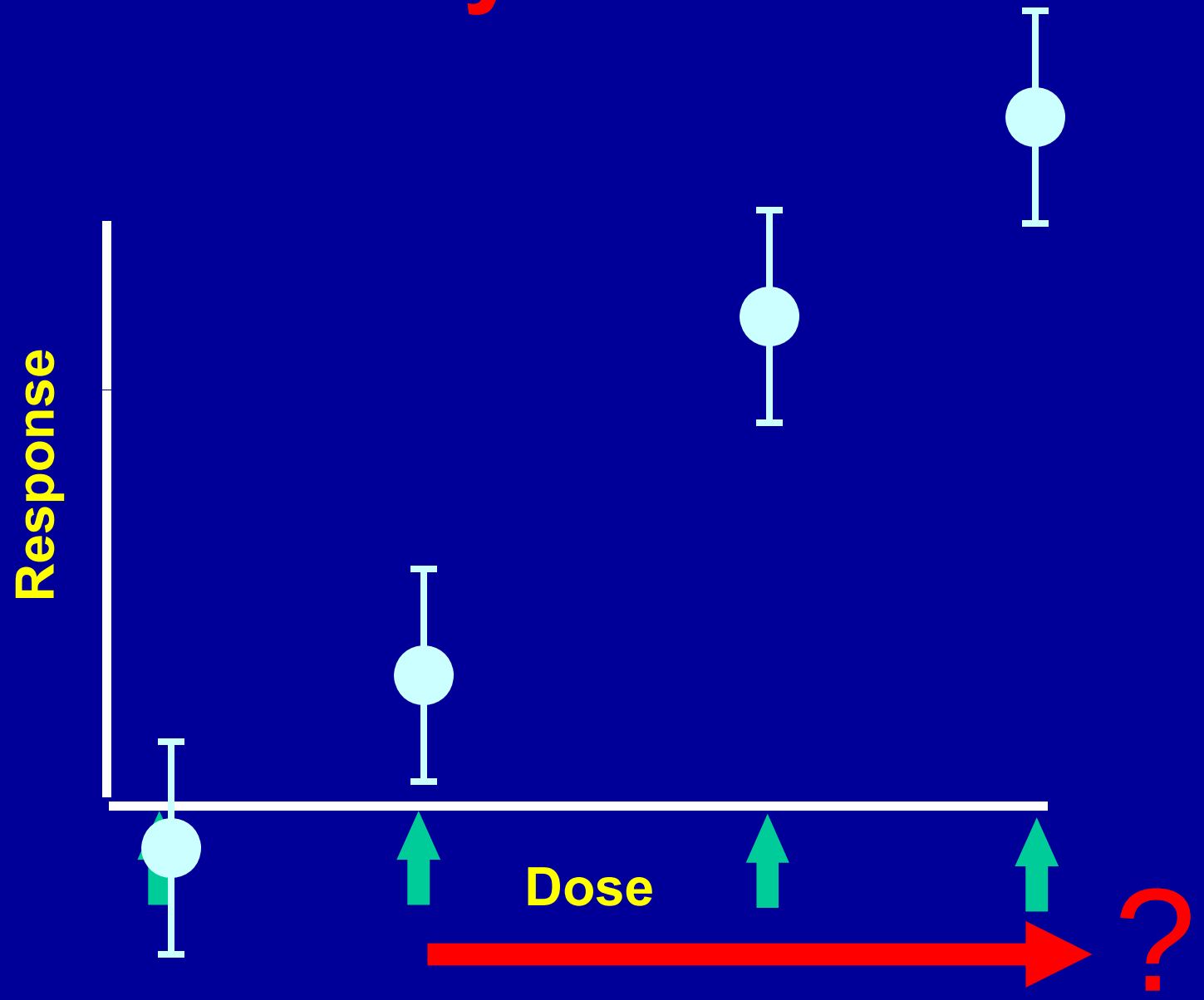
Dose at which 95% max effect



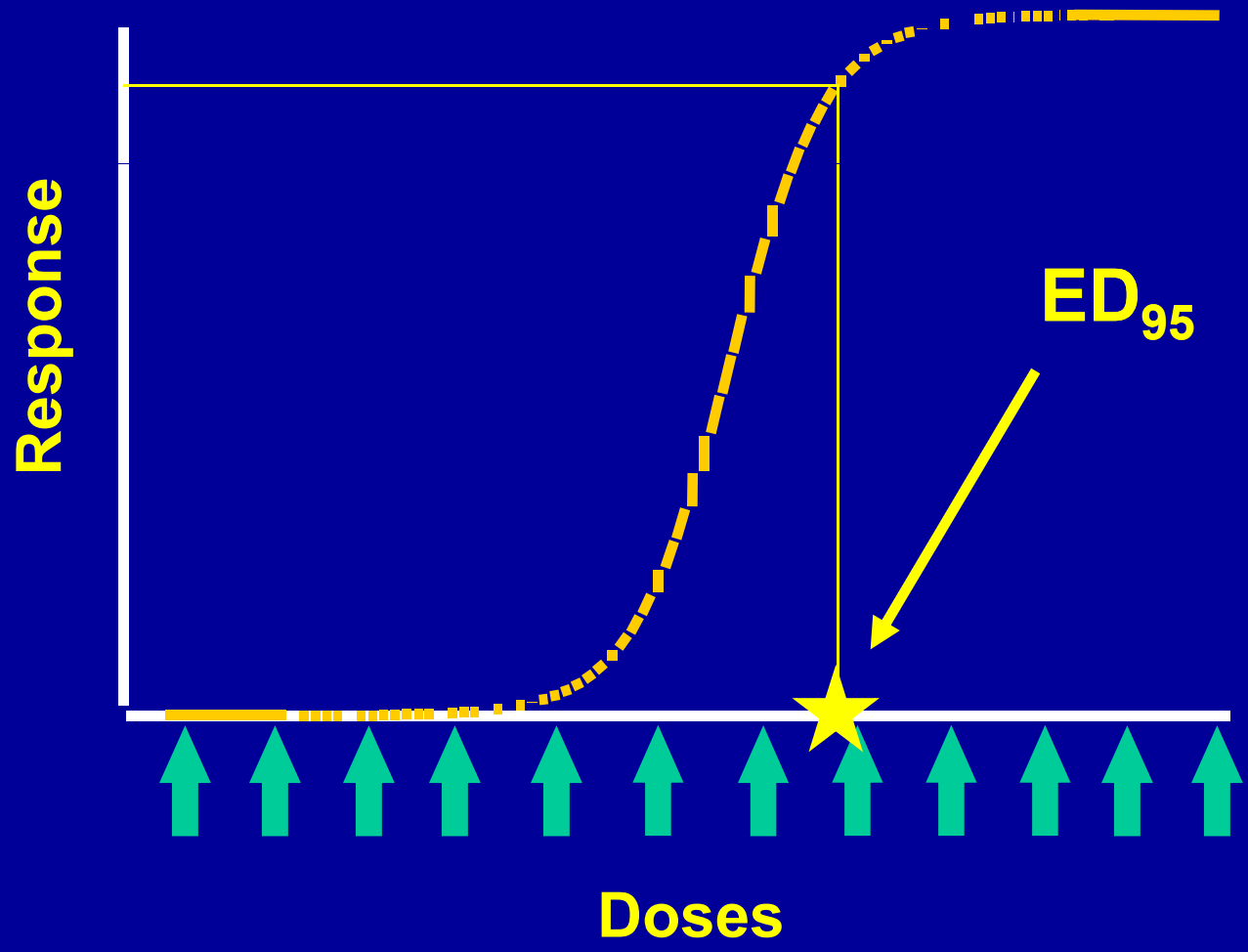
Uncertainty about ED95



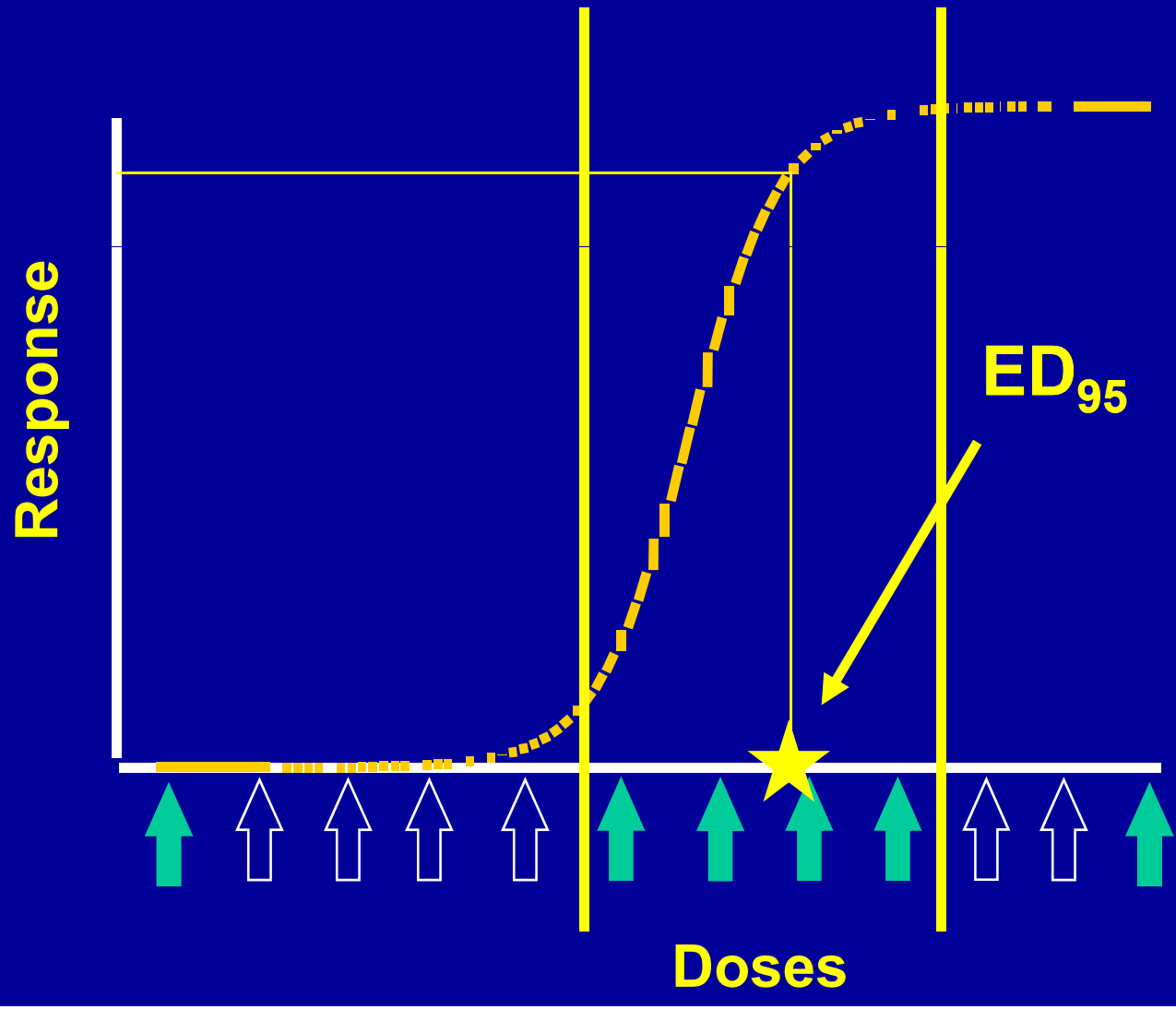
Uncertainty about ED95



Solution:
Increase number of doses



But, enormous sample size, and . . . wasted dose assignments—always!

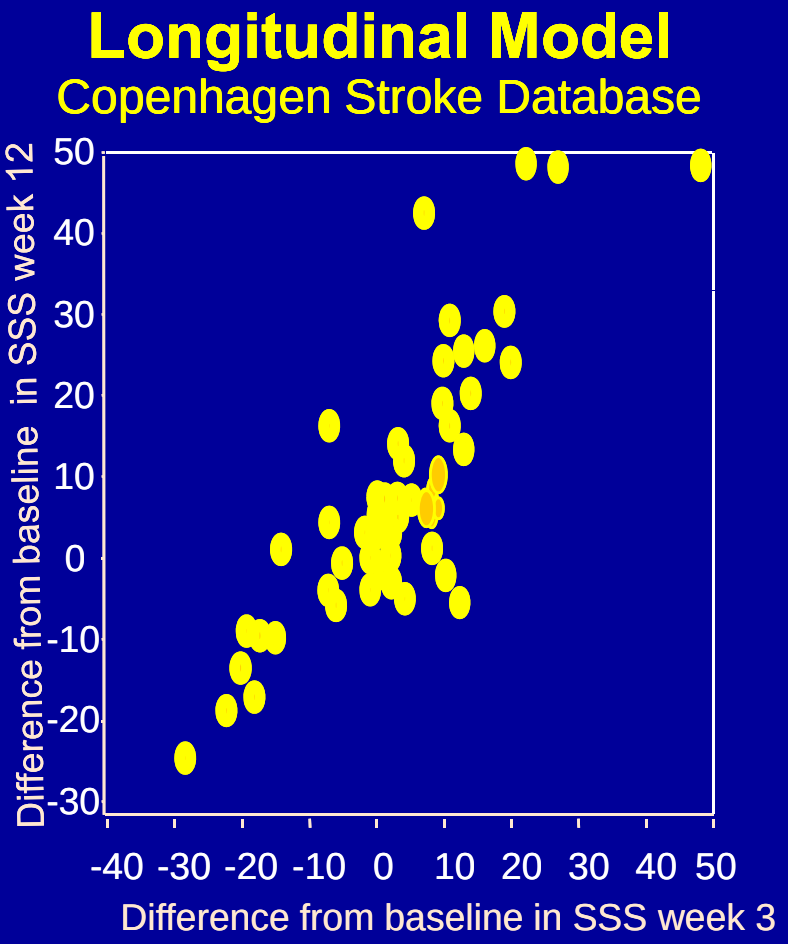


Our adaptive approach

- Observe data continuously
- Select next dose to maximize information about ED95, given available evidence
- Stop dose-ranging trial when know ED95 & response at ED95 "sufficiently well"

Our approach (cont'd)

Info accrues gradually about each patient; prediction using longitudinal model



Our approach (cont'd)

- Model dose-response (borrow strength from neighboring doses)
- Many doses (logistical issues)

Possible decisions each day:

- Stop trial and drug's development
- Stop and set up confirmatory trial
- Continue dose-finding (what dose?)

Size of confirmatory trial based on info from dose-ranging phase

Choices by decision analysis

(Human safeguard: DSMB)

Dose-response trial

- Learn efficiently and rapidly about dose-response; if + go to Phase III
- Assign dose to maximize info about dose-response parameters given current info
- Use predictive probabilities, based on early endpoints
- Doses in continuum, or preset grid

Dose-response trial (cont'd)

- Learn about SD on-line
- Halt dose-ranging when know dose sufficiently well
- Seamless switch from dose-ranging to confirmatory trial—2 trials in 1!

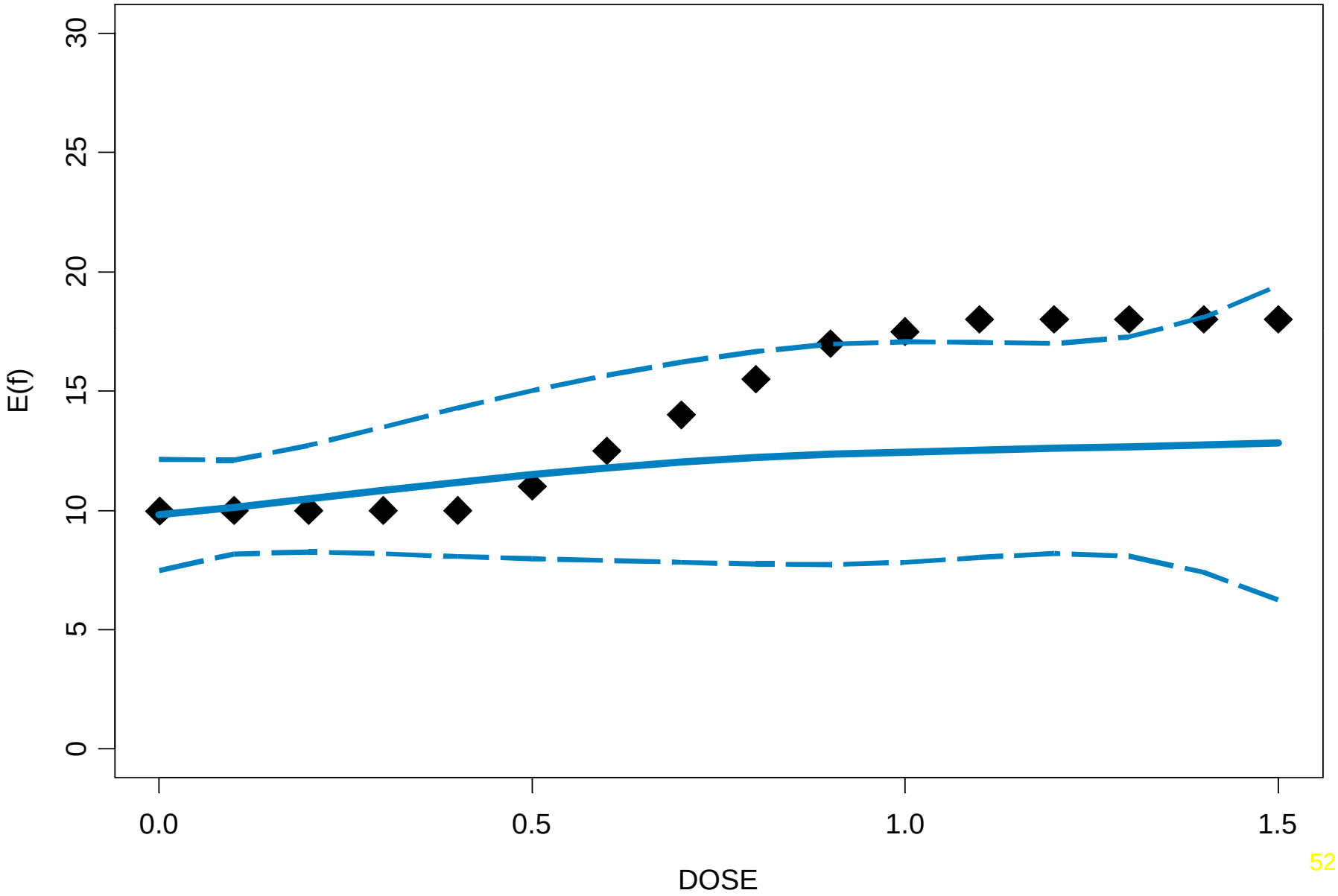
Advantages over standard design

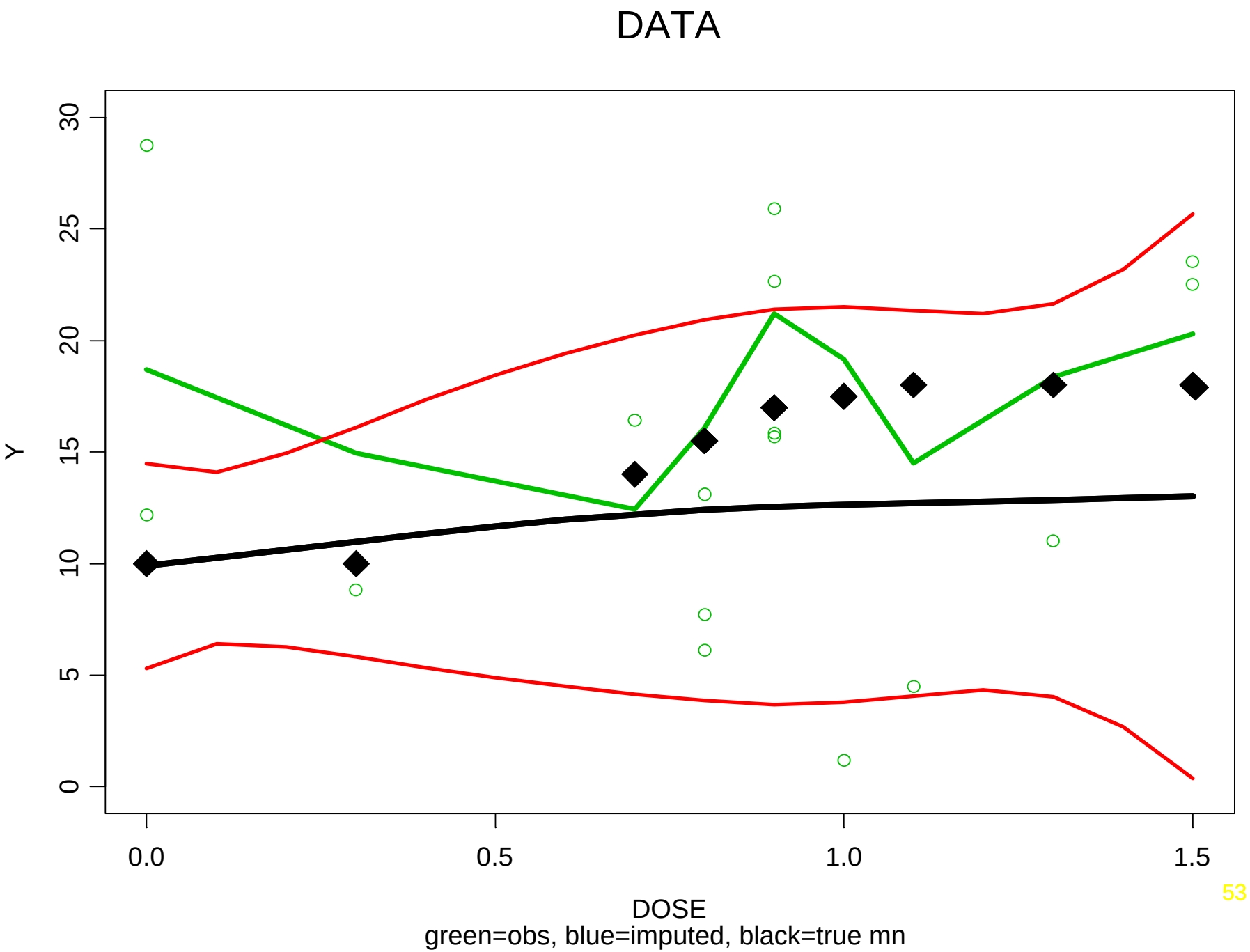
- Fewer patients (generally); faster & more effective learning
- Better at finding ED95
- Tends to treat patients in trial more effectively
- Drops duds early—actual trial!

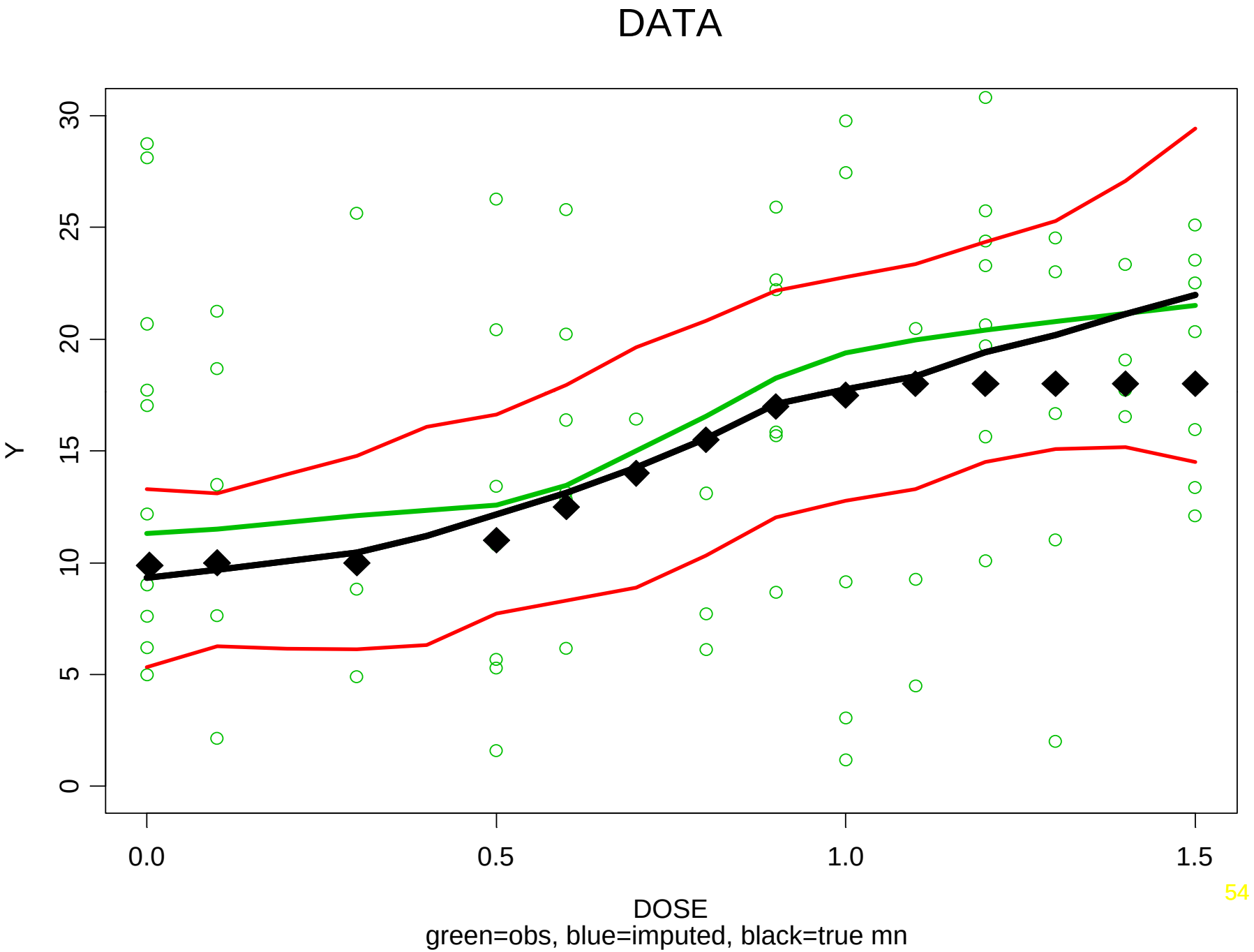
Dose-assignment simulation

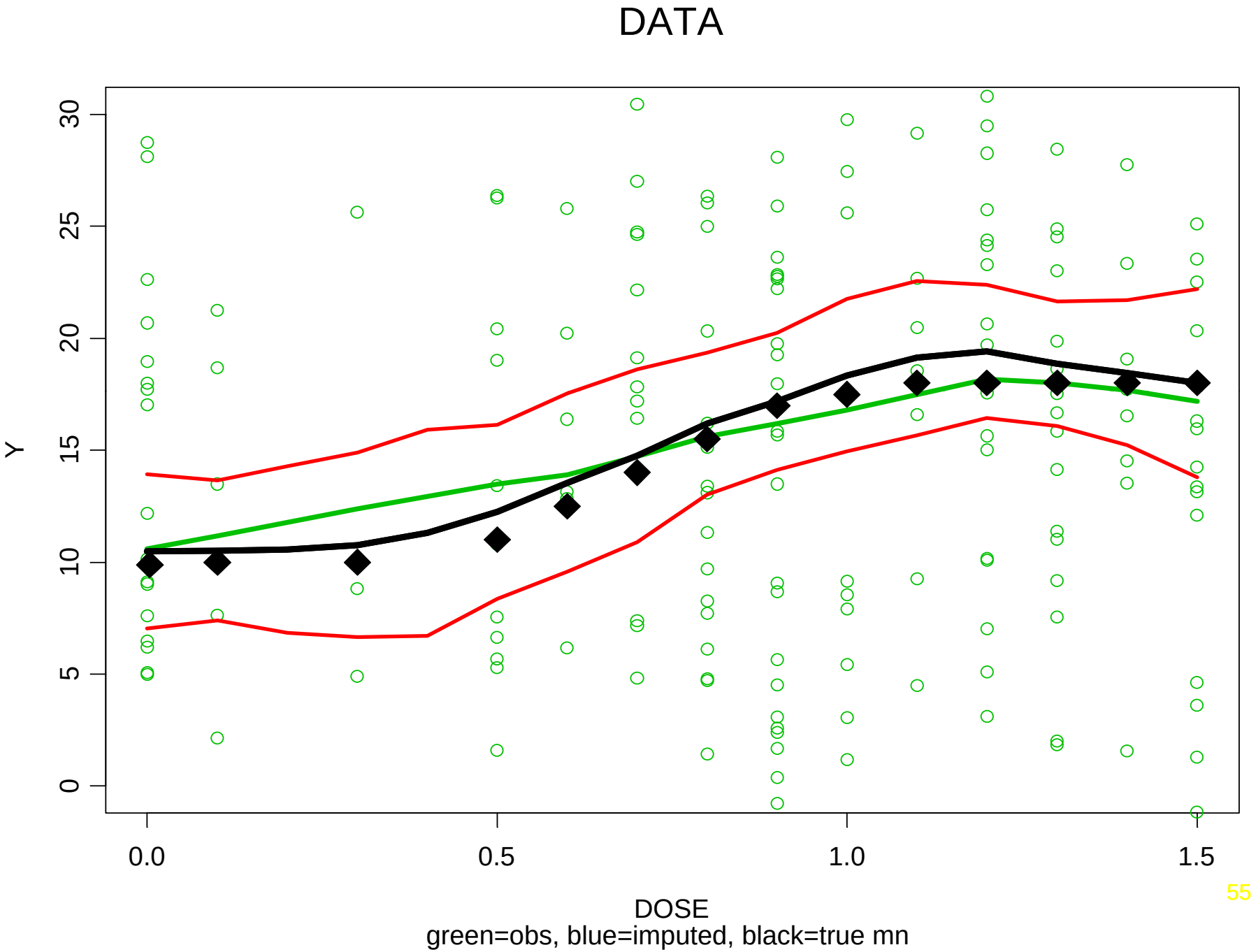
- Assumes particular dose-response curve
- Assumes SD = 12
- Shows weekly results, several patients at a time (green circles)

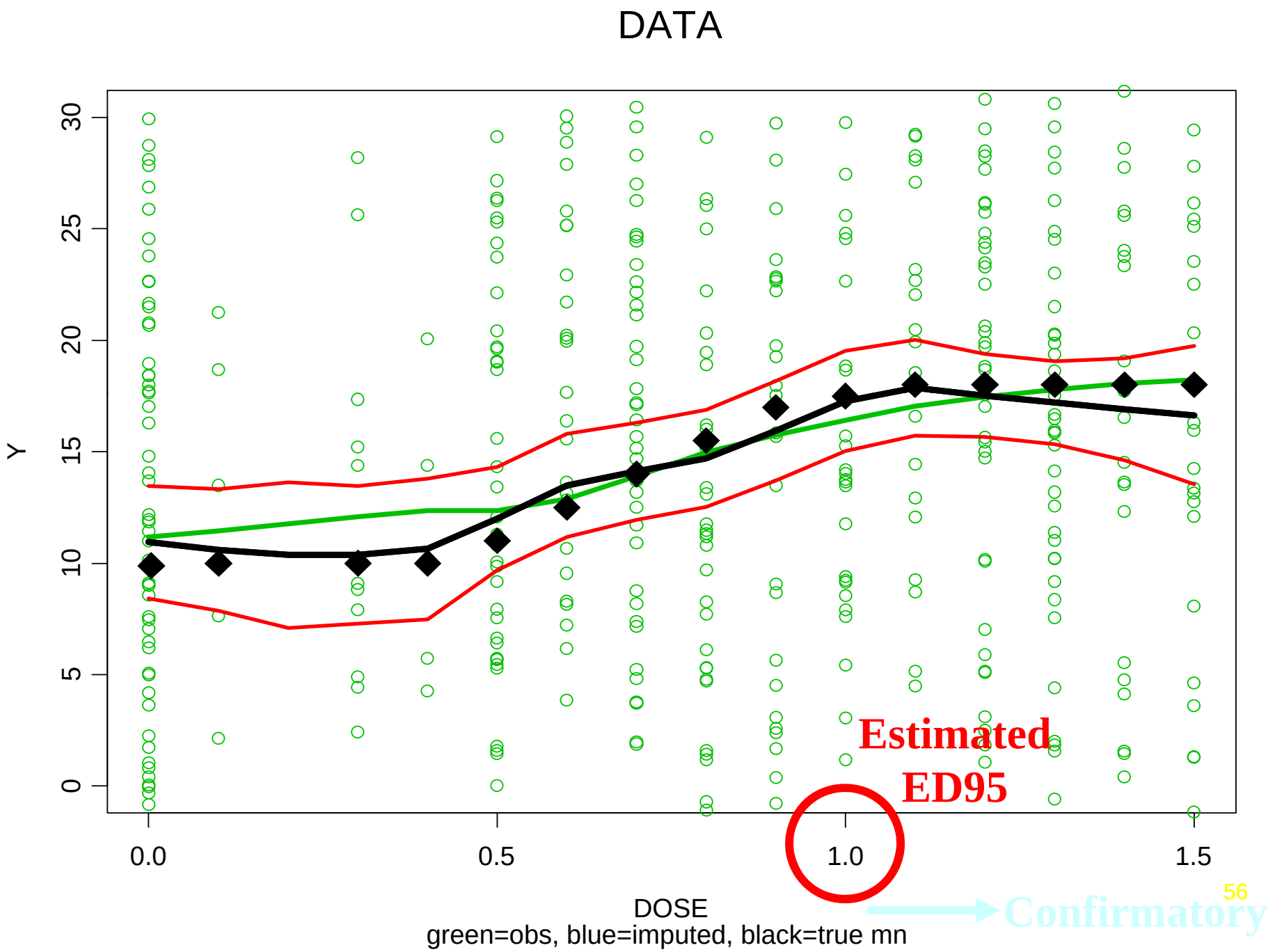
Prior



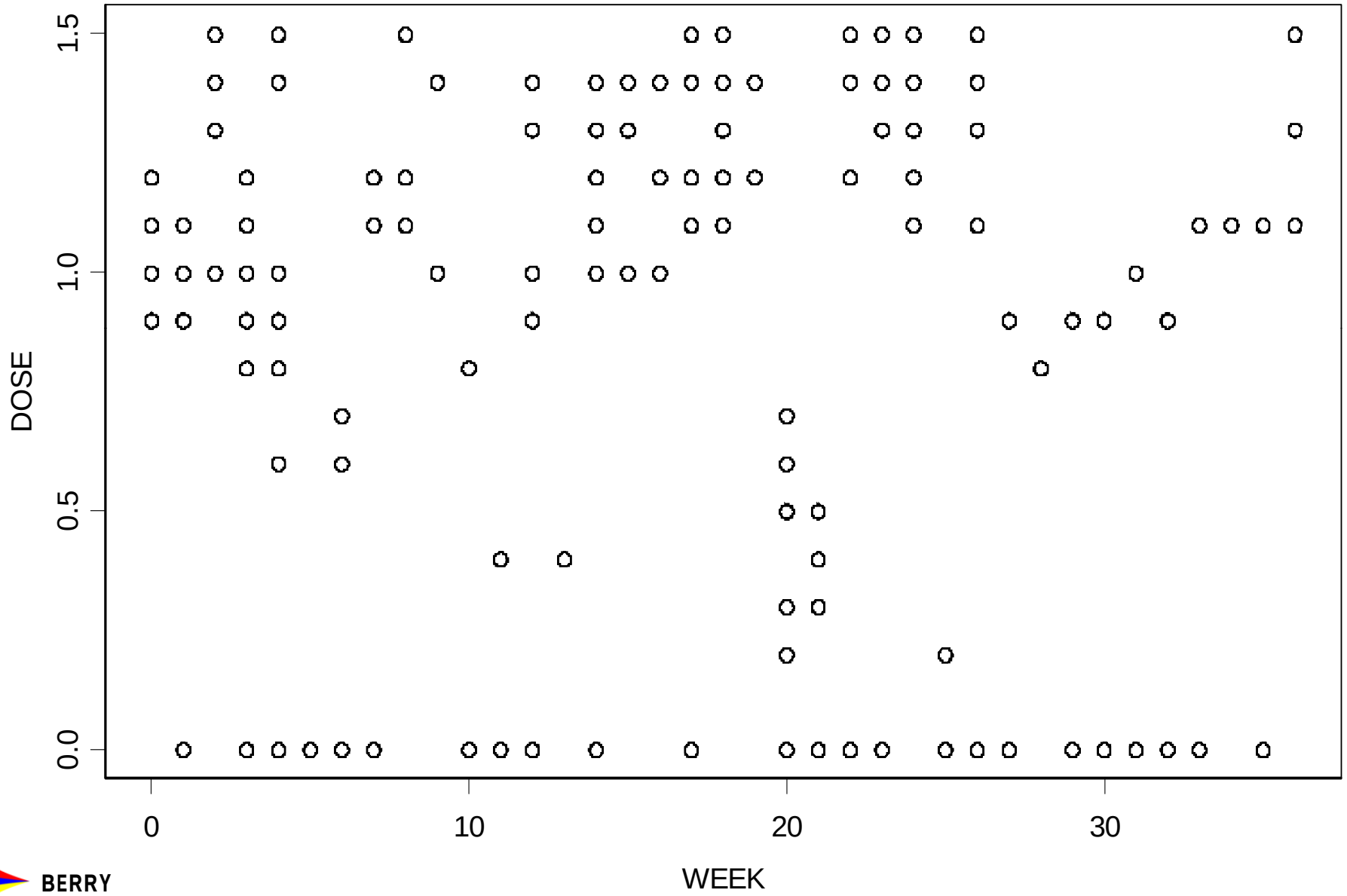

DATA
Y
0
5
10
15
20
25
30
0.0
0.5
1.0
1.5
DOSE
green=obs, blue=imputed, black=true mn


DATA
Y
0
5
10
15
20
25
30
0.0
0.5
1.0
1.5
DOSE
green=obs, blue=imputed, black=true mn

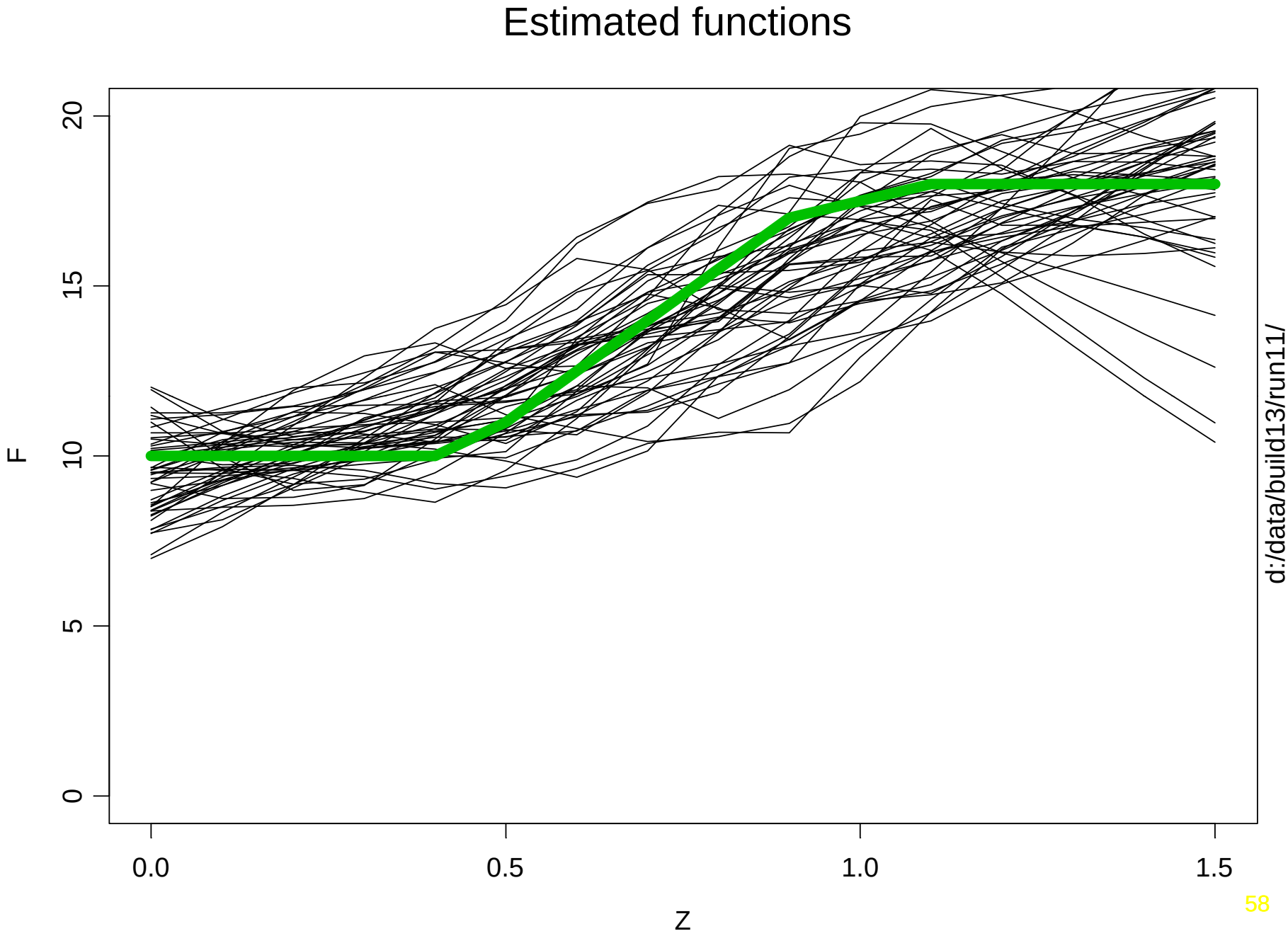

DATA
Y
0
5
10
15
20
25
30
0.0
0.5
1.0
1.5
DOSE
green=obs, blue=imputed, black=true mn

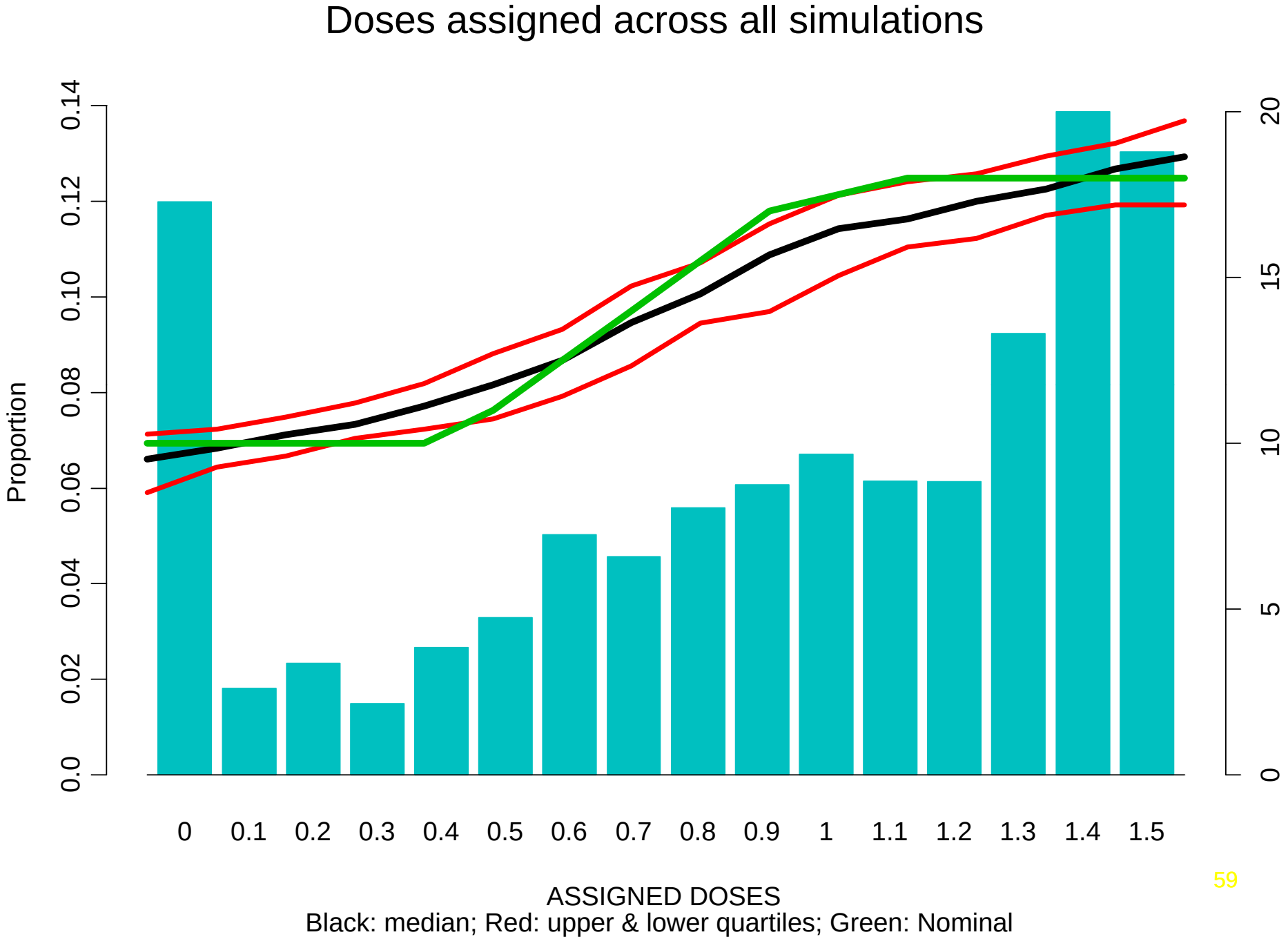

DATA
Y
0
5
10
15
20
25
30
0.0
0.5
1.0
1.5
DOSE
green=obs, blue=imputed, black=true mn
Estimated
ED95
Confirmatory

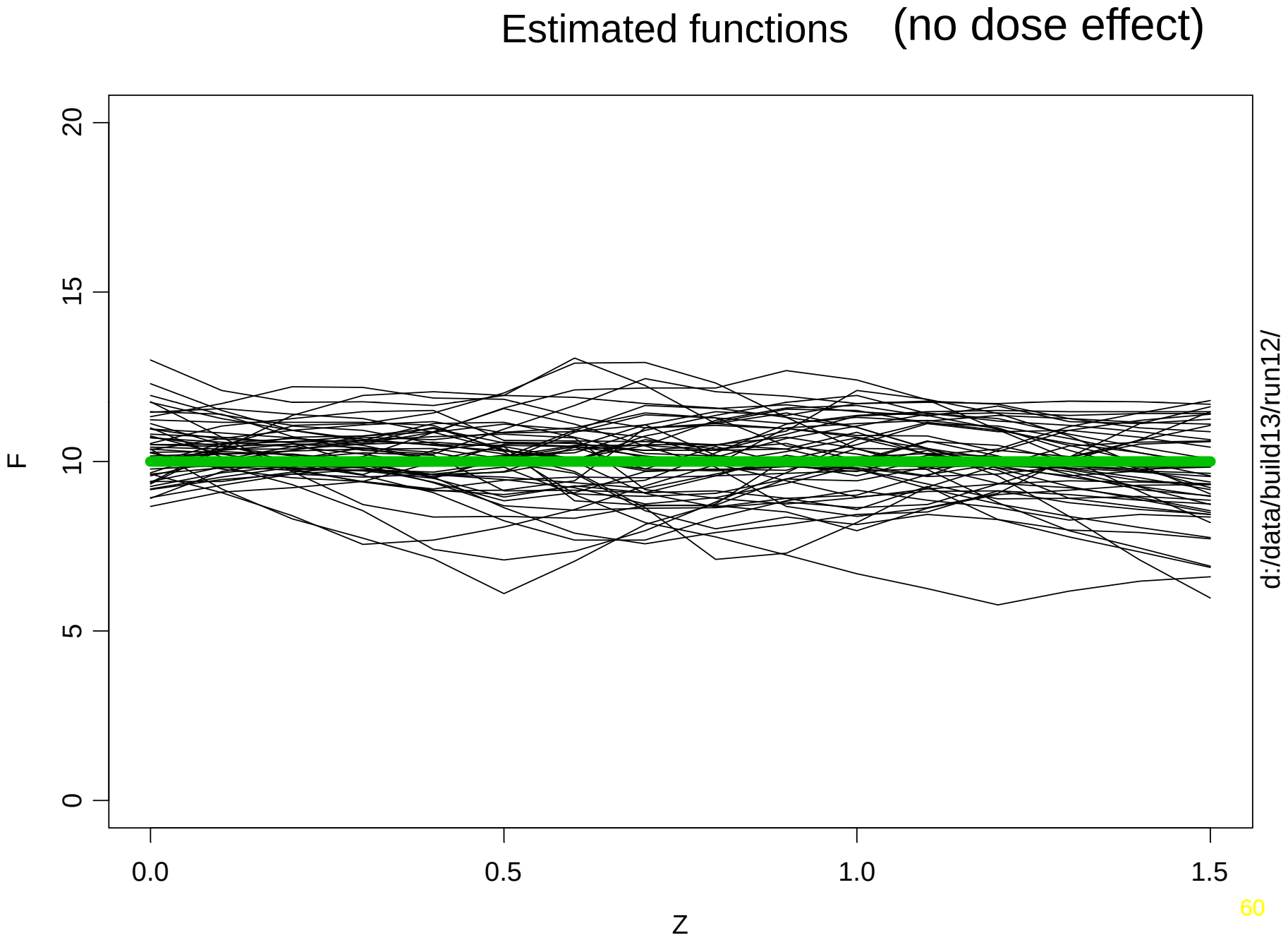
Assigned Doses by Week - one simulation



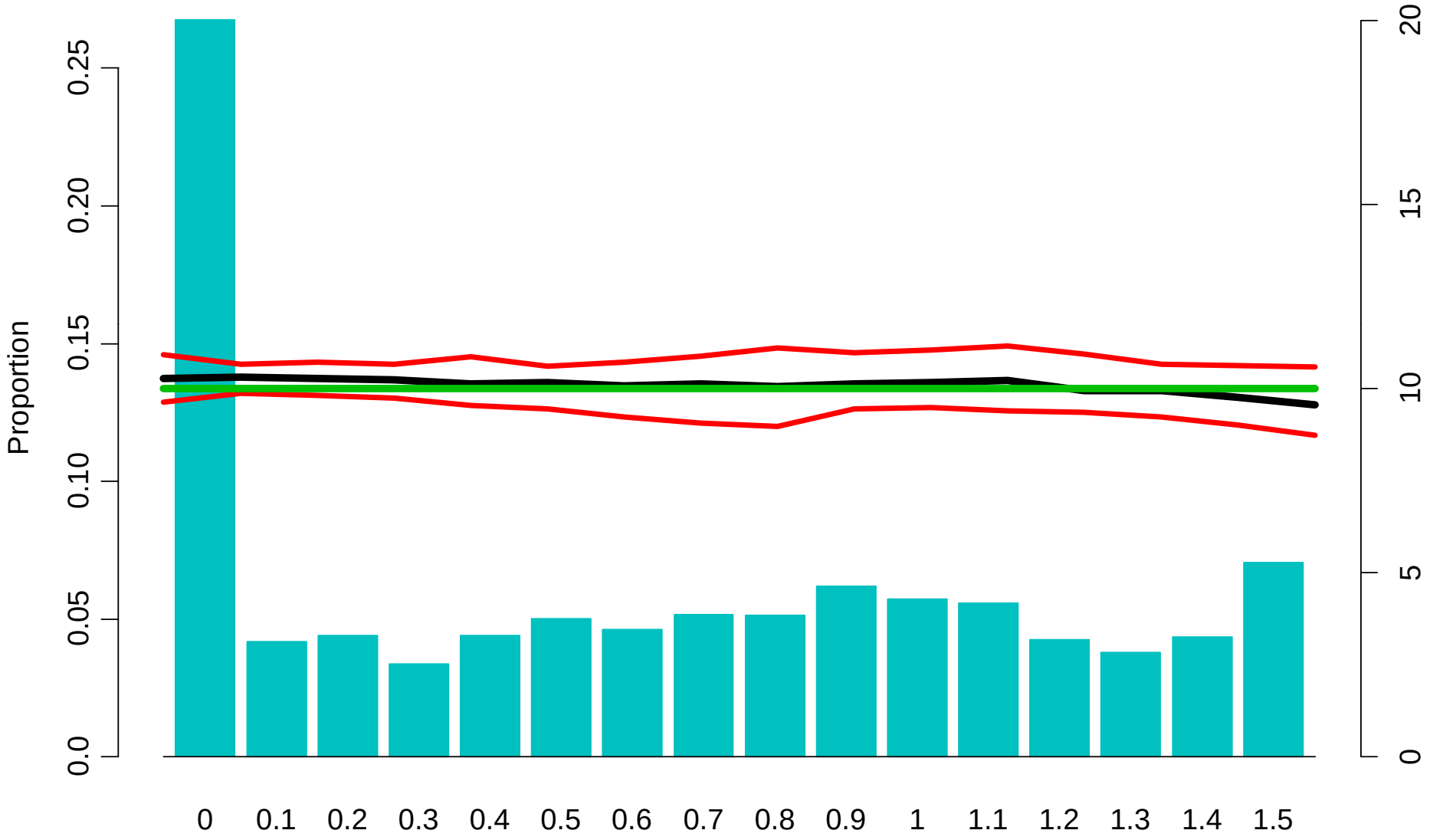
BERRY CONSULTANTS
STATISTICAL INNOVATION


Estimated functions
F
Z
0
5
10
15
20
0.0
0.5
1.0
1.5
d:/data/build13/run11/


Doses assigned across all simulations
Proportion
0.0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
5
10
15
20
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
ASSIGNED DOSES
Black: median; Red: upper & lower quartiles; Green: Nominal


Estimated functions (no dose effect)
F
Z
0
5
10
15
20
0.0
0.5
1.0
1.5
d:/data/build13/run12/
60

Doses assigned across all simulations



ASSIGNED DOSES

Black: median; Red: upper & lower quartiles; Green: Nominal

Deploying the design: ASTIN stroke study

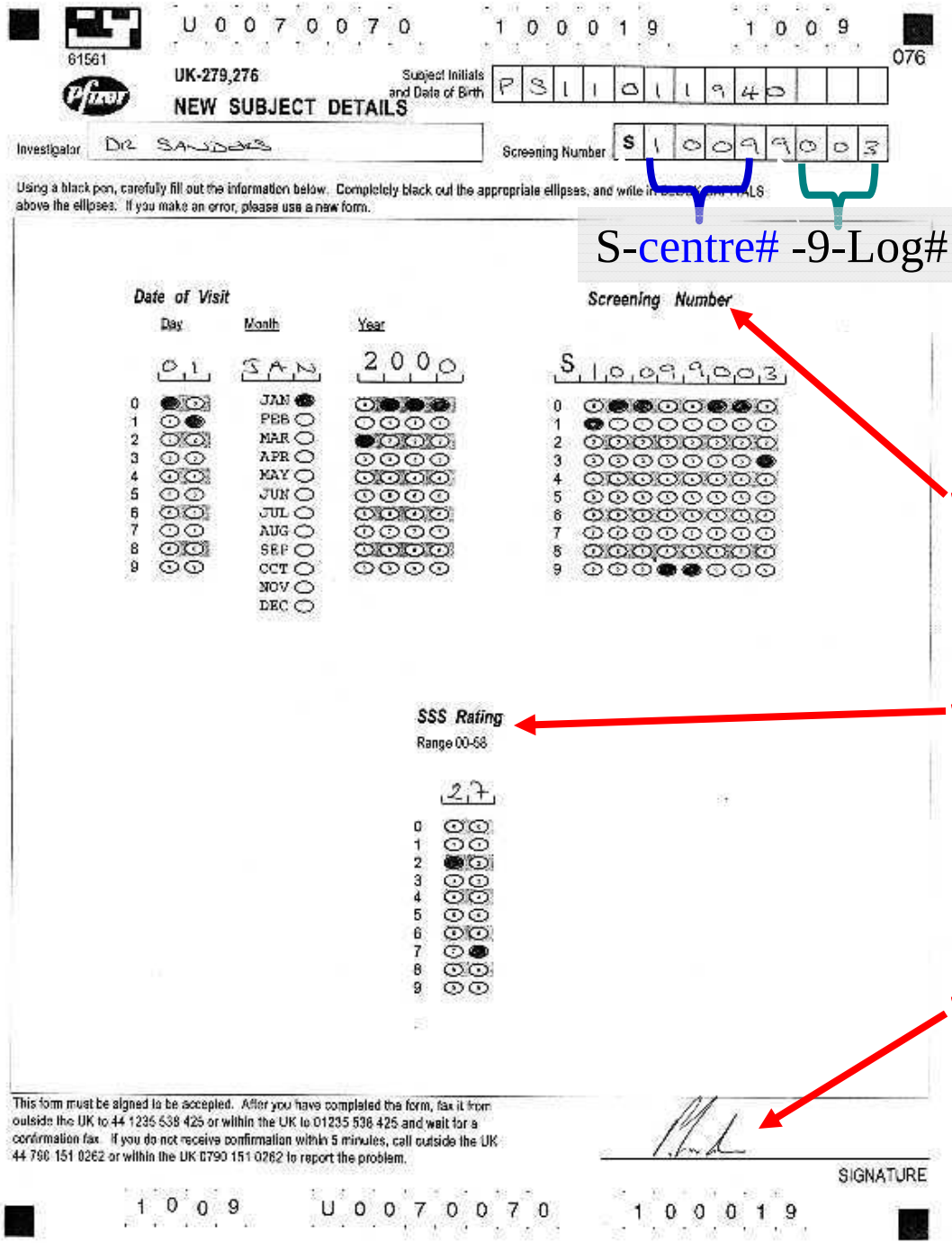
ASTIN Study

- Does UK-279,276 improve recovery in ischemic stroke? Dose effect?
- Placebo and 15 doses
- Primary efficacy endpoint: Scandinavian Stroke Scale, Mean change from baseline to day 90
- Patients with
  - Age ≥ 50 years
  - Acute ischemic stroke < 6 hours

ASTIN Study

- Executive Steering Committee
- Independent Data Monitoring Committee
- Independent statistician preparing reports for IDMC
- Computer system run by Tessella Ltd (UK)

NEW SUBJECT DETAILS



S-centre# -9-Log# of randomized patients in your centre

Screening number

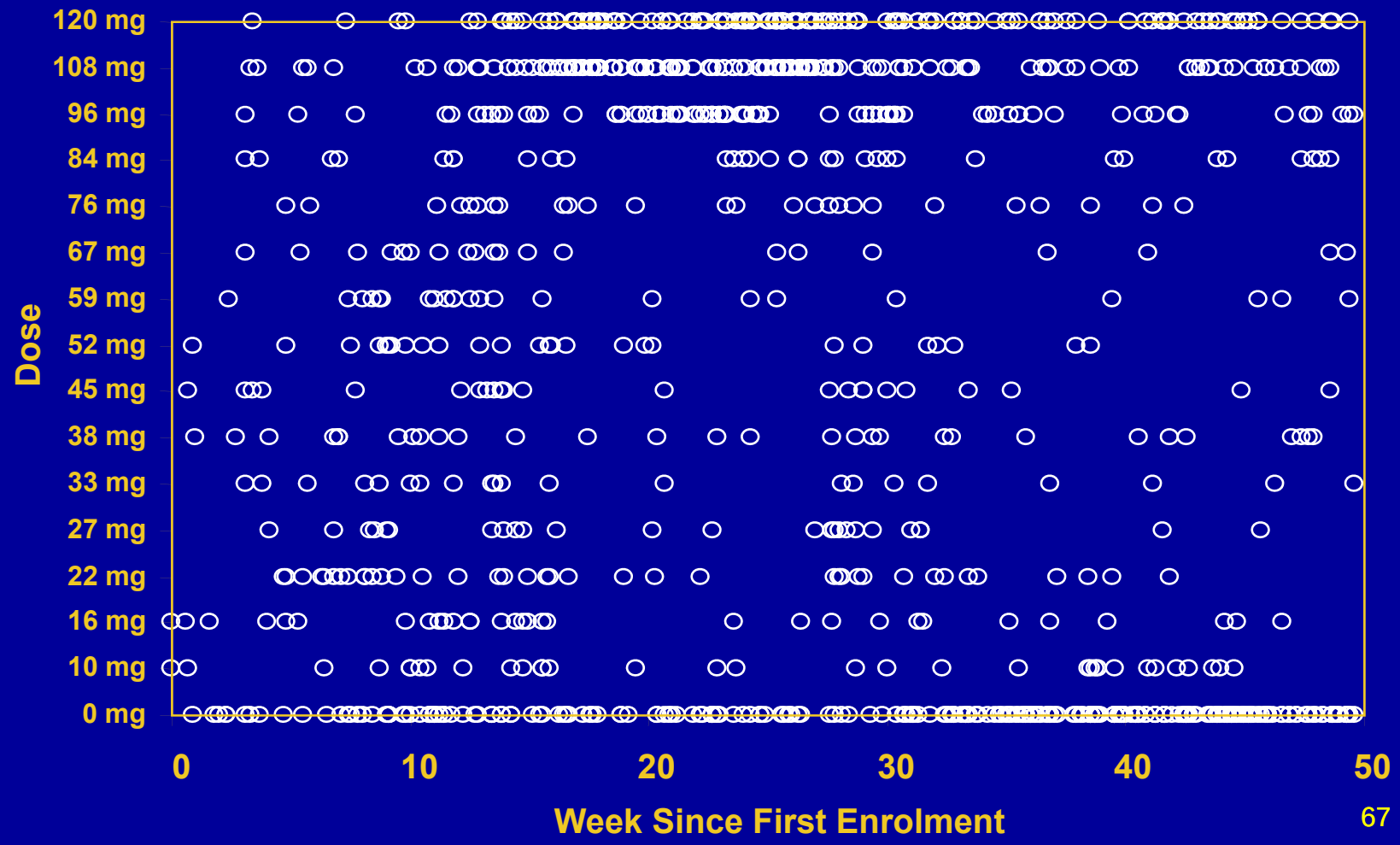
Scandinavian Stroke Scale

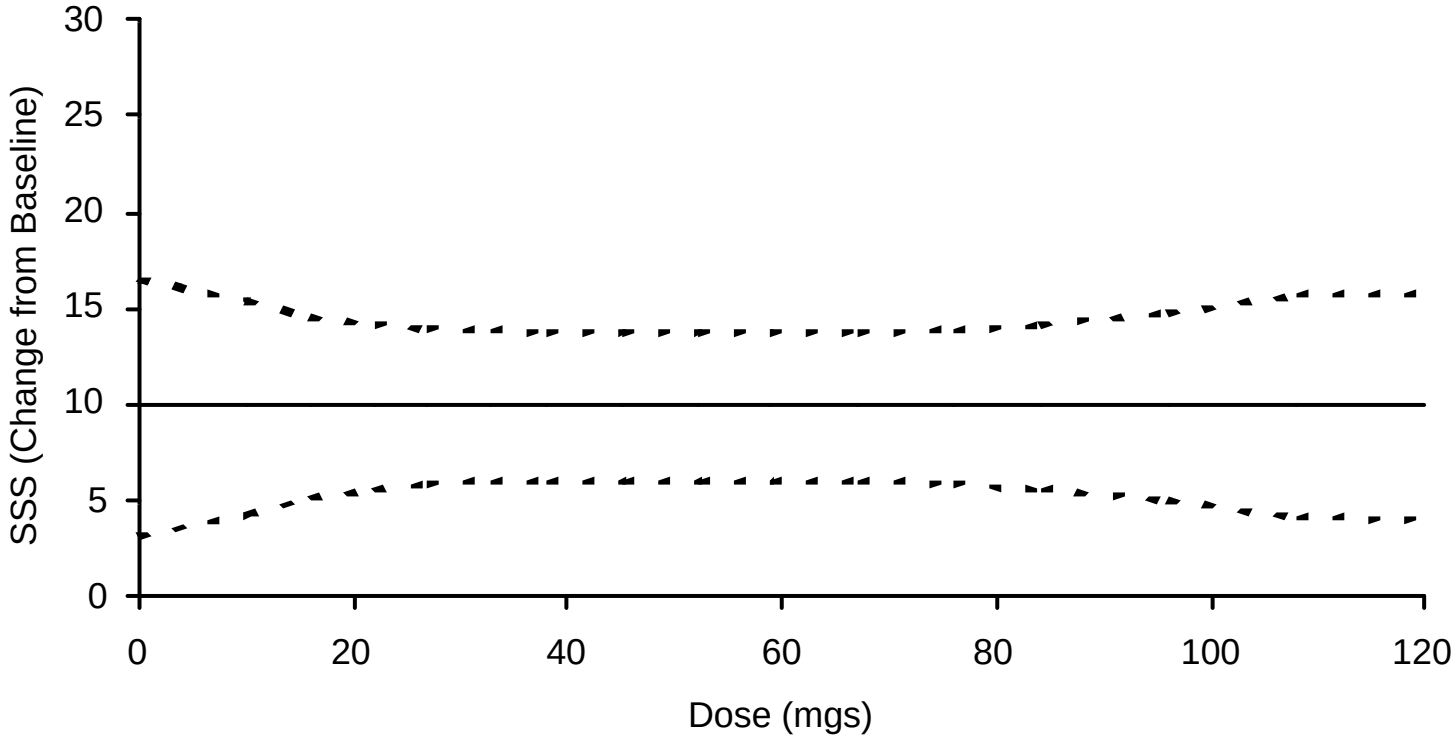
Signature

ASTIN Results

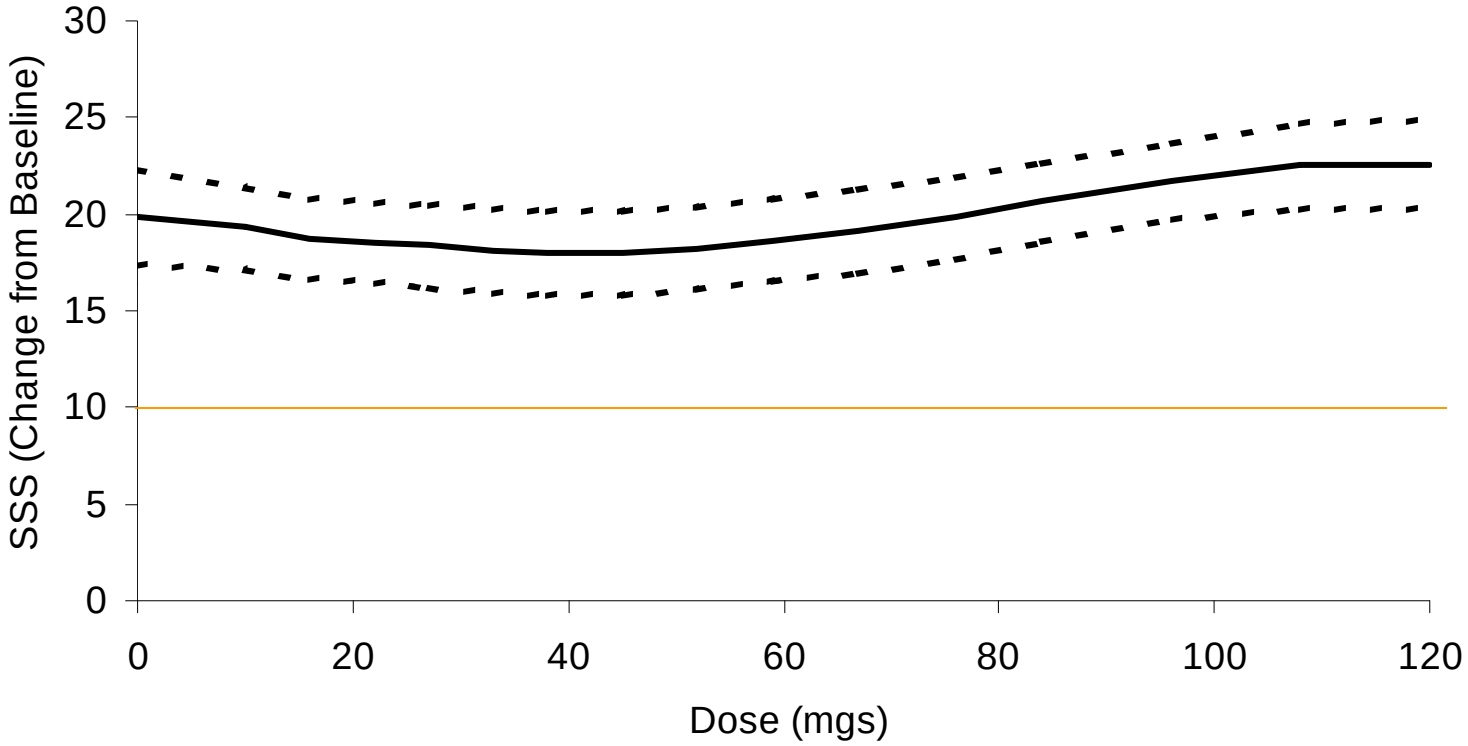
- Trial stopped for futility; 500 with week 13 SSS
- 966 patients randomized/treated
- 93% ischemic stroke
  - 21% cotreated with tPA
  - Mean baseline severity ~ 28 points Scand Stroke Scale
  - Demographics comparable across treatment arms

Accrual/dose over time

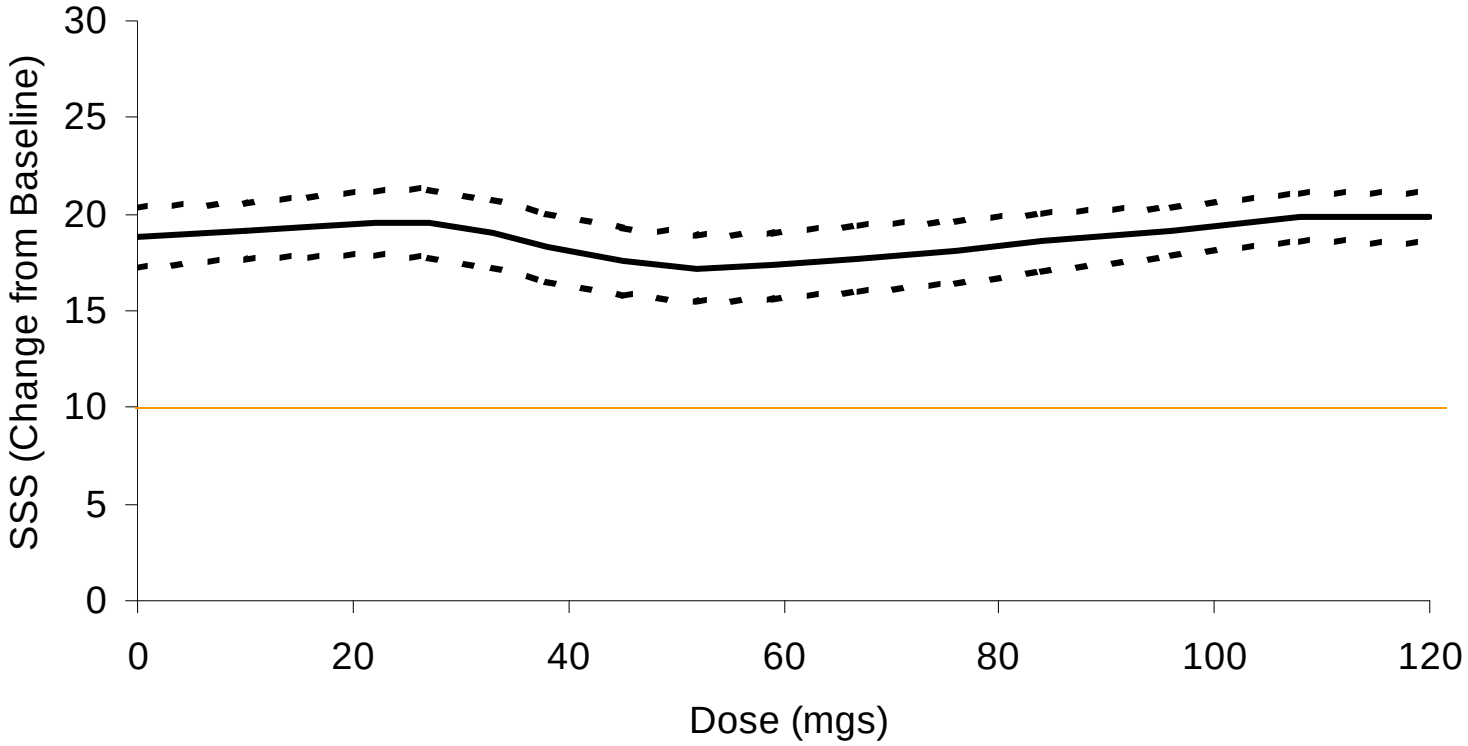



30
25
20
15
10
5
0
SSS (Change from Baseline)
0
20
40
60
80
100
120
Dose (mgs)

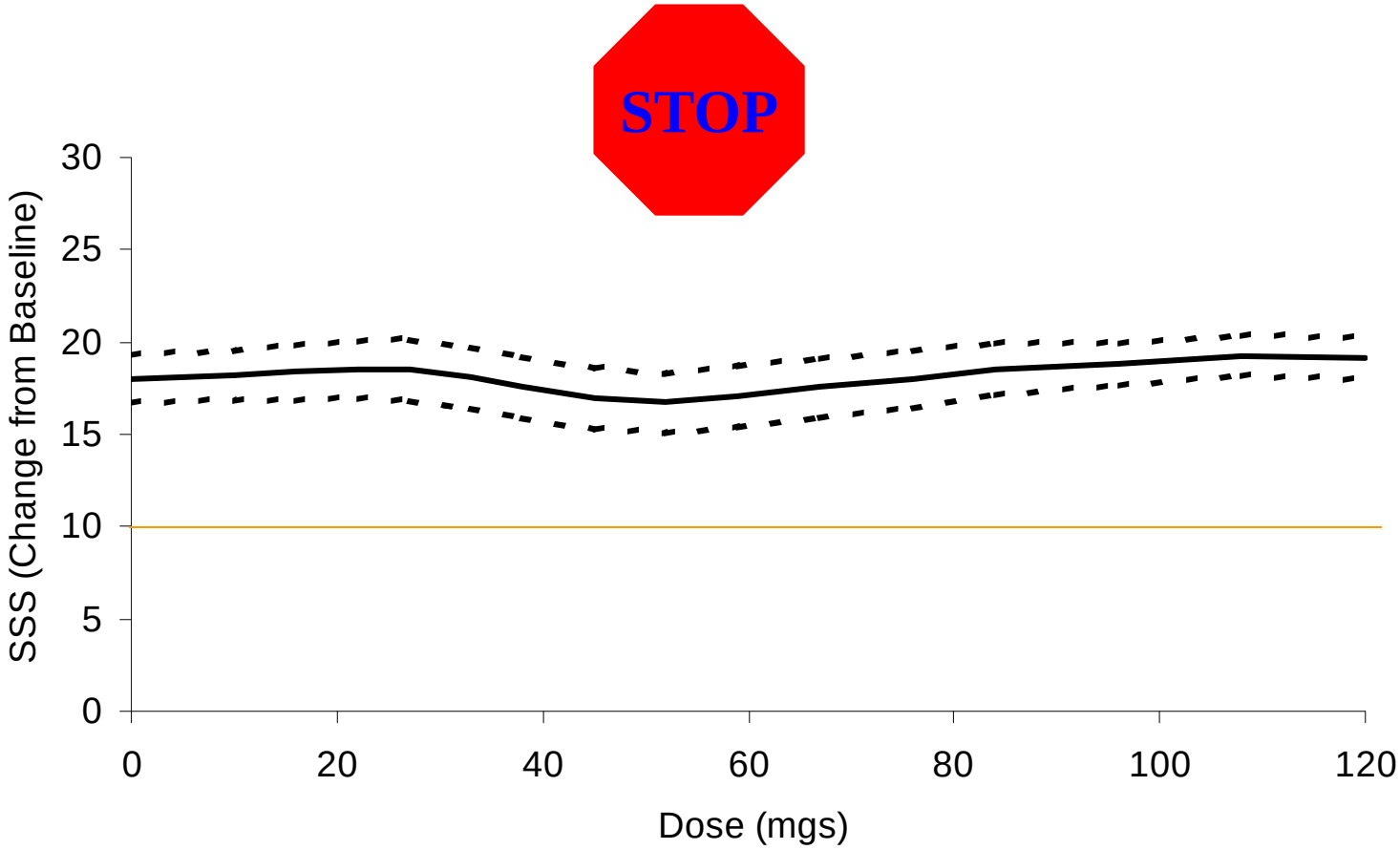
Week - 0


30
25
20
15
10
5
0
SSS (Change from Baseline)
0
20
40
60
80
100
120
Dose (mgs)

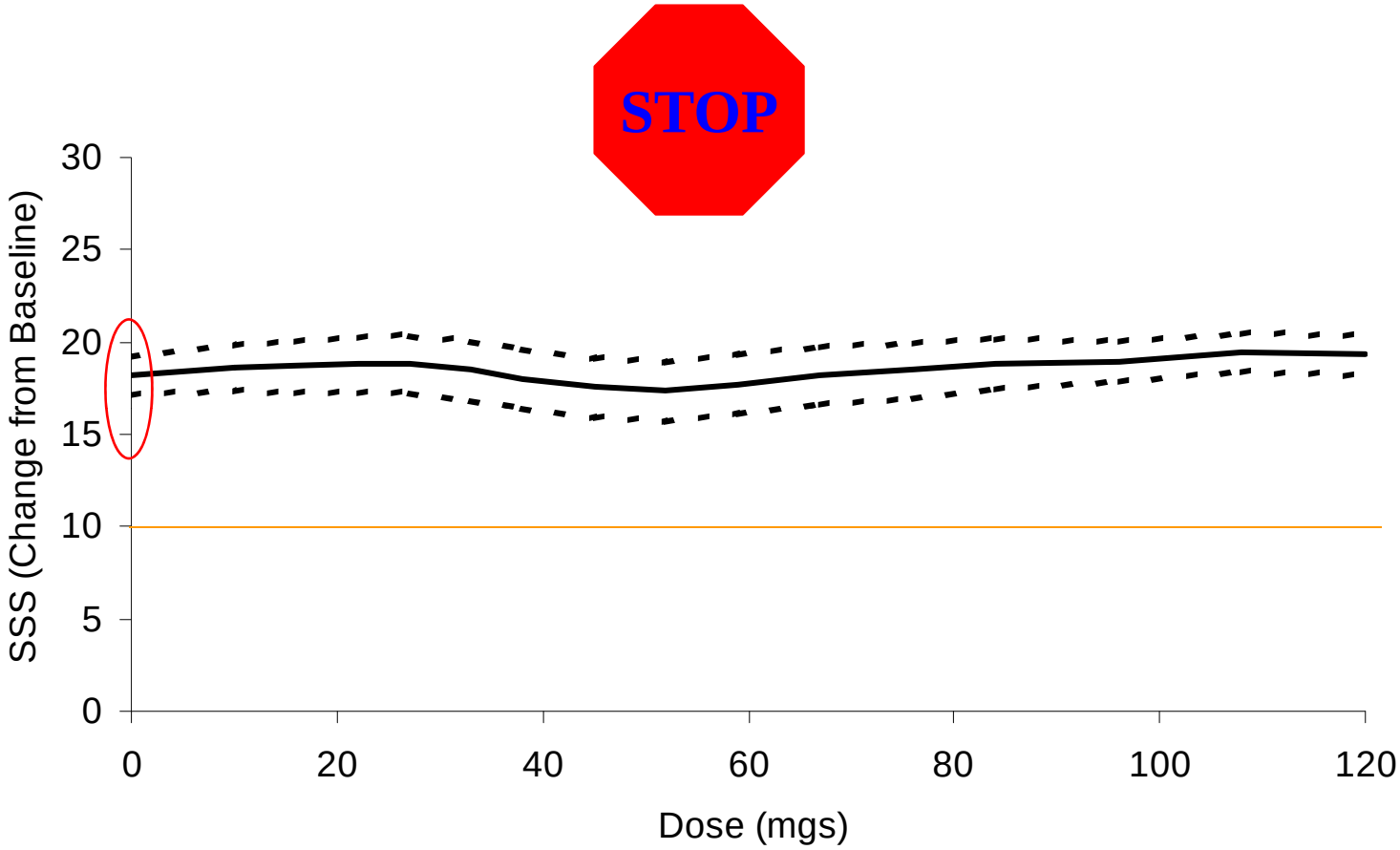
Week - 20


30
25
20
15
10
5
0
SSS (Change from Baseline)
0
20
40
60
80
100
120
Dose (mgs)

Week - 40

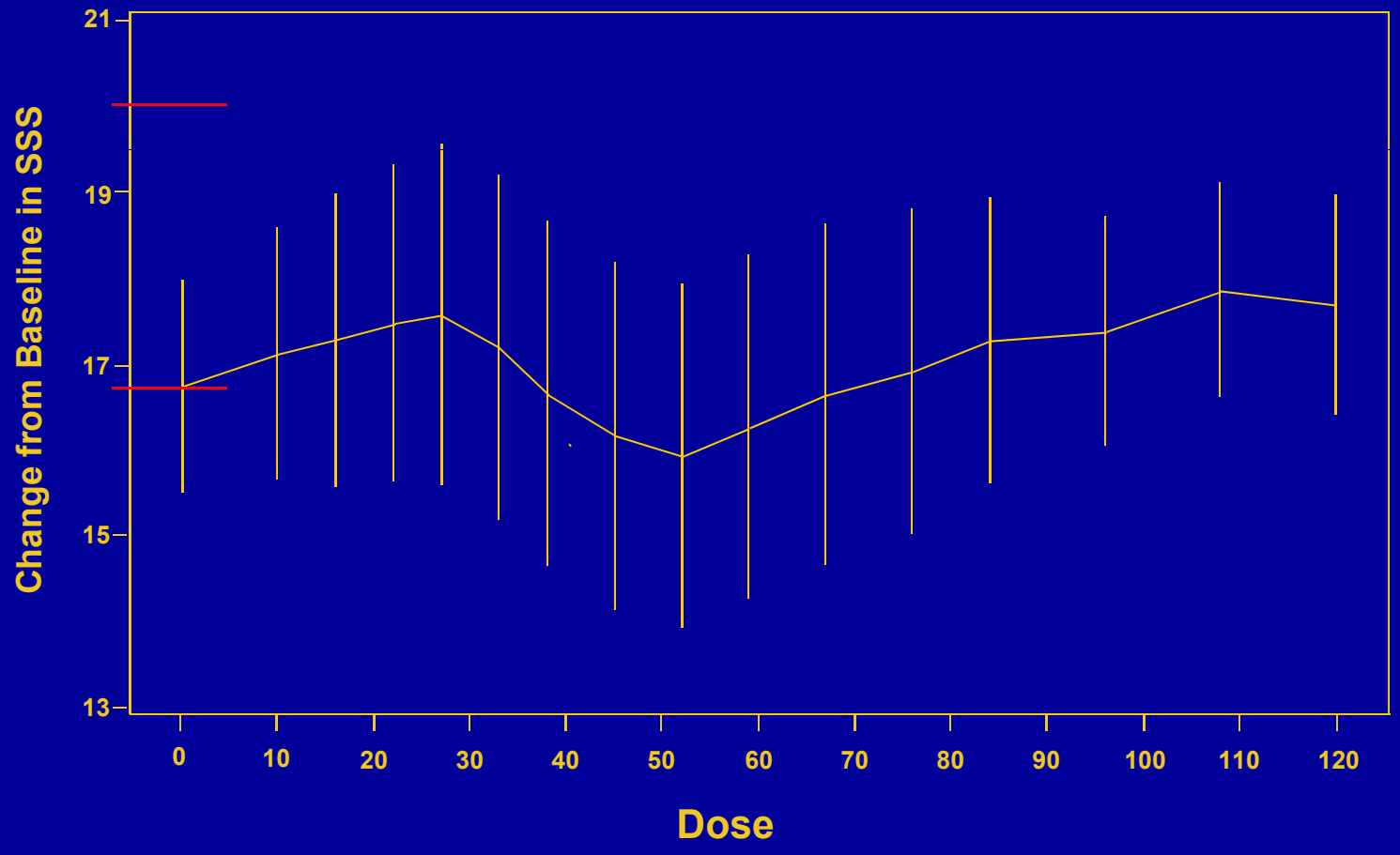

STOP
SSS (Change from Baseline)
30
25
20
15
10
5
0
0
20
40
60
80
100
120
Dose (mgs)

Week - 48

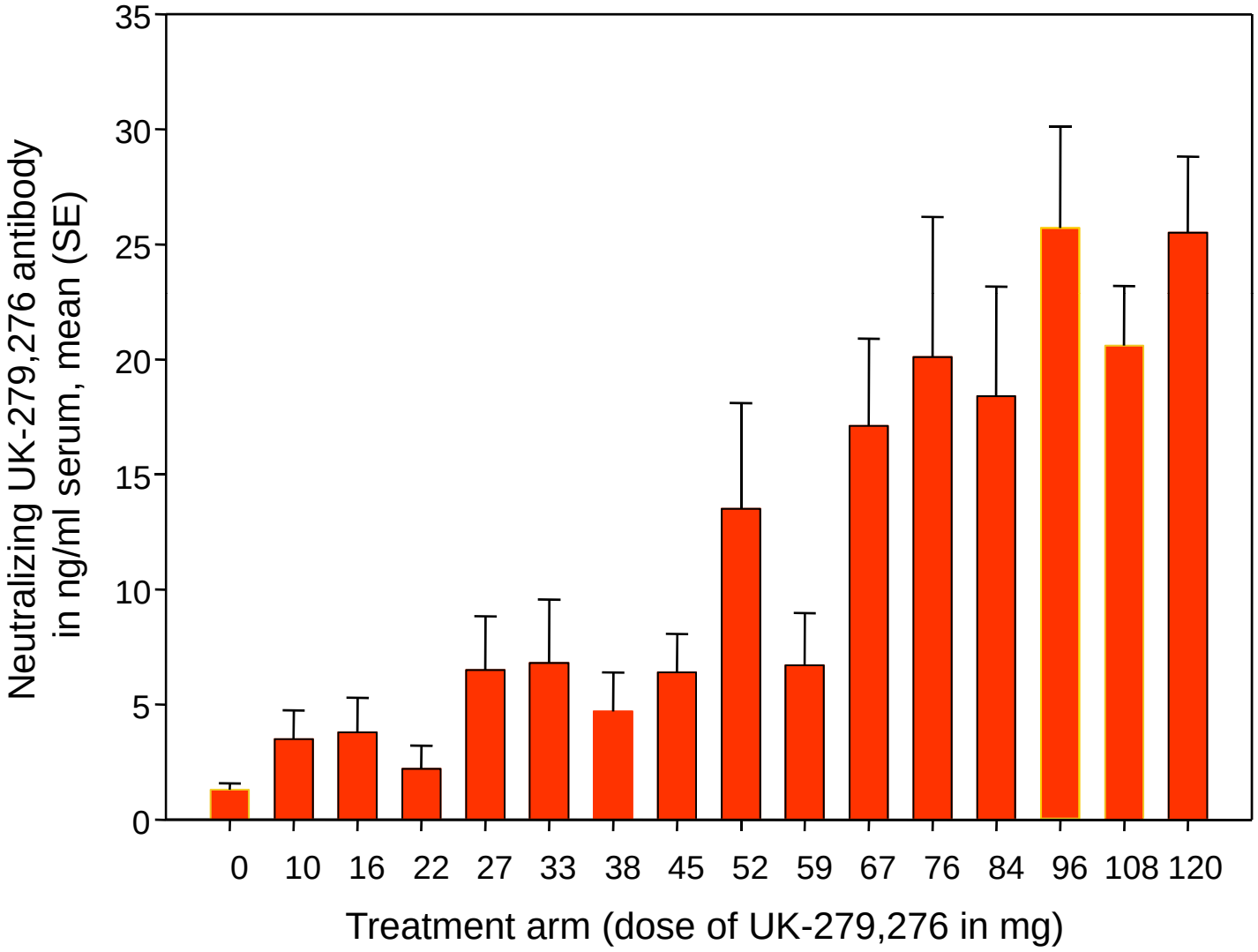

STOP
SSS (Change from Baseline)
30
25
20
15
10
5
0
0
20
40
60
80
100
120
Dose (mgs)

Week - 66

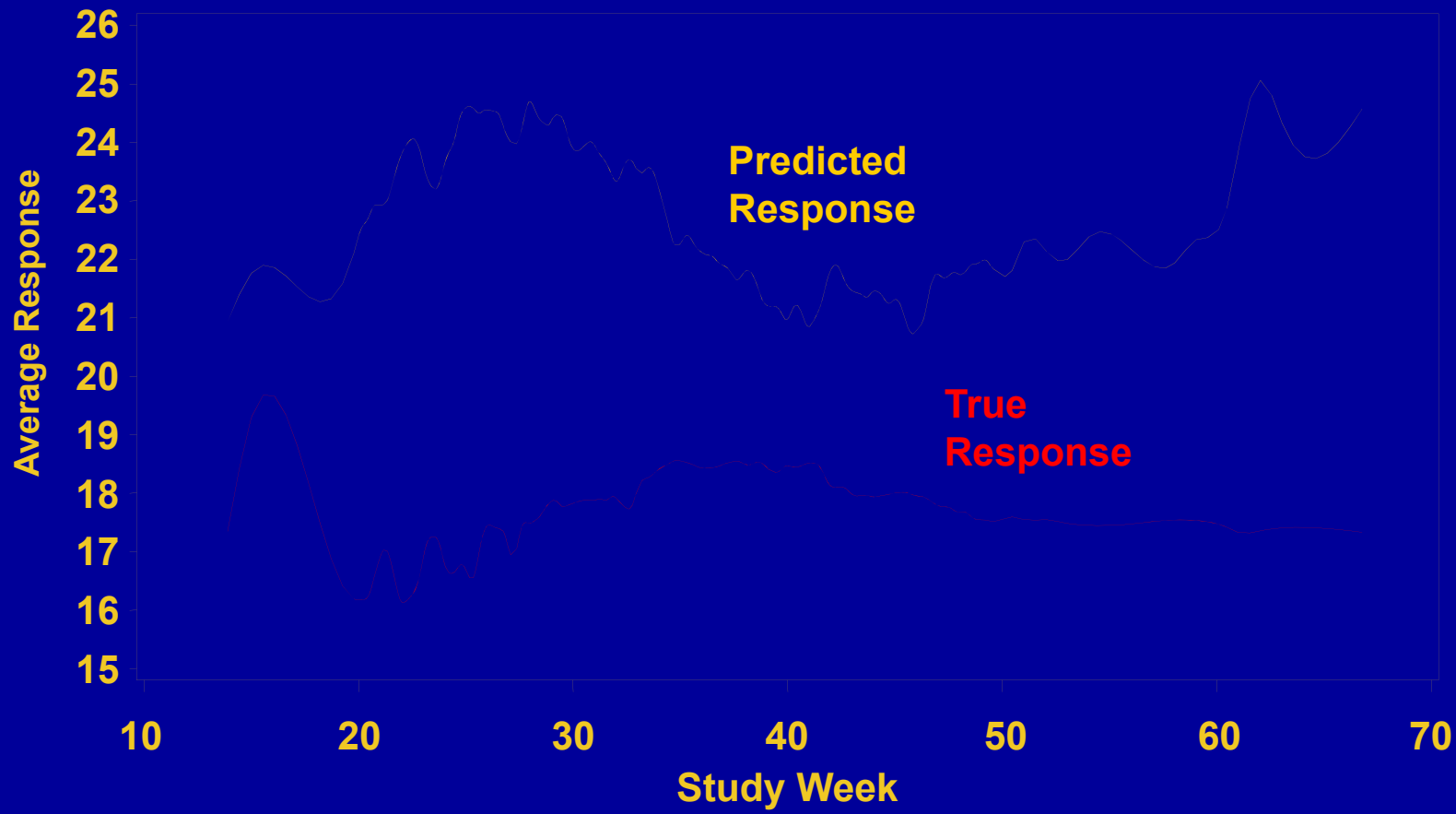
SSS Response Curve – Primary Analysis



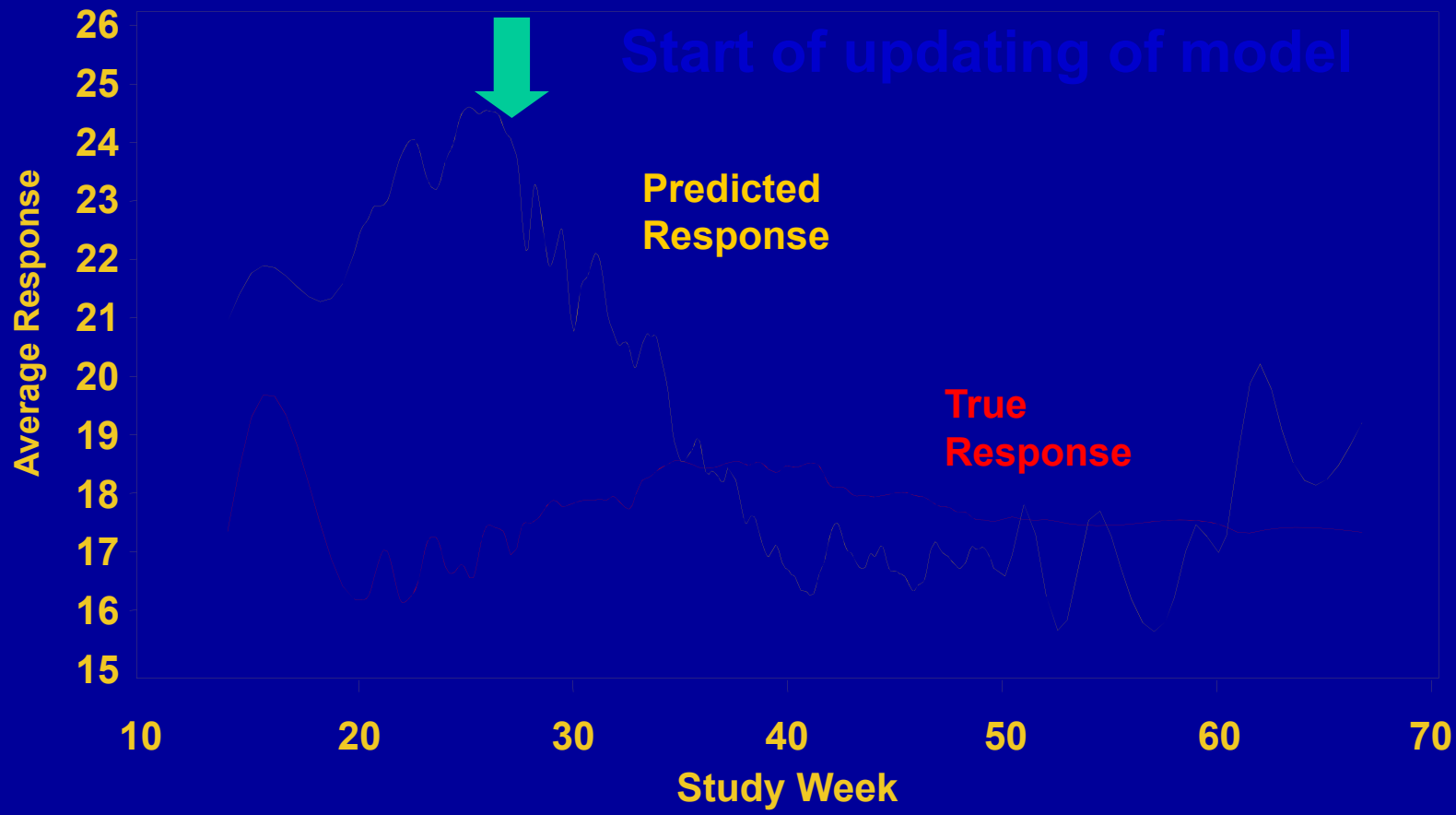
Neutralizing antibodies to UK-279,276



Comparison of Imputed and True Response—Initial longitudinal model



Comparison of Predicted and True Response—Updated longitudinal model



Mike Krams's conclusions

- Complex sequential design possible
- Requires up-front effort
  - Simulation
  - Validation
  - Regulators: US, EMEA, UK, Germany, Sweden, Canada
- Saves time & resources
- Algorithm perceived as success
  - Hunted effectively
  - Stopped as soon as allowed, clear-cut decision
  - Could have stopped sooner!

Consequences of Adaptive Approach

- Fundamental change in way we do medical research
- We'll get the dose right
- Better treatment of patients
- More rapid progress
- . . . at less development cost

OUTLINE

- Bayesian adaptive design?
- Predictive probabilities in design
- Adaptive dose-finding: A stroke trial