

The Promise and the Perils of Adaptive Designs

Donald A. Berry
dberry@mdanderson.org



Outline

- **Gaps**
- **Summary of promises/perils**
- **The future**
- **Fragmented presentation, but being adaptive ain't easy!**

Bayes/frequentist

- Both can be adaptive
- Frequentist: Sample space for the experiment
- Bayes: Learning via observation, maximize some objective, model
- Bayes as a tool to build good frequentist designs!
- Simulate (or other) to control type I error, power, etc.

Current use of Bayesian adaptive designs

- **MDACC**
(> 200 trials)
- **Device companies**
(> 14 PMAs, many IDEs)
- **Drug companies**
(Most of top 40 companies;
many small companies)

Current areas of application of adaptive drug trials

- Oncology
- Spinal Cord Inj
- Migraine
- HIV
- Lupus
- Hep C
- Sepsis
- Pre-term labor
- Diabetes
- Constipation
- Obesity
- Micturition
- Stroke
- Emesis

The Coming Bayesian Tsunami of Clinical Development

目次	
インタビュー特集	P1-3
臨床開発にベイズ統計学の「津波」が到来	
行政関連ニュース	P4-5
BMSのスプライセル、FDAが承認	
FDA、臨床試験とパイオリサーチ・モニタリングの規制改正を発表	
FDA、ジョンソン・エンド・ジョンソンのプレジスタを承認	
製薬企業ニュース	P6-7
J&J、PCHを166億ドルで買収	
ファイザー、ゾロフトのジェ	

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。

——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



ドナルド・A・ベリー博士

What types of designs?

Mostly Phase II, some I & III & IV

1. Dose-finding
2. Seamless phases
(phase III traditional)
3. Dropping doses
4. Identifying responders
5. Adaptive randomization

Pre-clinical experiments

- **Save animals, resources**
- **Training ground for learning how to run adaptive designs**

Some attitudes expressed here

- **“Most of this is not new”**
- **“There are serious problems with adaptive designs”**

What is not new?

- **Stagewise designs for variance estimation**
- **Sample size re-estimation**
- **Interim analysis for early stopping**
- **Adaptive randomization**

What is new?

- We're doing it!

Lack of knowledge/understanding

- Sponsor's drug team
- Bob O'Neill: "a lot of buzz, a lot of cluelessness out there"
- Even among statisticians

Critical importance

- **Prespecification**
- **Control Type I error rate**
- **Controls/randomization**
- **Address who knows what and the associated problems, especially in phase III**
- **Modeling, modeling, modeling**

Simulation

- Many scenarios; gazillions of runs
- Stu Pocock's question about bias in picking "best dose"
- Pluck out runs as examples for the protocol appendix
- Databases, data sharing issue

Some gaps

- **Accrual rate should be part of the design**
- **Not enough attention to modeling patient outcomes over time**

Role of DSMB

- DSMBs don't design or redesign trials
- Sponsor plans ahead, *makes all decisions*, writes protocol/DSMB charter; DSMB is an automaton
- Lots of work. Negative?
- Armin Koch's suggestion?

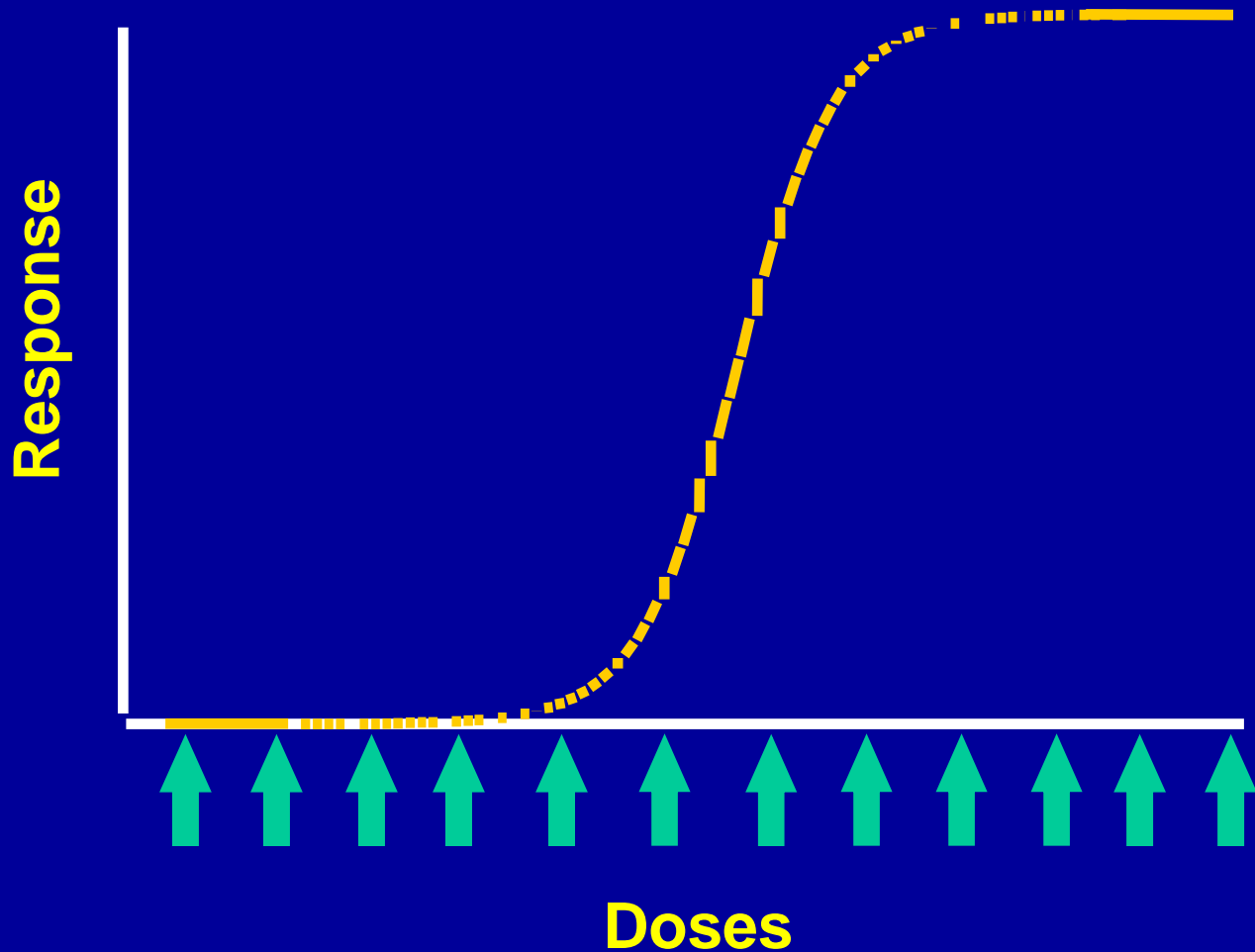
Perils

- **Who knows what and when**
- **Effect of info “leakage”**
- **Lack of knowledge/training (If you don’t know what you’re doing, don’t do it!)**
- **We may accrue (to the adaptive bandwagon) too fast!**

FDA/Industry

- Agreement as to promise and perils
- FDA is leading, statistically & otherwise
- FDA “principles” are well conceived and serve as important guidelines

Phase II dose-finding: where to put doses?



**What do you want to
learn from the trial?
A challenging example**

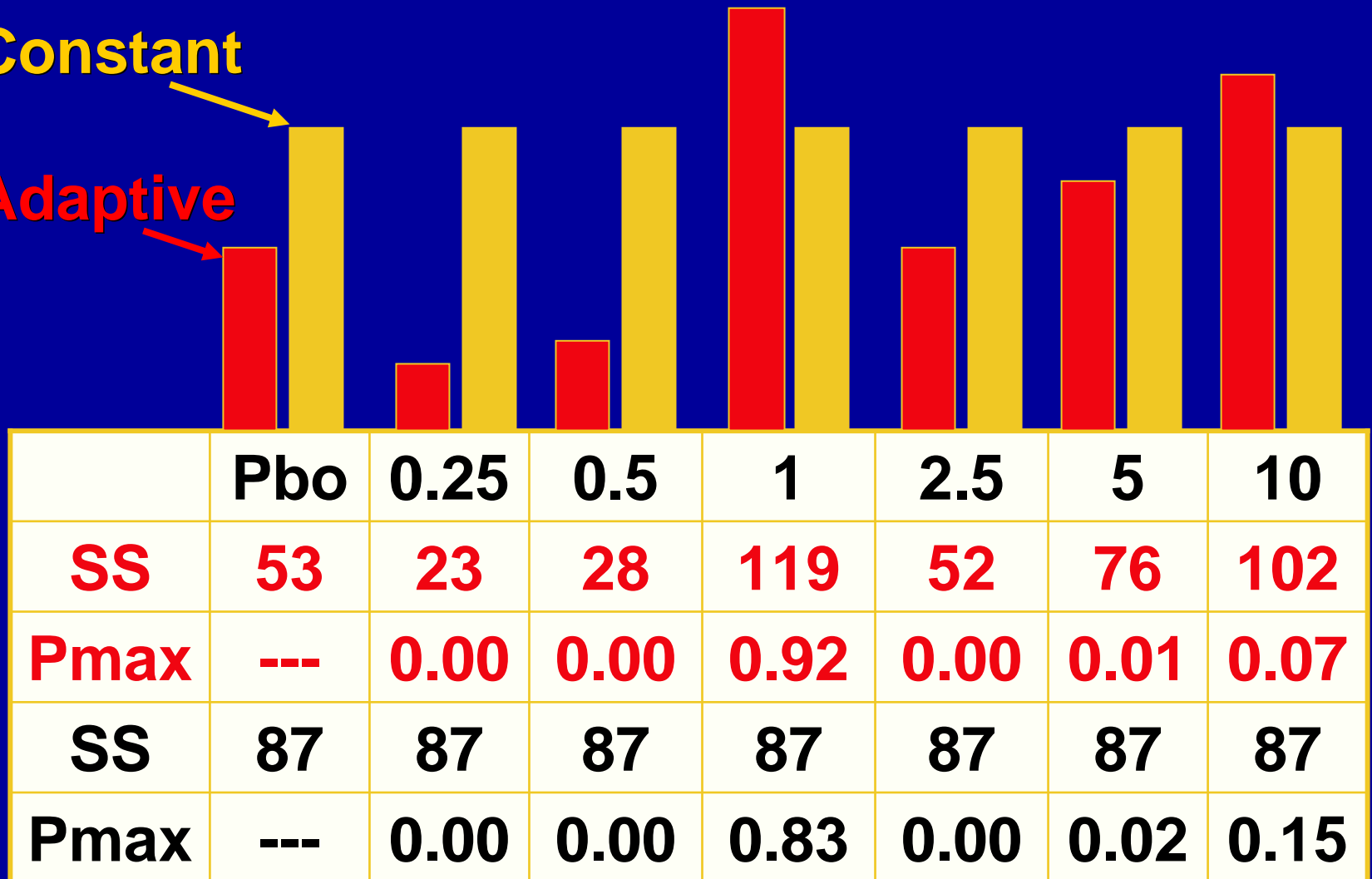
Simulation #3

Dose	P_1	P_2	P_3	P_4	UTIL
0	0.06	0.05	0.05	0	0
0.1	0.10	0.05	0.06	0	0
0.5	0.13	0.08	0.07	0	0.063
1	0.30	0.25	0.11	0	0.656
2.5	0.17	0.12	0.08	0	0.323
5	0.20	0.15	0.09	0	0.457
10	0.23	0.18	0.10	0	0.532

Sample size & Pmax (both employ early stopping)

Constant

Adaptive

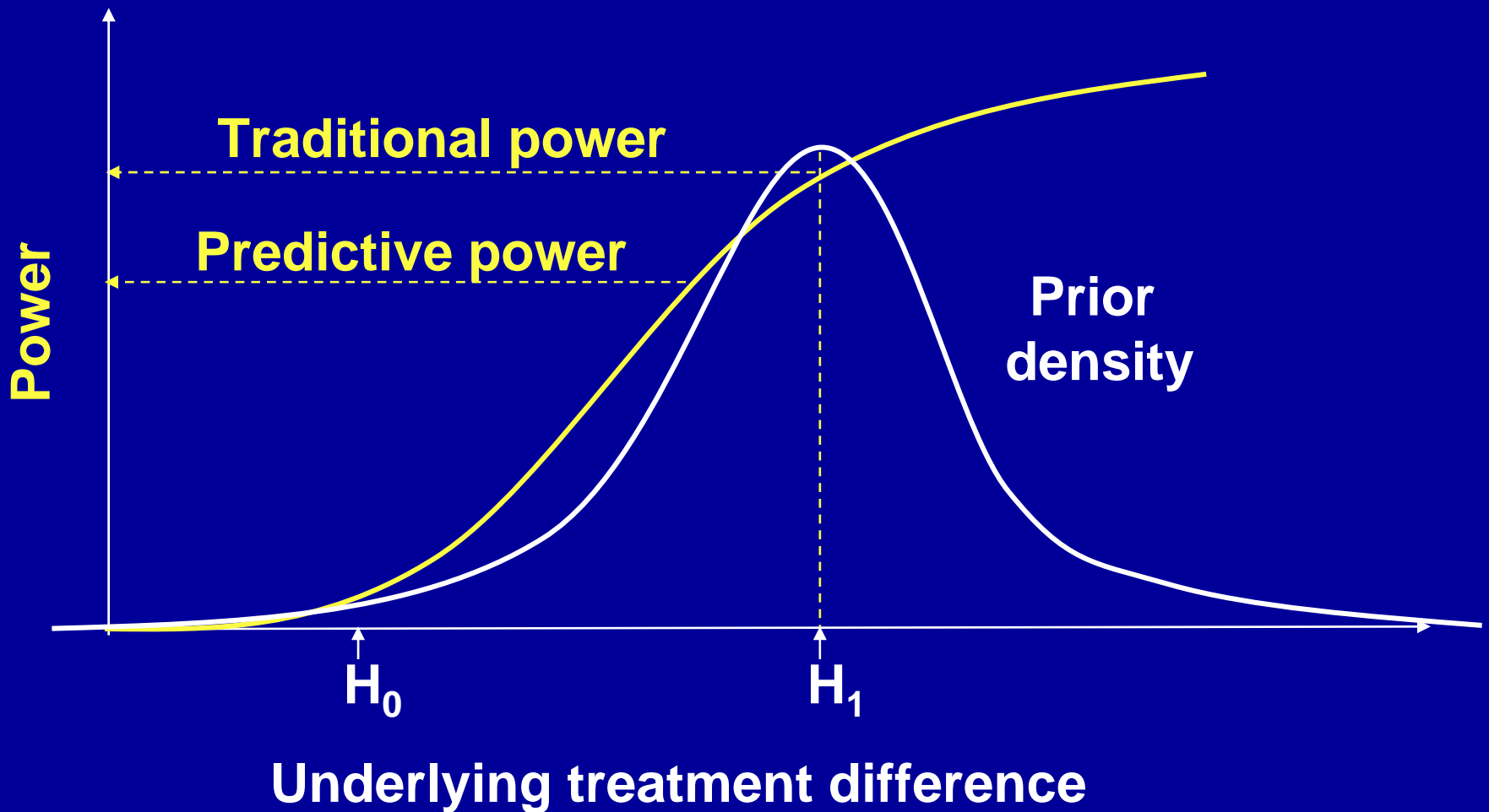


Scenario-specific operating characteristics

	Adaptive	Constant
P(Success)	0.906	0.596
P(Cap)	0.092	0.404
P(Futility)	0.002	0.000
Mean SS	453	606
SD SS	187	205
Mean TDose	1663	1662
Max TDose	3771	2384.25

**Sorry performance in
phase III**

True (predictive) power



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

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).

SAINT II:

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

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

Outcome Variable	Placebo Group	NXV-059 Group	Difference between NXV-059 and Placebo† % or score (95% CI)	P Value
Modified Rankin scale score (primary end point)				
No. of patients	849	850		
Score — no. (%)				
0	93 (11.0)	131 (15.4)	1.4	
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**



**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 \ll “power”
- Simulate the trial!
- Use adaptive design
 - Start trial cautiously
 - Look at the data!!
 - Ask “where we are going”
 - Adapt (in this case, abort)

The Future

- **Unknown**
- **Promising: Who wants to be non-adaptive?**
- **Pet rock? Hula hoop?**
- **Progress slow while we train statisticians and “get up to speed”**
- **If we’re not careful, we’ll crash this ship into the rocks**