The Power of Designed Experiments

Nancy Flournoy

University of Missouri

Conference on Applied Statistics for African Development
Cotonou, Benin
March 6, 2013

This talk has three parts

1. A simple example of the power of designed experiments
2. A story of discovery intertwining hypothesis generation and planned experiments
3. Some examples of adaptive designs for dose-finding

Outline

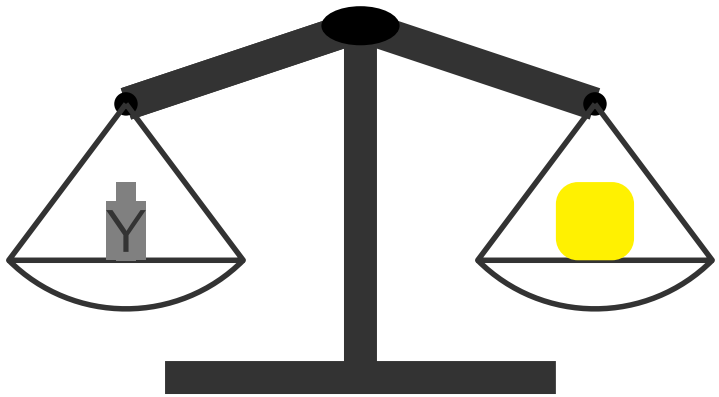
Design Matters, Gold Mass Example

- Four gold bars: A, B, C, D.
- \$1 charge each time the scale is used.
- The scale may not be very accurate: error$\sim N(0, \sigma^2)$.

How would you weight the gold bars if you have only \$4?

Method 1

Weight each gold bar individually:



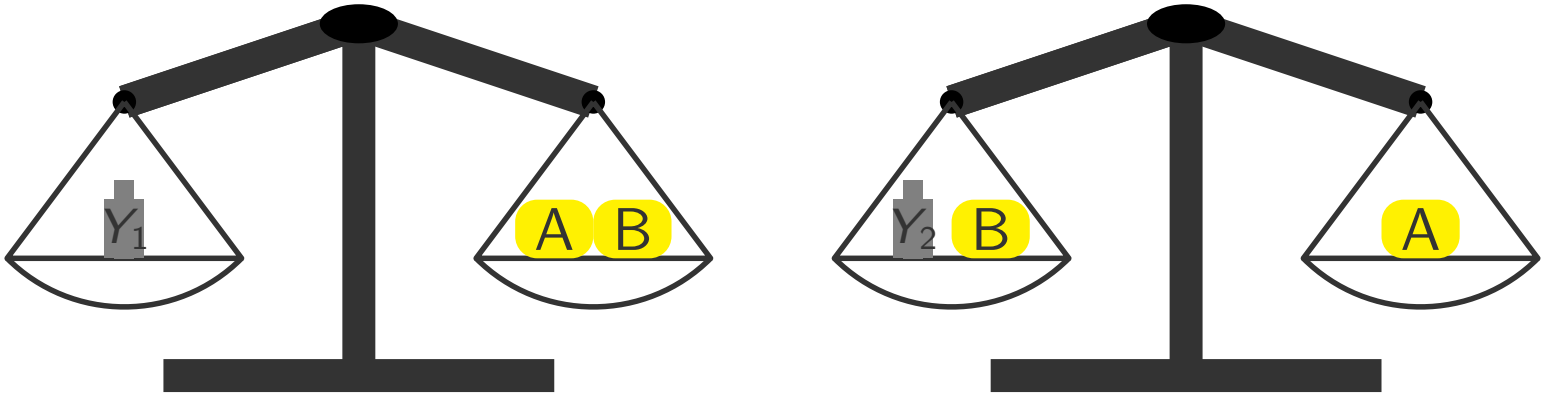
$$Y_1 = A + \varepsilon_1; \qquad Y_2 = B + \varepsilon_2;$$
$$Y_3 = C + \varepsilon_3; \qquad Y_4 = D + \varepsilon_4.$$

How precise is this method?

- $\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \sigma^2$

Method 2

Weight $A + B$, $A - B$, $C + D$, and $C - D$:



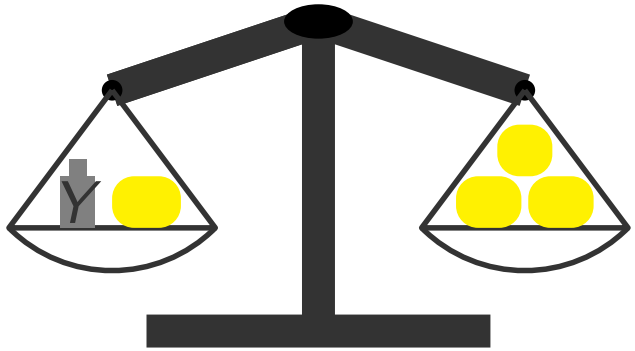
$$Y_1 = A + B + \varepsilon_1; \qquad Y_2 = A - B + \varepsilon_2;$$
$$Y_3 = C + D + \varepsilon_3; \qquad Y_4 = C - D + \varepsilon_4.$$

How precise is this method?

- $\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \frac{1}{2}\sigma^2$

Method 3

Weight the difference between one gold bar and the sum of the other three:



$$Y_1 = A + B + C - D + \varepsilon_1; \qquad Y_2 = A + B + D - C + \varepsilon_2$$
$$Y_3 = A + C + D - B + \varepsilon_3; \qquad Y_4 = B + C + D - A + \varepsilon_4$$

How precise is this method?

- $\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \frac{1}{4}\sigma^2$

Comparison

The three methods corresponded to three designs.

- All methods give unbiased estimates of the weights.
- Method 3 has smallest variances $\frac{1}{4}\sigma^2$.
- To achieve the same precision
  - Method 1 needs \$16
  - Method 2 needs \$8
  - Method 3 needs \$4

Are there any designs better than Method 3?

No. It can be shown that Method 3 is the optimal design.

Optimal Design

Optimality criterion: specify treatments to allocate to subjects, for example, so as to

- minimize variance of estimator
- minimize variance of estimator given constraint on cost or subject risk
- minimize confidence ellipsoid for multiple estimators
- minimize maximum eigenvalue of information matrix

Typical Goals:

- Reduce sample size and/or cost and/or subject risk needed to achieve a specific precision for a desired estimator.
- Increase precision for given cost and/or sample size and/or limit to subject risk.

Outline

A Powerful Cycle

1. Observe and imagine
2. Generate hypotheses
3. Design experiment to inform hypothesis
4. Repeat

Early Days of Bone Marrow Transplantation

Big Concern

High risk of bacterial infection due to patients' long time without granulocytes

Resultant Action

Prophylactic granulocyte transfusions

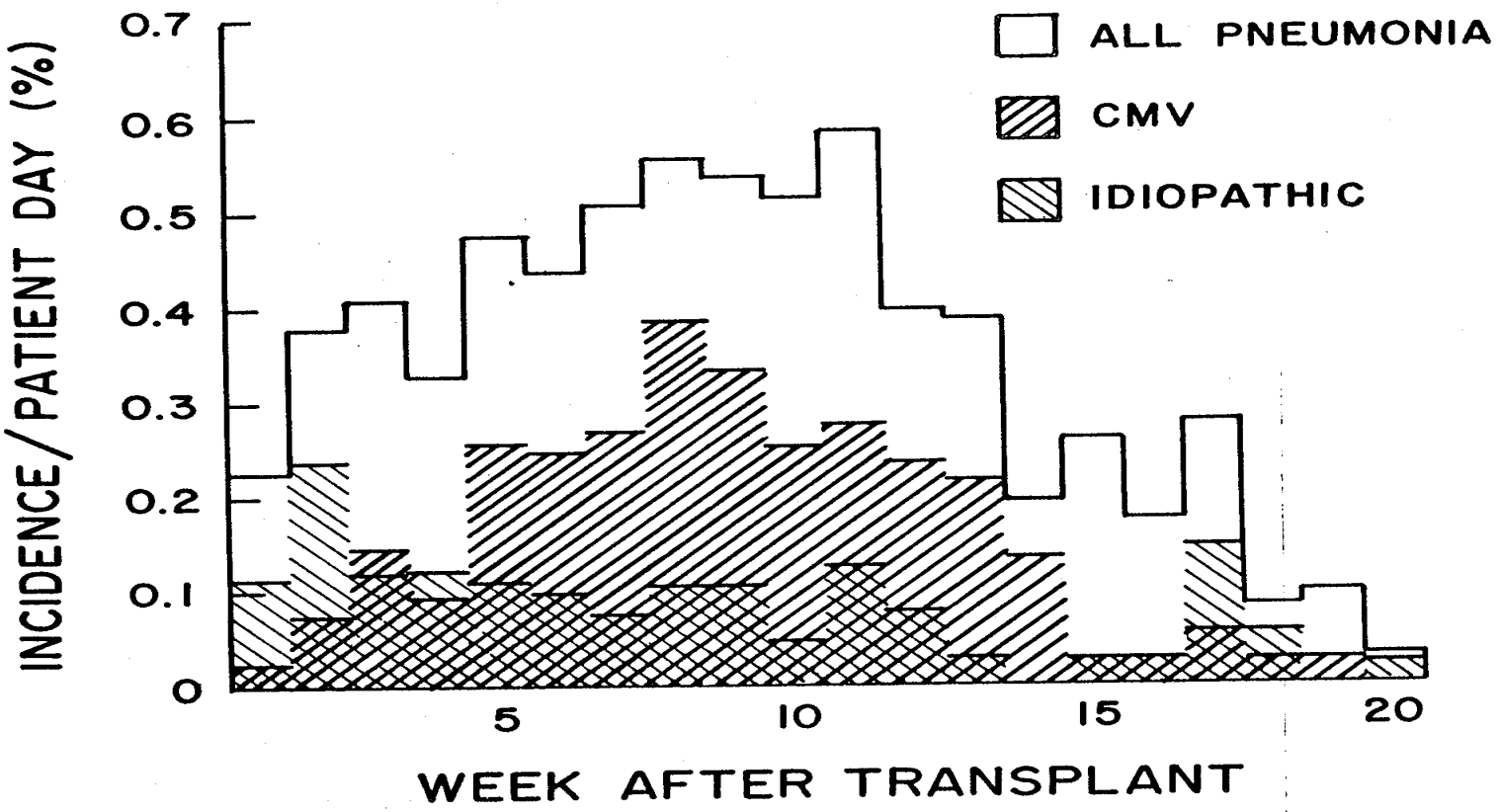
Ten Years of Observations About Pneumonia

525 AMONG PATIENTS TRANSPLANTED WITHOUT AN IDENTICAL TWIN DONOR:

Cases of Pneumonia	215	(38% of 525)
Nonbacterial With CMV Identified	40%	
Other with Viruses Identified	29%	
Other with Viruses Not Identified	31%	

84% of the 215 cases were fatal.

Incidence of pneumonias from 10 years of observation



Speculation on why twins had no CMV pneumonia

- Less pretransplant chemotherapy and irradiation
- Less suppression of the immune system

Failure to see cause

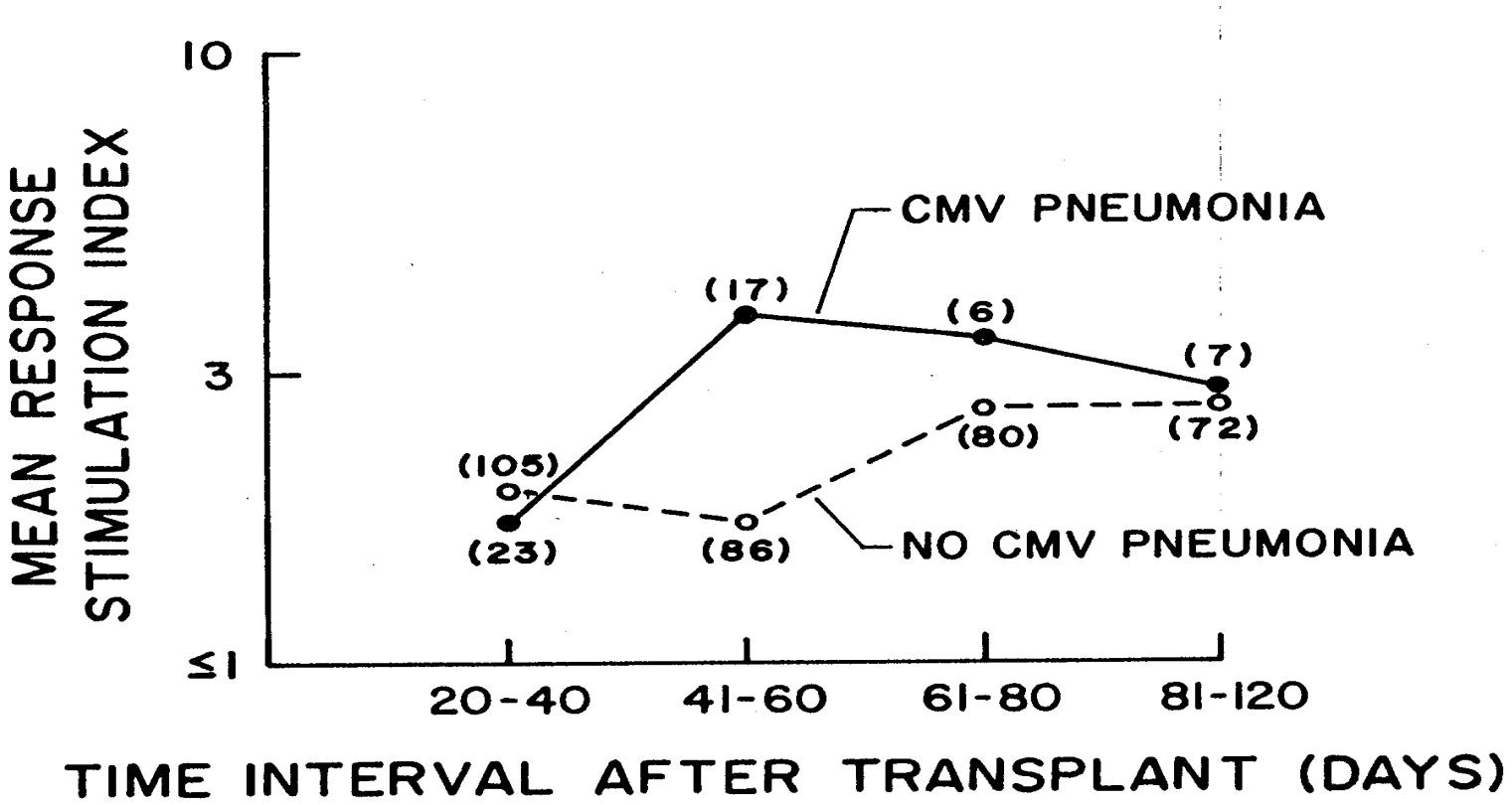
- Twins had a reduced need for blood transfusions because their immune system was not so suppressed.
- Twins had no prophylactic granulocyte transfusions

Designed Surveillance

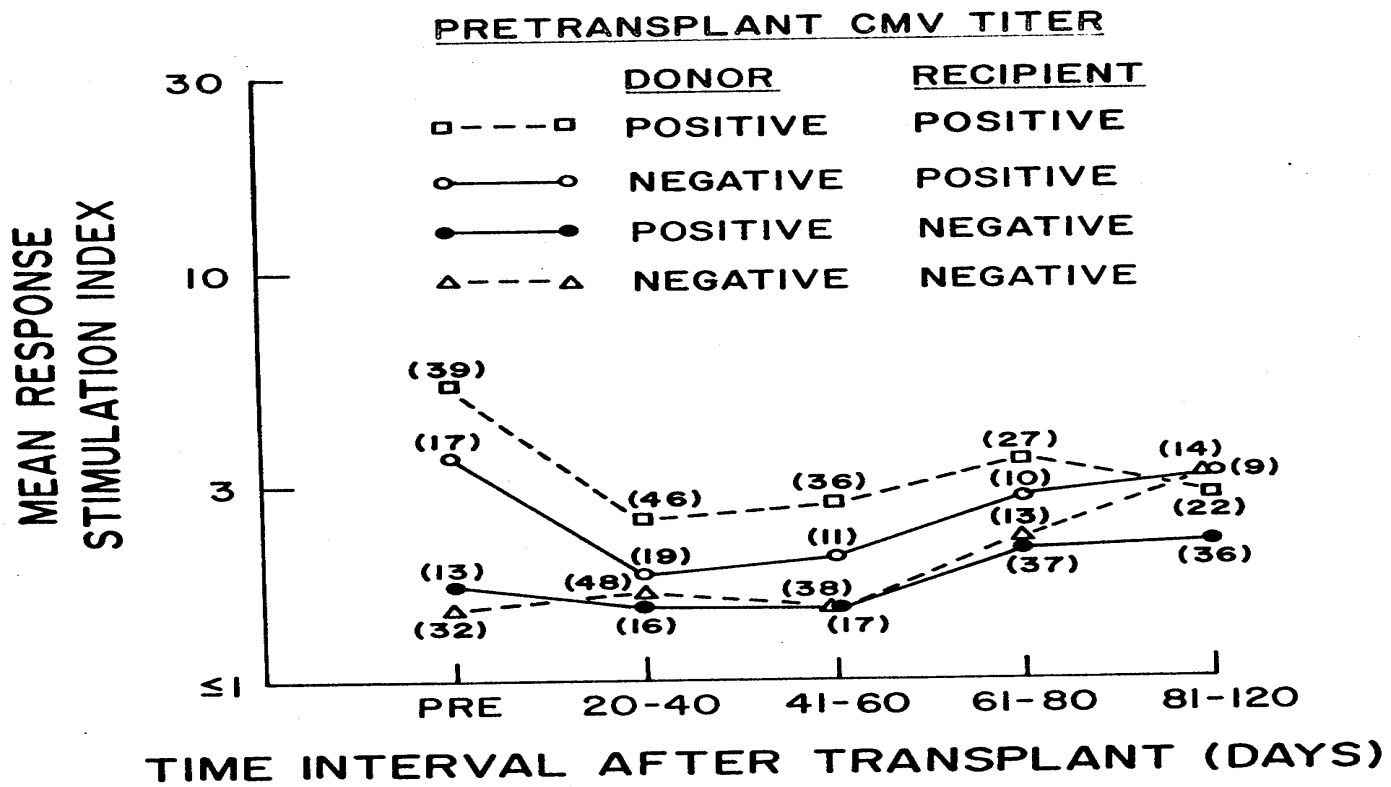
To standardize diagnostic procedures and remove bias

CMV antibodies were measured on a regular schedule on 158 patients over 2 years

Infection Rate from Designed Surveillance: Lymphocyte Response to CMV Antigen



Infection Rate from Designed Surveillance: Lymphocyte Response to CMV Antigen



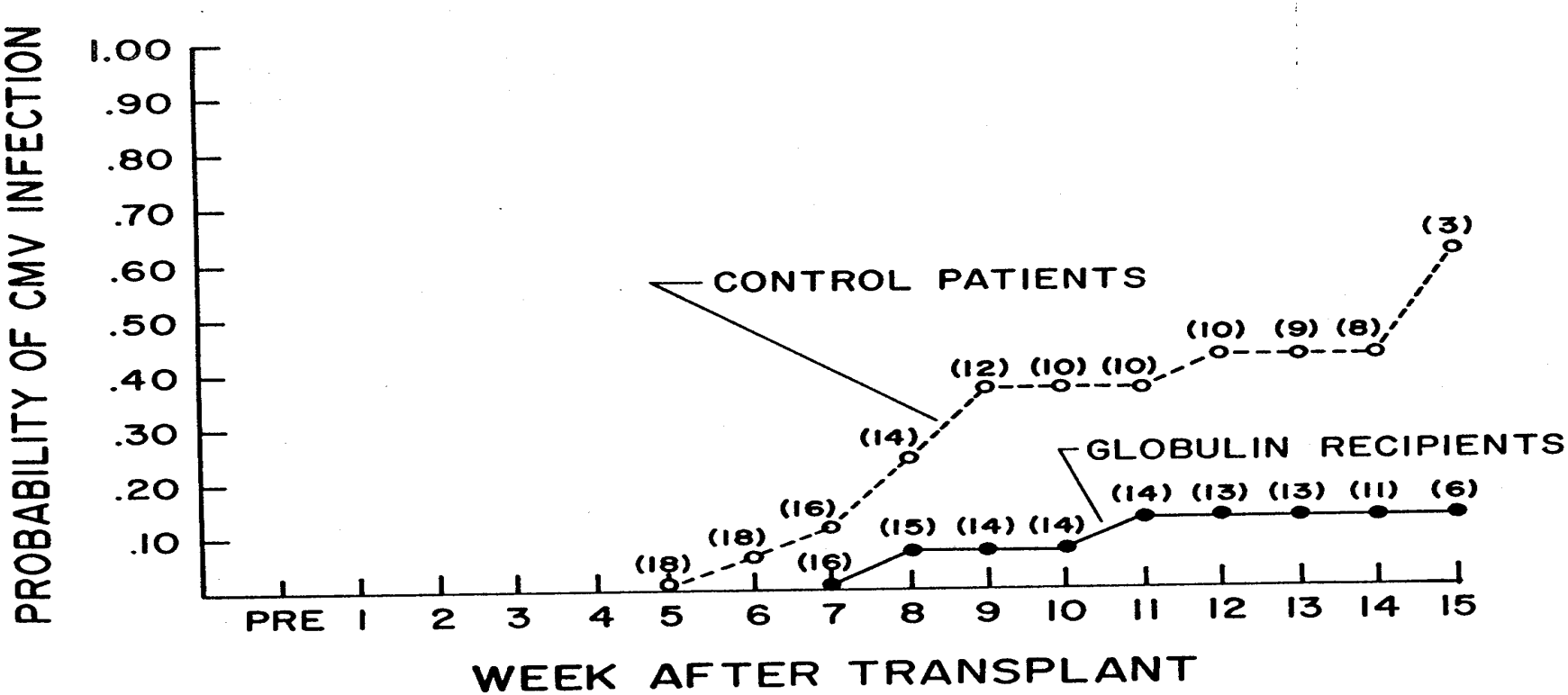
Pretransplant CMV and Subsequent Infection

Pretransplant Status	**CMV Infection Rate Post Transplant**
CMV in patient blood pretransplant	69%
No CMV in donor blood pretransplant	
Donor with CMV in blood pretransplant	57%
Donor with NO CMV in blood pretransplant	28%

Randomized Globulin Trial: Incidence of CMV Infection

Patients receiving	**GLOBULIN**	**NO GLOBULIN**
Granulocytes from seropositive donors	7/8 (88.5%)	6/7 (85.7%)
Granulocytes from seronegative donors	1/5 (20.0%)	0/6 (00.0%)
No granulocytes	2/17 (11.8%)	8/19 (42.1%)

Unplanned comparison of CMV infection among patients who did not receive prophylactic granulocyte transfusions



Randomized Blood Screening × Globulin Study

- Spent one year to develop reliable CMV screening techniques
- Included only patients who were sero–negative pretransplant
- Conducted randomized 2×2 clinical trial
- Stratified on donor's CMV status

Incidence of CMV Infection among 85 patients studied for at least 62 days after transplantation

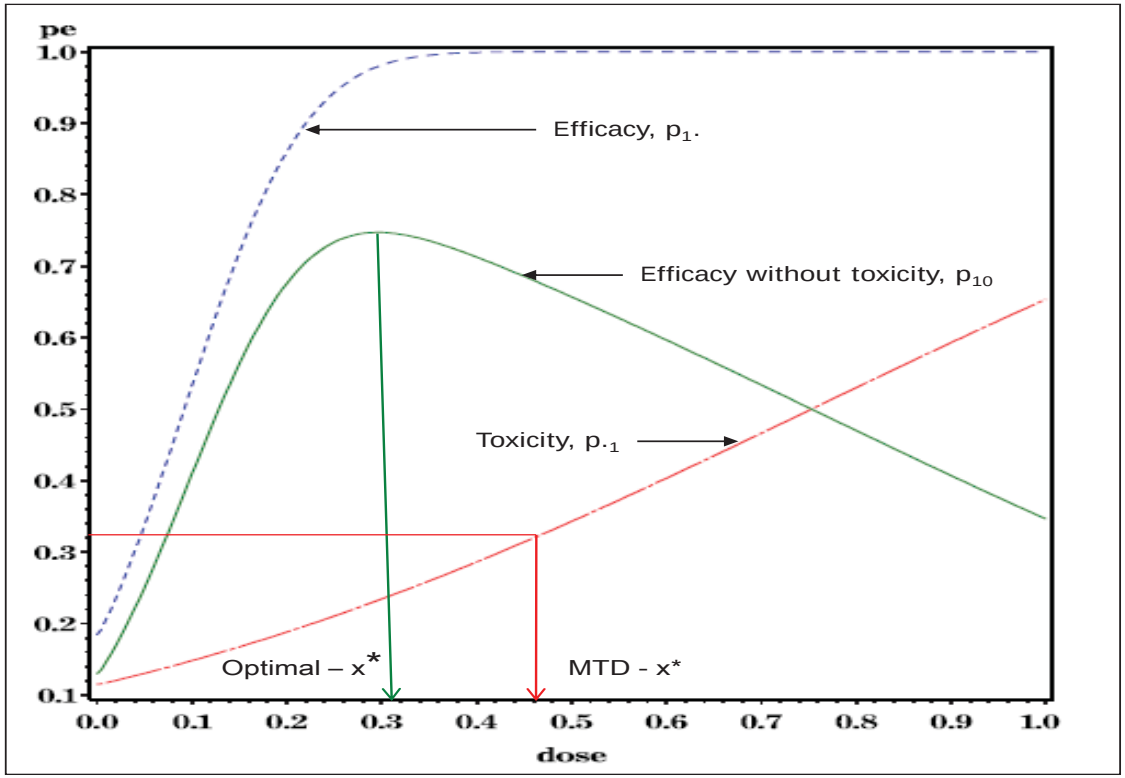
Donor's CMV status	BLOOD TRANSFUSIONS SERO–NEGATIVE	SERO–POSITIVE
Seropositive	1/22 (04.5%)	8/25 (32.0%)
Seronegative	3/12 (25.0%)	5/16 (31.3%)

Can't See the Forest for the Trees

- It may take years to get the question right.
- It may take years to develop biological, chemical, physical and/or statistical methods that address the question.
- Few arm randomized trials are good for confirmation.
- Few arm randomized trials are not good for exploration.

Outline

Basic Response Function Shapes For Dose-finding



[0] curtesy of Valerii Fedorov and Yuehui Wu

Consider Binary Responses: $F = \mathrm{P}(y = 1|x, \theta)$ is a probit

Type I best dose

$$x^*(\theta) = \arg\min_{x \in \mathcal{X}} \{ \mid \mathrm{P}(y = 1|x, \theta) - U \mid \},$$

where U is a predefined constant.

$$\mathrm{P}(y = 1|x, \theta) = F(\theta_1 + \theta_2 x).$$

Type II best dose

$$x^*(\theta) = \arg\max_{x \in \mathcal{X}} \left[\mathrm{P}(y = 1|x, \theta) \right].$$

$$\mathrm{P}(y = 1|x, \theta) = F(\theta_1 + \theta_2 x + \theta_3 x^2).$$

Best Intention Designs

1. Estimate the target dose
2. Treat next subject at that dose

Bayesian and frequentist best intention designs are proliferating:

Recent examples include Li, Durham and Flournoy (1995), West (1997), Conaway (2004), Smith, et al.(2006), Bartroff and Lai (2010), Cheung (2010) and Thall (2010).

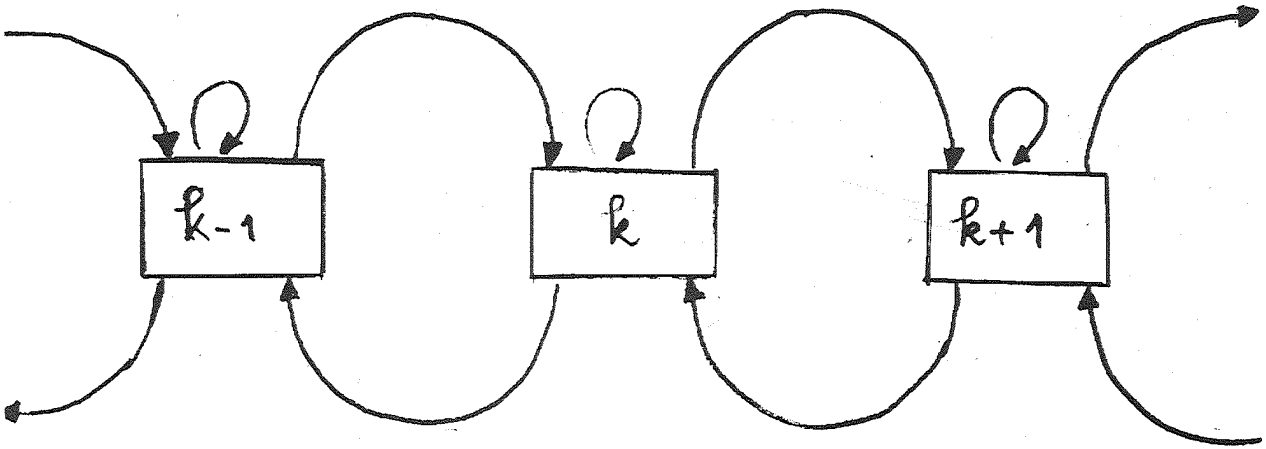
Up and Down Designs

Treat the next subject at the same, or an adjacent dose

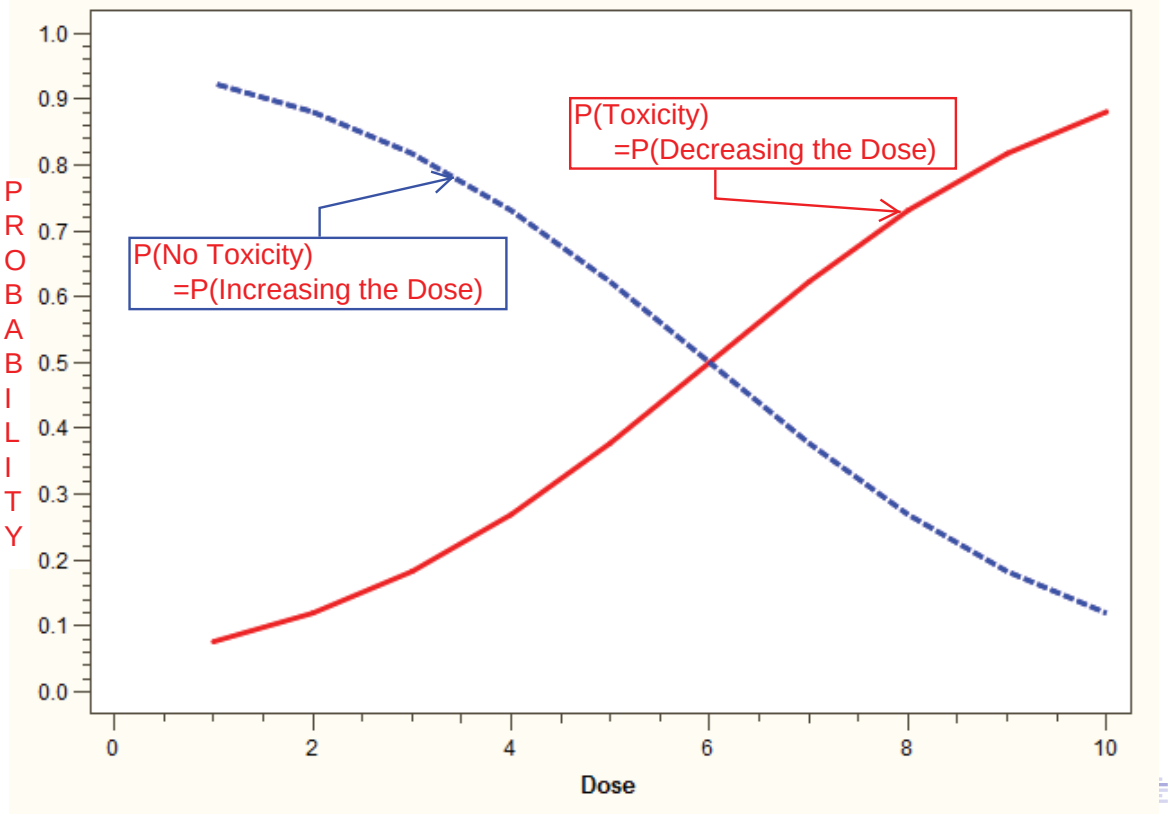
1. Select the next dose based on the outcome of the last trial and a random bias
2. Select the bias to cause treatment allocations to converge in distribution, centered around the target dose

Up and Down Design

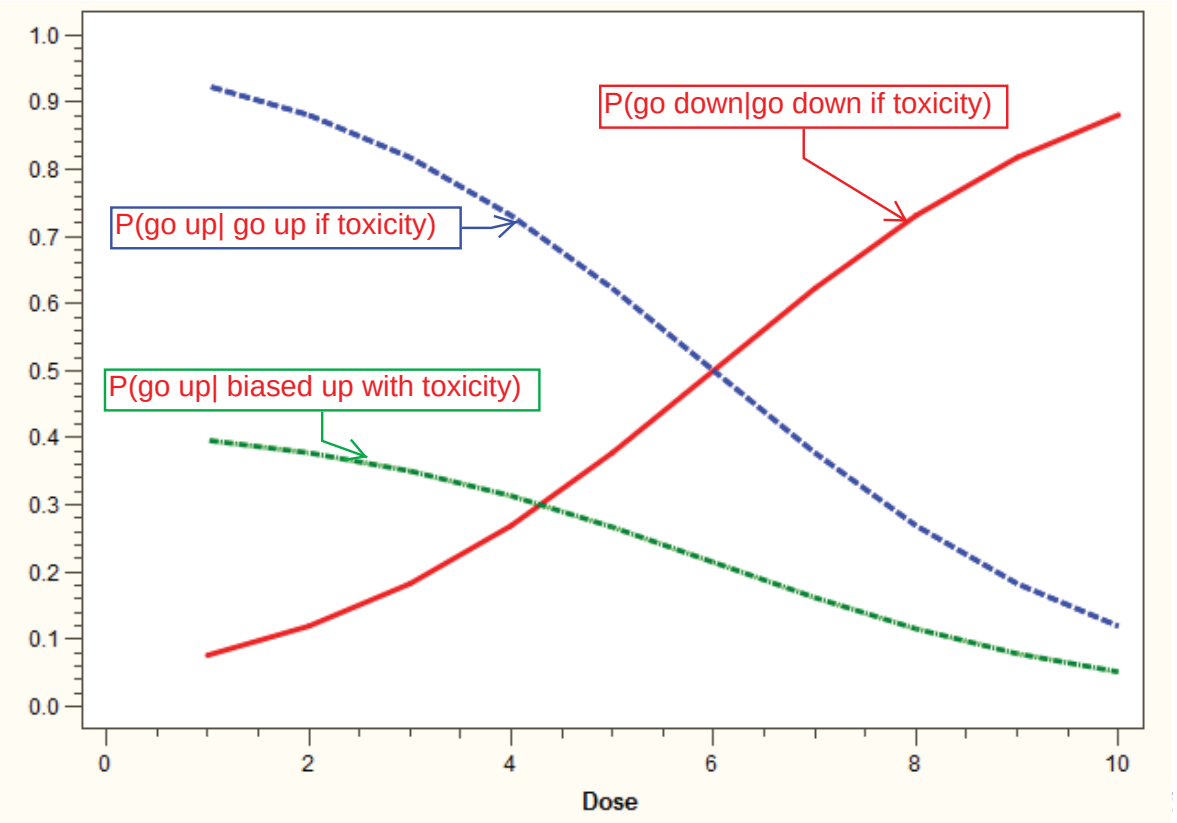
Allocations follow a random walk



Allocations Cluster Around the Balance Point, here the 50th percentile



Move the balance point to the 30% toxicity rate to cluster treatments there



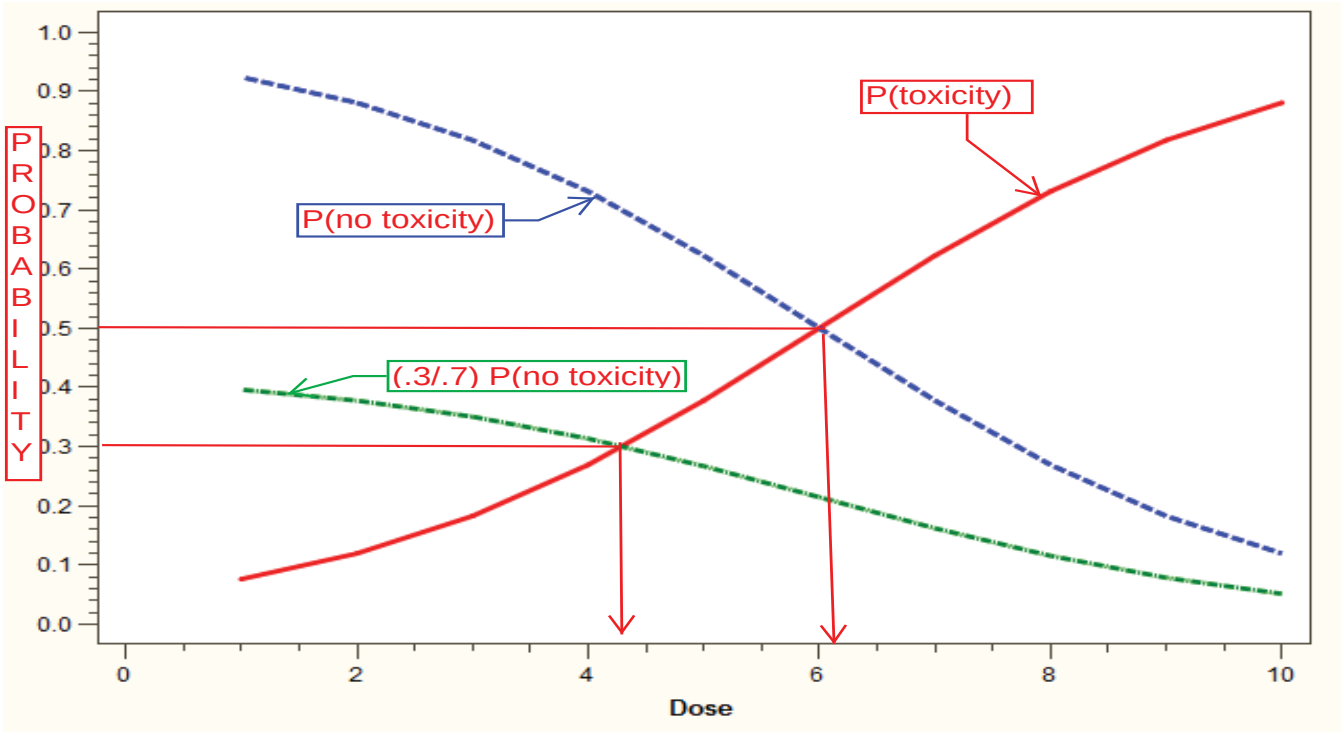
Cluster Treatment Allocations Around Dose with .3 Toxicity Rate

One way to Move the balance point is to change P(UP) using a biased coin, e.g.,

1. go down if toxicity
2. go up if 'no toxicity' and heads wp .3/.7
3. else, stay put.

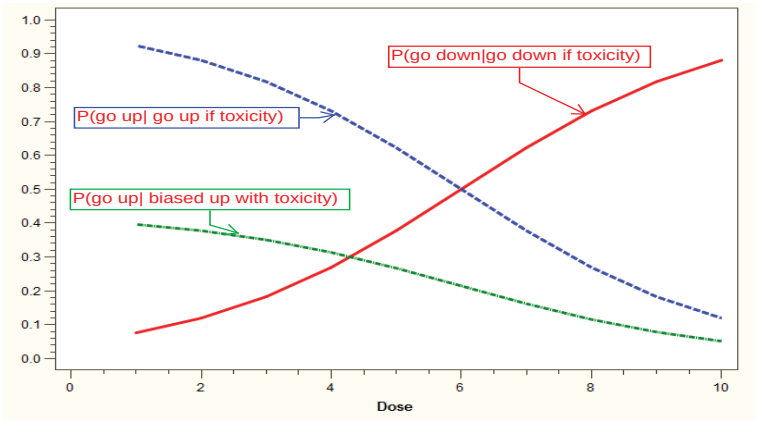
[0] (1) Durham & Flournoy, 1994.
Stylianou and Flournoy (2002) recommend estimating target by isotonic regression.

$P(\text{up}) = (.3/.7)P(\text{No Toxicity})$ clusters allocations around the 30% toxicity rate



[0] Dorman, 1957; Durham & Flournoy, 1994; Durham & Flournoy, 1995; Durham, Flournoy and Haghighi

Cohorts Provide Another Way to Cluster Allocations Around Dose with .3 Toxicity Rate



Change P(up) by allocating in cohorts: e.g.,

1. go up if 0 toxicity in group of size 2
2. go down if 1–2 toxicities
3. Allocation distribution clusters at dose where $P(\uparrow) = P(\downarrow)$

[0] (1) Wetherill, 1963 (2) Tsutakawa, (3) Gezmu & Flournoy, 2006.

Now Consider Interest in the MTD

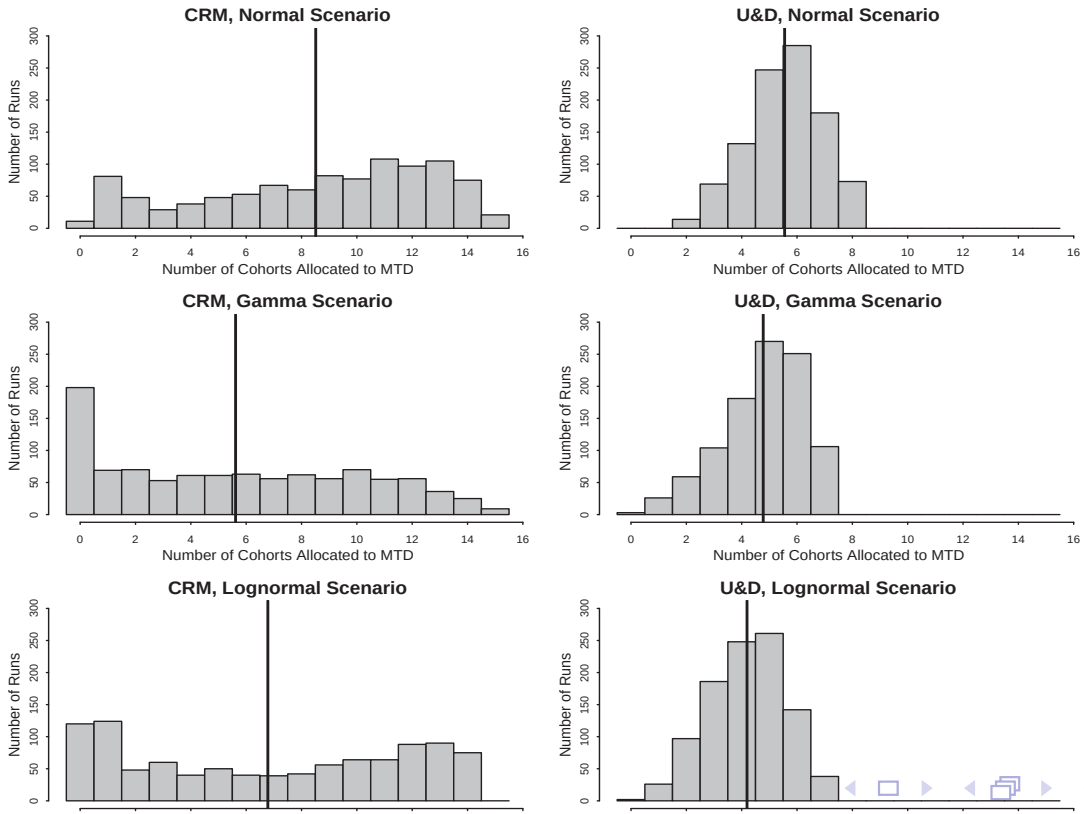
Proportion of Correct Selection of MTD
by CRM and
by U&D with groups of size 2 after 8 cohorts

Senario	MTD	CRM	U&D
Uniform	1	50.2	54.0
Gamma	2	36.6	40.8
Normal	3	57.8	56.4
Lognormal	4	46.7	33.0
Weibull	5	39.0	38.1
Logistic	6	26.0	32.2

From Oron & Hoff (2012)

Oron & Hoff give similar results for 16 cohorts.

Number of Cohorts Allocated to MTD: CRM left; U&D right; Oron, et al., 2012



Impossibility Theorem by Azriel, Mandel & Rinott, 2011

No sequence of best intention designs can guarantee almost-sure convergence to the MTD on the class of all monotone increasing dose-toxicity functions.

Azriel (2011) proves Cheung & Chappell (2002) conjecture for CRM consistency

Oron et al. (2012) sampled dose-response curves and found conditions met 26% of the time.

Azriel sampled curves that satisfy the commonly used power function prior and found conditions satisfied 43% of the time.

How to use this information?

Cheung (2011) provides instructions for constructing priors that will satisfy his conditions.

Is constructing priors to make an algorithm work consistent with Bayesian philosophy?

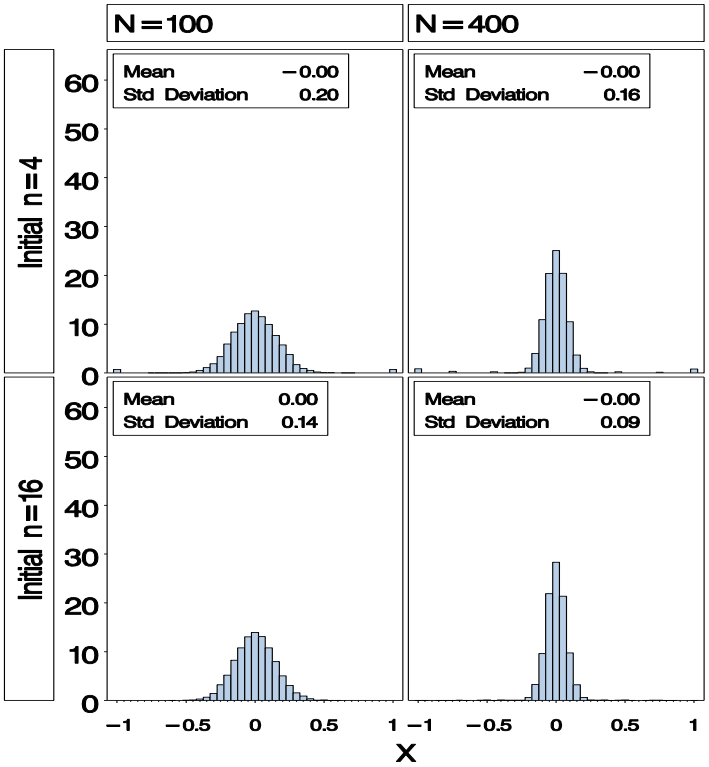
Most Informative Designs

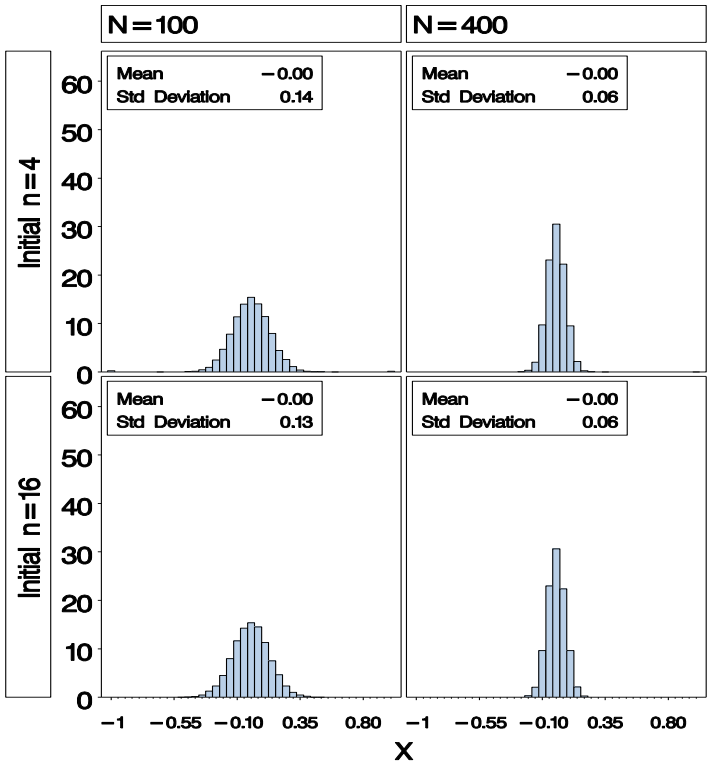
- Goal: allocate to maximize estimation accuracy
  Allocate doses to gain maximum information.
- $\mathcal{D}$–optimality criterion$\Rightarrow$ allocate to maximize the determinant of the information matrix; (approx) minimize the variance–covariance matrix.
- Dose allocations for each new cohort are based on updated estimates of the covariance matrix.

Compound Criteria, Penalties and Costs

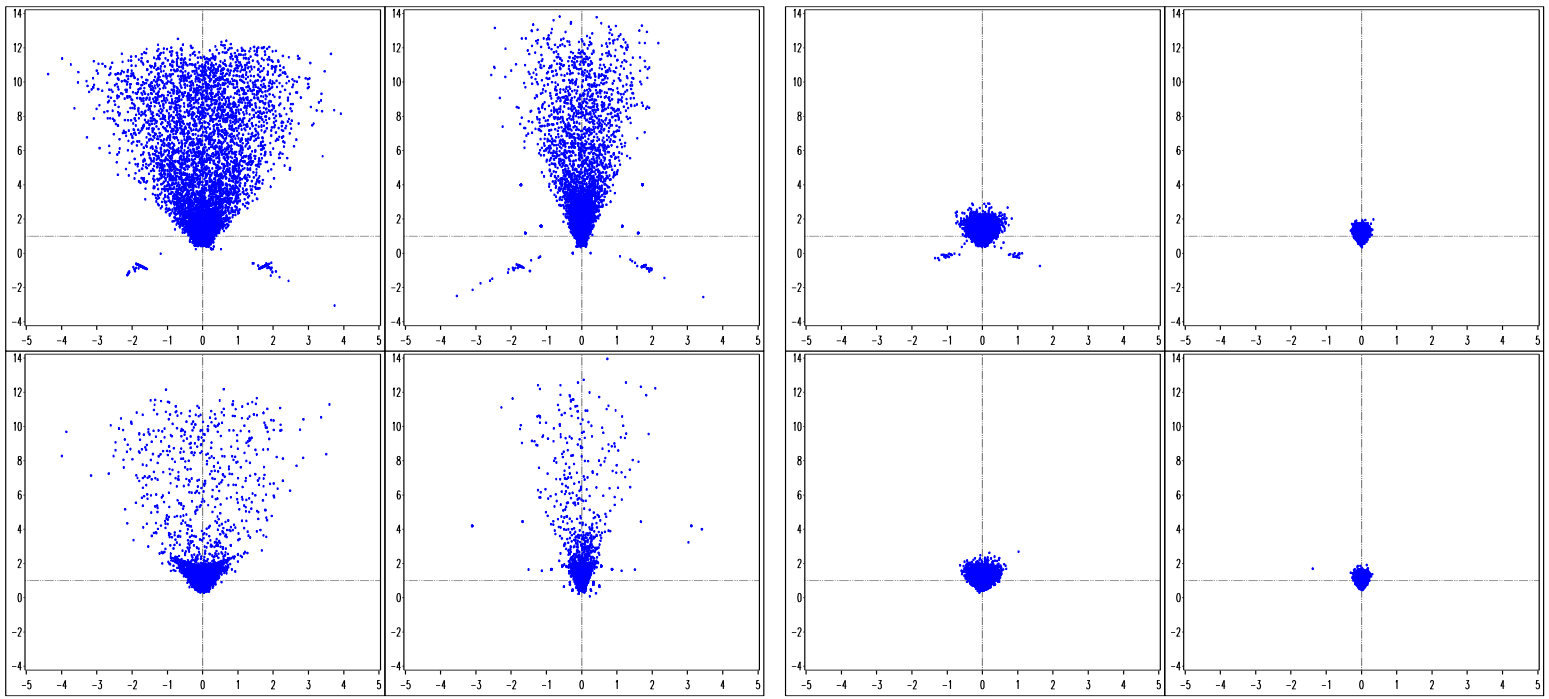
- Goals: good prediction for future subjects;
  limited harm to in-study patients;
  include cost constraint.
- Optimal dose allocation may be constrained by a penalty function.

Predicted Best Dose from naive BI & Adaptive D-Optimal design with cost $C = 1$ under the linear probit model.

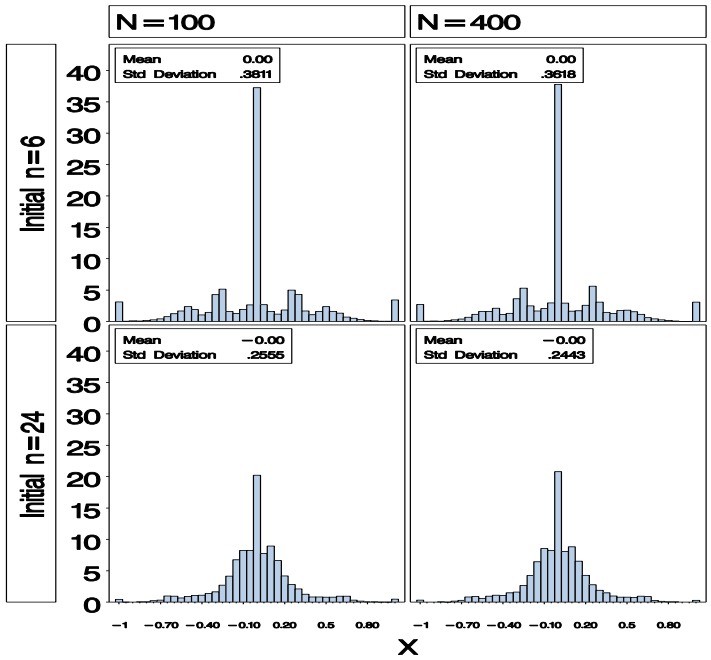


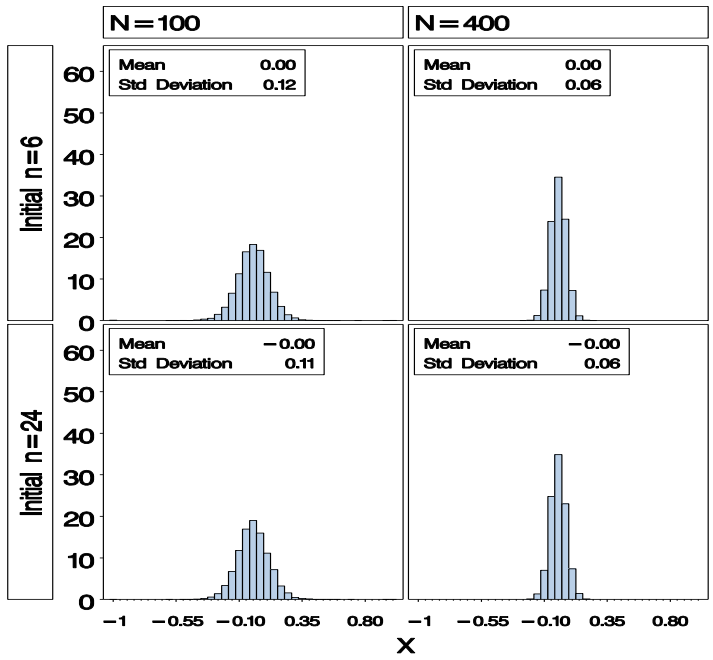


Parameter estimates ($n = 100$) from naive BI and Adaptive D-Optimal design with cost $C = 1$ under linear probit model.

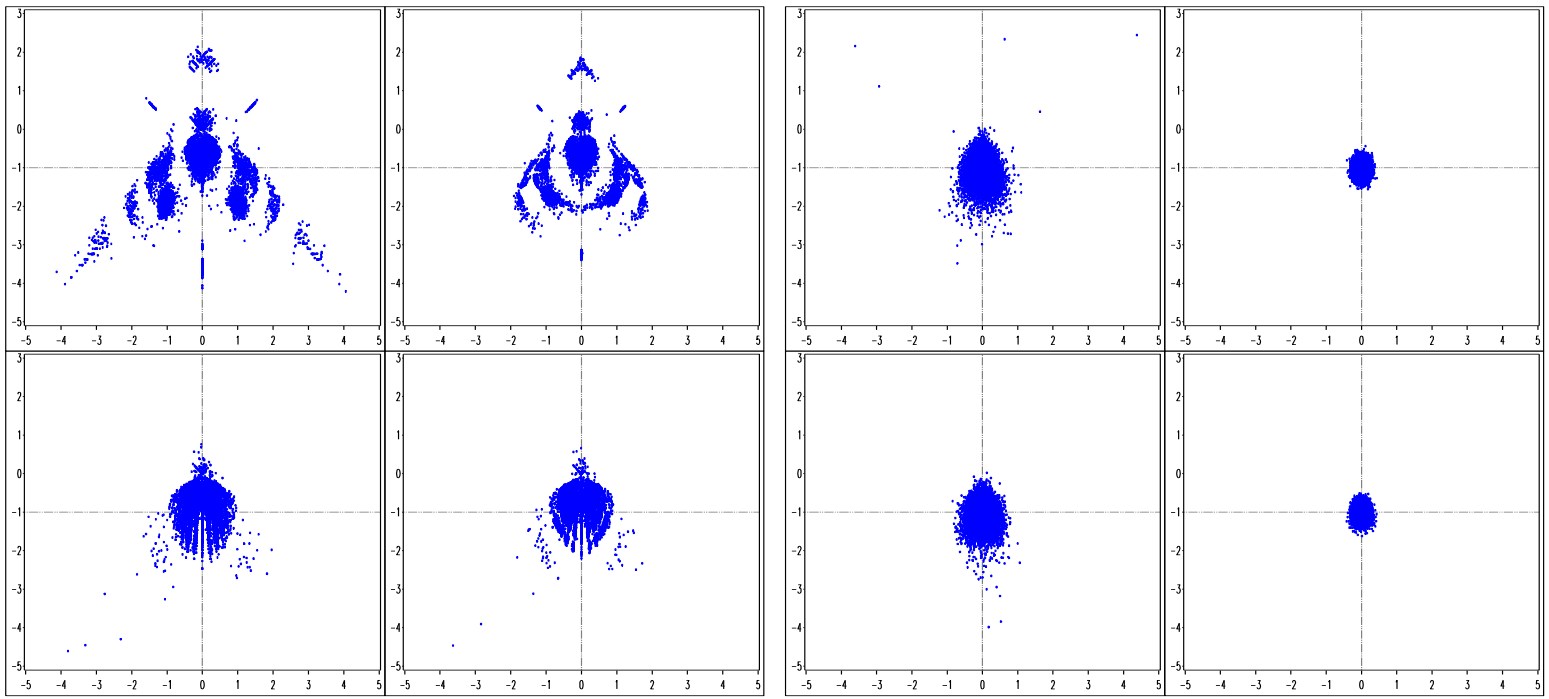


Predicted best dose ($n = 100$) from naive BI and Adaptive D-Optimal design with cost $C = 1$ under quadratic probit model.





Parameter estimates ($n = 100$) from naive BI and Adaptive D-Optimal design with cost $C = 1$ under probit quadratic model.



- planned observations are useful for hypothesis generation
- a good question is worth a thousand experiments
- designed experiments reduce costs, reduce time, increase precision and provide confirmation of hypotheses.
- two arm trials performed without good knowledge of dose response space underlying the definition of those arms is wasteful of time and resources and may lead to erroneous conclusions about the treatment".
- dose-finding, or more generally, exploration, is critical to scientific progress

Most importantly, make exploration and experimental design part of your life.

Thank you!