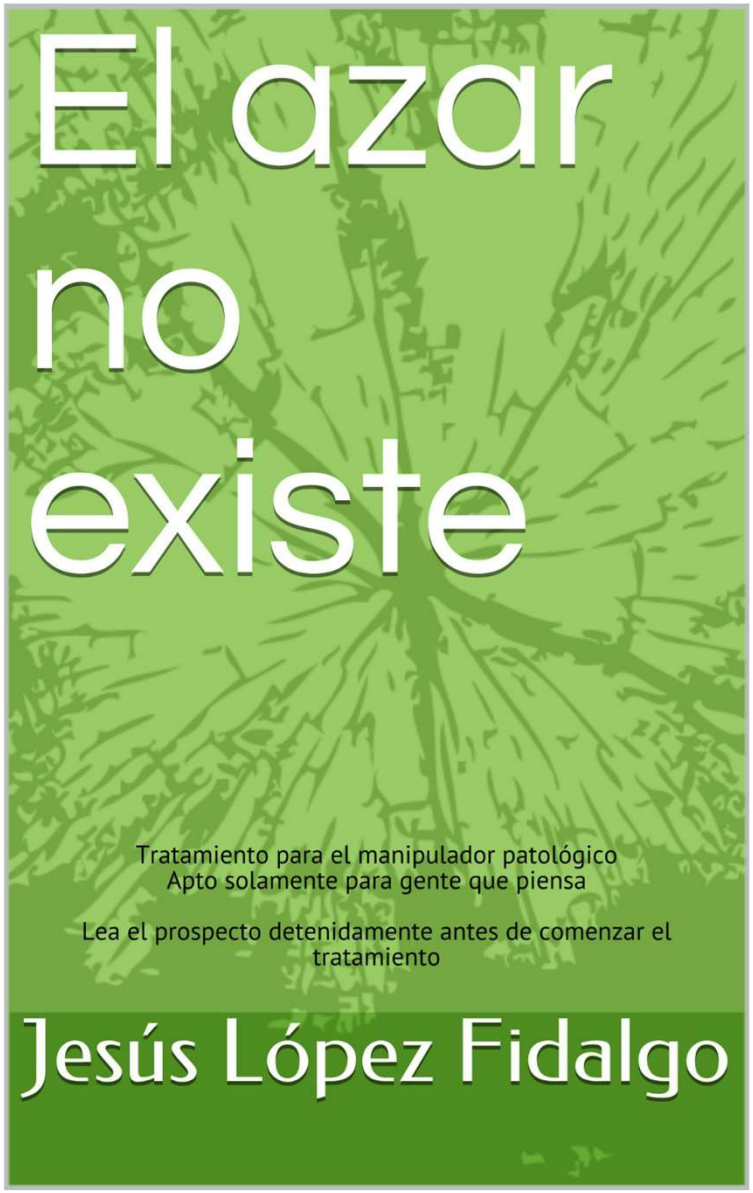
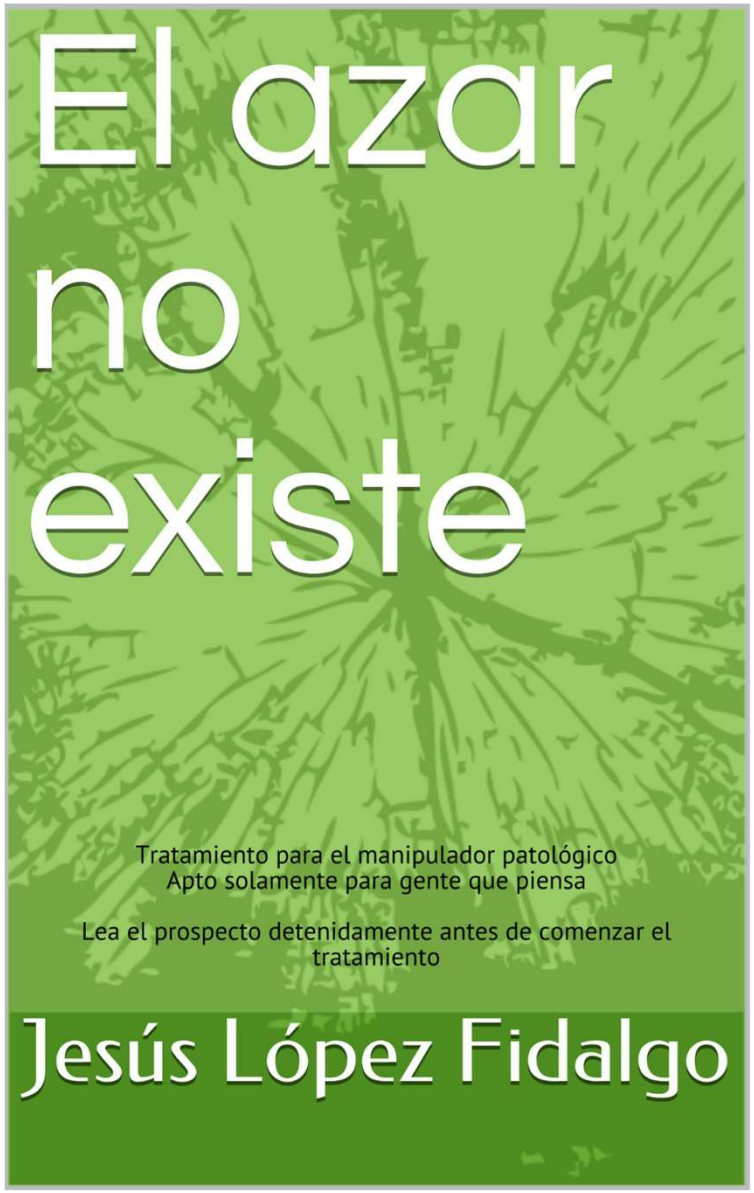
Best intention designs in dose-finding studies

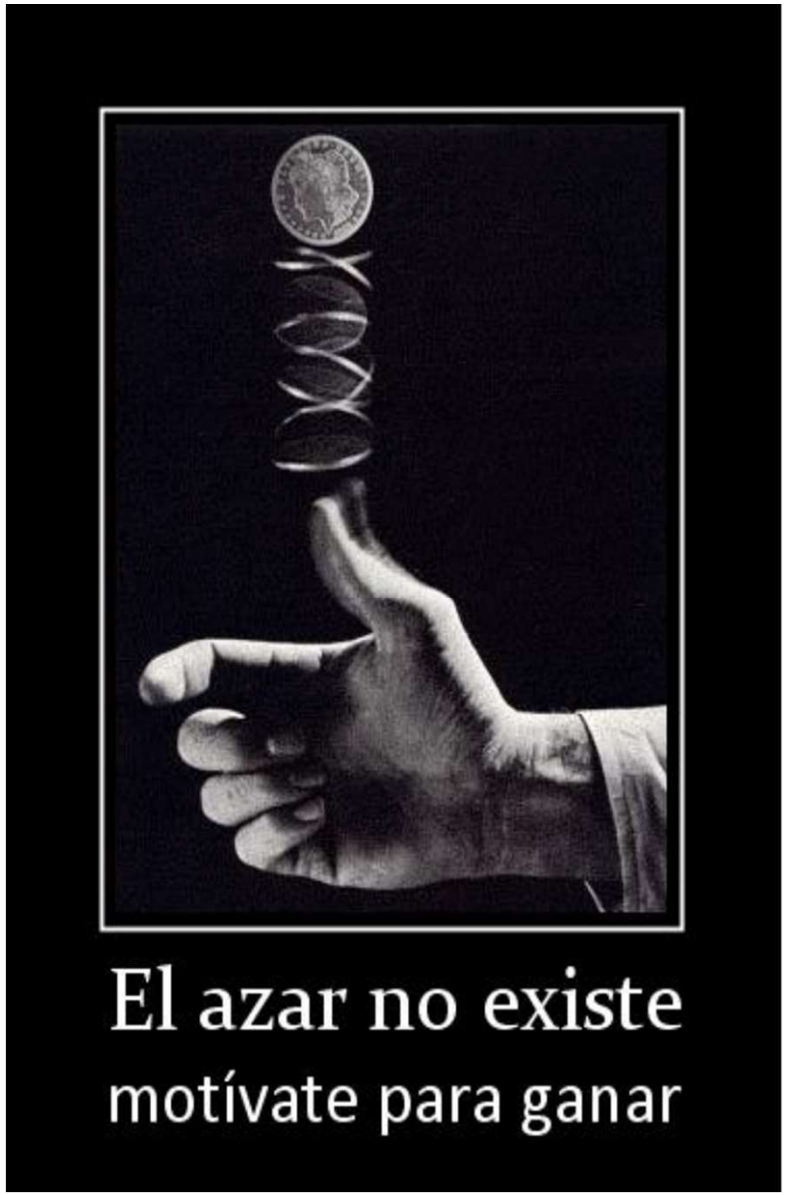
V. V. Fedorov

(JEDE 6, June 5, 2023)

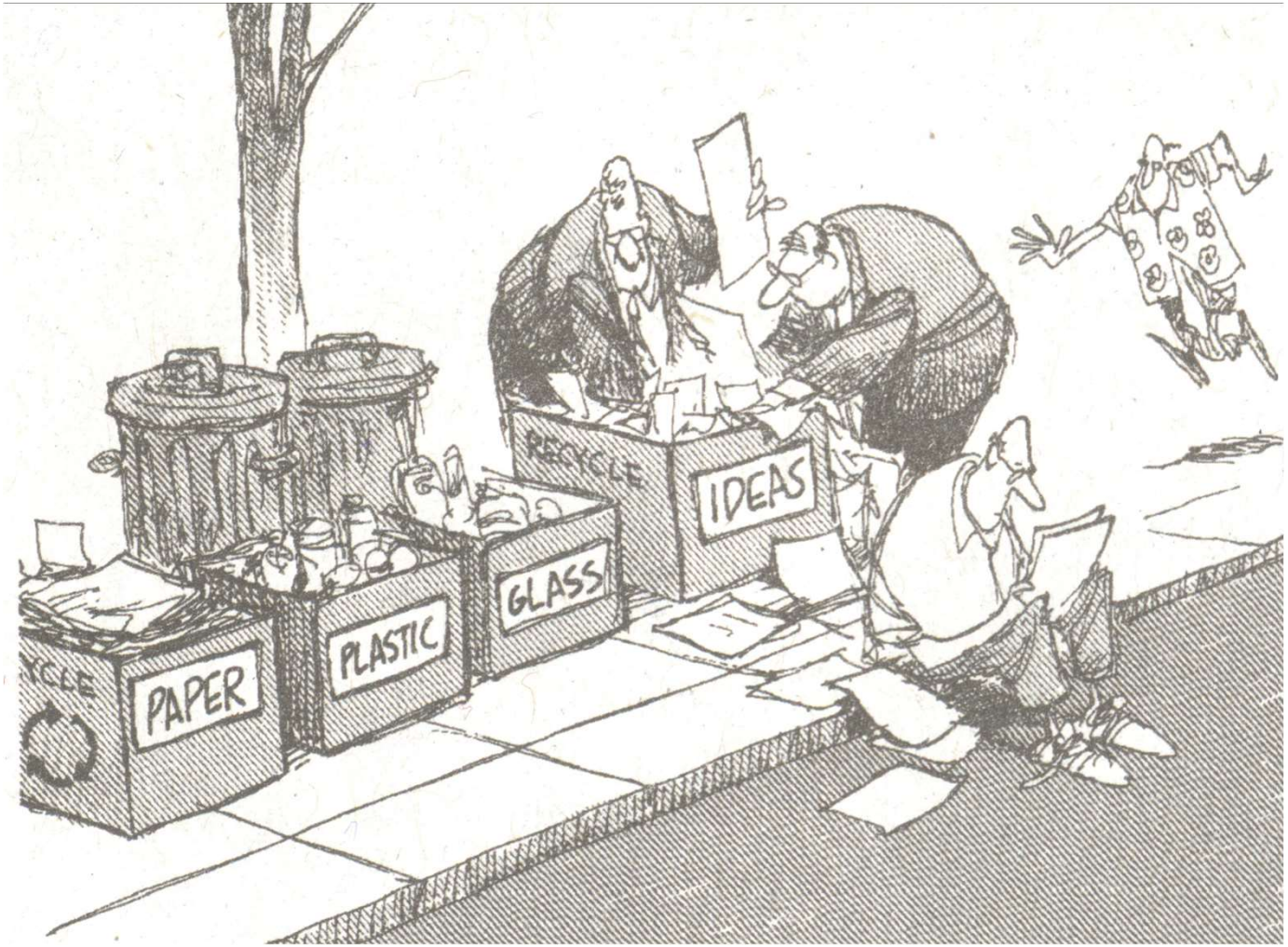

Jesús López Fidalgo
El azar no existe


El azar no existe
Tratamiento para el manipulador patológico
Apto solamente para gente que piensa
Lea el prospecto detenidamente antes de comenzar el tratamiento
Jesús López Fidalgo


El azar no existe
Tratamiento para el manipulador patológico
Apto solamente para gente que piensa
Lea el prospecto detenidamente antes de comenzar el tratamiento
Jesús López Fidalgo


El azar no existe
motívate para ganar

My approach today



Two major motivating sources for this presentation

Sequential Estimation of Quantal Response Curves

By G. B. WETHERILL

Birkbeck College, University of London

[Read at a RESEARCH METHODS MEETING of the SOCIETY, October 10th, 1962, Professor D. R. COX in the Chair]

Wetherill G. Sequential estimation of quantal response curves. Royal Statistical Society B 1963; 25:1–48

Adaptive design in regression and control

(iterated least squares/adaptive stochastic approximation/nonlinear regression/control theory/optimal dosage estimation)

T. L. LAI AND HERBERT ROBBINS

Department of Mathematical Statistics, Columbia University, New York, New York 10027

Contributed by Herbert Robbins, November 18, 1977

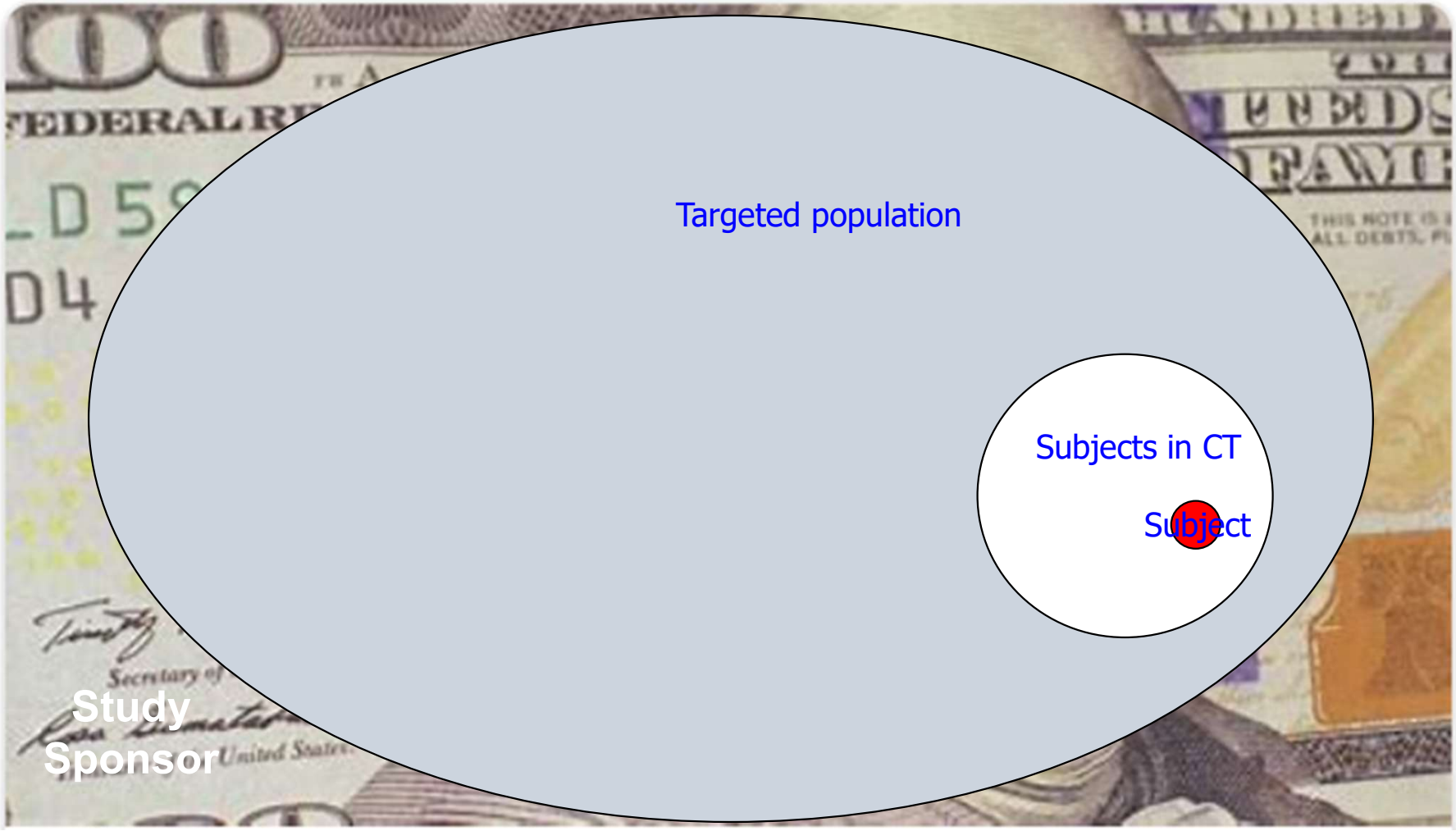
ABSTRACT When $y = M(x) + \epsilon$, where M may be nonlinear, adaptive regression designs of the levels $x_1, x_2, \ldots$ at which $y_1, y_2, \ldots$ are observed lead to asymptotically efficient estimates of the value θ of x for which $M(\theta)$ is equal to any desired value y^*. More importantly, these designs also make the "cost" of the observations, defined at the nth stage to be $\Sigma_1^n (x_i - \theta)^2$, to be of the order of log n instead of n, an obvious advantage in medical and other applications.

Lai TL, Robbins H. Adaptive design in regression and control. Proceedings of the National Academy of Sciences of the U.S.A. 1978; 75:586–587

Outline

- Dose-finding clinical studies: major players and basic models
- Setting two types of statistical problems in dose-finding
- BIDs in Type I problems and ARM
- BIDs in Type II problems, self-tuning optimizers, and model based designs
- Conclusion

Major players in the design of clinical studies



Probit dose-response model

	Toxicity Z		
Efficacy Y	1	0	
1	p_{11}	p_{10}	$p_{1\bullet}$
0	p_{01}	p_{00}	$p_{0\bullet}$
	$p_{\bullet 1}$	$p_{\bullet 0}$	

$$p_{00}(x,\boldsymbol{\theta}) = 1 - \Phi[\eta_1(x,\boldsymbol{\theta}), \eta_2(x,\boldsymbol{\theta}); \rho] \ ,$$

$$p_{0\bullet} = \Phi[\eta_1(x,\boldsymbol{\theta})] \ ,$$

$$p_{\bullet 0} = \Phi[\eta_2(x,\boldsymbol{\theta})] \ ,$$

$$p_{10} = 1 - p_{\bullet 1} - p_{00}, \quad p_{01} = 1 - p_{1\bullet} - p_{00}$$

$$\Phi[\eta_1, \eta_2; \rho] = \int_{-\infty}^{\eta_1}\int_{-\infty}^{\eta_2} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1-\rho^2)}} dz_1 dz_2 \ ,$$

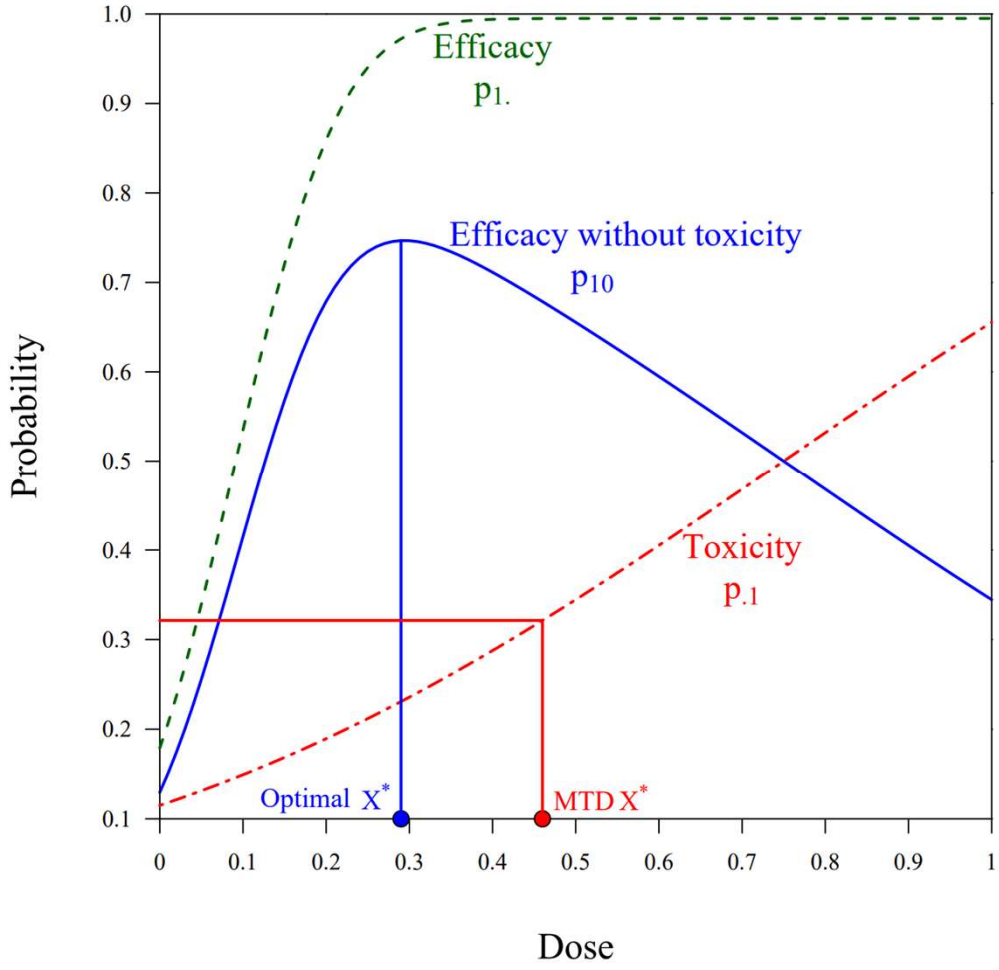
$$\Phi[\eta] = \int_{-\infty}^{\eta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \ .$$

Dragalin V, Fedorov V. Adaptive designs for dose-finding based on efficacy–toxicity response. Journal of Statistical Planning and Inference 2006; 136:1800–1823.
Dragalin V, Fedorov V, Wu Y. Two–stage design for dose–finding that accounts for both efficacy and safety. Statistics in Medicine 2008; 27:5156–5176

An interesting remark was made by Dr. C.C. Spicer during the vote of thanks at the JRSS meeting after Wetherill's presentation

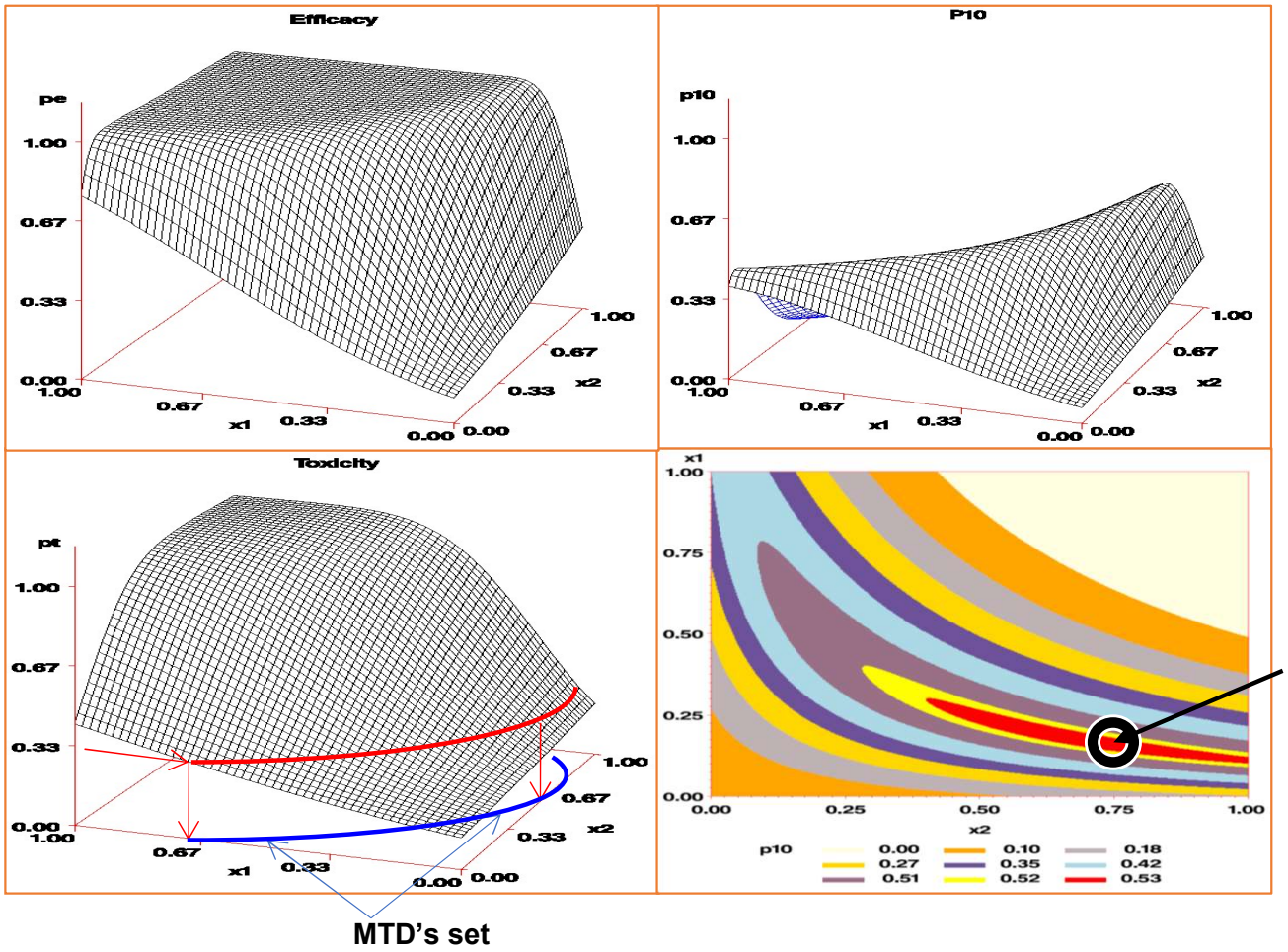
A point that seems to me a sign of the times is that Dr Wetherill goes fairly strongly for the logit curve rather than the probit curve. The widespread introduction of computers has put another nail in the coffin of the probit curve, as the logit is very much quicker and simpler to calculate than the existing approximations for the cumulative normal curve.

Example of a dose-response curves - I



Example of a dose-response model – II

Combination of two drugs



Two problems in dose finding

Let:

$$y \sim F(y|\eta) \quad \text{and} \quad E[y|x] = \eta(x, \theta)$$

Type I:

$$x^*(\theta) = \text{Arg}[\eta(x, \theta) = C]$$

Type II:

$$x^*(\theta) = \text{Arg}\max_x \eta(x, \theta)$$

Best intention vs most informative design

- Best intention designs:

Allocate every new patient to the BEST (accordingly to the current knowledge) dose

$$x_{n+1} = x^*(\hat{\theta}_{n+1})$$

- Most informative designs:

Allocate every new patient to the dose that is most informative (accordingly to the current knowledge)

$$x_{n+1} = \operatorname{Arg}\max_{x \in \mathbf{X}} \{\text{Increment of Information } (x, \hat{\theta}_n)\}$$

For D-criterion: $$x_{n+1} = \operatorname{Arg}\max_{x \in \mathbf{X}} \{\operatorname{Var}[\eta(x, \hat{\theta}_n)]\}$$

Adaptive Robbins-Monro (ARM) procedure

$$y = \eta(x, \theta) + \varepsilon, \quad \eta(x, \theta) = \theta_1 + \theta_2 x, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$x_{n+1} = \frac{C - \hat{\theta}_{1,n}}{\hat{\theta}_{2,n}}$$

The (n+1)-th design point = the MLE/LSE of the "best/target" dose after n observation

One can verify that if the slope is known then:

$$x_{n+1} = \bar{x}_n - \frac{1}{\theta_2}(\bar{y}_n - C)$$

and

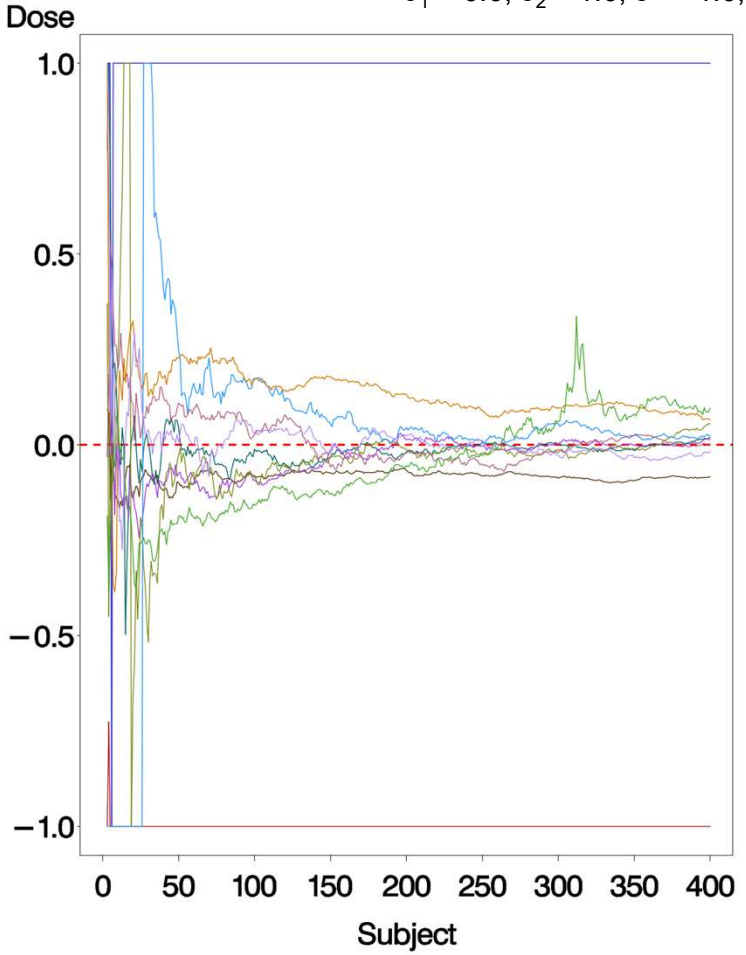
$$x_{n+1} = x_n - \frac{1}{n\theta_2}(y_n - C)$$

If the slope is unknown replace it with estimator

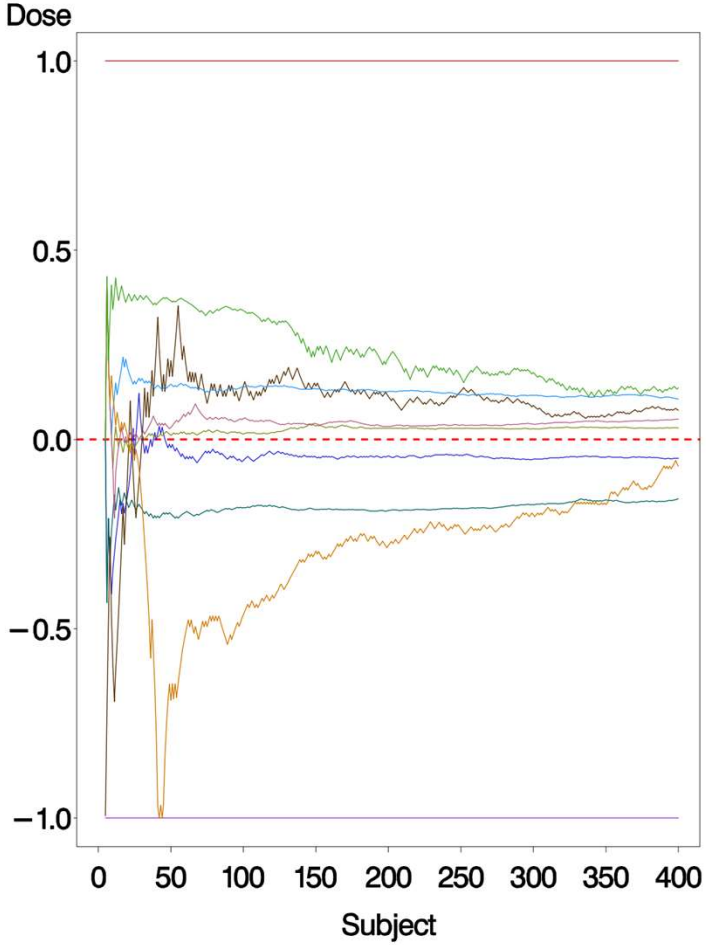
Robbins H, Monro S. A stochastic approximation method. Ann. Math. Statistics 1951; 22:400–407
T. Lai and H. Robbins (1978) Adaptive design in regression and control. Proc. Natl. Acad. Sci. 75 (2), pp.586-587
H. Kushner and G. Yin (1997) Stochastic approximation algorithms and applications. Springer, pp. i-xii, 1-417, Ch.1

ARM procedure

$\theta_1 = 0.0$, $\theta_2 = 1.0$, $\sigma^2 = 1.0$, C= 0.0, i.e., $x^* = 0.0$; doses were restricted to lie in [−1, 1].



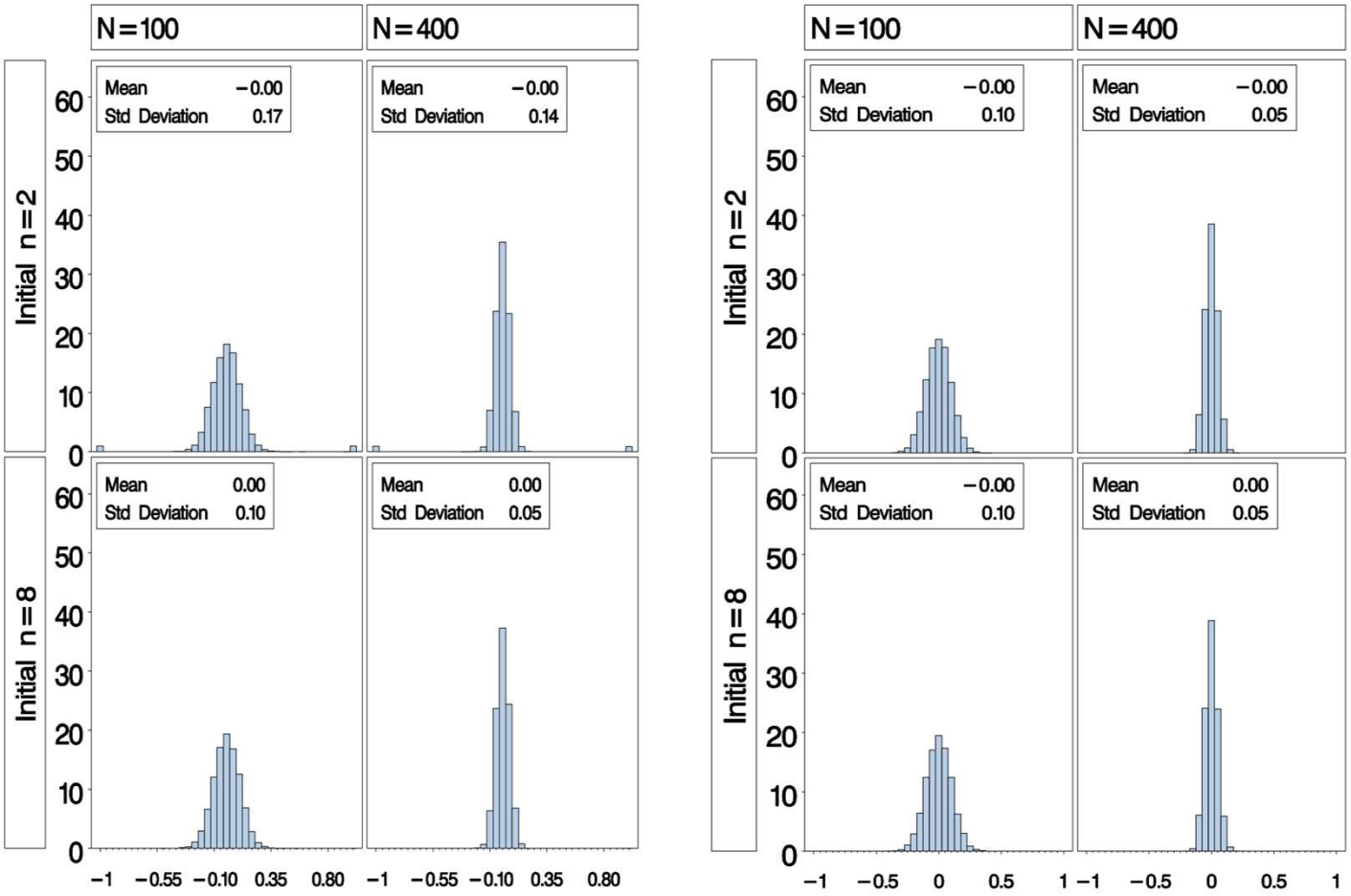
Type I ARM predicted the best dose sequences under the continuous linear model $\theta_1 + \theta_2 x$.



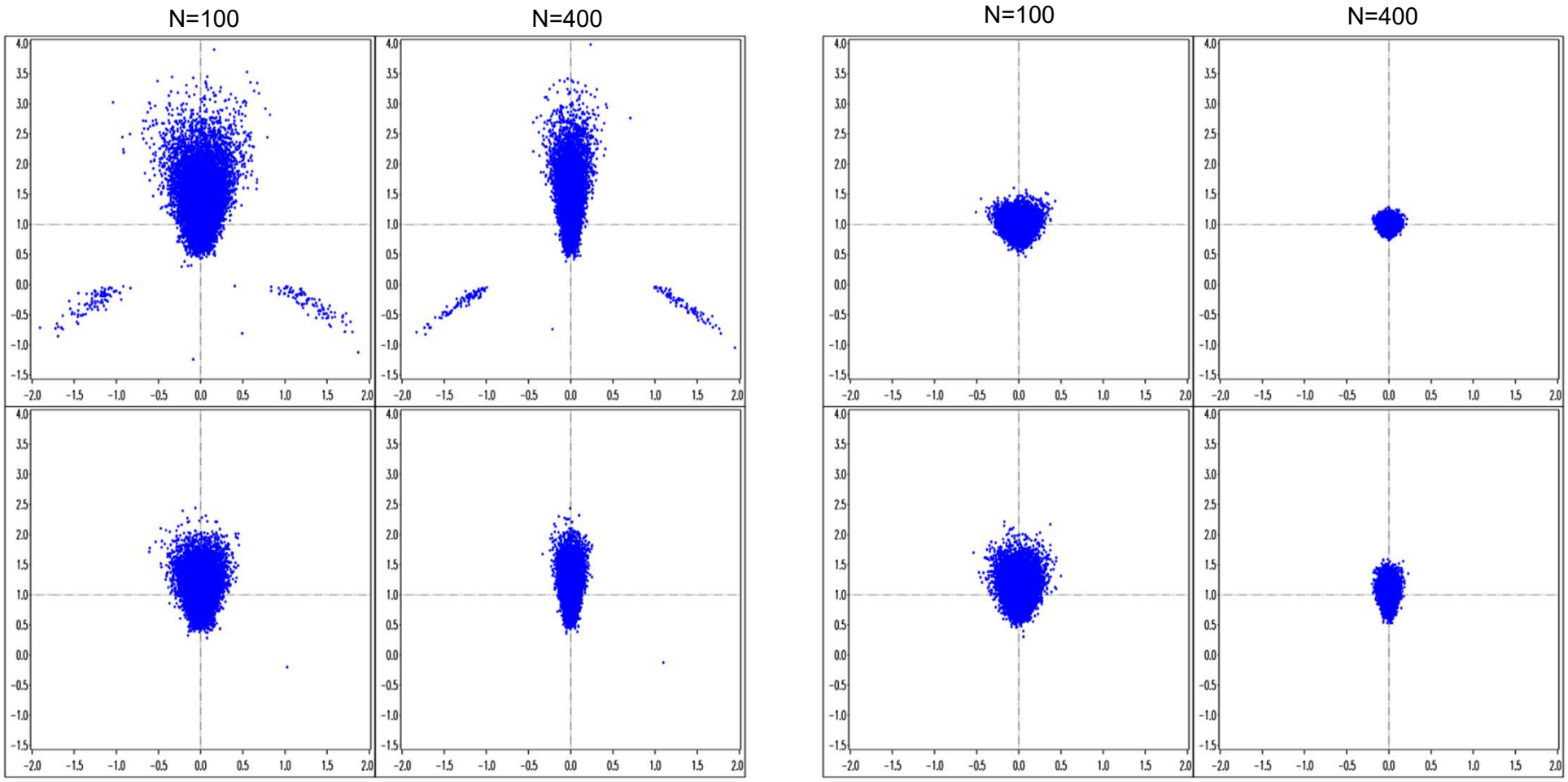
Type I ARM predicted the best dose sequences under the probit model with F ($\theta_1 + \theta_2 x$).

ARM and penalized D-adaptive; M-C simulations for Type I

$$\theta_1 = 0, \quad \theta_2 = 1, \quad \sigma^2 = 1$$



Predicted best doses from the naive ARM and PAD (with cost(x) =0.1) designs. Columns show frequencies of x_{100} and x_{400}, respectively.



Plots of θˆ2n by θˆ1n under the continuous linear model with θ1 = 0.0 and θ2 = 1.0; x ∗ = 0.0.
The top and bottom rows had initial cohorts of size 2 and 8, equally allocated to ±1

Risks for different customers

Customers	Risk	Tuned ARM	Locally $\mathscr{D}-$optimal
Targeted Population	$\mathrm{E}\left(x_N^* - x^*\right)^2$	$\sim (\sigma/\theta_2)^2 N^{-1}$	$(\sigma/\theta_2)^2 N^{-1}$
Patient Sample	$\mathrm{E}\left[\sum_{i=1}^N \left(x_i^* - x^*\right)^2\right]$	$\sim (\sigma/\theta_2)^2 \ ln\ N$	N
nth Patient	$\mathrm{E}\left(x_n^* - x^*\right)^2$	$\sim (\sigma/\theta_2)^2 n^{-1}$	1
Sponsor	N	$Q + qN$	$Q + qN$

- Character "$\sim$" denotes "asymptotically", or more loosely, for large N.
- Q is the cost of a trial initiation; q is the cost of patient enrollment, patient treatment, administrative expenses, etc.
- Recall that $\sum_{i=1}^N n^{-1}\ 1/i \sim 0.577 + \ln N$.

Summary I

- ARM works for type I problems (almost always – some tuning is needed).
- Start with a reasonable pilot sample.
- ARM is good for estimating x* but the slope cannot be separately estimated.
- ARM inspired the development of a few methods like CRM, 3+3 (more generally A+B), and the Bayesian dose escalation are rather popular in clinical studies.
- ARM is a logistic challenge.
- Extension to multivariate x is not obvious.
- Penalized adaptive D-optimal is a competitive alternative to ARM and allows estimating x*, and the slope.
- Two-three stage PD-optimal designs often outperform the fully adaptive and logistically are more attractive.

- *V. Fedorov, N. Flournoy, Y. Wu, R. Zhang. Best Intention Designs in Dose-Finding Studies, TR of Isaac Newton Institute for Mathematical Sciences, 2011 pp. 1-37, http://www.newton.ac.uk/preprints/NI11065.pdf*
- *P. Thall, J. Cook (2004), Dose-Finding Based on Efficacy-Toxicity Trade-offs, Biometrics, 60,684-693*
- *Berry, D. A., Mueller, P., Grieve, A. P., Smith, M. K., Parke, T., Krams, M. (2002). Bayesian designs for dose-ranging drug trials. In: Gatsonis, C., Kass, R. E., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., West, M., eds. Case Studies in Bayesian Statistics. Vol. 5. New York: Springer-Verlag, pp. 99–181.*

Response optimization: self-tuning optimizer

$$y = \eta(x, \theta) + \varepsilon$$

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2$$

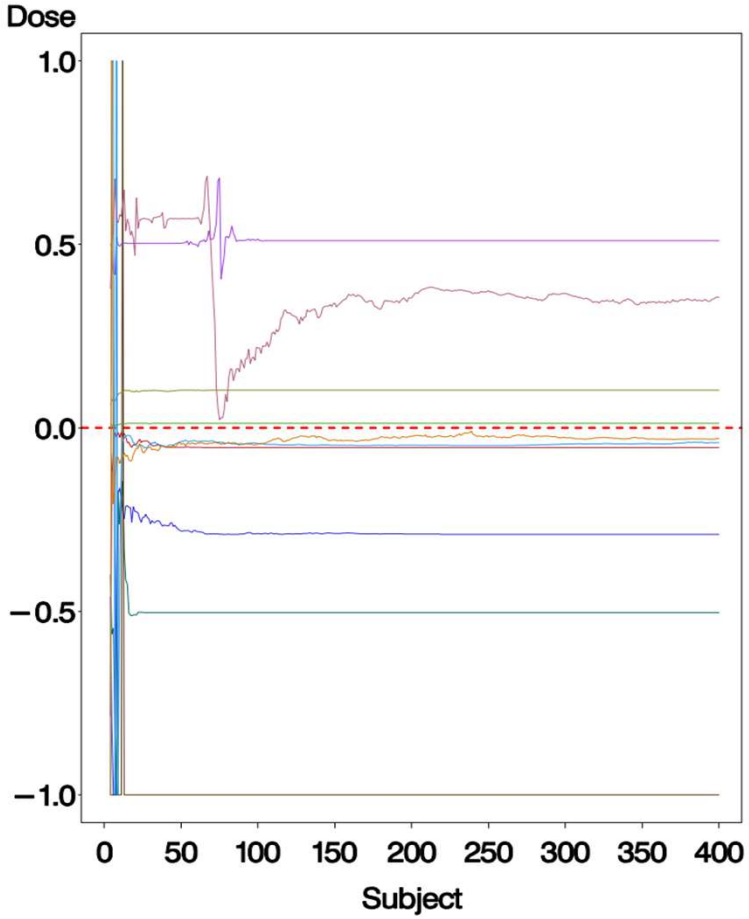
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Best intention design: allocate the next patient to the “best” dose

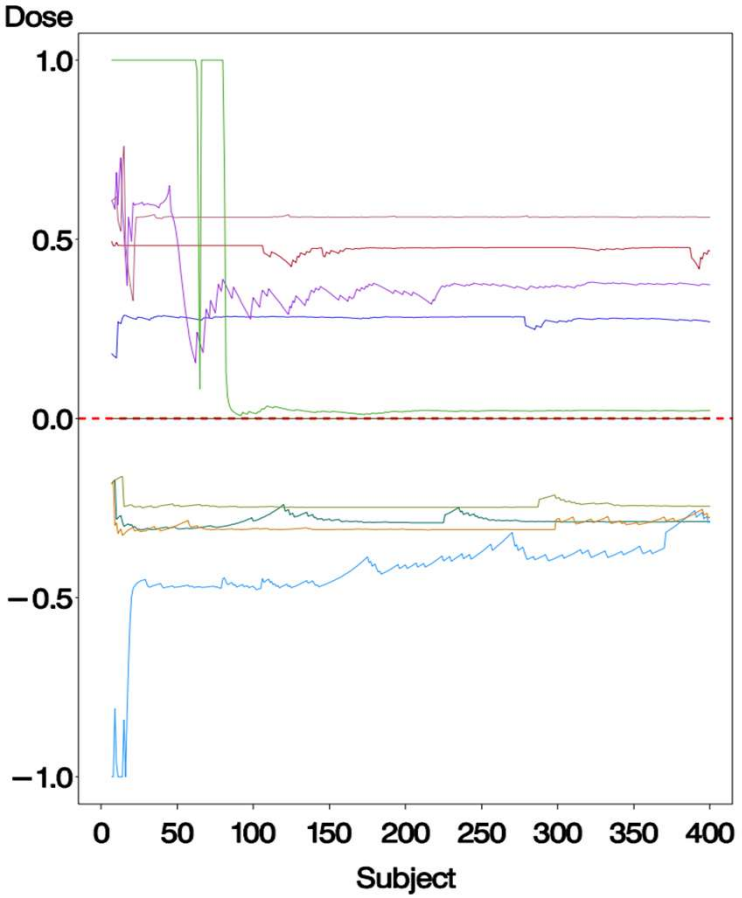
$$x_{n+1} = -\frac{\hat{\theta}_{2,n}}{2\hat{\theta}_{3,n}}$$

Pronzato L. Optimal experimental design and some related control problems. Automatica 2008; 44:303–325

ARM procedure



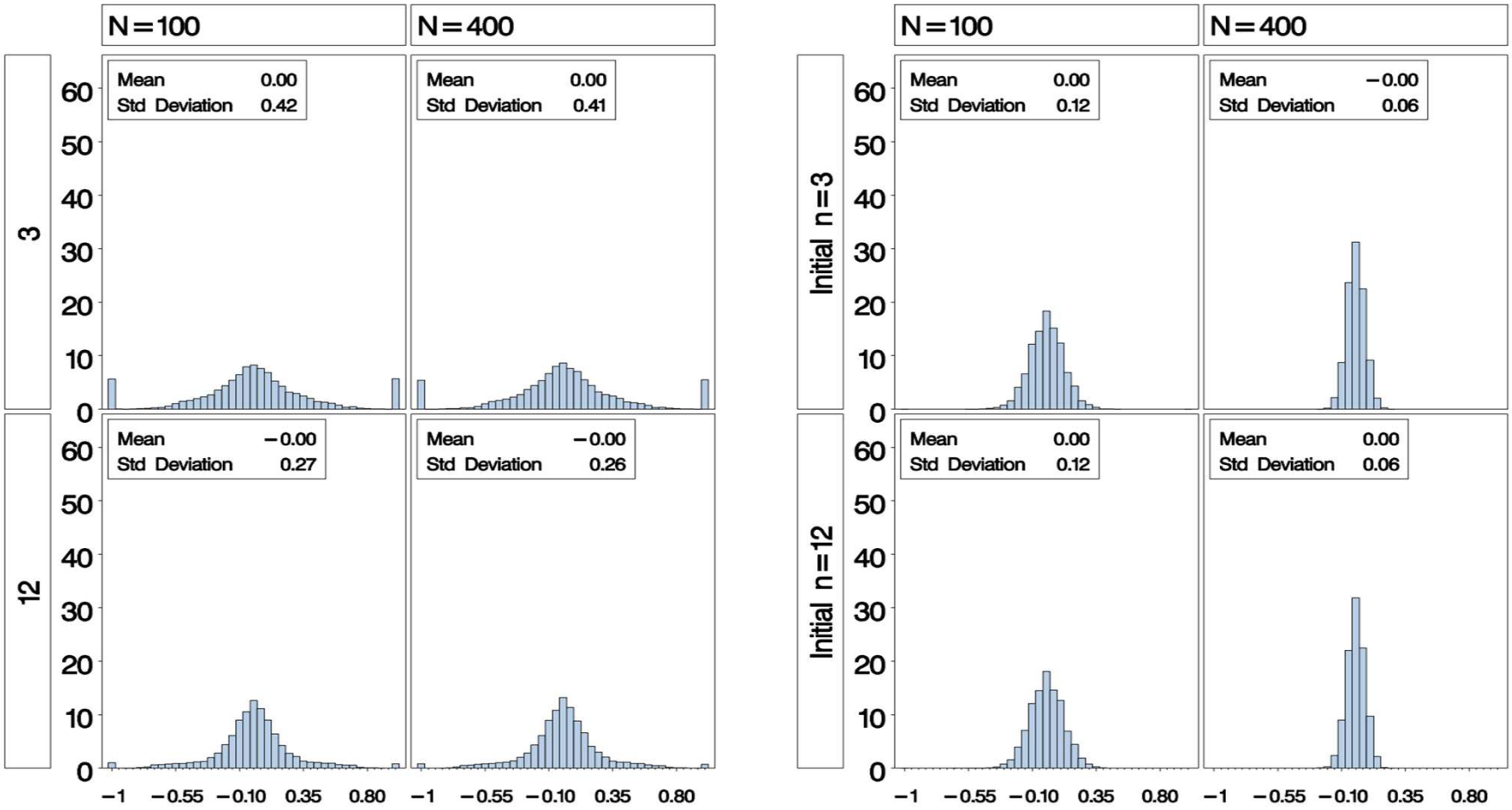
BI predicted the “best” dose sequences under the continuous quadratic model



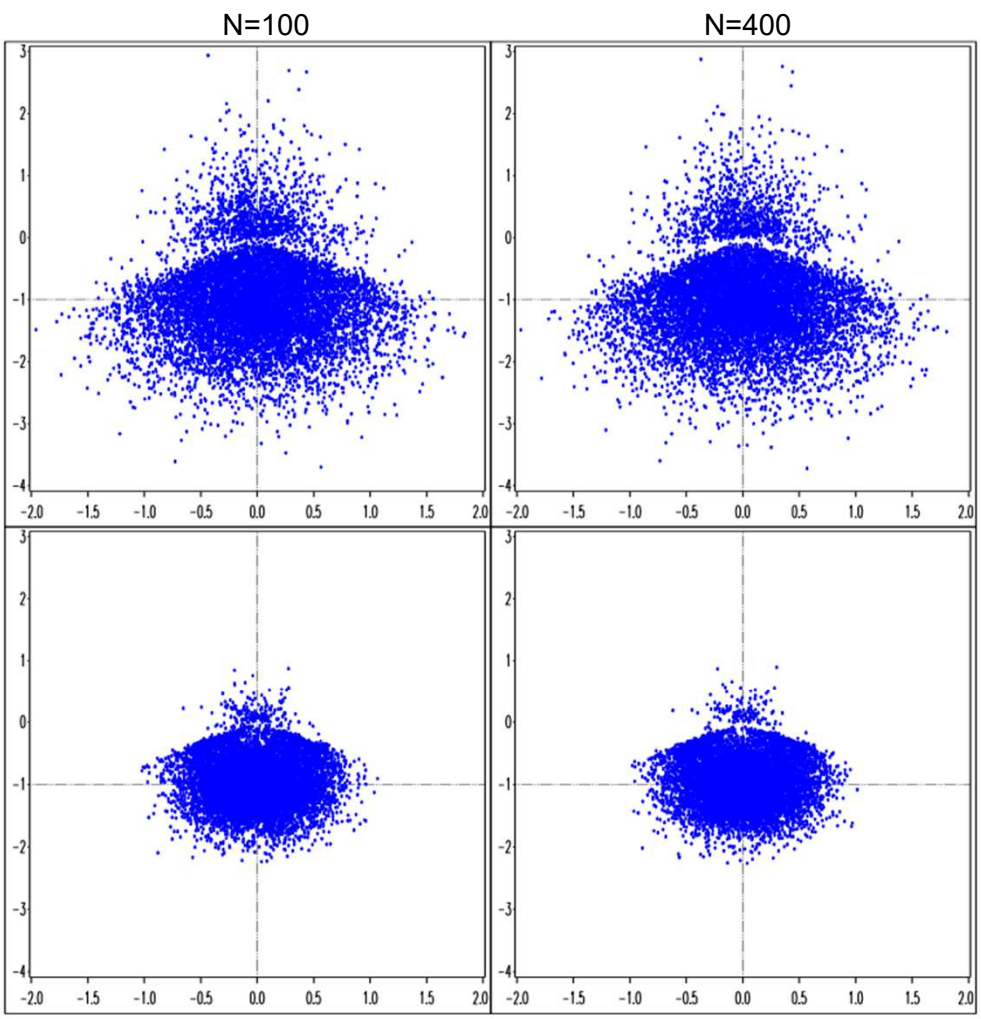
BI predicted the “best” dose sequences under the probit model with F ($\theta_1 + \theta_2 x + \theta_3 x^2$)

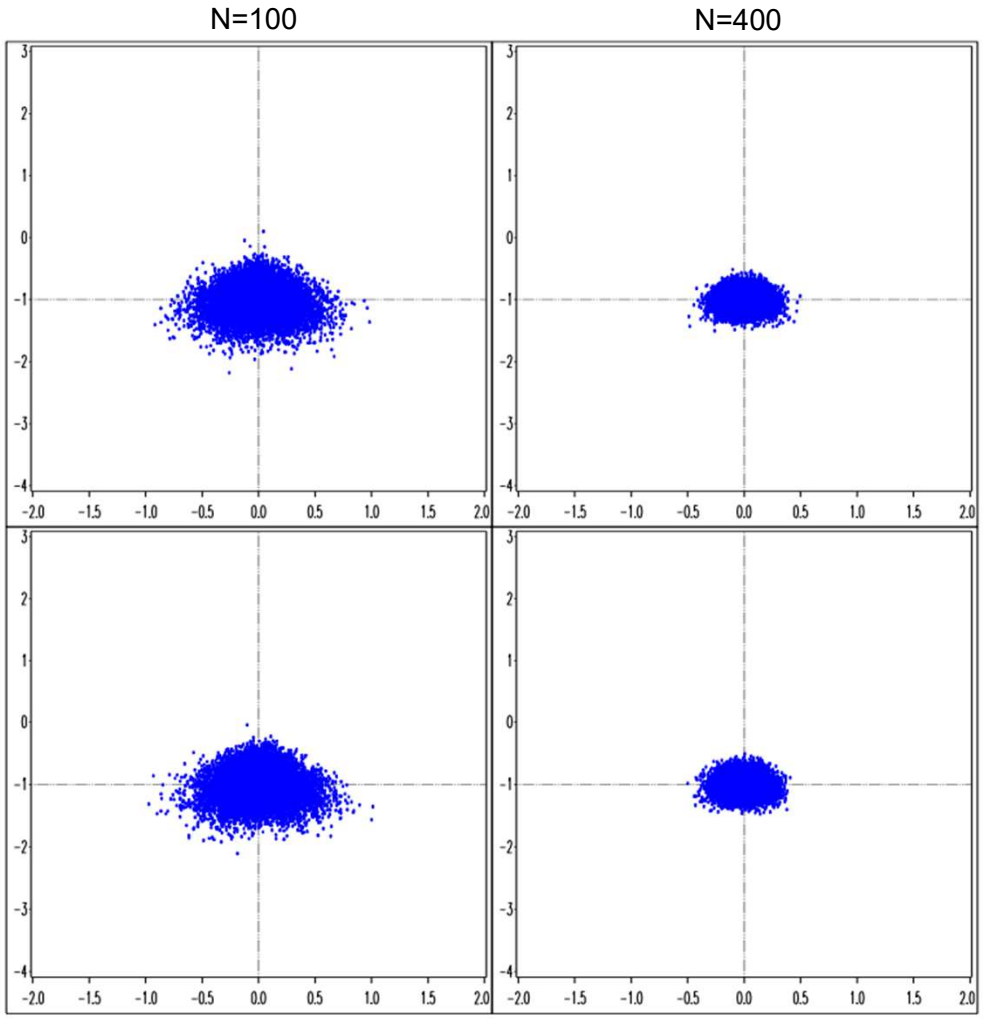
ARM and penalized D-adaptive; M-C simulations for Type II

$\theta_1 = 1.0$, $\theta_2 = 0.0$, $\theta_3 = -1.0$, and $x^* = 0.0$



Predicted best doses from the naive ARM and PAD (with cost(x) =0.1) designs. Columns show frequencies of x_{100} and x_{400}, respectively.





Plots of θˆ2n by θˆ1n under the continuous quadratic model with θ1 = 1.0, θ2 = 0.0 and θ3 = −1.0; x ∗ = 0.0.
Plots in the top and bottom rows had initial cohort sizes of 3 and 12, equally allocated at −1, 0, 1.

Why do the best intention designs perform poorly for Type II problems?

- Let BID perform at least as well as any regular design, i.e. an estimator x_N for x*=0 is at least consistent and

$$x_N \sim 1/\sqrt{N}$$

- Foor the MLE or LSE consistency can be achieved when the moment matrix

$$M_{\alpha\beta}(N) = \sum_{n=1}^{N} x_n^{\alpha+\beta-2}$$

is regular and its diagonal elements tend to infinity.

- Observing that

$$M_{11}(N) = N \Rightarrow \infty$$

$$M_{22}(N) = \sum_{n=1}^{N} x_n^2 \sim \sum_{n=1}^{N} 1/n \sim 0.577 + \ln N \Rightarrow \infty$$

$$M_{33}(N) = \sum_{n=1}^{N} x_n^4 \sim \sum_{n=1}^{N} 1/n^2 \Rightarrow \pi^2/6 < \infty$$

We may conclude that BID may lead to reasonable estimators of x* only for Type I problems and for Type II problems more sophisticated designs are needed, for instance, K-W adaptive procedure* and its numerous modifications.

*Kiefer, J.; Wolfowitz, J. (1952). "Stochastic Estimation of the Maximum of a Regression Function". *The Annals of Mathematical Statistics*. **23** (3): 462-466

Adaptive D-optimal designs that maximize information gained per unit of penalty/cost

Model: $y(x) = \eta(x,\theta) + \varepsilon = \theta^T f(x) + \varepsilon, \quad \mathrm{E}[\varepsilon] = 0, \mathrm{Var}[\varepsilon = \sigma^2]$

Popular penalty/loss functions

$\phi(x,\theta) = A[\eta(x^*,\theta) - \eta(x,\theta)] + C$ and $\phi(x,\theta) = A(x^* - x)^2 + C$

Adaptive procedure

$$x_{n+1} = \arg\max_{x \in X} \left[\frac{\mathrm{Var}[\eta(x,\hat{\theta}_n)]}{\sigma^2} - m\frac{\phi(x,\hat{\theta}_n)}{\Phi_n} \right]$$

$\Phi_n = \sum_{i=1}^{n} \phi(x_i,\hat{\theta}_n)$ - cumulative penalty

$$\hat{x}_n^* = \arg\max_{x \in X} \eta(x,\hat{\theta}_n)$$

•*Pronzato L. (2000) Adaptive optimization and D-optimum experimental design, The Annals of Statistics, 28, pp. 1743-1761*

•*V. Fedorov, N. Flournoy, Y. Wu, R. Zhang. Best Intention Designs in Dose-Finding Studies, TR of Isaac Newton Institute for Mathematical Sciences, 2011 pp. 1-37, http://www.newton.ac.uk/preprints/NI11065.pdf*

Summary II

- There is a serious problem with the convergence BID.
- BIDs converge but to the wrong doses.
- BIDs fail for simple models. Will they work for more complicated ones?
- Apply penalized adaptive designs.
- Use penalized locally optimal designs as benchmarks for more “practical” designs and the MC simulations to explore their properties
- Use the Kiefer-Wolfowitz procedure and its numerous modifications, including cases when the response functions are unknown.

Conclusion

- The road to bad doses is paved with good intention designs
- The road to good doses is paved with good adaptive designs

JEDE ⇨ JEDI

AI

Alien Intelligence

Artificial Intelligence