Design of clinical trials with multiple end points

Valerii Fedorov

August, 2011

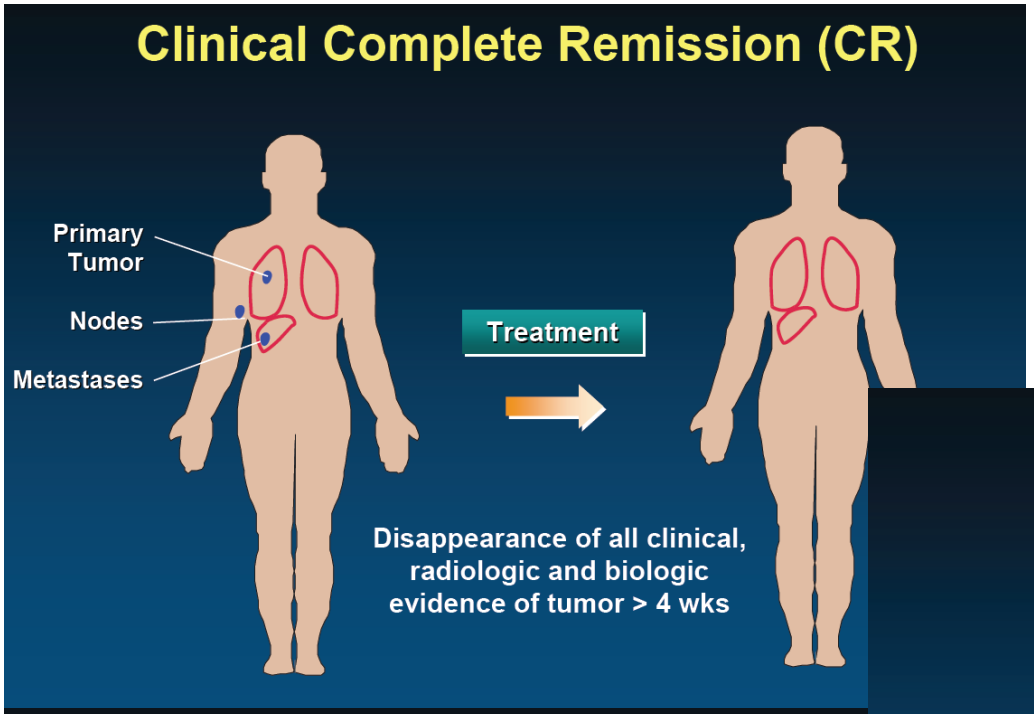
Workshop at the Isaac Newton Institute of Mathematics

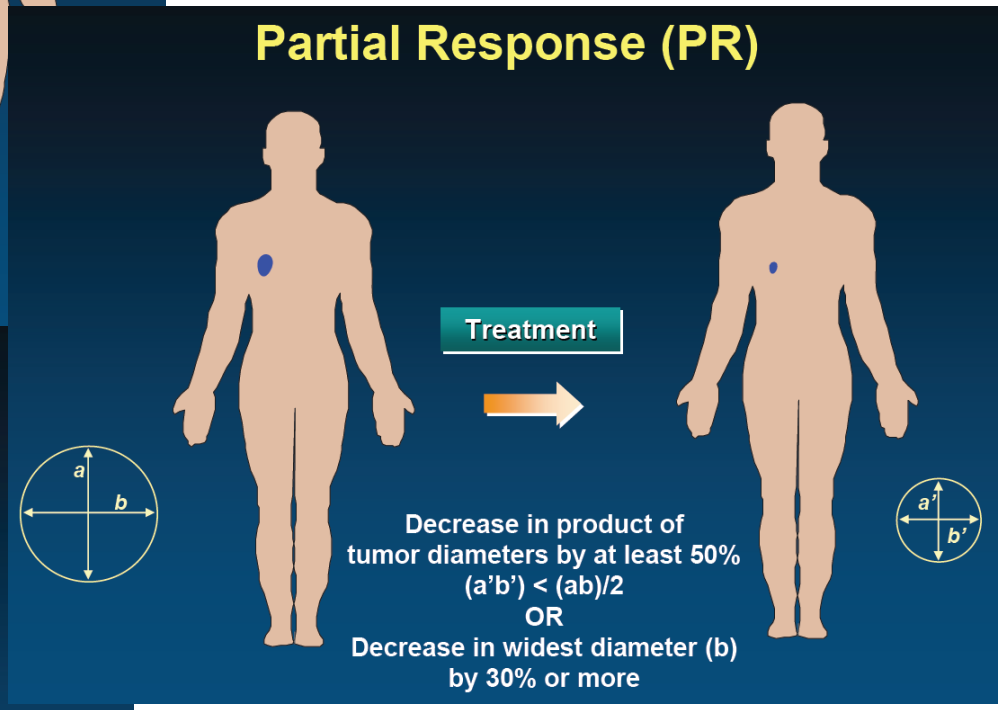
Major steps in design

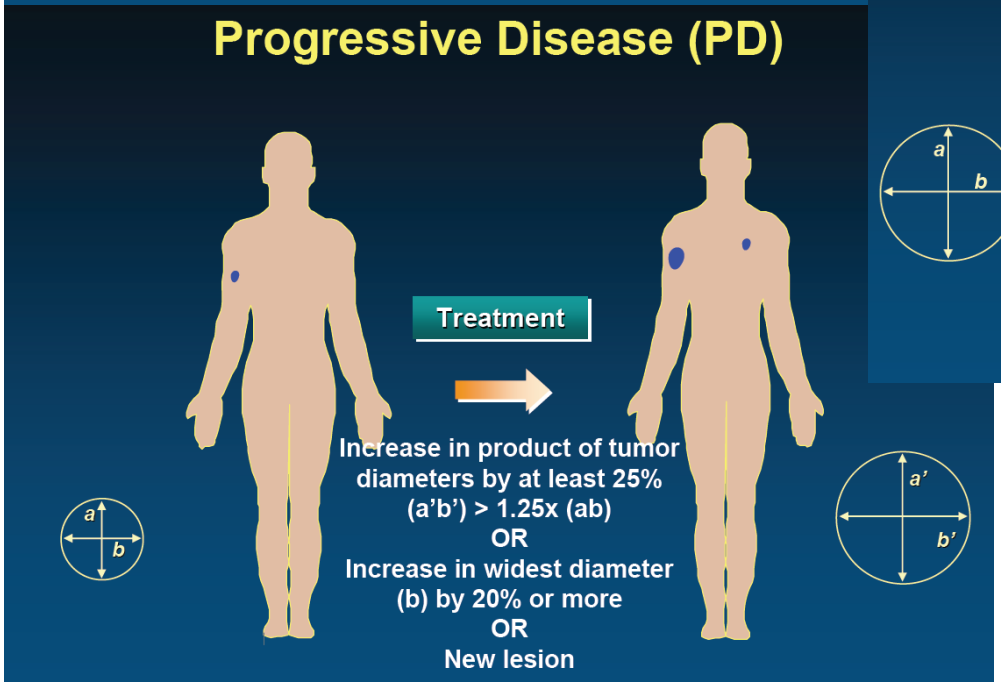
- Select a model
- Derive the information matrix of a single observation
- Define utility function
- Select a penalty (cost) function
- Select a criterion of optimality
- Analyze prior information, get guesstimates
- Compute locally optimal designs – they are our benchmarks
- Build more “practical” designs and compare them with the benchmarks

Thanks to Peter Ho and Alka Preston

Efficacy







Often (both for reporting and analysis):

Y = 1 if CR or PR, otherwise Y = 0.

Toxicity

Measurements and Reported End-Points

	Typical study	
	Placebo	**Treatment**
Any SAE	30%	70%
Diarrhea	0	8%
Cardiac disorders	2%	2%
Nausea/Vomiting	5% / 3%	10% / 10%
Anemia	3%	5%
Hepatotoxicity	3%	10%
Blood and lymphatic system	4%	6%
Gastrointestinal	2%	4%
Hypertension	3%	12%
Metabolism and nutrition disorders	5%	8%

Y = 1, if there is any SAE, Y = 0 otherwise

Dichotomized analysis and reporting

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

Probit model

$$\mathbf{Z} \sim N(\boldsymbol{\eta}, \boldsymbol{\Sigma})$$

$$p_{00} = P(Y_1 = 0, Y_2 = 0) = F(V; \boldsymbol{\Sigma}^*) = \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{1}{2\pi |\boldsymbol{\Sigma}^*|^{1/2}} \exp\left\{ -\frac{1}{2} \mathbf{v}^T \boldsymbol{\Sigma}^{*-1} \mathbf{v} \right\} d\mathbf{v}$$

$$\mathbf{v} = (v_1, v_2)^T,\ v_k = (c_k - \eta_k)/\sigma_k,\ Y_k = I(Z_k > c_k)$$

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

More about probit model can be found in:

III. *Mathematical Contributions to the Theory of Evolution.*—VIII. *On the Inheritance of Characters not capable of Exact Quantitative Measurement.*—Part I. *Introductory.* Part II. *On the Inheritance of Coat-colour in Horses.* Part III. *On the Inheritance of Eye-colour in Man.*

By KARL PEARSON, *F.R.S., with the assistance of* ALICE LEE, *D.Sc., University College, London.*

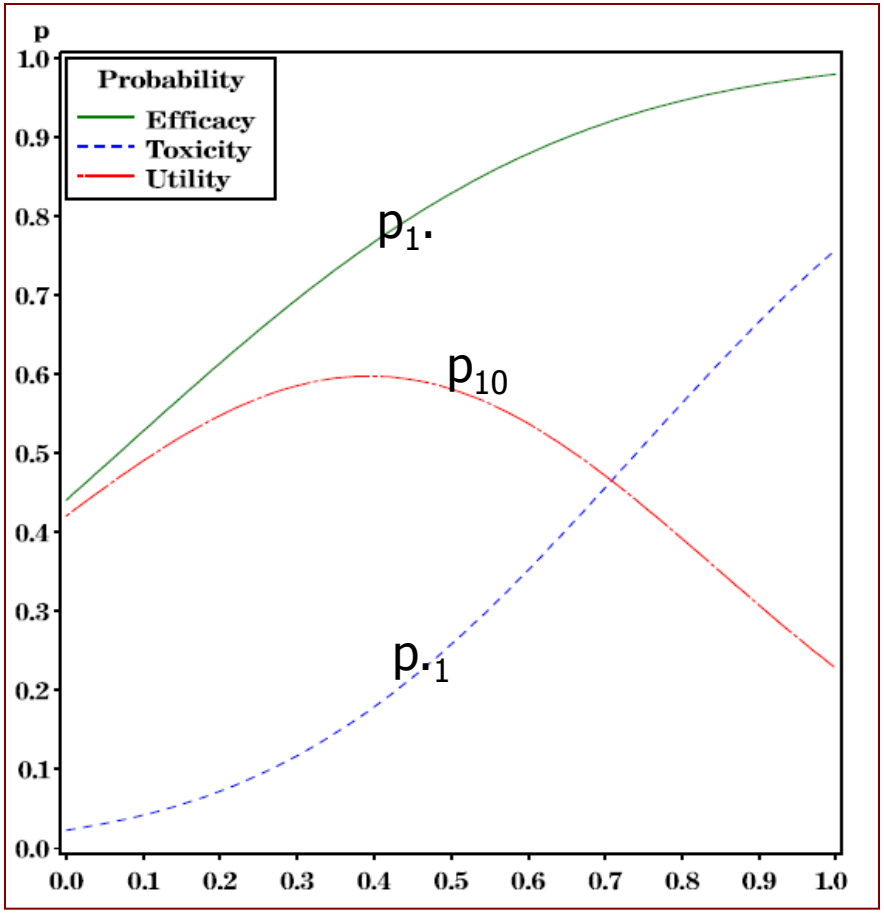
Received August 5,—Read November 16, 1899 ; withdrawn, rewritten, and again received March 5, 1900.

On the Probable Error of a Coefficient of Correlation as Found from a Fourfold Table

Karl Pearson

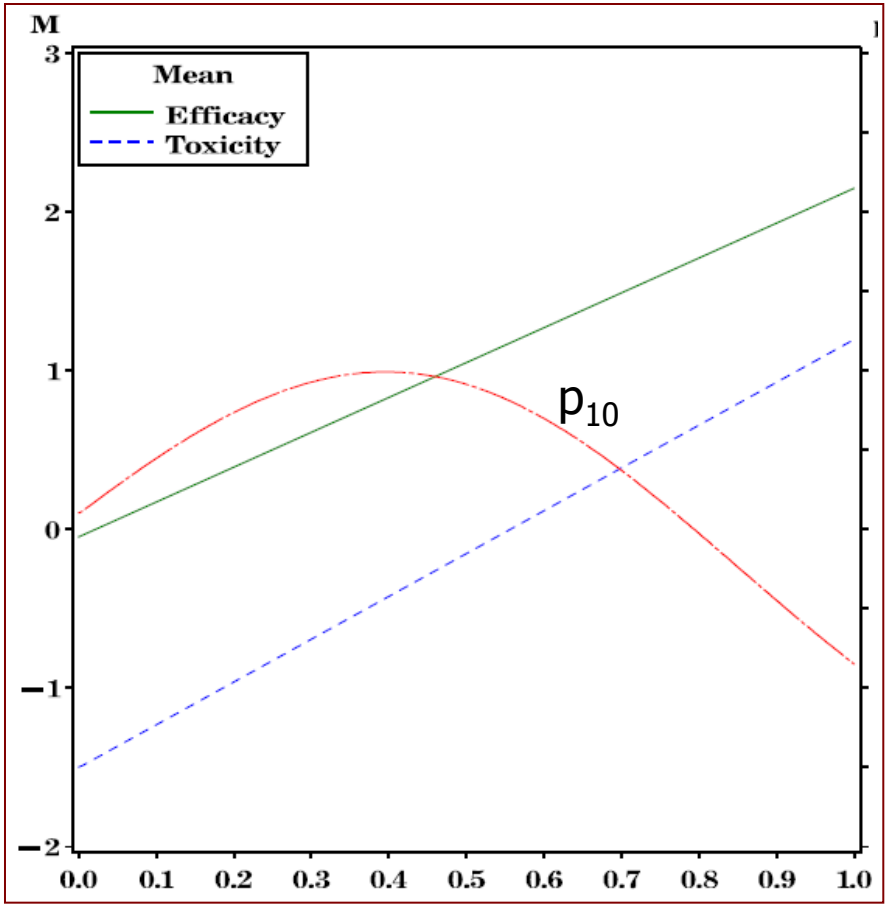
Biometrika, Vol. 9, No. 1/2. (Mar., 1913), pp. 22-33.

Typical dose-response settings



Responses and reported end-point are dichotomized

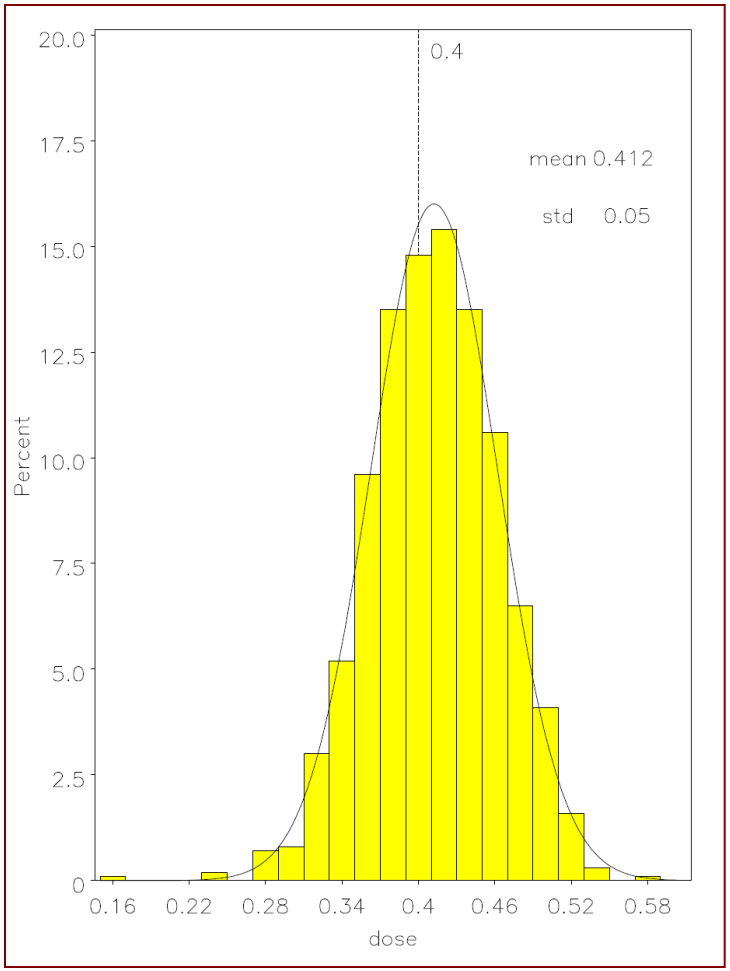
Dragalin. V., and Fedorov V. (2006) *J. Stat.Plann. Inf*, **136**, 1800-1823.

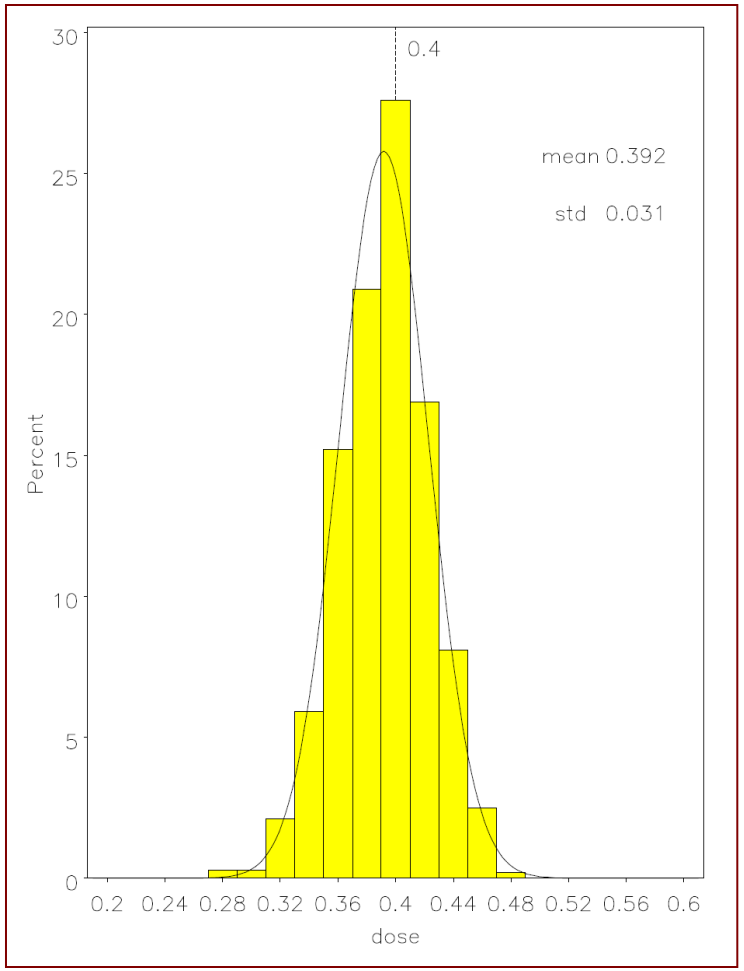


Responses are continuous, reported end-point is dichotomized

Fedorov V., Wu Y. (2007) *JBS* **17**:1085–1096.

Estimation of the best dose with and without dichotomization





Simulation results for two stage D-optimal designs

Responses of different types

$$Y_1 = Z_1, \quad Y_2 = \begin{cases} 1, & \text{if } Z_2 \geq c_2 \\ 0, & \text{otherwise.} \end{cases}$$

$$l(y_1, y_2; \boldsymbol{\vartheta}) \quad \propto \quad y_2 \log\{1 - F(u_2)\} + (1 - y_2) \log\{F(u_2)\} - \log \sigma_1 - \frac{(y_1 - \eta_1)^2}{2\sigma_1^2}$$

$$u_2 = (v_2 - \rho(y_1 - \eta_1)/\sigma_1)/\sqrt{1 - \rho^2}$$

$$v_2 = (c_2 - \eta_2)/\sigma_2$$

$$\boldsymbol{\vartheta} = (\eta_1, v_2, \rho, \sigma_1)^T$$

Elemental information matrix

- $\mu(\vartheta) =$

$$\begin{pmatrix} \frac{1-\rho^2+\rho^2 a_0}{\sigma_1^2(1-\rho^2)} & \frac{\rho a_0}{\sigma_1(1-\rho^2)} & \frac{\rho(\rho v_2 a_0 - a_1)}{\sigma_1(1-\rho^2)^2} & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} \\ - & \frac{a_0}{(1-\rho^2)} & \frac{\rho v_2 a_0 - a_1}{(1-\rho^2)^2} & \frac{\rho a_1}{\sigma_1(1-\rho^2)} \\ - & - & \frac{a_2 - 2\rho v_2 a_1 + \rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{\rho(\rho v_2 a_1 - a_2)}{\sigma_1(1-\rho^2)^2} \\ - & - & - & \frac{2(1-\rho^2)+\rho^2 a_2}{\sigma_1^2(1-\rho^2)} \end{pmatrix}$$

$$a_k(v_2, \rho) = \int_{-\infty}^{+\infty} \frac{t^k \varphi(t) \varphi^2\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)}{F\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)\left[1 - F\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)\right]} \, dt$$

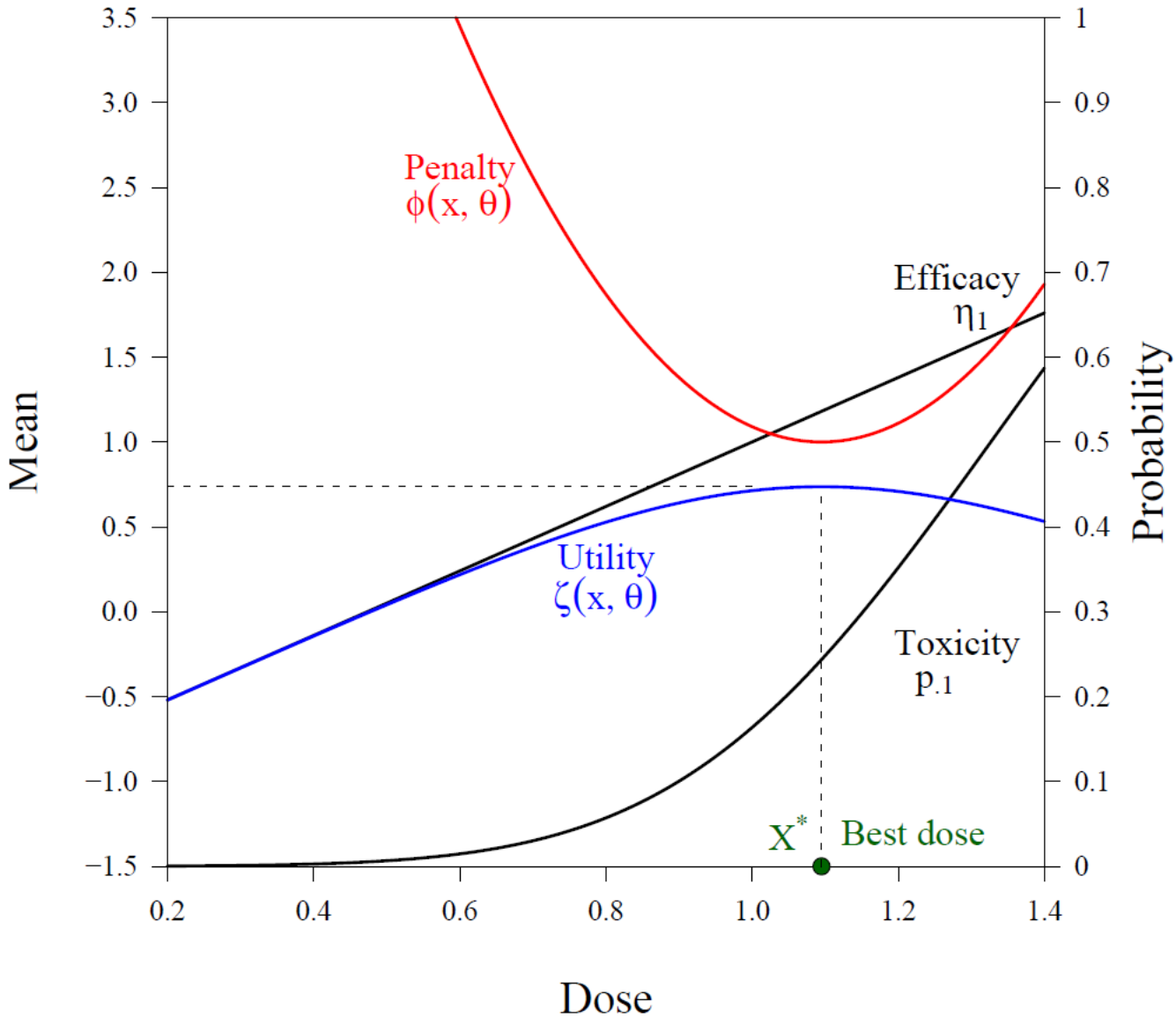
Elemental information matrix

- $\mu(\vartheta) =$

$$\begin{pmatrix} \frac{1-\rho^2+\rho^2 a_0}{\sigma_1^2(1-\rho^2)} & \frac{\rho a_0}{\sigma_1(1-\rho^2)} & \frac{\rho(\rho v_2 a_0 - a_1)}{\sigma_1(1-\rho^2)^2} & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} \\ - & \frac{a_0}{(1-\rho^2)} & \frac{\rho v_2 a_0 - a_1}{(1-\rho^2)^2} & \frac{\rho a_1}{\sigma_1(1-\rho^2)} \\ - & - & \frac{a_2 - 2\rho v_2 a_1 + \rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{\rho(\rho v_2 a_1 - a_2)}{\sigma_1(1-\rho^2)^2} \\ - & - & - & \frac{2(1-\rho^2)+\rho^2 a_2}{\sigma_1^2(1-\rho^2)} \end{pmatrix}$$

$$a_k(v_2, \rho) = \int_{-\infty}^{+\infty} \frac{t^k \varphi(t) \varphi^2\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)}{F\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)\left[1 - F\left(\frac{v_2 - \rho t}{\sqrt{1-\rho^2}}\right)\right]} \, dt$$

Tate R. The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, 1955; 42: 205-216.



$\boldsymbol{\theta} = (-0.9, 1.9, 3.98, -3)$, $\sigma_1 = 1 \quad \rho = 0.5$

Information matrix of a single observation I

if $\boldsymbol{\vartheta} \in R_{m'}$ is a continuous function of $\boldsymbol{\theta} \in R_m$ then

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{J}\boldsymbol{\mu}[\boldsymbol{\vartheta}(\boldsymbol{\theta})]\mathbf{J}^T, \qquad \mathbf{J} = \frac{\partial \boldsymbol{\vartheta}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left|\left| \frac{\partial \vartheta_\beta(\theta)}{\partial \theta_\alpha} \right|\right|_{\alpha=1,\beta=1}^{m,m'}.$$

If we assume that $\eta_1 = \boldsymbol{\theta}_1^T \mathbf{f}_1(x)$ and $v_2 = \frac{c_2 - \eta_2}{\sigma_2} = \boldsymbol{\theta}_2^T \mathbf{f}_2(x)$
i.e. $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \rho, \sigma_1)^T$, then

$$\mathbf{J} = \begin{pmatrix} \mathbf{f}_1(x) & 0 & 0 & 0 \\ 0 & \mathbf{f}_2(x) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Information matrix of a single observation II

$$\boldsymbol{\mu}(\theta) = \begin{pmatrix} \frac{1-\rho^2+\rho^2 a_0}{\sigma_1^2(1-\rho^2)}\mathbf{f}_1\mathbf{f}_1^T & \frac{\rho a_0}{\sigma_1(1-\rho^2)}\mathbf{f}_1\mathbf{f}_2^T & \frac{\rho(\rho v_2 a_0 - a_1)}{\sigma_1(1-\rho^2)^2}\mathbf{f}_1 & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)}\mathbf{f}_1 \\ - & \frac{a_0}{(1-\rho^2)}\mathbf{f}_2\mathbf{f}_2^T & \frac{\rho v_2 a_0 - a_1}{(1-\rho^2)^2}\mathbf{f}_2 & \frac{\rho a_1}{\sigma_1(1-\rho^2)}\mathbf{f}_2 \\ - & - & \frac{a_2 - 2\rho v_2 a_1 + \rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{\rho(\rho v_2 a_1 - a_2)}{\sigma_1(1-\rho^2)^2} \\ - & - & - & \frac{2(1-\rho^2)+\rho^2 a_2}{\sigma_1^2(1-\rho^2)} \end{pmatrix}$$

Note that dim $\boldsymbol{\mu}(\vartheta)$ = 4x4, dim $\boldsymbol{\mu}(\boldsymbol{\Theta})$ = (m1+m2+2)x(m1+m2+2)

Utility and penalty

- Our utility function is the mean efficacy multiplied by the probability of no-toxicity:

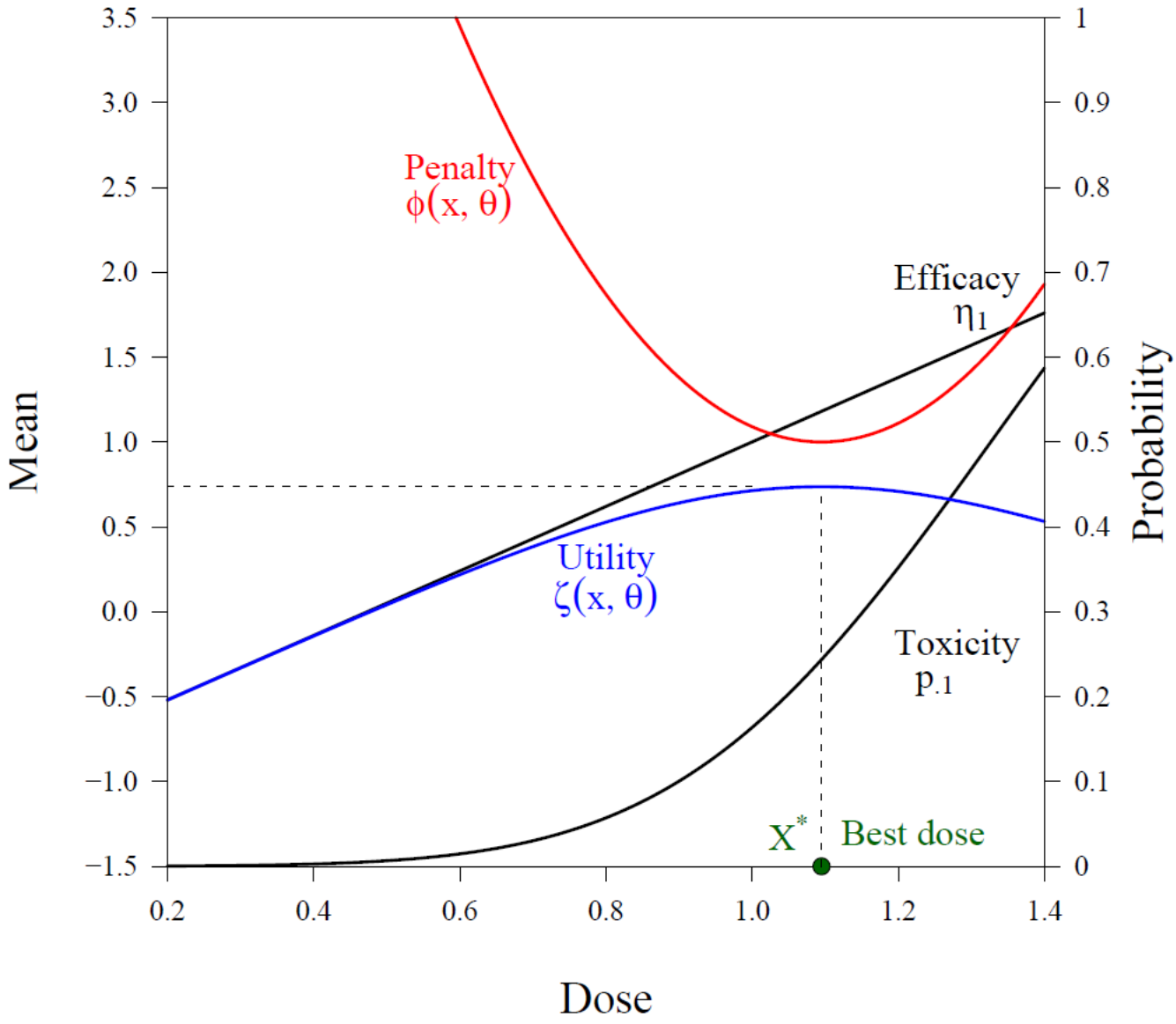
$$\begin{aligned}\zeta(x,\boldsymbol{\theta}) &= E(Y_1|Y_2=0)P(Y_2=0)\\ &= \eta_1 F(v_2) - \rho\sigma_1\varphi(v_2) = \boldsymbol{\theta}_1^T\mathbf{f}_1(x)F(\boldsymbol{\theta}_2^T\mathbf{f}_2(x)) - \rho\sigma_1\varphi(\boldsymbol{\theta}_2^T\mathbf{f}_2(x))\end{aligned}$$

- Penalty

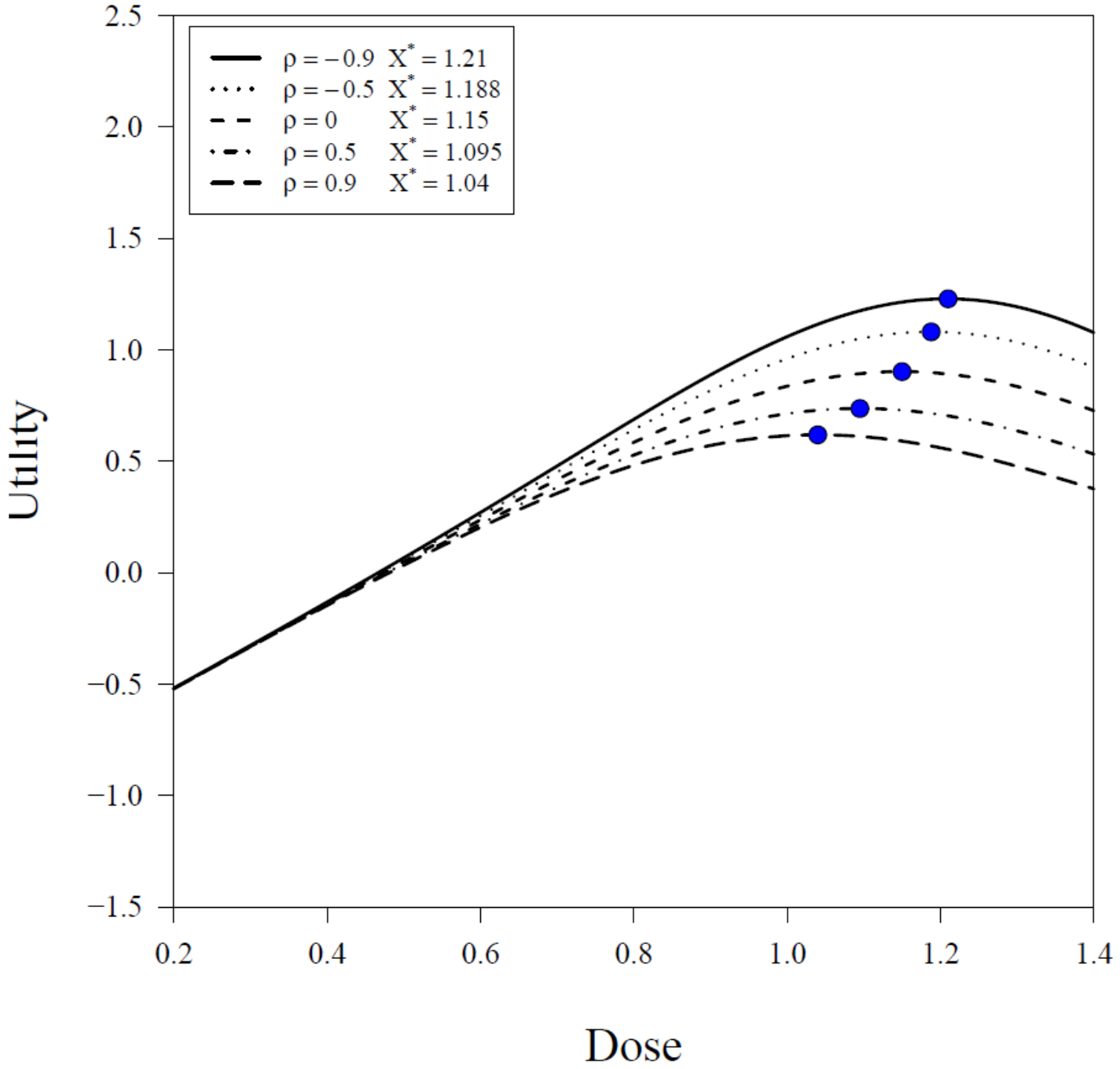
$$\phi(x,\boldsymbol{\theta}) = r(x - x^*(\boldsymbol{\theta}))^2 + c$$

where

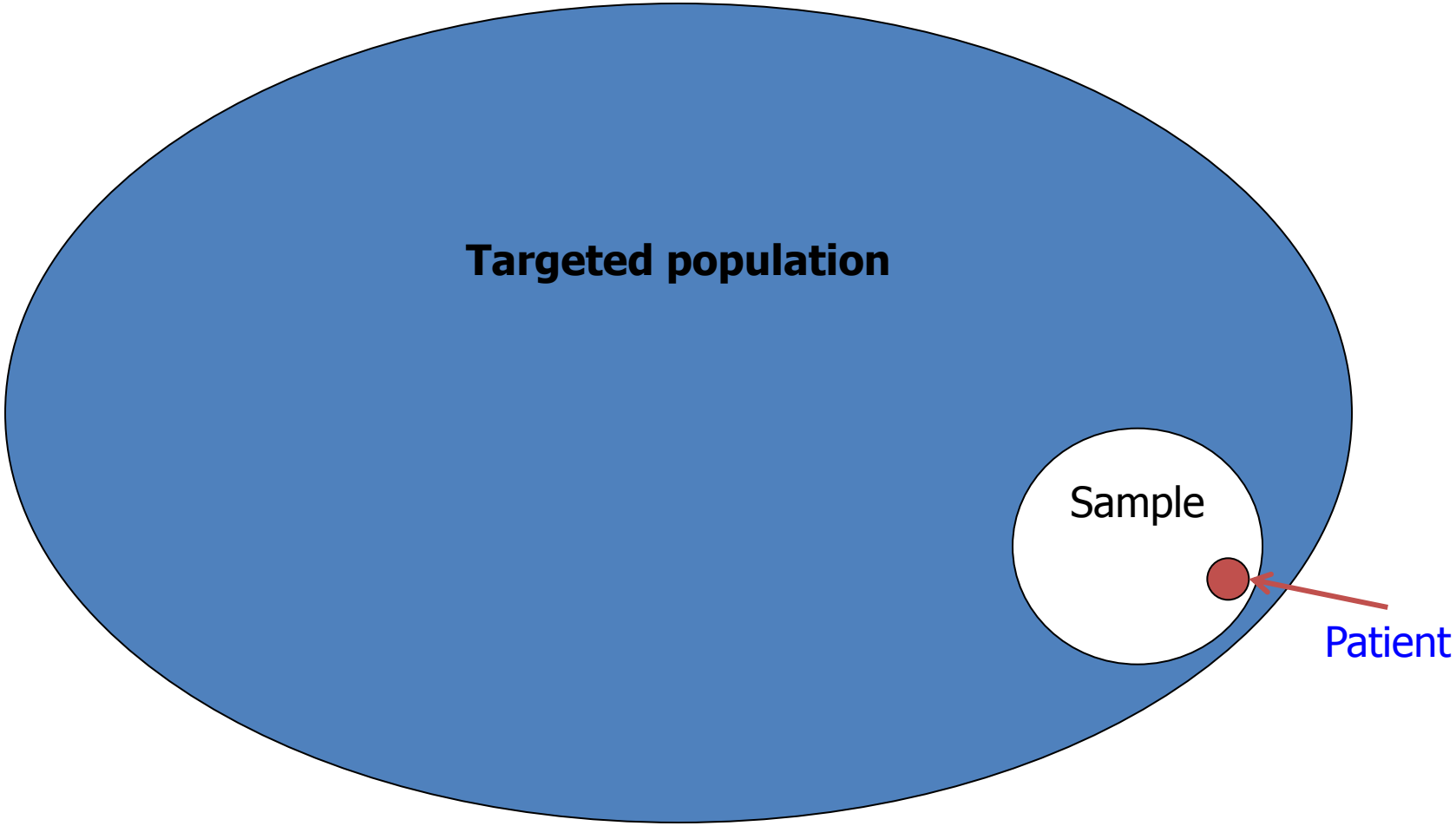
$$x^*(\boldsymbol{\theta}) = \arg\max_{x\in\mathfrak{X}} \zeta(x,\boldsymbol{\theta}).$$



$\boldsymbol{\theta} = (-0.9, 1.9, 3.98, -3)$, $\sigma_1 = 1 \quad \rho = 0.5$


ρ = −0.9 X* = 1.21
ρ = −0.5 X* = 1.188
ρ = 0 X* = 1.15
ρ = 0.5 X* = 1.095
ρ = 0.9 X* = 1.04
Utility
Dose
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.2
0.4
0.6
0.8
1.0
1.2
1.4

Population, sample, patient



Cost/penalty averaged across a sample: $Cost = N\Phi(\xi) = N\sum_{i=1}^{n} w_i \varphi(\mathbf{x}_i)$

Main optimization problem

Optimal design:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi(N(\xi) M(\xi, \theta))$$
$$\text{s.t.} \quad N(\xi)\Phi(\xi) \leq C$$

Equivalently:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi\left(\frac{M(\xi, \theta)}{\Phi(\xi)}\right)$$

The same but with “prior” information

Optimal design:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi(\mathbf{M}_0 + N(\xi) M(\xi, \theta))$$
$$\text{s.t.} \quad N(\xi)\Phi(\xi) \leq C$$

Equivalently:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi\left(\frac{\mathbf{M}_0}{C} + \frac{M(\xi, \theta)}{\Phi(\xi)}\right)$$

Basic “design” formulae

Necessary and sufficient conditions (D-criterion):

$$\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi^*,\theta)] \le m\varphi(\mathbf{x})/\Phi(\xi^*)$$

First order algorithm (D-criterion)

Step forward:

$$\mathbf{x}_{s+1}^{\oplus} = \arg\max_{\mathbf{x}\in\mathcal{X}} \{\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi_s,\theta)] - m\varphi(\mathbf{x})/\Phi(\xi_s)\}$$

Step backward:

$$\mathbf{x}_{s+1}^{\ominus} = \arg\min_{\mathbf{x}\in\mathcal{X}_s} \{\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi_s,\theta)] - m\varphi(\mathbf{x})/\Phi(\xi_s)\}$$

Note. If there is a prior information then use:

$$\mathrm{tr}\{[\mu(\mathbf{x},\theta)+\mathbf{M}_0\varphi(\mathbf{x})/C][M(\xi^*,\theta)+\mathbf{M}_0\Phi(\xi^*)/C]^{-1}\} \le m\varphi(\mathbf{x})/\Phi(\xi^*)$$

Numerical construction of D-optimal design

- Step N + : add more observations to the dose (design point) where the variance of the estimated response divided by the penalty function is maximal, i.e. place more observation where the knowledge about response standardized by penalty is worst
- Step N - : remove observations from the dose (design point) where the variance of the estimated response divided by the penalty function is minimal, i.e. exclude observations from design points where the knowledge about response standardized by penalty is better than in any other design point

Note: Replace “variance of the estimated response” by “sensitivity function” in general case.

Two stage design

- Original optimization problem:

$$\xi^*(\boldsymbol{\theta}) = \arg\min_{\xi} \Psi[N_0\mathbf{M}(\xi_0) + N_1\mathbf{M}(\xi)] \quad \text{s.t. } N_0\Phi(\xi_0) + N_1\Phi(\xi) \leq C$$

- Continuous approximation:

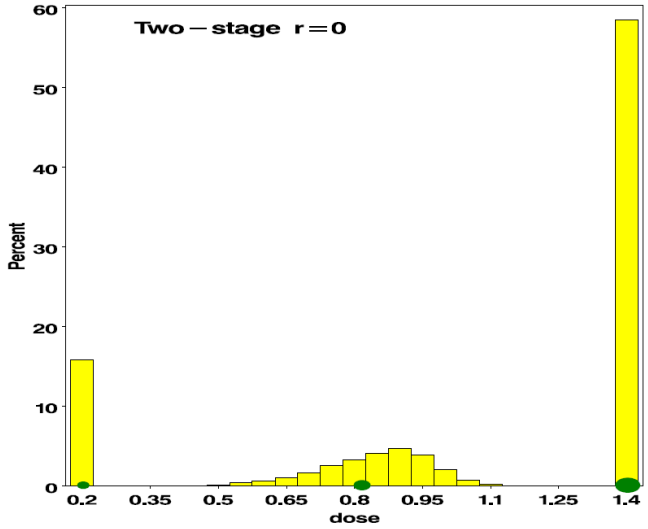
$$\xi^*(\boldsymbol{\theta}) = \arg\min_{\xi} \Psi\left[\frac{\pi\mathbf{M}(\xi_0) + (1-\pi)\mathbf{M}(\xi)}{\pi\Phi(\xi_0) + (1-\pi)\Phi(\xi)}\right]$$

- Sensitivity function:

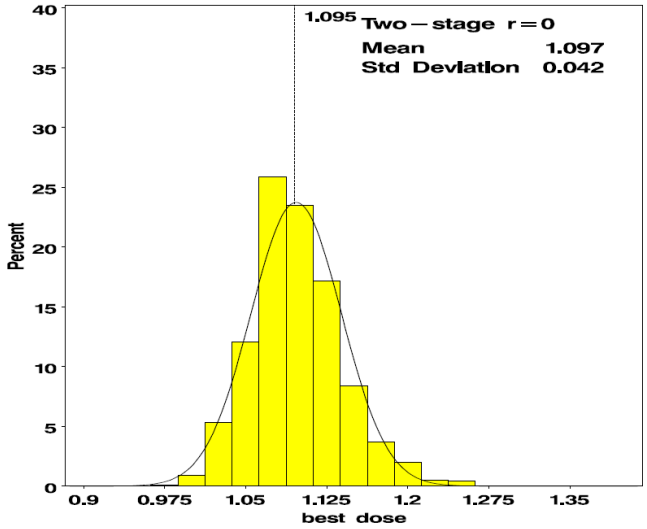
$$\psi(x, \xi, \boldsymbol{\theta}) = \text{tr}\left\{\boldsymbol{\mu}(x, \boldsymbol{\theta})\left[\pi\mathbf{M}(\xi_0, \boldsymbol{\theta}) + (1-\pi)\mathbf{M}(\xi, \boldsymbol{\theta})\right]^{-1}\right\} - \frac{m\phi(x, \boldsymbol{\theta})}{\pi\Phi(\xi_0, \boldsymbol{\theta}) + (1-\pi)\Phi(\xi, \boldsymbol{\theta})}$$

- In practice the second stage is optimized with the “plugged in” estimates of parameters. How to select π ?

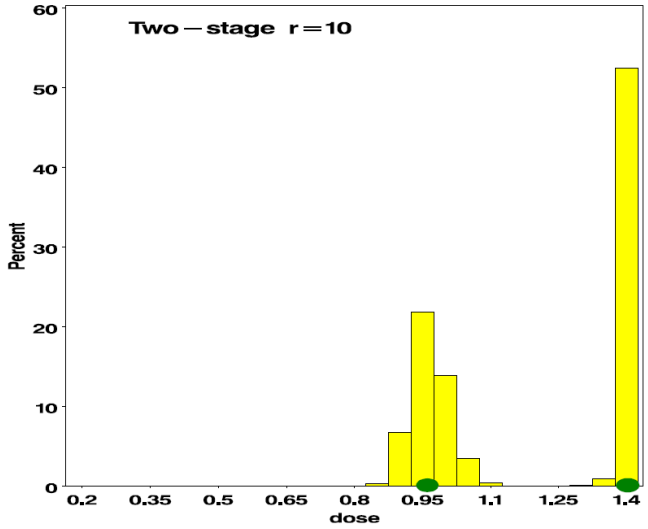
Comparison



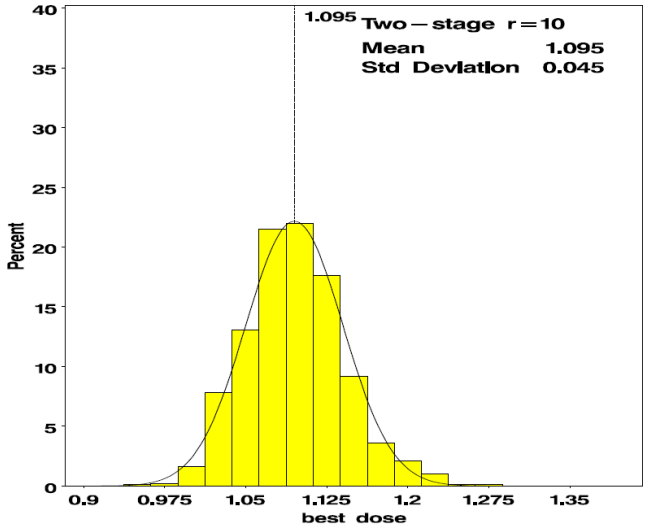
(a)



(b)

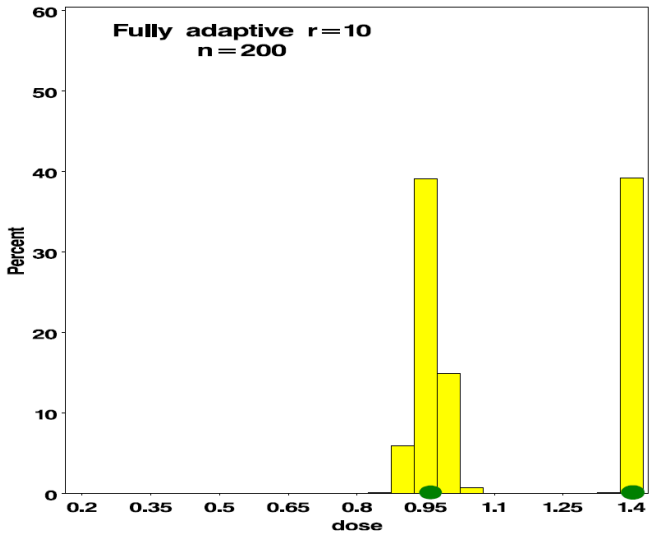


(c)

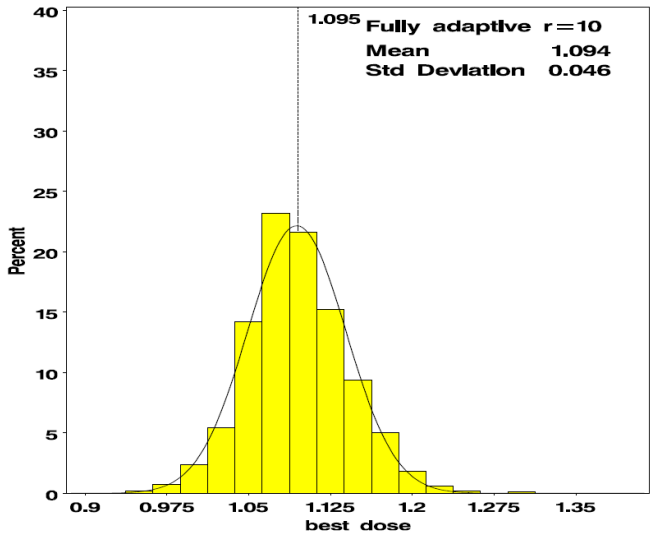


(d)

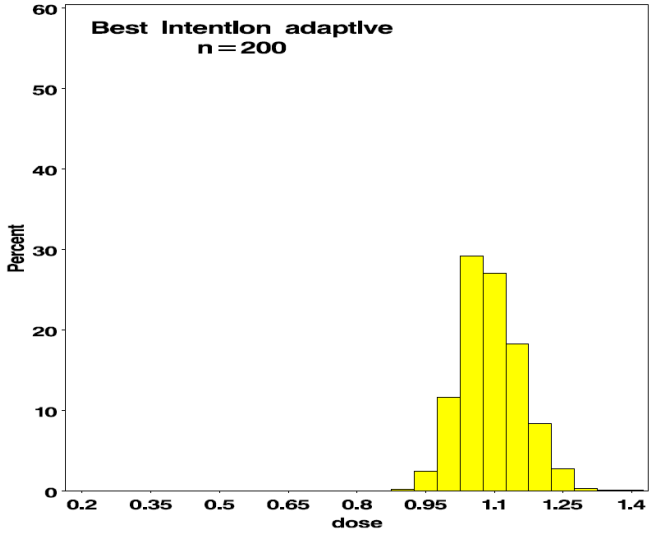
Comparison



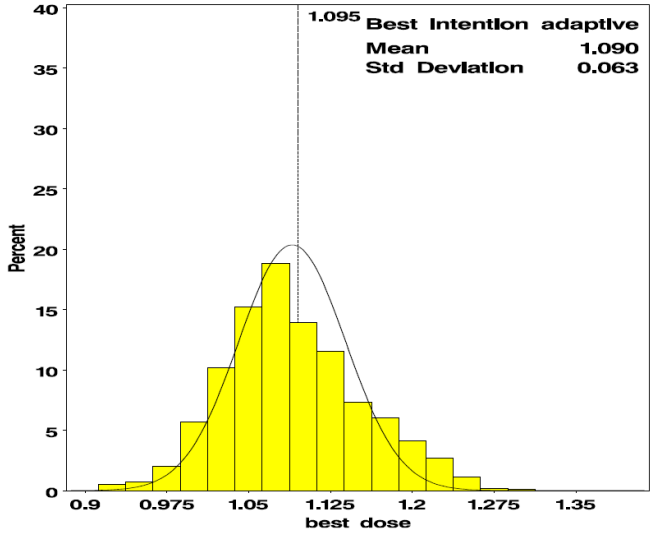
(e)



(f)

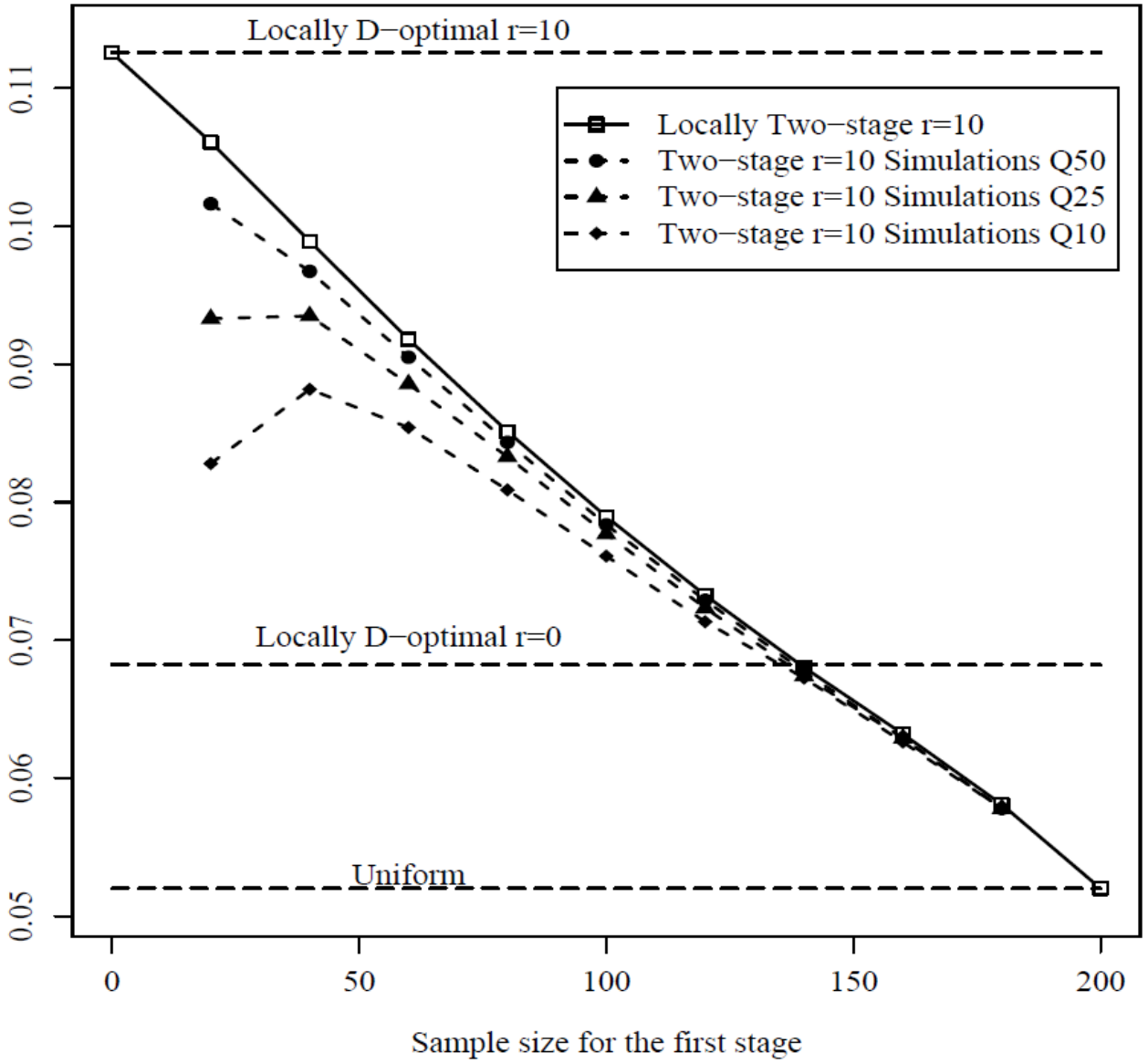


(g)



(h)

Selection of π



Summary

- The use of a background (latent*) multivariate normal distribution allows to build models containing correlated responses of various types
- The traditional eight step dance can be routinely performed
- Two stage experiments often yield more information than fully adaptive ones. Operationally/logistically they are always superior.

* http://videolectures.net/slsfs05_titterington_salsa/

References

1. Tate R. (1955) The Theory of Correlation Between Two Continuous Variables when One is Dichotomized, *Biometrika,* 42, 205-216.
2. Fedorov V., Hackl P. (1997). *Model-Oriented Design of Experiments,* Lecture Notes in Statistics, New York: Springer.
3. Dragalin V., Fedorov V. (2006). Adaptive designs for dose finding, Journal of Statistical Planning and Inference, 136, 1800-1823.
4. Fedorov V, Wu Y. (2007). Dose-finding designs for continuous responses and binary utility. *Journal of Biopharmaceutical Statistics,* 17, 1085–1096.
5. Fedorov V, Wu Y. (2007). Generalized probit model in design of dose finding experiments. In: Lopez-Fidalgo J, Rodriguez-Diaz JM, Torsney B (Eds), *MODA 8 – Advances in Model-Oriented Design and Analysis,* Physica-Verlag, 67–74.
6. Dragalin V., Fedorov V, Wu Y. (2008). Adaptive designs for selecting drug combinations based on efficacy-toxicity response. *Journal of Statistical Planning and Inference,* 138, 352-373
7. Dragalin V, Fedorov V, Wu Y (2008). Two-stage design for dose-finding that accounts for both efficacy and toxicity, *Statistics in Medicine,* 27, 5156–5176
8. Fedorov V., Wu Y. and Zhang R. (2011) Optimal dose finding designs with correlated continuous and discrete responses, Statistics in Medicine, …