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
Efficacy

Clinical Complete Remission (CR)
- Disappearance of all clinical, radiologic and biologic evidence of tumor > 4 wks
- Treatment

Partial Response (PR)
- Decrease in product of tumor diameters by at least 50% 
  \((a'b') < (ab)/2\)
  OR
  Decrease in widest diameter \((b)\) by 30% or more
- Treatment

Progressive Disease (PD)
- Increase in product of tumor diameters by at least 25%
  \((a'b') > 1.25 \times (ab)\)
  OR
  Increase in widest diameter \((b)\) by 20% or more
  OR
  New lesion
- Treatment

Often (both for reporting and analysis):
\[ Y = 1 \text{ if CR or PR, otherwise } Y = 0. \]

Measurements and Reported End-Points
## Toxicity

Measurements and Reported End-Points

<table>
<thead>
<tr>
<th>Condition</th>
<th>Placebo</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any SAE</td>
<td>30%</td>
<td>70%</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>0</td>
<td>8%</td>
</tr>
<tr>
<td>Cardiac disorders</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Nausea/Vomiting</td>
<td>5% / 3%</td>
<td>10% / 10%</td>
</tr>
<tr>
<td>Anemia</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>Hepatotoxicity</td>
<td>3%</td>
<td>10%</td>
</tr>
<tr>
<td>Blood and lymphatic system</td>
<td>4%</td>
<td>6%</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3%</td>
<td>12%</td>
</tr>
<tr>
<td>Metabolism and nutrition disorders</td>
<td>5%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Y = 1, if there is any SAE, Y = 0 otherwise
Dichotomized analysis and reporting

<table>
<thead>
<tr>
<th>Efficacy Y</th>
<th>Toxicity Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>$p_{11}$</td>
</tr>
<tr>
<td>0</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>$p_{\bullet1}$</td>
<td>$p_{\bullet0}$</td>
</tr>
</tbody>
</table>

Probit model

\[
Z \sim N(\eta, \Sigma)
\]

\[
p_{00} = P(Y_1 = 0, Y_2 = 0) = F(V; \Sigma^*) = \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{1}{2\pi|\Sigma_1|^{1/2}} \exp \left\{ -\frac{1}{2} V^T \Sigma_1^{-1} V \right\} \, dv_1 \, dv_2
\]

\[
v = (v_1, v_2)^T, \quad v_k = (c_k - \eta_k)/\sigma_k, \quad Y_k = I(Z_k > c_k)
\]

\[
\Sigma^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}
\]
More about probit model can be found in:


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

Typical dose-response settings

Responses and reported end-point are dichotomized

Responses are continuous, reported end-point is dichotomized


Estimation of the best dose with and without dichotomization

Simulation results for two stage D-optimal designs
Responses of different types

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

\[
\ell(y_1, y_2; \vartheta) \propto y_2 \log \{1 - F(u_2)\} + (1 - y_2) \log \{F(u_2)\} - \log \sigma_1 - \frac{(y_1 - \eta_1)^2}{2\sigma_1^2}
\]

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

\[
u_2 = (c_2 - \eta_2)/\sigma_2
\]

\[
\vartheta = (\eta_1, \nu_2, \rho, \sigma_1)^T
\]
Elemental information matrix

\[ \mu(\mathcal{I}) = \begin{pmatrix} \frac{1-\rho^2+\rho^2a_0}{\sigma_1^2(1-\rho^2)} & \frac{\rho a_0}{\sigma_1(1-\rho^2)} & \frac{\rho(\rho v_2 a_0-a_1)}{\sigma_1(1-\rho^2)^2} & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} \\ \frac{a_0}{(1-\rho^2)} & \frac{\rho v_2 a_0-a_1}{(1-\rho^2)^2} & \frac{\rho a_0}{\sigma_1(1-\rho^2)} & \frac{\rho(\rho v_2 a_1-a_2)}{\sigma_1(1-\rho^2)^2} \\ \frac{1}{\rho v_2 a_0-a_1} & \frac{2\rho v_2 a_0-a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{a_2-\rho v_2 a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^2} & \frac{1}{\sigma_1(1-\rho^2)^2} \\ \frac{1}{\rho v_2 a_0-a_1} & \frac{2\rho v_2 a_0-a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{a_2-\rho v_2 a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^2} & \frac{1}{\sigma_1(1-\rho^2)^2} \end{pmatrix} \]

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

\[ \mu(\vartheta) = \begin{pmatrix} \frac{1-\rho^2+\rho^2 a_0}{\sigma_1^2(1-\rho^2)} & \frac{\rho a_0}{\sigma_1(1-\rho^2)} & \frac{\rho(\rho v_2 a_0-a_1)}{\sigma_1(1-\rho^2)^2} & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} \\ \frac{-}{\sigma_1(1-\rho^2)} & \frac{a_0}{(1-\rho^2)} & \frac{\rho v_2 a_0-a_1}{(1-\rho^2)^2} & \frac{\rho a_1}{\sigma_1(1-\rho^2)} \\ \frac{-}{-} & \frac{-}{\frac{a_2-2\rho v_2 a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^3}} & \frac{\rho(\rho v_2 a_1-a_2)}{\sigma_1(1-\rho^2)^2} & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} \\ \frac{-}{-} & \frac{-}{2(1-\rho^2)+\rho^2 a_2} & \frac{-}{\frac{\sigma_1(1-\rho^2)^2}{\rho^2 a_2}} & \frac{\sigma_1^2(1-\rho^2)}{\rho^2 a_2} \end{pmatrix} \]

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

Tate R. The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, 1955; 42: 205-216.
\[ \theta = (-0.9, 1.9, 3.98, -3), \quad \sigma_1 = 1 \quad \rho = 0.5 \]
Information matrix of a single observation \( I \)

If \( \mathbf{v} \in R_{m'} \) is a continuous function of \( \theta \in R_m \) then

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

If we assume that \( \eta_1 = \theta_1^T \mathbf{f}_1(x) \) and \( \nu_2 = \frac{c_2-\eta_2}{\sigma_2} = \theta_2^T \mathbf{f}_2(x) \)

i.e. \( \theta = (\theta_1^T, \theta_2^T, \rho, \sigma_1)^T \), then

\[
\mathbf{J} = \\
\begin{pmatrix}
\mathbf{f}_1(x) & 0 & 0 & 0 & 0 \\
0 & \mathbf{f}_2(x) & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
Information matrix of a single observation II

\[
\mu(\theta) = \begin{pmatrix}
\frac{1-\rho^2+\rho^2a_0}{\sigma_1^2(1-\rho^2)} f_1 f_1^T & \frac{\rho a_0}{\sigma_1(1-\rho^2)} f_1 f_2^T & \frac{\rho(\rho v_2 a_0-a_1)}{\sigma_1(1-\rho^2)^2} f_1 & \frac{\rho^2 a_1}{\sigma_1^2(1-\rho^2)} f_1 \\
- & \frac{a_0}{(1-\rho^2)} f_2 f_2^T & \frac{\rho v_2 a_0-a_1}{(1-\rho^2)^2} f_2 & \frac{\rho a_1}{\sigma_1(1-\rho^2)} f_2 \\
- & - & \frac{a_2-2\rho v_2 a_1+\rho^2 v_2^2 a_0}{(1-\rho^2)^3} & \frac{\rho(\rho v_2 a_1-a_2)}{\sigma_1(1-\rho^2)^2} \\
- & - & - & \frac{2(1-\rho^2)+\rho^2 a_2}{\sigma_1^2(1-\rho^2)}
\end{pmatrix}
\]

Note that \( \dim \mu(\mathcal{D}) = 4 \times 4 \), \( \dim \mu(\Theta) = (m_1+m_2+2) \times (m_1+m_2+2) \).
Utility and penalty

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

\[
\zeta(x, \theta) = E(Y_1|Y_2 = 0)P(Y_2 = 0) = \eta_1 F(v_2) - \rho \sigma_1 \phi(v_2) = \theta_1^T f_1(x) F(\theta_2^T f_2(x)) - \rho \sigma_1 \phi(\theta_2^T f_2(x))
\]

• Penalty

\[
\phi(x, \theta) = r(x - x^*(\theta))^2 + c
\]

where

\[
x^*(\theta) = \arg \max_{x \in \mathcal{X}} \zeta(x, \theta).
\]
\[ \theta = (-0.9, 1.9, 3.98, -3), \sigma_1 = 1 \quad \rho = 0.5 \]
Cost/penalty averaged across a sample: 

$$\text{Cost} = N \Phi(\xi) = N \sum_{i=1}^{n} w_i \phi(x_i)$$
Main optimization problem

Optimal design:

\[
\xi^* = \arg \min_{\xi \in \Xi(\mathcal{X})} \Psi(N(\xi)M(\xi, \theta))
\]

s.t.  \(N(\xi)\Phi(\xi) \leq C\)

Equivalently:

\[
\xi^* = \arg \min_{\xi \in \Xi(\mathcal{X})} \Psi\left(\frac{M(\xi, \theta)}{\Phi(\xi)}\right)
\]
The same but with “prior” information

Optimal design:

\[
\xi^* = \arg\min_{\xi \in \Xi(\mathcal{X})} \Psi((M_0 + N(\xi)M(\xi, \theta))) \\
\text{s.t. } N(\xi)\Phi(\xi) \leq C
\]

Equivalently:

\[
\xi^* = \arg\min_{\xi \in \Xi(\mathcal{X})} \Psi\left(\frac{M_0}{C} + \frac{M(\xi, \theta)}{\Phi(\xi)}\right)
\]
Basic “design” formulae

Necessary and sufficient conditions (D-criterion):

\[ \text{tr}[\mu(x, \theta)M^{-1}(\xi^*, \theta)] \leq m\varphi(x)/\Phi(\xi^*) \]

First order algorithm (D-criterion)

Step forward:

\[ x_{s+1}^\oplus = \arg \max_{x \in \mathcal{X}} \left\{ \text{tr}[\mu(x, \theta)M^{-1}(\xi_s, \theta)] - m\varphi(x)/\Phi(\xi_s) \right\} \]

Step backward:

\[ x_{s+1}^\ominus = \arg \min_{x \in \mathcal{X}_s} \left\{ \text{tr}[\mu(x, \theta)M^{-1}(\xi_s, \theta)] - m\varphi(x)/\Phi(\xi_s) \right\} \]

Note. If there is a prior information then use:

\[ \text{tr}\left\{[\mu(x, \theta) + M_0\varphi(x)/C][M(\xi^*, \theta) + M_0\Phi(\xi^*)/C]^{-1}\right\} \leq m\varphi(x)/\Phi(\xi^*) \]
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 observations 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:

\[ \xi^*(\theta) = \arg \min_{\xi} \psi \left[ N_0 M(\xi_0) + N_1 M(\xi) \right] \quad \text{s.t.} \quad N_0 \Phi(\xi_0) + N_1 \Phi(\xi) \leq C \]

• Continuous approximation:

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

• Sensitivity function:

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

• In practice the second stage is optimized with the “plugged in” estimates of parameters. How to select \( \pi \)?
Comparison

(a) Two-stage $r=0$

(b) Two-stage $r=0$

(c) Two-stage $r=10$

(d) Two-stage $r=10$
Comparison

![Graph e](image)

![Graph f](image)

![Graph g](image)

![Graph h](image)
Selection of $\pi$
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

8. Fedorov V., Wu Y. and Zhang R. (2011) Optimal dose finding designs with correlated continuous and discrete responses, Statistics in Medicine, ...