Budapesti Műszaki és Gazdaságtudományi Egyetem
Matematikai Modellalkotás Szeminárium

Curvature, Optimal Design and Statistical Modelling: Intersections of Mathematics and Statistics, Theory and Practice

Timothy E. O'Brien
Department of Mathematics and Statistics
Institute of Environmental Sustainability
Loyola University Chicago

Budapest and Chicago **29 September 2020**

Talk Outline

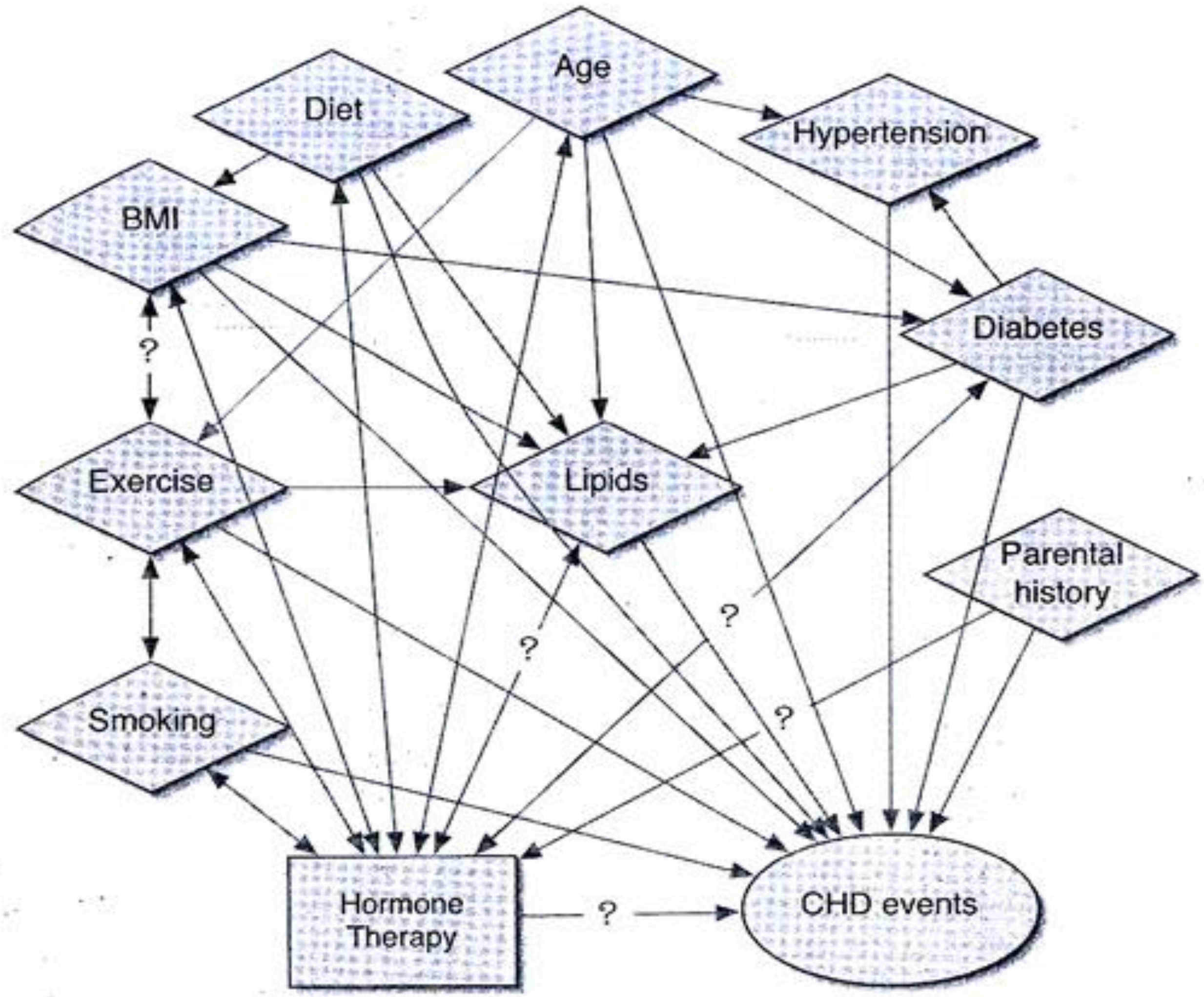
A. Modelling Overview
B. Confidence Regions and Intervals
C. Optimal Design Topics
D. Two Examples from Toxicology
E. Robust Design Strategies for Toxicology Examples
F. Strategy Applied to Examples
G. General Robust Design Strategies
H. Bayesian and Geometric Design Strategies
I. Conclusion and Comments

Acknowledgements/Thanks to Professor Domokos Szasz, Dr. Gergely Madi-Nagy, and Roland Molontay

A. Modelling Overview

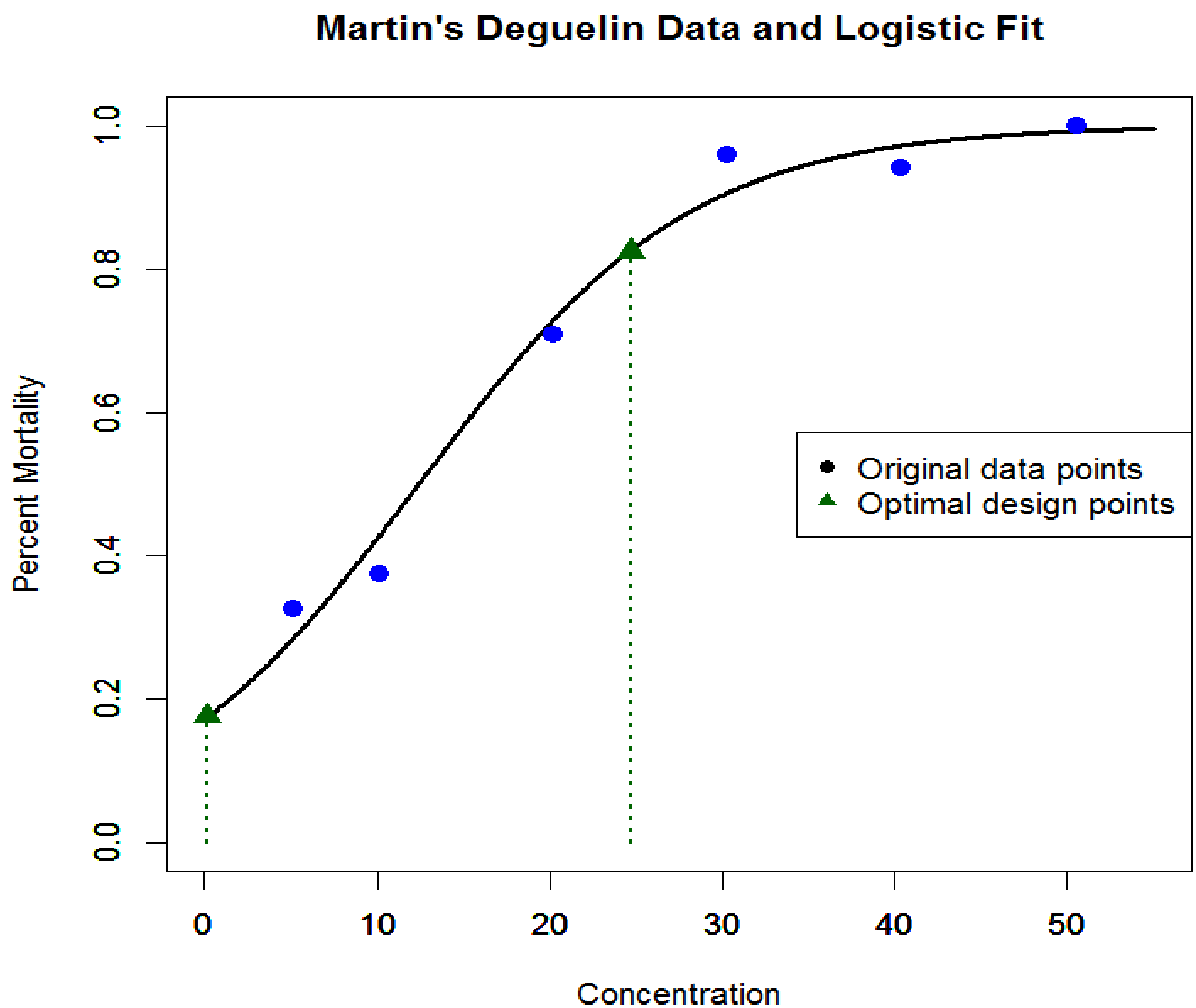
- ***"The model"*** **includes four components:**
  (a) assumed distribution for the response variable (Y) – usually chosen from the exponential family
  (b) the link function connecting $\mu = E(Y)$ with the explanatory variable(s) and model parameters
  (c) the [mean] model function $\eta(x, \theta)$, which combines the explanatory variable(s) with the model parameters
  (d) the variance or variance function (perhaps depending on θ and/or additional parameter(s))

Modelling: finding relationships – but mindful of potential confounding, mediating, and/or interacting variables

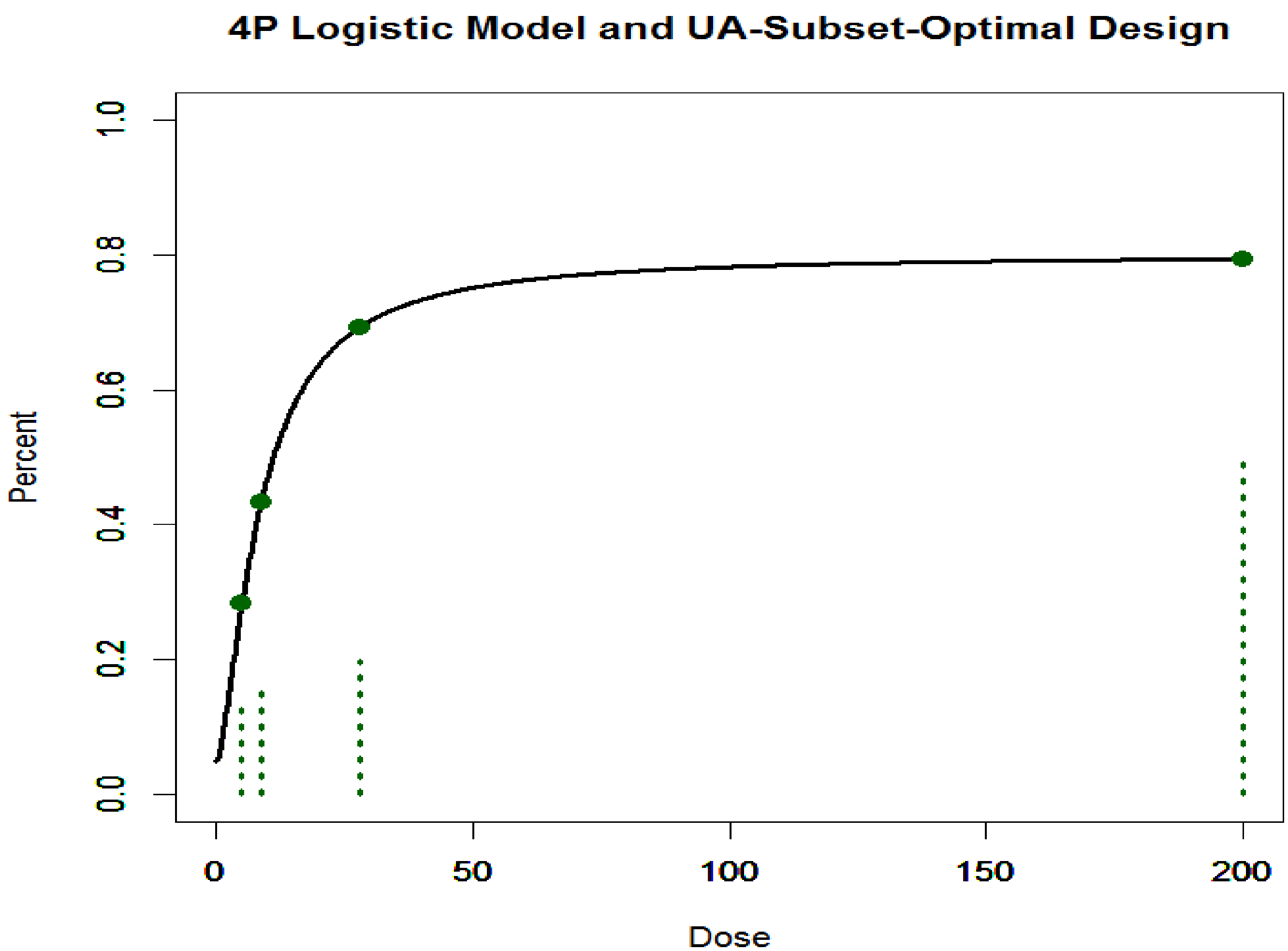


We’ll look only at *experimental studies* here

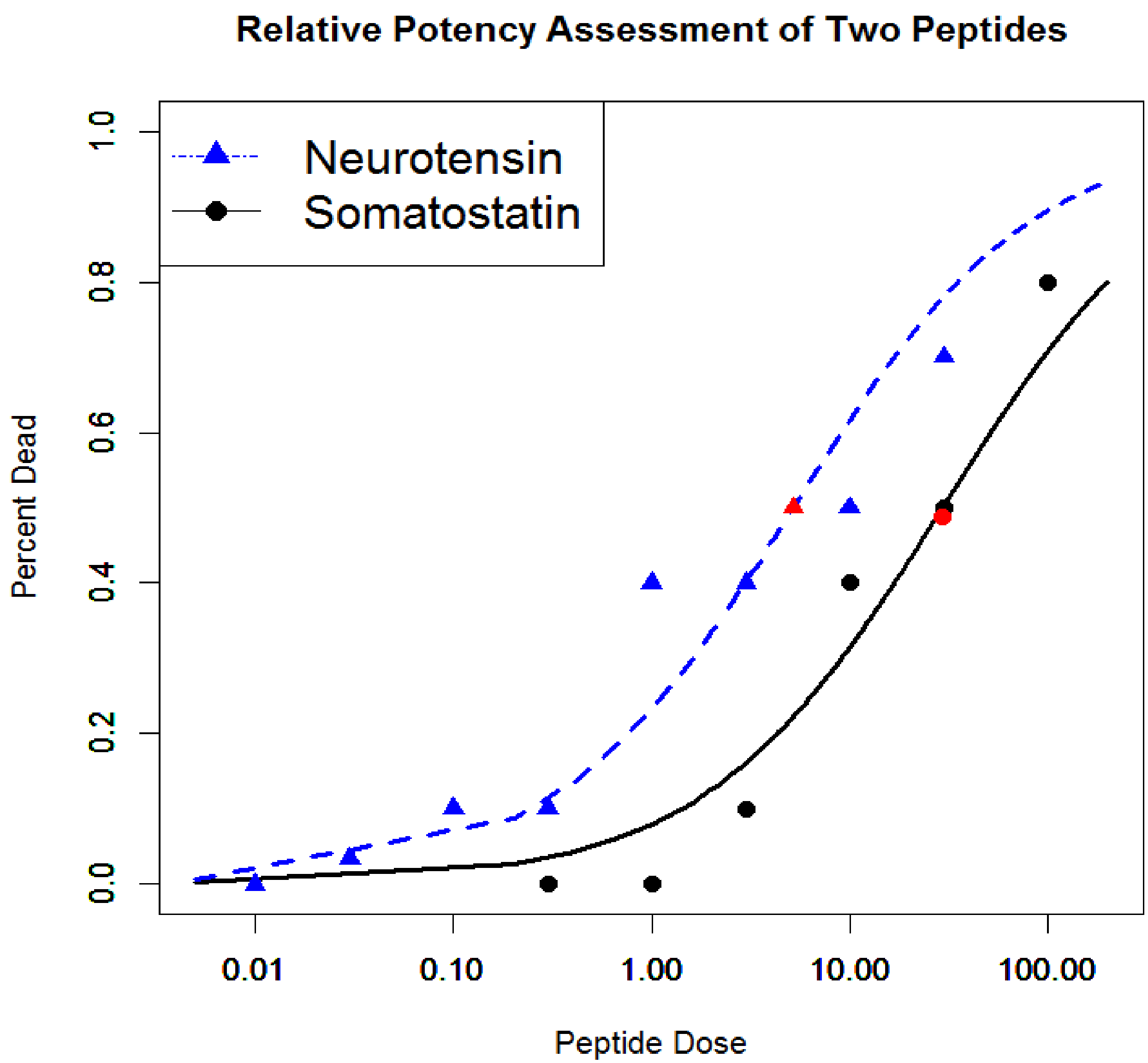
- **Simple linear regression (SLR) entails the Normal distribution/likelihood, the identity link, the model function $\eta(x,\theta) = \alpha + \beta x$, and constant variance**
- **The Binary/Binomial Logistic Model is an illustration of a *generalized linear model*; it assumes the Binomial or Bernoulli distribution and (usually) the logit link: $log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$ or equivalently $\pi = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$. No new variance parameter is introduced here (it comes from the binomial distribution)**
- **In this Logistic Model, sometimes we write $\pi = \frac{e^{\beta(x-\gamma)}}{1+e^{\beta(x-\gamma)}}$ so that γ is the LD50/EC50 parameter, and this is a *nonlinear model* since now $log\left(\frac{\pi}{1-\pi}\right) = \beta(x-\gamma)$**


Martin's Deguelin Data and Logistic Fit
Percent Mortality
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
Concentration
Original data points
Optimal design points

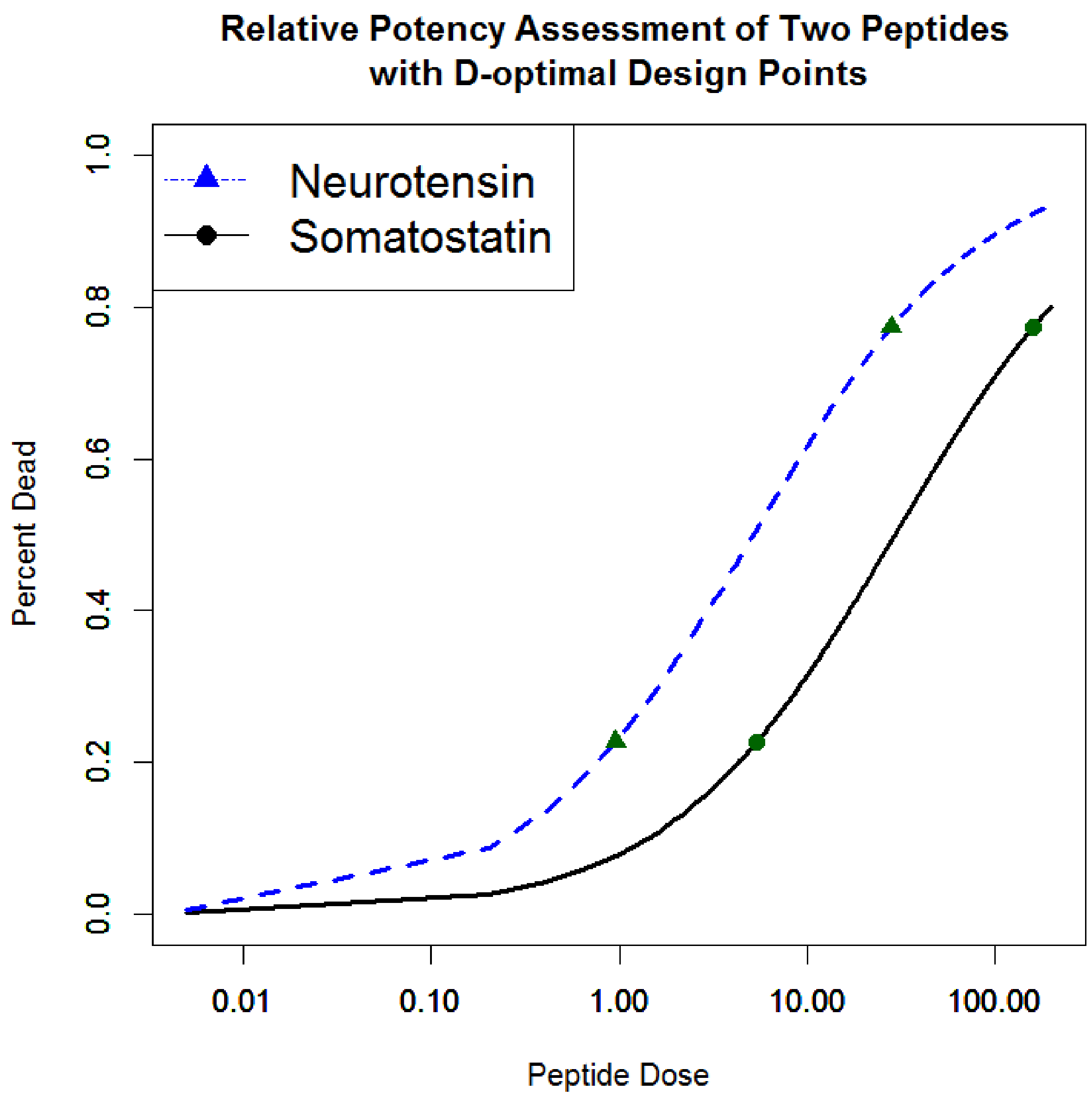
- **In the applied sciences, we often use a nonlinear model function (i.e., nonlinear in the parameters), e.g.** $\eta(x, \theta) = \theta_2 + \frac{\theta_1 - \theta_2}{1+(x/\theta_3)^{\theta_4}}$**, and sometimes, this can be combined with the above Logistic Model to produce** $\pi = \theta_2 + \frac{\theta_1 - \theta_2}{1+(x/\theta_3)^{\theta_4}}$ **(see next page)**
- **In bioassay, researchers often focus on relative potency, wherein two parallel Logistic Regression Models are fit – for example for two viruses (WT and Mutant) or two different peptides – and where the relative potency parameter is the ratio of the two EC50's:** $\rho = \frac{\gamma_M}{\gamma_{WT}}$



Design points (dose, mg):	**5.00**	**8.77**	**27.79**	**200.00**
Sample sizes:	**39**	**45**	**56**	**140**



Relative Potency Assessment of Two Peptides
Neurotensin
Somatostatin
Percent Dead
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.10
1.00
10.00
100.00
Peptide Dose



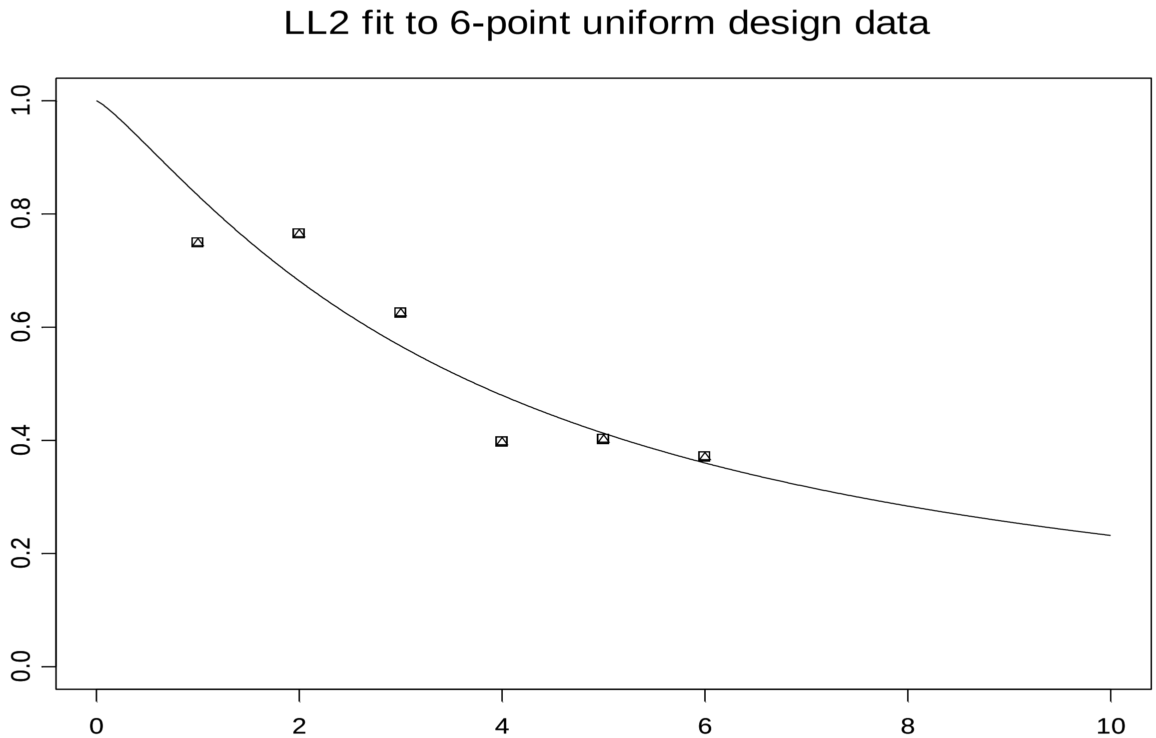
Relative Potency Assessment of Two Peptides
with D-optimal Design Points
Neurotensin
Somatostatin
Percent Dead
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.10
1.00
10.00
100.00
Peptide Dose

To illustrate a nonlinear model, the homoskedastic two-parameter log-logistic ($LL2$) model function is written

$$\eta(x, \vec{\theta}) = \frac{1}{1+t} = \frac{1}{1 + (x/\theta_2)^{\theta_3}}$$

In this expression, θ_2 is the LD_{50} parameter and θ_3 is the slope parameter. Here, the 'upper asymptote' – the expected response for $x = 0$ – is one; the $LL3$ model puts θ_1 in place of "1" in the numerator so that the upper asymptote would then need to be estimated as well.

Here is a graph of some data and the fitted $LL2$ curve.



For these data, the LD_{50} is estimated to be 3.740 and the slope is estimated to be 1.217.

Some [rival] models can be complex – e.g., AIDS models:

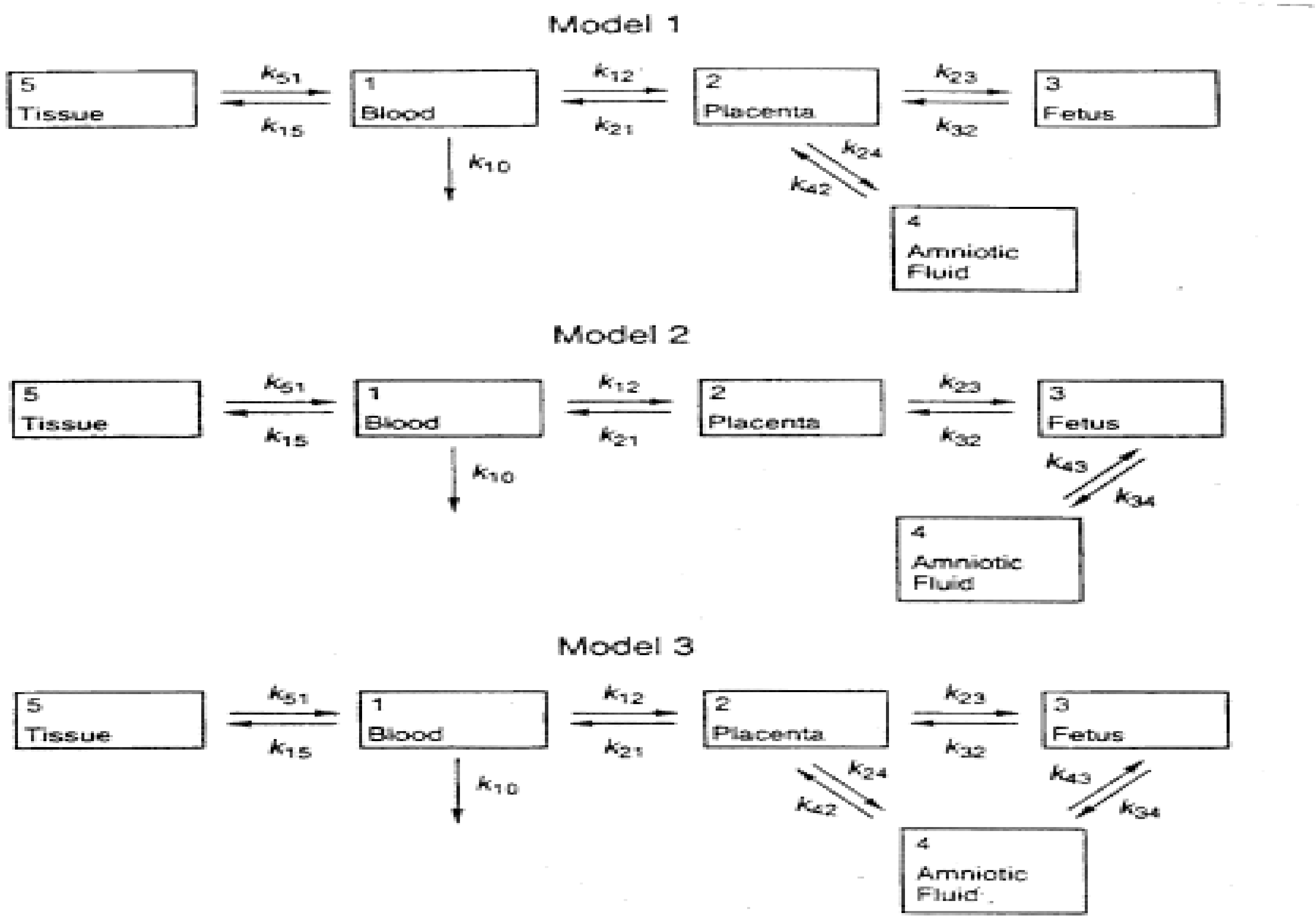


Figure 1—Schematic representation of pharmacokinetic models 1, 2, and 3 for the distribution of AZT to placenta, fetus, amniotic fluid, and tissue after intravenous administration to pregnant rats.

Model 3 (eqs 11−15) incorporated bidirectional transfer between maternal plasma ⇌ placenta, placenta ⇌ fetus, placenta ⇌ amniotic fluid, fetus ⇌ amniotic fluid, and maternal plasma ⇌ tissue compartment.

$$\frac{dC_1}{dt} = \frac{k_{21}V_p}{V_b}C_2 + \frac{k_{51}V_t}{V_b}C_5 - (k_{12} + k_{15} + k_{10})C_1 \qquad (11)$$

$$\frac{dC_2}{dt} = \frac{k_{12}V_b}{V_p}C_1 + \frac{k_{32}V_f}{V_p}C_3 + \frac{k_{42}V_a}{V_p}C_4 - (k_{21} + k_{23} + k_{24})C_2 \qquad (12)$$

$$\frac{dC_3}{dt} = \frac{k_{23}V_p}{V_f}C_2 + \frac{k_{43}V_a}{V_f}C_4 - (k_{32} + k_{34})C_3 \qquad (13)$$

$$\frac{dC_4}{dt} = \frac{k_{24}V_p}{V_a}C_2 + \frac{k_{34}V_f}{V_a}C_3 - (k_{42} + k_{43})C_4 \qquad (14)$$

$$\frac{dC_5}{dt} = \frac{k_{15}V_b}{V_t}C_1 - k_{51}C_5 \qquad (15)$$

- **Compartmental models are [multivariate] nonlinear models defined by a system of differential equations, containing parameters to be estimated.**

- **In the Binomial Logistic case above, the outcome variable had 2 choices (S and F). When it has > 2 outcomes, there are at least *four* model function choices (all with the Multinomial distribution) and each in the multi-category logit (MCL) class of models:**

Continuation Ratio A (CRA) Logit	**Un-Proportional Odds (UPO) Logit**
$\begin{cases} \log\left(\dfrac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x \\ \log\left(\dfrac{\pi_1 + \pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$	$\begin{cases} \log\left(\dfrac{\pi_1}{\pi_2 + \pi_3}\right) = \alpha_1 + \beta_1 x \\ \log\left(\dfrac{\pi_1 + \pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$
Adjacent Category (AC) Logit	**Continuation Ratio B (CRB) Logit**
$\begin{cases} \log\left(\dfrac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x \\ \log\left(\dfrac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$	$\begin{cases} \log\left(\dfrac{\pi_1}{\pi_2 + \pi_3}\right) = \alpha_1 + \beta_1 x \\ \log\left(\dfrac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$

Pregnant Mice Example (data here and used below)

Response variable is: Non-live, Malformation, Normal

Concentration (mg/kg/day)	0	62.5	125	250	500
n_i	297	242	312	299	285
Non-live	15	17	22	38	144
Malformation	1	0	7	59	132
Normal	281	225	283	202	9

B. Confidence Regions/Intervals Review

For Normal Linear and Nonlinear Models:

- **Wald CR (WCR): with $V_{n\times p}$ = Jacobian matrix**

$$\left\{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)^T \widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \leq \boldsymbol{p}\boldsymbol{s}^2 \boldsymbol{F}_\alpha\right\}$$

- **Likelihood Ratio CR (LRCR):**

$$\left\{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \boldsymbol{S}(\boldsymbol{\theta}) - \boldsymbol{S}\left(\widehat{\boldsymbol{\theta}}\right) \leq \boldsymbol{p}\boldsymbol{s}^2 \boldsymbol{F}_\alpha\right\}$$

where $\boldsymbol{S}(\boldsymbol{\theta}) = \left(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})\right)^T \left(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})\right) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$

- **Connection: if** $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta}) \approx \boldsymbol{\eta}\left(\boldsymbol{x}, \widehat{\boldsymbol{\theta}}\right) + \widehat{\boldsymbol{V}}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)$**, then**

$$\boldsymbol{\varepsilon} \approx \widehat{\boldsymbol{\varepsilon}} - \widehat{\boldsymbol{V}}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right), \boldsymbol{S}(\boldsymbol{\theta}) \approx \boldsymbol{S}\left(\widehat{\boldsymbol{\theta}}\right) + \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)^T \widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)$$

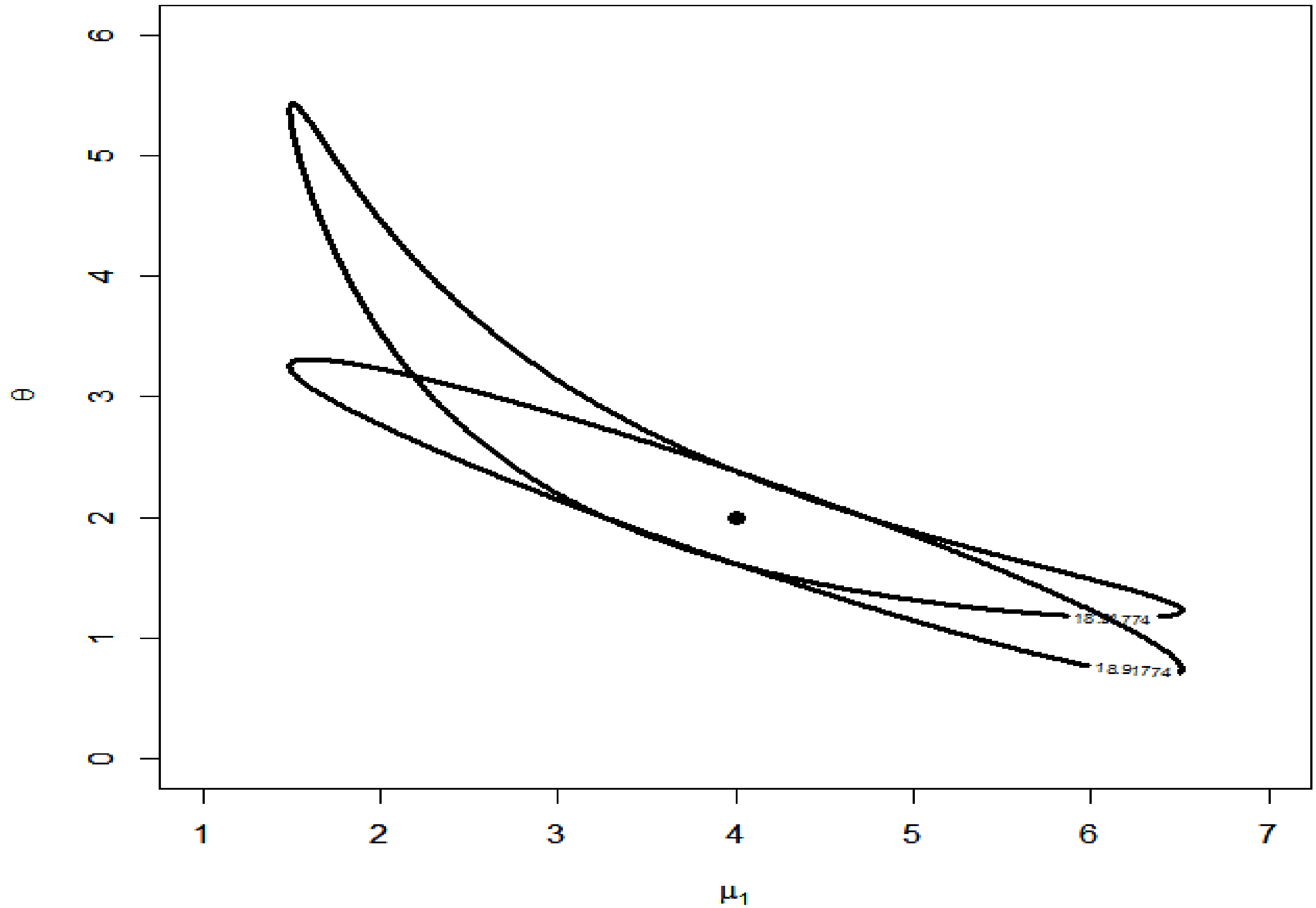
- **For non-Normal (asymptotic result) we use:**

$$\left\{\boldsymbol{\theta} \in \boldsymbol{\Theta}: 2\left[\boldsymbol{LL}(\boldsymbol{\theta}) - \boldsymbol{LL}\left(\widehat{\boldsymbol{\theta}}\right)\right] \leq \chi^2_\alpha\right\}$$

Confidence intervals (CIs) are obtained by projection

Illustration for a Normal nonlinear model – WCR is elliptical

95% WCR (ellipse) and LRCR for FC Example



C. Optimal Design Topics

- **For Normal Linear Models, the (Fisher) Information is $M(\xi) = X^T X$ – since variance is proportional to $(X^T X)^{-1}$ – and for Normal Nonlinear Models this is $M(\xi, \theta) = V^T V$; D-optimal designs minimize the area/volume of the Wald confidence region**
- **An n-point design is**

$$\xi = \begin{Bmatrix} x_1 & x_2 & \cdots & x_n \\ \omega_1 & \omega_2 & \cdots & \omega_n \end{Bmatrix}$$

Here the ω_k are non-negative 'design weights' which sum to one; the x_k which may indeed be vectors (but called 'design points'), belong to the design space, and are *not necessarily distinct*

- **For the Normal case and model function $\eta(x,\theta)$, the $n \times p$ Jacobian matrix is $V = \frac{\partial\eta}{\partial\theta}$ and the $p \times p$ information matrix is $M(\xi,\theta) = V^T\Omega V$, with $\Omega = diag\{\omega_1,\omega_2,\ldots,\omega_n\}$**
- **For the non-Normal case, $M(\xi,\theta) = -E\left[\frac{\partial^2 LL}{\partial\theta\partial\theta^T}\right]$**
  The first-order/asymptotic variance of the LS estimator of θ, $\hat{\theta}$, is proportional to $M^{-1}(\xi,\theta)$, so designs are often chosen to minimize some convex function of M^{-1}. Designs which minimize its determinant are called D-optimal; those that minimize its trace are A-optimal.
- **Variance function – the (first-order) variance of the predicted response at $X = x$ is given by**

$$d(x,\xi,\theta) = \left(\frac{\partial\eta(x,\theta)}{\partial x}\right)^T M^{-1}(\xi,\theta)\frac{\partial\eta(x,\theta)}{\partial x}$$

- **Designs that minimize (over ξ) the maximum (over x) of $d(x, \xi, \theta)$ are called G-optimal.**
- **The General Equivalence Theorem (GET) of Kiefer and Wolfowitz (1960) proves that D- and G-optimal designs are equivalent, that the variance function evaluated using the D-/G-optimal design does not exceed the line $y = p$ – but that it will exceed this line for *all* other designs. A corollary of the GET establishes that the maximum of the variance function is achieved for the D/G-optimal design at the support points of this design.**
- **For the Binomial Logistic model with $\alpha = 0, \beta = 1$, the optimal design is $\begin{Bmatrix} -1.5434 & 1.5434 \\ 1/2 & 1/2 \end{Bmatrix}$.**

This design is problematic since it has only 2 support points.

D. Two Examples from Toxicology

- **Pregnant Mice ($n = 1435$) from Price *et al* 1987, *Fund. Appl. Toxicology*, and Agresti p.192; chosen design is**

$$\xi = \begin{Bmatrix} \mathbf{0} & \mathbf{62.5} & \mathbf{125} & \mathbf{250} & \mathbf{500} \\ \frac{\mathbf{297}}{\mathbf{1435}} & \frac{\mathbf{242}}{\mathbf{1435}} & \frac{\mathbf{312}}{\mathbf{1435}} & \frac{\mathbf{299}}{\mathbf{1435}} & \frac{\mathbf{285}}{\mathbf{1435}} \end{Bmatrix}$$

Outcome variable Y: dead, malformed, normal; x is concentration of ether in mg/kg per day

- **Emergence of House Flies ($n = 3500$) from Itepan 1995 and Zocchi & Atkinson 1999 *Biometrics*; chosen design is**

$$\xi = \begin{Bmatrix} \mathbf{80} & \mathbf{100} & \mathbf{120} & \mathbf{140} & \mathbf{160} & \mathbf{180} & \mathbf{200} \\ \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} & \frac{\mathbf{500}}{\mathbf{3500}} \end{Bmatrix}$$

Y: dead, opened but died before complete emergence, complete emergence; x is dose of gamma radiation

E. Robust Design Strategies for Toxicology Examples

Local optimal design strategies:

- **PO model: Perevozskaya, Rosenberger, Haines, 2003**
- **CRA model: Fan & Chaloner, 2001**
- **CRB model: Zocchi & Atkinson, 1999**

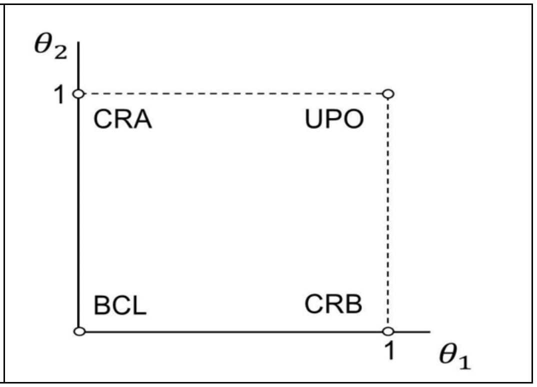
If model has 4 parameters, these designs often have ≤ 4 support points and are only 'optimal' for the assumed model.

For Pregnant Mice example, these designs are:

$\xi_{CRA}^{*} = \begin{Bmatrix} 194.5 & 428.1 & 1682.0 \\ 0.3023 & 0.4531 & 0.2445 \end{Bmatrix}$	$\xi_{UPO}^{*} = \begin{Bmatrix} 0 & 353.2 & 678.2 \\ 0.3575 & 0.4066 & 0.2359 \end{Bmatrix}$
$\xi_{AC}^{*} = \begin{Bmatrix} 193.5 & 425.5 & 1554.8 \\ 0.3037 & 0.4527 & 0.2435 \end{Bmatrix}$	$\xi_{CRB}^{*} = \begin{Bmatrix} 222.6 & 401.3 & 767.9 \\ 0.4058 & 0.3805 & 0.2136 \end{Bmatrix}$

- Often need a way to find near-optimal designs with extra support points. One way to do this is via model nesting as suggested by Atkinson 1972; in the current case, note that each of the UPO, CRA, AC and CRB models are nested within the Generalized Ordinal Logit (GOL) model introduced and discussed in Jamroenpinyo *et al* 2012.
- The GOL model with 3 categories is written as follows:

$$\begin{cases} \log\left(\dfrac{\pi_1}{\pi_2 + \theta_1\pi_3}\right) = \alpha_1 + \beta_1 x \\ \log\left(\dfrac{\theta_2\pi_1 + \pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$$



- This model has 6 different parameters, including 2 hyper-parameters (nuisance parameters: θ_1 and θ_2)

- **The 'distance' between two designs (ξ and ξ^*) can be measured by the so-called D-efficiency**

$$D_{EFF} = \left(\frac{|M(\xi)|}{|M(\xi^*)|} \right)^{1/p}$$

- **In this expression, $M(\cdot)$ is the $p \times p$ information matrix associated with the given model and p is the number of model parameters (e.g., $p = 4$ for the UPO, CRB models)**
- **The reason for the exponent is to adjust for the fact that $|M(\xi)|$ increases with the dimension (p)**
- **Can think of D_{EFF} in terms of sample size: if $D_{EFF} = 0.80$, since $\frac{1}{D_{EFF}} = 1.25$, then sample size if we use design ξ needs to be 25% higher versus had we used the design ξ^* (e.g., $n = 50$ instead of $n = 40$).**

- **For Normal nested models, let $V_{n\times p} = [V_1|V_2]$ where $n \times p_1$ V_1 corresponds to the sub-model and $n \times p_2$ V_2 corresponds to the extra parameters ($p = p_1 + p_2$), then**

$$M = V^T \Omega V = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

- **Note that $|M| = |M_{11}|\left|M_{22} - M_{21}M_{11}^{-1}M_{12}\right|$**
- **New objective function (to maximize over ξ):**

$$\psi(\xi) = \frac{1-\phi}{p_1}\log|M_{11}| + \frac{\phi}{p_2}\log\left|M_{22} - M_{21}M_{11}^{-1}M_{12}\right|$$

- **In this expression, $\phi = 0$ corresponds to just the sub-model, $\phi = 1$ corresponds to only the additional terms, and $\frac{1-\phi}{p_1} = \frac{\phi}{p_2}$ or $\phi = \frac{p_2}{p}$ corresponds to all parameters.**

F. Robust Strategy Applied to Toxicology Examples

Pregnant Mice: For this example, best fitting model is CRB ($-LL = 730.4$) and second is UPO ($-LL = 743.5$), but note from above the ODEs for these two models are quite different (e.g., latter includes 0 but not former). Also, D-efficiency of the original chosen design (with respect to ξ^*_{CRB}) is only 62.8%. Yet, the ODE has only 3 support points and little/no ability to test for lack of fit.

To provide for LOF, we now nest the CRB model in the UPOCRB (which is GOL with $\theta_1 = 1$) and with the CRB parameter estimates and $\theta_2 = 0$. For $\phi = 0.05$ ($D_{EFF} = 95.3\%$), the ODE is

$$\xi = \begin{Bmatrix} 0 & 230.5 & 405.8 & 760.9 \\ 0.0856 & 0.3635 & 0.3572 & 0.1937 \end{Bmatrix}$$

Next, we insist on a geometric-like design with x-points of the form $0, a, ab, ab^2, ab^3$ and weights $\omega^*, \frac{1-\omega^*}{4}, \frac{1-\omega^*}{4}, \frac{1-\omega^*}{4}$, and $\frac{1-\omega^*}{4}$. The "weighted" optimal design again with $\phi = 0.05$ ($D_{EFF} = 90.7\%$) is $a = 169.7, b = 1.58, \omega^* = 0.064$

$$\xi = \begin{Bmatrix} 0 & 169.7 & 268.6 & 424.9 & 672.3 \\ \frac{}{}91 & 336 & 336 & 336 & 336 \end{Bmatrix}$$

Finally, if we nest the CRB model in the GOL, and search for a geometric-like design as above, the "weighted" ODE with $\phi = 0.10$ ($D_{EFF} = 90.6\%$) is $a = 160.2, b = 1.65, \omega^* = 0.054$ is

$$\xi = \begin{Bmatrix} 0 & 160.2 & 264.2 & 435.8 & 718.9 \\ 79 & 339 & 339 & 339 & 339 \end{Bmatrix}$$

Flies: These data show significant nonlinearity, so we fit quadratic models – for the CRA model, this is

$$\begin{cases} \log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x + \gamma_1 x^2 \\ \log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x + \gamma_2 x^2 \end{cases}$$

The best fitting (quadratic) model is CRA ($-LL = 1782.04$) and second is AC ($-LL = 1782.15$), but all $-LL$ values very similar (fits are similar). Over $[8, 20]$, the ODE's are:

$\xi^*_{CRA} = \begin{Bmatrix} 8 & 12.52 & 16.36 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$	$\xi^*_{AC} = \begin{Bmatrix} 8 & 12.49 & 16.31 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$

Since all fits and designs are similar, we can just use the design for the CRA model (i.e., no nesting needed here). Next, consider equal-weight uniform designs of the form A, $A + B$, $A + 2B$, … $A + 6B$. In this case, the D-efficiency for the optimal *uniform* design exceeds that of the optimal *geometric* design, and we obtain the "weighted" seven-point design ($D_{EFF} = 92.6\%$) with $A = 8$, $B = 1.58$:

$$\xi = \begin{Bmatrix} 8 & 9.58 & 11.16 & 12.75 & 14.33 & 15.91 & 17.49 \\ 500 & 500 & 500 & 500 & 500 & 500 & 500 \end{Bmatrix}$$

This design represents an improvement over the chosen seven-point design (with $A = 8$ and $B = 2$), which results in a value $D_{EFF} = 84.1\%$).

G. General Robust Design Strategies

We have introduced the following design strategies:

- **G.1. Q-optimality approach**

In contrast with D-optimality, Hamilton and Watts (1985) use a quadratic approximation to show that for an exact design the volume of the above likelihood-based confidence region is approximately equal to

$$\boldsymbol{\nu} = \boldsymbol{c}\left|\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}\right|^{-\frac{1}{2}}|\boldsymbol{D}|^{-\frac{1}{2}}\{\mathbf{1} + \boldsymbol{k}^{2} \times \boldsymbol{tr}(\boldsymbol{D}^{-1}\boldsymbol{C})\}$$

Here, '$\boldsymbol{c}$' and '$\boldsymbol{k}$' are constants relative to the design, $\boldsymbol{C}$ is a function of the parameter-effects curvature and $\boldsymbol{D}$ measures the intrinsic curvature in the direction of the residual vector; we write $\boldsymbol{D} = \boldsymbol{I}_{\boldsymbol{p}} - \boldsymbol{B}$ and $\boldsymbol{B} = \boldsymbol{L}^{\boldsymbol{T}}[\boldsymbol{e}^{\boldsymbol{T}}][\boldsymbol{W}]\boldsymbol{L}$. Claiming that this

volume approximation cannot be used as a design criterion since the residual vector (e) is not known at the design stage, Hamilton & Watts suggest obtaining designs to minimize the further volume approximation

$$v' = c\left|V^T V\right|^{-1/2}\{1 + k^2 \times tr(C)\}$$

These designs are called Q´-optimal. Hamilton & Watts also note that for all their examples, the (local) Q´-optimal designs have only $n = p$ support points.

Since the residual vector is always orthogonal to the tangent plane at the least-squares estimate, O'Brien (1992) points out that the residual vector can be written $e = N\alpha$; from the QR decomposition we write $V = QR = [U|N]R = UL^{-1}$. Thus, when we have the sample size $n = p + 1$, α is a scalar, and we make the expected squared length of e equal σ^2.

We thus seek (local) Q-optimal designs which minimize the original second-order volume approximation (ν) above.
To illustrate, we use the two-parameter intermediate product ($IP2$) model function given by the expression

$$\eta(x, \vec{\theta}) = \frac{\theta_1}{\theta_1 - \theta_2}\{e^{-\theta_2 x} - e^{-\theta_1 x}\}$$

This model can be reparameterized in three ways, and the corresponding 3-point local Q-optimal designs follow:

Table 1. *Quadratic optimal designs for varying parameterizations and noise levels*

σ	Original	Ratio	Logarithm	Peak
0·00	(1·23, 6·86, 6·86)	(1·23, 6·86, 6·86)	(1·23, 6·86, 6·86)	(1·23, 6·86, 6·86)
0·025	(1·21, 6·41, 7·08)	(1·22, 6·37, 7·06)	(1·23, 6·51, 7·17)	(1·25, 6·55, 7·21)
0·05	(1·15, 5·79, 7·11)	(1·20, 5·60, 7·09)	(1·25, 6·11, 7·43)	(1·30, 6·28, 7·59)
0·075	(1·09, 5·17, 7·01)	(1·16, 4·58, 6·97)	(1·26, 5·67, 7·66)	(1·38, 6·02, 7·97)
0·10	(1·02, 4·72, 6·81)	(1·09, 3·59, 6·55)	(1·28, 5·12, 7·84)	(1·47, 5·69, 8·32)

Q-optimal designs also depend upon the assumed value of σ.

We use the term 'local' here since this design is chosen using the *a priori* estimates $\theta_1 = 0.7$ and $\theta_2 = 0.2$; the Bayesian strategy (discussed below) extends this strategy.

The above was extended in O'Brien *et al* (2010) in two ways: (1) to incorporate exact designs for $n = p + s$ support points ($s > 1$) and (2) to incorporate continuous designs. In these cases, Q-optimal designs minimize the expected volume, $E[\nu(\phi)]$, and use polar or spherical coordinates. Design algorithms used to obtain these designs are computationally intensive: we use the GAUSS and SAS/IML programming languages and associated minimization routines.

For the above $IP2$ model function, 4-point exact Q-optimal designs are easily obtained; designs with a greater number of support points 'collapse' to just 4 support point designs.

- **G.2. Discrimination-Estimation approach**

Suppose that we have a given model function in mind, and can obtain a reasonable [non-nested] rival model function (or maybe several). To illustrate, the $SE1$ model function, $\eta_1 = e^{-\theta_1 x}$, may model a given process; a good rival model function is the $MM1$ model function, $\eta_2 = \theta_2/(\theta_2 + x)$.

In general, for the class of m rival model functions with respective parameter numbers $p_1, p_2, \ldots p_m$, the estimation measure is

$$E(\xi) = \sum (\pi_k/p_k) log|M_k(\xi)|$$

The weights $\pi_1, \pi_2, \ldots \pi_m$ (which sum to one) control the emphasis placed upon each of the m rival models. O'Brien & Rawlings (1996) introduces an analogous term so-called discrimination measure, $D(\xi)$, and these two terms are

combined into the single estimation-discrimination measure

$$\boldsymbol{B(\xi) = \alpha \times E(\xi) + (1-\alpha) \times D(\xi)}$$
$$\boldsymbol{= b \times log|M(\xi)| + \sum c_k \times log|M_k(\xi)|}$$

We also provide a GET; $\boldsymbol{D_B}$-optimality of an estimation-discrimination design can then be verified. For example, the following graph confirms optimality as it lies below $\boldsymbol{y = 1}$.

T.E. O'Brien, J.O. Rawlings / Journal of Statistical Planning and Inference 55 (1996) 77–93 85

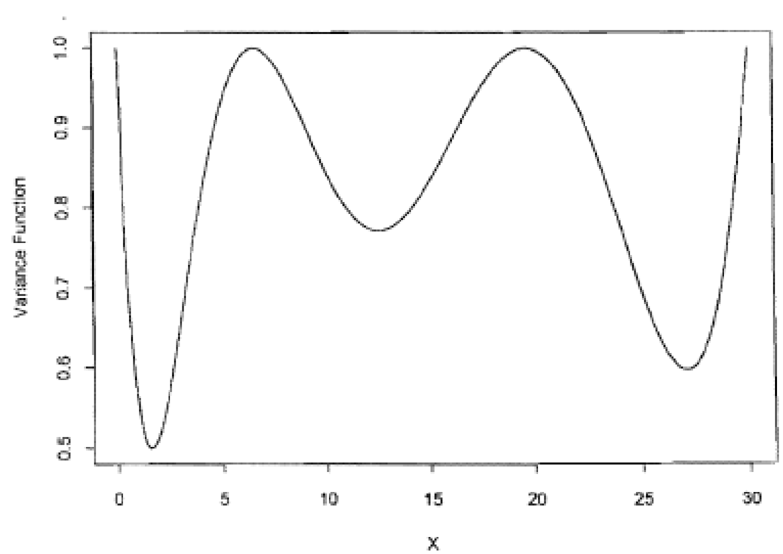


Fig. 1. Variance function for Example 2.

- **G.3. Model Nesting approach**

In the context of linear models, suppose that we have faith that the true model is $\boldsymbol{\eta_A = F_1\beta_1}$ but we feel that a better model may be the larger 'super-model' $\boldsymbol{\eta_B = F_1\beta_1 + F_2\beta_2}$. We seek a design that is efficient for $\boldsymbol{\eta_A}$ but that be used to check for departures "in the direction of" $\boldsymbol{\eta_B}$. To illustrate, we might feel that a quadratic two-factor RSM or mixture model may well fit a given process, but may want to check if cubic term(s) may be significant. In this case, a compound design criterion function (objective function) is

$$\boldsymbol{\Phi(\xi) = \frac{\kappa}{r} log|M_{11}| + \frac{1-\kappa}{s} log\left|M_{22} - M_{21}M_{11}^{-1}M_{12}\right|}$$

In this expression, the sub-model contains 'r' parameters, the super-model contains 's' additional parameters, and κ (which lies between 0 and 1) controls the emphasis placed on the

original versus the additional parameters. Again, D-optimality is confirmed by plotting the corresponding variance function plot and noting whether or not the graph exceeds the relevant horizontal line.

Nonlinear cases are more complicated, so we restrict our attention to just sigmoidal models, and these are of the logistic/Richards, Weibull, and log-logistic families. A super-model which generalizes the latter two families is the three-parameter Eclectic (EC3) model function

$$\eta(x, \theta) = \frac{1}{\left(1 + \frac{(x/\theta_2)^{\theta_3}}{\theta_5}\right)^{\theta_5}}$$

We obtain the LL2 model function when $\theta_5 = 1$, and the WEIB2 model for $\theta_5 \to \infty$. The strategy here is analogous to

the above linear case since we are using D-optimality (linear approximation), but since the models are nonlinear, we use either a ‘local’ approach or a Bayesian one. For example, we can find a three-point design with high efficiency for the $LL2$ model but which protects for departures in the direction of all other sigmoidal functions. Basic results are given in O’Brien (1994) and extensions in O’Brien *et al* (2009).

G.4. A General Departures approach

In our 1995 publication, we took another robust design approach akin to space-filling designs, which were then being explored by Randy Tobias at SAS. We focused instead on the general replicated-point LOF (lack of fit) test from an assumed model function, but in the direction of *any* departure from this model. Similar results are given a 2008 JSPI article by Bogacka *et al*.

We proposed the following three-step approach:

- **The D-optimal design (ξ_D) is obtained (often with only p support points) and the variance function is obtained;**
- **To this D-optimal design are added (perhaps on a second replication, perhaps not) the (t) points where the variance**

function cuts the line $y = p\left\{\left[\frac{(p+1)\delta}{p}\right]^{p} - 1\right\}$ with say $\delta = 0.90$ and $p = \#$ of model function parameters);

- **Choose as the final design r_1 replicates of the p D-optimal support points and r_2 replicates of the t support points from the previous step.**

The justification for the above strategy as follows: Let ξ_x represent the one-point design which puts all weight at the one support point x, and suppose that the minimal D-optimal design (ξ_D) is D-optimal. Then the design

$$\xi_N = \frac{p}{p+1}\xi_D + \frac{1}{p+1}\xi_x$$

associates the weight $1/(p+1)$ with each of the p D-optimal support points and x. A measure of the 'distance' between ξ_D and ξ_N is the D-efficiency, $\delta = \left[\frac{|M(\xi_N)|}{|M(\xi_D)|}\right]^{1/p}$

In this case, this is equal to $\frac{p}{p+1}\left[1 + \frac{1}{p} d(x, \xi_D)\right]^{1/p}$.

When this is solved, we obtain the above equation for cutting the variance function.

To illustrate, consider again the $IP2$ model function given in §E.1 again with the prior estimates $\theta_1 = 0.7$ and $\theta_2 = 0.2$; in this case, the local D-optimal support points are $x = 1.229$ and $x = 6.858$, which are to be replicated r_1 times. With $p = 2$ and $\delta = 0.9$, the above cut-line is $y = 1.645$. This gives the $t = 4$ additional points $x = 0.761, 1.909, 4.890, 9.366$ (to be replicated r_2 times). In terms of final efficiencies, note that for $r_1 = r_2 = 1$, the final D-efficiency is 88%, meaning that only 12% efficiency has been sacrificed in order that we can test for LOF of the assumed model function.

In general, final D-efficiencies can be obtained by the expression

$$DE_F = \frac{(r_1 p)^{1-t/p}}{r_1 p + r_2 t} \left| r_1 p I_t + r_2 D(\vec{x}, \xi_D, \hat{\theta}) \right|^{1/p}$$

H. Bayesian and Geometric Design Strategies

Further robust design strategies for nonlinear models include:

- **H.1. <u>A Bayesian approach</u>**

Local designs for nonlinear model suffer from the criticism that they are only *efficient* for the required *a priori* guess of (some or all of) the model parameters. Several authors have explored how to obtain designs that are '*efficient*' regardless of the parameter values. Although several alternatives are possible and important, when we have at our disposal a prior distribution of parameter values p(θ), a reasonable strategy is to obtain designs to minimize the expected log-generalized variance,

$$E\{log\left|M^{-1}(\xi,\theta)\right|\} = \int log\left|M^{-1}(\xi,\theta)\right| p(\theta)\, d\theta$$

References include Chaloner & Larntz (1989) and Atkinson *et al* (2007, Chap.18); these sources point out that the above is only one of several criteria function which may be chosen. In general, Bayesian optimal designs have at least p support points (often more), and the number of support points usually increases with the dispersion of the assumed prior. Of course, like 'local' designs, Bayesian designs suffer from criticism too: (a) obtaining a reasonable prior is often difficult in practice, and (b) these designs are often not robust to the assumed prior distribution.

- **H.2. <u>Strategies for Geometric and Uniform Designs</u>**

Geometric designs of the form: $x_1 = a, x_2 = ab, x_3 = ab^2 \ldots x_k = ab^{k-1}$ (as well as uniform designs): optimal choices for a and b are explored in O'Brien *et al* (2009b).

I. Conclusion and Comments

- **Designs that are "optimal" only for the assumed model are obviously of limited use**
- **Often several models can fit a given dataset/situation, and we need the ability to judge which is 'best'**
- **Robust designs – such as those obtained by model nesting (provided one can find an adequate larger model) – are often more useful in practice**
- **Could also consider robustness to the chosen parameter values, link function, etc.**
- **Future work: robust optimal design useful to test for drug or similar compounds synergy, and developing easy-to use computational tools for the practitioner.**

Thank You!