Optimal design for generalised linear mixed models

Tim Waterhouse and John Eccleston

School of Physical Sciences, University of Queensland

Motivation

- Pharmacodynamic study
  - Drug with binary response
  - Need to design cross-over trials to test effect of various dose levels
  - Logistic regression
  - D-optimal designs
- What about subject effect?
- Include random effects: GLMM

Overview

- Need information matrix for logistic regression model with random coefficients
- Longford (1994) presents an approximation based on Taylor expansion of log-likelihood, in the context of estimation
- Numerical investigation showed large discrepancies
- Aim: to develop a more accurate approximation and investigate its use in optimal design

Generalised linear model

- Response vector: Y, $\mathrm{E}(Y_j) = \mu_j$
- Linear predictor: $\eta = X\theta$
- Link function: $\eta_j = g(\mu_j)$
- eg. Logistic regression: Bernoulli response, logit link function

$$\eta_j = \mathrm{logit}\{P(y_j = 1)\} = \log\left\{\frac{p_j}{1 - p_j}\right\}$$

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} \xi_1 & \xi_2 & \dots & \xi_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}$$

- Exact designs: where $w_i = r_i/N$, with r_i replications at ξ_i
- Approximate designs: where weights $w_i \in [0, 1]$ sum to 1, represent proportion of experimental effort at each point

Optimal design

- Given the form of the model, how do we best select design ξ to efficiently estimate parameter vector θ?
- Cramer-Rao Lower Bound:
  Var$(\hat{\theta}) \geq$ (Fisher information matrix)$^{-1}$
- D-optimality : maximise determinant of Fisher information matrix

D-optimality

- Fisher information matrix:

$$M_F(\theta,\,\xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\theta;y)}{\partial\theta\partial\theta^T}\right]$$

- For nonlinear models (including logistic regression), $M_F(\theta,\,\xi)$ involves unknown θ
- So D-optimal design depends on θ, such designs are *locally* optimal

Logistic regression

- Fixed effects model:

$$\eta_j = \text{logit}\{P(y_j = 1)\} = \log\{p_j/(1-p_j)\} = x_j\theta$$

so that

$$\eta = X\theta$$

- Fisher information matrix:

$$M_F(\theta, \xi) = X^T W X$$

where

$$W = \text{diag}\{w_j p_j (1-p_j)\}$$

Logistic regression example

- Model (Atkinson & Haines, 1996):

$$\eta = \mathrm{logit}\{p\} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$(x_1, x_2) \in [-1,\ 1] \times [-1,\ 1]$

- Locally D-optimal design for $\theta = (3, 3, 1)^{\mathrm{T}}$:

$$\xi = \left\{ \begin{array}{ccc} (-1, -1) & (-1, 1) & (-0.068, -1) \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}$$

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$
- Alternatively, $\theta_j = \beta + b_j$, where $b_j \sim \mathrm{N}(0, \Sigma)$
- New model, p parameters, $q \leq p$ of these random:

$$\eta_{ij} = \mathrm{logit}\{p_{ij}\} = x_{ij}\beta + z_{ij}b_j$$

  where $b_j \sim \mathrm{N}(0, \Sigma)$ is vector of length q,
  z_{ij} are the relevant subsets of vectors x_{ij}
- Assume elements of b_j are uncorrelated, i.e.
  $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_q^2\}$
- Parameters: $\psi = (\beta^{\mathrm{T}}, \sigma_1^2, \ldots, \sigma_q^2)^{\mathrm{T}}$

Log-likelihood

$$\ell = \sum_j \log \int \underset{\mathbb{R}^q}{\cdots} \int P_j(b_j)\Phi_q(b_j)\mathrm{d}b_j$$

where

$$\begin{aligned} P_j(b_j) &= \prod_{i=1}^{n_j} \{p_{ij}(b_j)\}^{y_{ij}} \{1 - p_{ij}(b_j)\}^{1-y_{ij}} \\ p_{ij}(b_j) &= \mathrm{logit}^{-1}\{x_{ij}\beta + z_{ij}b_j\} \end{aligned}$$

- Gaussian quadrature
- Taylor expansion

Taylor approximation to log-likelihood

From Longford (1994),

$$\ell \approx \sum_{j=1}^{N} \left[\log\{P_j(0)\} + \frac{1}{2} e_j^{\mathrm{T}} H_j e_j - \frac{1}{2} \log |G_j| \right]$$

where

$$\begin{aligned} G_j &= I_q + \Sigma^{1/2} Z_j^{\mathrm{T}} W_j Z_j \Sigma^{1/2}, \quad \text{and} \\ H_j &= W_j Z_j \Sigma^{1/2} G_j^{-1} \Sigma^{1/2} Z_j^{\mathrm{T}} W_j \\ W_j &= \mathrm{diag}\{w_j\}, \quad w_j = (w_{1j}, \ldots, w_{n_j j})^{\mathrm{T}} \\ w_{ij} &= p_{ij}(0)(1 - p_{ij}(0)) \end{aligned}$$

etc.

Approximate information matrix

- Longford (1994) ignores dependence of w_{ij} on on β to give

$$M_F(\psi,\, \xi_j) \approx \begin{bmatrix} A_j & 0 \\ 0 & B_j \end{bmatrix},$$

where

$$A_j = X_j^{\mathrm{T}} V_j^{-1} X_j,$$

and the $(u,\, v)$ element of B_j is

$$B_j^{(u,v)} = \frac{1}{2}\operatorname{tr}\left(V_j^{-1} \frac{\partial V_j}{\partial \sigma_u^2} V_j^{-1} \frac{\partial V_j}{\partial \sigma_v^2} \right)$$

where $V_j = W_j^{-1} + Z_j \Sigma Z_j^{\mathrm{T}}$

More precise information matrix?

- What if we acknowledge the dependence of w_{ij} on on β?
-

$$\begin{aligned} M_F(\psi, \xi_j) &\approx \begin{bmatrix} A_j & C_j \\ C_j^{\mathrm{T}} & B_j \end{bmatrix} \\ &\approx \text{10 pages of \LaTeX} \end{aligned}$$

Brute force numerical methods

- Recall $M_F(\psi, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\psi; y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right]$
- For an exact design, we can calculate the expectation by the following sum:

$$\mathrm{E}\left[\frac{\partial^2 \ell(\psi; y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right] = \sum_{y_{ij}} \frac{\partial^2 \ell(\psi; y_{ij})}{\partial\psi\partial\psi^{\mathrm{T}}} \ell(\psi; y_{ij})$$

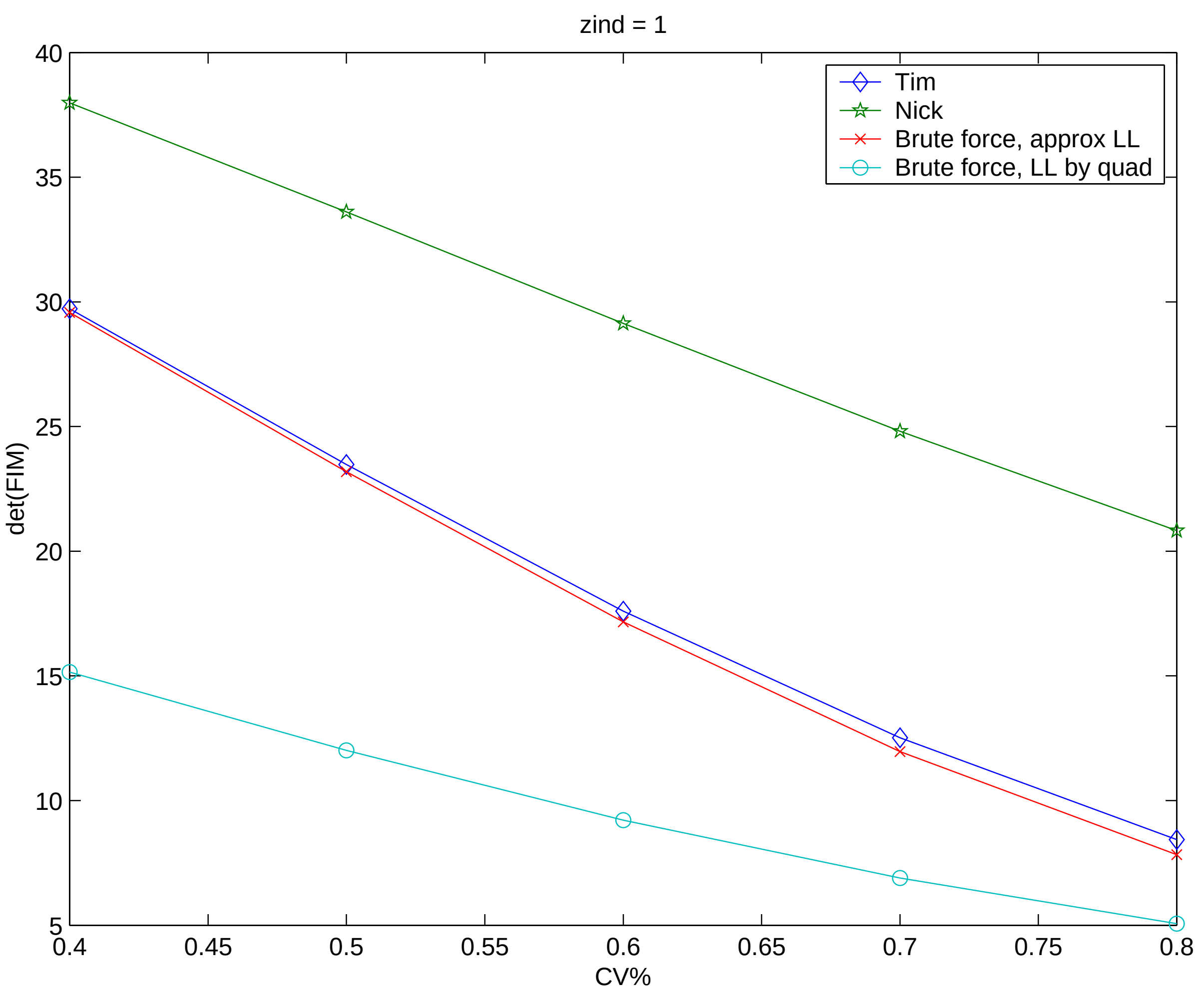
  where the derivatives are evaluated numerically
- ℓ may be evaluated
  - by Gaussian quadrature, or
  - Taylor approximation

Summary of methods

	Derivatives	Log-likelihood	w_{ij} depends on β?
Nick [Longford]	Analytic	Approx.	No
Tim [me]	Analytic	Approx.	Yes
Brute force 1	Numerical	Approx.	N/A
Brute force 2	Numerical	Quadrature	N/A

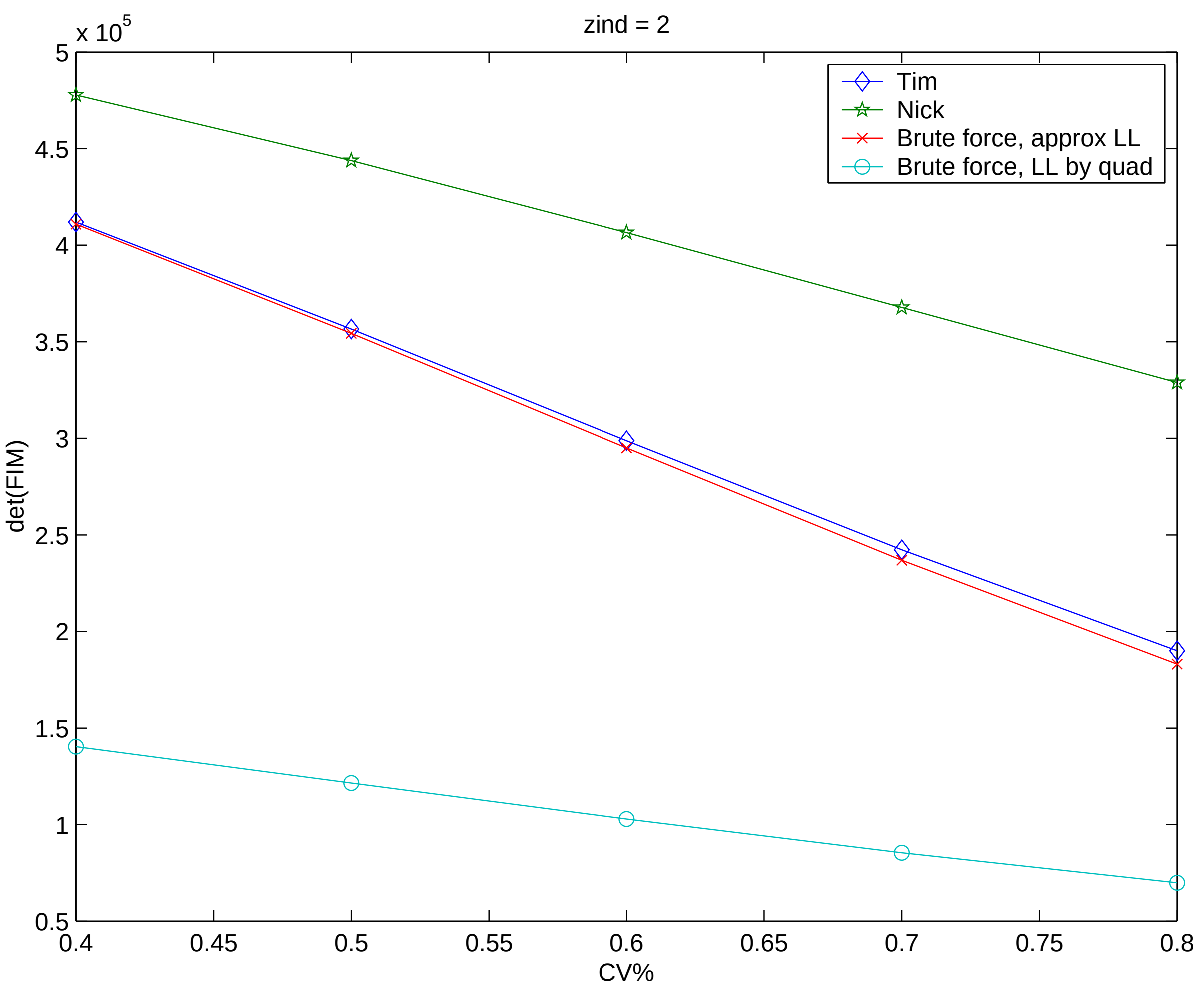
Testing. . .

$$\text{logit}\{p_j\} = (\beta_0 + b_j) + \beta_1 x_1 + \beta_2 x_2$$



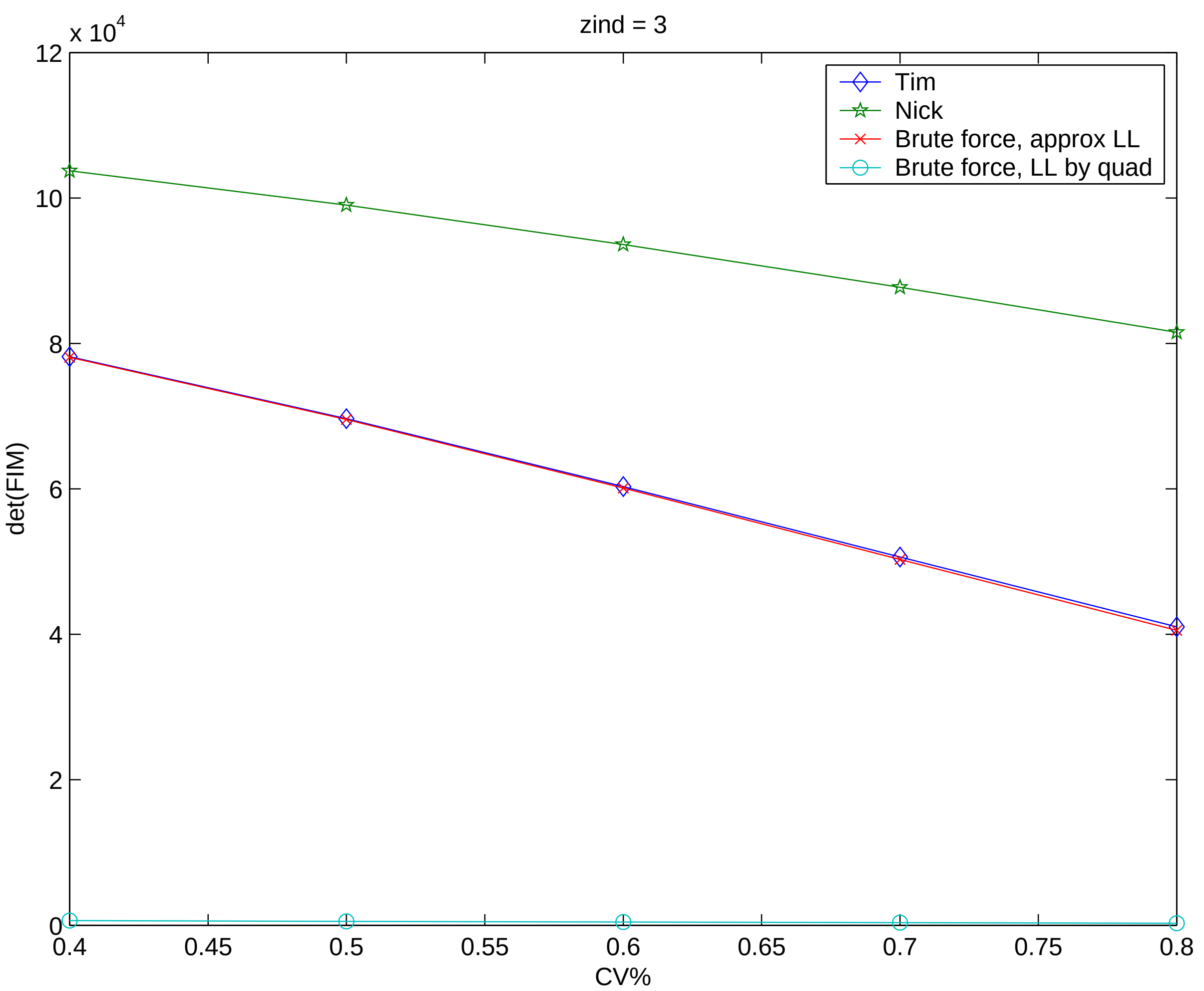
Testing. . .

$$\mathrm{logit}\{p_j\} = \beta_0 + (\beta_1 + b_j)x_1 + \beta_2 x_2$$



Testing. . .

$$\operatorname{logit}\{p_j\} = \beta_0 + \beta_1 x_1 + (\beta_2 + b_j)x_2$$



Example: Cross-over study

- Test influence of dose level on binary (success/failure) response over two periods

$$\begin{aligned} \operatorname{logit}\{P(y_{ij}=1)|\boldsymbol{\theta}_j\} &= \theta_{j0}+\theta_{j1}d_{ij}+\theta_{j2}d_{(i-1)j}, \\ & \qquad i=1,2, \quad j=1,\ldots,n \end{aligned}$$

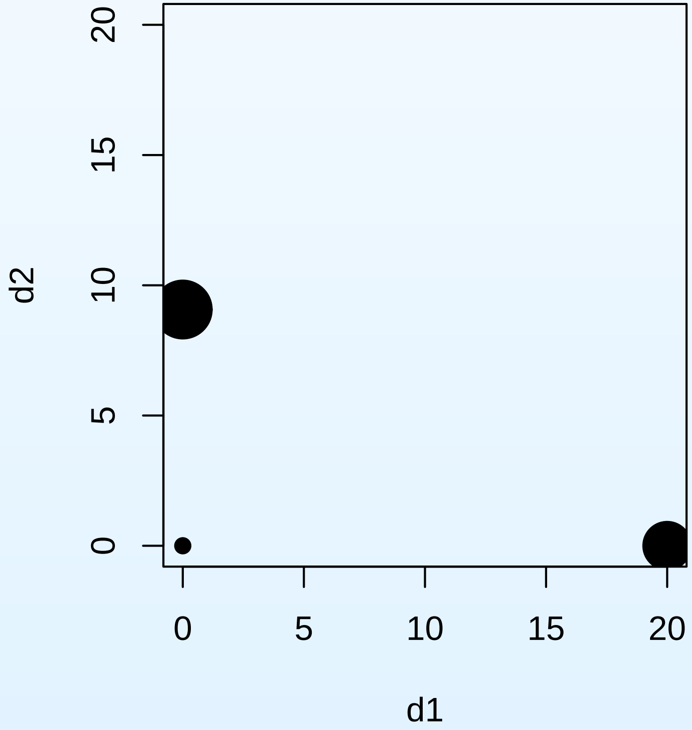
  where $d_{0j}=0$

- Dose levels may range between 0 units (placebo) and 20 units
- For locally D-optimal designs, we assume $\mathrm{E}(\boldsymbol{\theta}_j)=(-1,0.3,0.25)^{\mathrm{T}}$

Example: Cross-over study

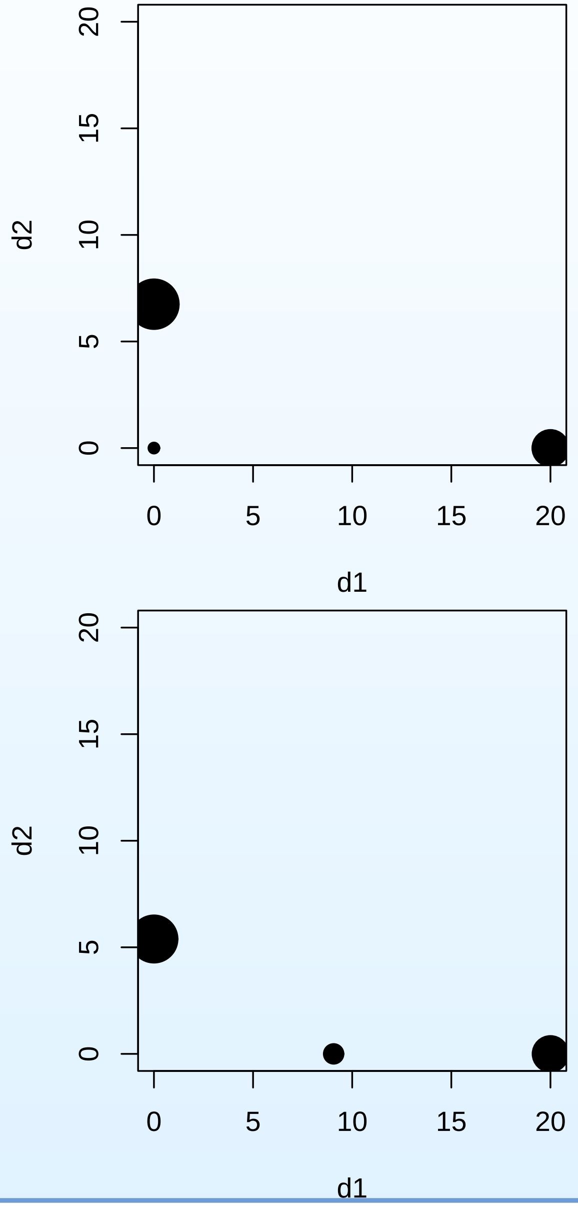
Optimal design for fixed effects model:

$$\xi^* = \left\{ \begin{array}{ccc} (0.00, 0.00) & (0.00, 9.07) & (20.00, 0.00) \\ 0.039 & 0.572 & 0.389 \end{array} \right\}$$



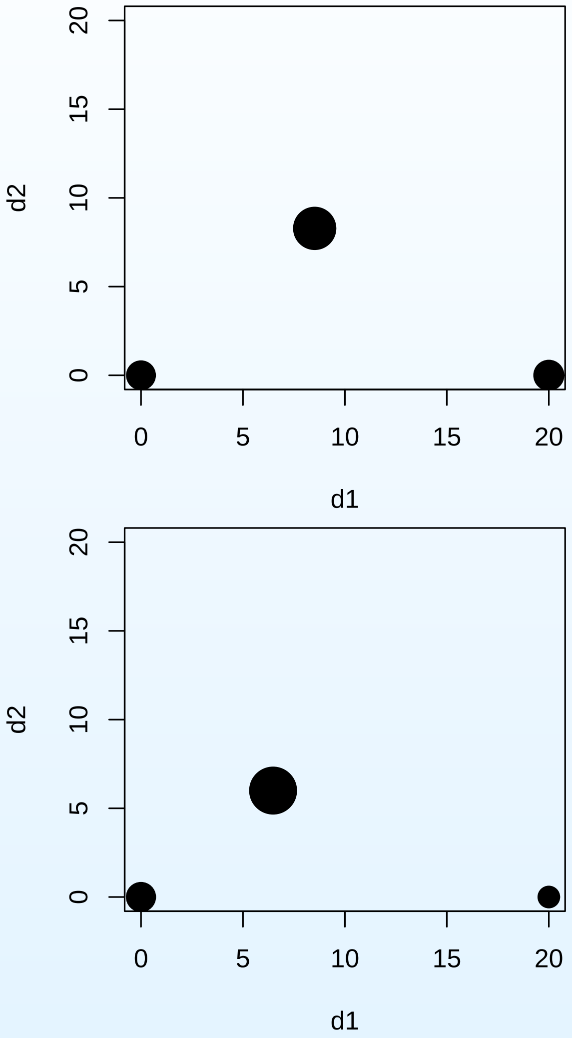
‘Optimal’ designs: intercept (θ_0) random, CV = 10%

Method	d_1	d_2	Criterion
Nick	0.00	6.67	5.6228
	0.00	6.67	
	0.00	6.67	
	0.00	6.67	
	20.00	0.00	
	20.00	0.00	
Tim	0.00	5.08	5.5333
	0.00	5.08	
	0.00	5.08	
	8.45	0.00	
	20.00	0.00	
	20.00	0.00	
Fixed effects design			5.2431



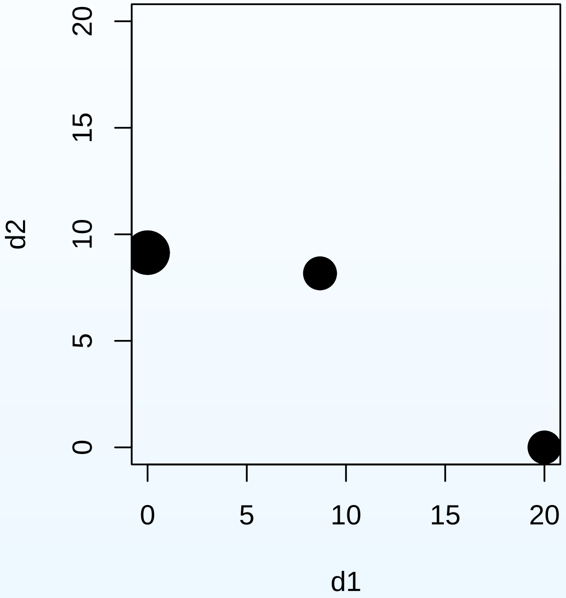
‘Optimal’ designs: direct effect (θ_1) random, CV = 10%

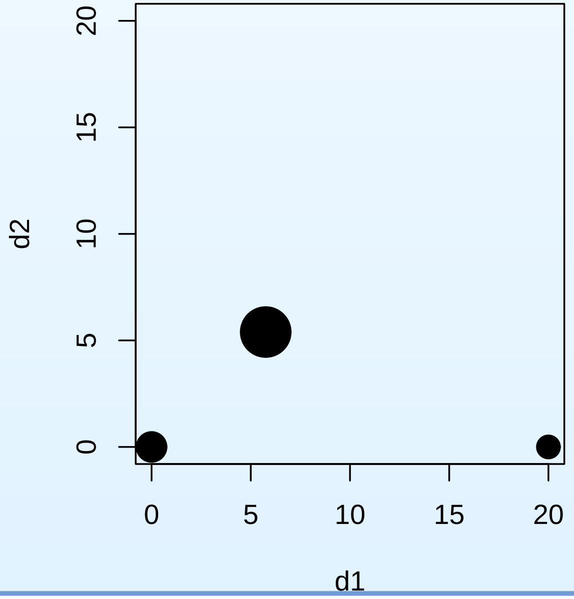
Method	d_1	d_2	Criterion
Nick	0.00	0.00	34.9416
	0.00	9.08	
	8.82	7.89	
	8.82	7.89	
	8.82	7.89	
	20.00	0.00	
Tim	0.00	0.00	32.7066
	6.37	6.04	
	6.37	6.04	
	6.37	6.04	
	6.37	6.04	
	20.00	0.00	
Fixed effects design			36.3835



‘Optimal’ designs: direct effect (θ_1) random, CV = 30%

Method	d_1	d_2	Criterion
Nick	0.00	0.00	42.8165
	0.00	9.10	
	8.51	8.33	
	8.51	8.33	
	20.00	0.00	
	20.00	0.00	
Tim	0.00	0.00	31.2136
	5.66	5.39	
	5.66	5.39	
	5.66	5.39	
	5.66	5.39	
	20.00	0.00	
Fixed effects design			38.5235





Discussion

- Acknowledging the dependence of w_{ij} on β gives a closer approximation to the true information matrix
- The Taylor approximation to the log-likelihood still adds a large degree of error
- Optimal designs based on Longford's less precise information matrix seem to be more efficient than designs found using the improved approximation
- Designs optimised using either of the methods which incorporate random effects may produce *less* efficient designs than the fixed effects design
- Future work: investigate different models/parameter values

References

1. Atkinson A.C. and Haines L.M. (1996). Designs for nonlinear and generalized linear models. In *Design and analysis of experiments*, vol. 13 of *Handbook of Statistics*, pp. 437–475, North-Holland, Amsterdam.
2. Longford N.T. (1994). Logistic regression with random coefficients. *Computational Statistics & Data Analysis* **17**, 1–15.