RANDOMIZATION, REGULARIZATION AND COVARIATE BALANCE IN RESPONSE-ADAPTIVE DESIGNS FOR CLINICAL TRIALS

Anthony C. Atkinson and Atanu Biswas

The London School of Economics, London WC2A 2AE, UK

Indian Statistical Institute, Kolkota

Department of Statistics
London School of Economics
London WC2A 2AE, UK

a.c.atkinson@lse.ac.uk

PHASE III CLINICAL TRIAL

- Patients arrive sequentially
- Each is immediately given one of t treatments
- Patient i also has a vector of prognostic factors x_i
- Main purpose is to find the best treatment
- Subsidiary purpose is to estimate the treatment effect
- Responses on patients before the ith may be available
- May stop for any n

REQUIREMENTS

- Adaptation: treat more frequently with the best treatment
- Balance across prognostic factors x_i: all kinds of patients should get all treatments
- Would like balance for all n
- Randomization: "objectivity", avoidance of conscious or unconscious biases (Lanarkshire)
- Blindedness

A LINEAR MODEL

$$E(y_i) = g_i^T \omega = h_i^T \alpha + z_i^T \theta$$

- α - the vector of unknown treatment effects
- h_i - vector of t indicator variables, the one non-zero element indicating which treatment the patient received.
- z_i a vector including any powers or interactions of the elements of the prognostic factors x_i. θ are nuisance parameters
- Analyse data using least squares, perhaps after transformation - log survival times
- Also for GLM with small variation in response - iterative weights are sensibly constant.

OPTIMUM EXPERIMENTAL DESIGN

- For n patients
$$E(Y_n) = G_n\omega$$
- The D-optimum design for ω maximizes
$$|G_n^T G_n|.$$
- When patient $n+1$ arrives, the allocation of the first n patients is already fixed.
- If treatment j is allocated to patient $n+1$, new row of design matrix is $g_{j,n+1}$.
- Sequential D-optimum design. Choose j for which
$$d(j, n, x_{n+1}) = g_{j,n+1}^T (G_n^T G_n)^{-1} g_{j,n+1}$$
is a maximum.

SEQUENTIAL OPTIMUM DESIGN

Linear Combinations of Parameters

- If interest is in s linear combinations of the parameters $A^T\hat{\omega}$,
- D_A-optimum design minimizes

$$|A^T(G_{n+1}^T G_{n+1})^{-1}A|.$$

- Sequentially experiment where

$$\begin{aligned} d_A(g_{n+1}, n) &= g_{n+1}^T(G_n^T G_n)^{-1}A\{A^T(G_n^T G_n)^{-1}A\}^{-1} \\ &\quad A^T(G_n^T G_n)^{-1}g_{n+1}, \end{aligned}$$

  is a maximum.

LINEAR COMBINATIONS OF PARAMETERS

$$E(y_i) = g_i^T \omega = h_i^T \alpha + z_i^T \theta.$$

- One combination ($s = 1$). Minimize var $l^T \hat{\omega}$
- **Two treatments**, prognostic factors, skewed (unequal) allocation

$$l = (p \quad -(1-p) \;\; 0 \quad \ldots \quad 0)^T$$

  For $0 < p < 1$, treatment 1 asymptotically allocated to a proportion p of the patients. $p = 0.5$ for treatment difference.

- Treatment mean and θ are the q nuisance parameters
- **Three treatments** with prognostic factors: proportions p_1, p_2, p_3

$$l = (p_1 \quad -p_2 \;\; 1 - \sum p_j \;\; 0 \quad \ldots \quad 0)^T$$

- **Adaptive:** use the data to choose p.

OPTIMUM DESIGN

- The design region $\mathcal{X}$ is the choice of treatment for the next patient
- Could sequentially allocate treatments to give D_A-optimum design minimising var $(l^T\hat{\alpha})$
- **Sequential Design Construction.** Allocate that j for which $d_A(j, n, x_{n+1})$ is maximum.
- BUT need to introduce randomness
- How do we measure the properties of the designs?

PROPERTIES: LOSS 1

- Interest is in one linear combination of the parameters

$$l^T\omega = l_1^T\alpha + l_2^T\theta.$$

- No covariates: var $l_1^T\hat{\alpha}$ is minimized when the proportion r_j of the patients receiving treatment j is p_j.
- The same is true for the design balanced across covariates.
- The effect of randomization is slightly to unbalance the trial and increase variance

LOSS 2

- **Two treatments.** No covariates – optimum design assigns treatment 1 to a proportion p of the patients and

$$\text{var}\, l^T(\hat{\alpha}) = \sigma^2/n,$$

  which is also the variance for a balanced trial with covariates.

- For other designs

$$\text{var}\,\{l^T\hat{\alpha}\} = \sigma^2 l^T(G_n^T G_n)^{-1} l$$

- The efficiency of a design is

$$E_n = 1/\left\{n l^T(G_n^T G_n)^{-1} l\right\}.$$

- The loss L_n is defined by writing the variance as

$$\text{var}\,\{l^T\hat{\alpha}\} = \frac{\sigma^2}{n - L_n}$$

BIAS

- Selection bias arises because the clinician can guess which treatment is to be applied next and allocate accordingly.
- Define bias as

$$\begin{aligned}\mathcal{B}_n &= \{E(\text{number of correct guesses}) \\ &\quad -E(\text{number of incorrect guesses})\}.\end{aligned}$$

- 1 for perfect guessing is bad, 0 good.
- Estimate this by simulation - depends on guessing rule.

SOME BIASED-COIN DESIGNS

- **Rule D - Deterministic.** Sequential Design Construction: allocate that j for which $p_j d_A(j, n, x_{n+1})$ is maximum.
- **Rule R - Randomized Design.** Allocate treatment j with probability p_j
- **Rule A.**
$$\pi_A(j|x_{n+1}) = \frac{p_j d_A(j, n, x_{n+1})}{\sum_{k=1}^{t} p_k d_c(k, n, x_{n+1})}.$$
- **Rule E - Efron's Biased Coin.** Probabilities proportional to $t : t-1 : ... : 1$ for treatments ordered by values of $p_j d_A(j, n, x_{n+1})$.

  2/3 for two treatments.

LOSS 3

$$\text{var}\left\{l^T\hat{\alpha}\right\} = \frac{\sigma^2}{n - L_n}$$

- Loss – number of patients on whom information is lost due to the lack of optimality of the design
- With a random element in treatment allocation, L_n is a random variable
- Let $E(L_n) = \mathcal{L}_n$
- There are asymptotic results for $\mathcal{L}_\infty$

Random (R)	q	DRC
A (above)	$q/5$	
Deterministic	0	

- It is easier to guess the next allocation for rules with a low value of $\mathcal{L}_\infty$

SIMULATIONS OF LOSS & BIAS

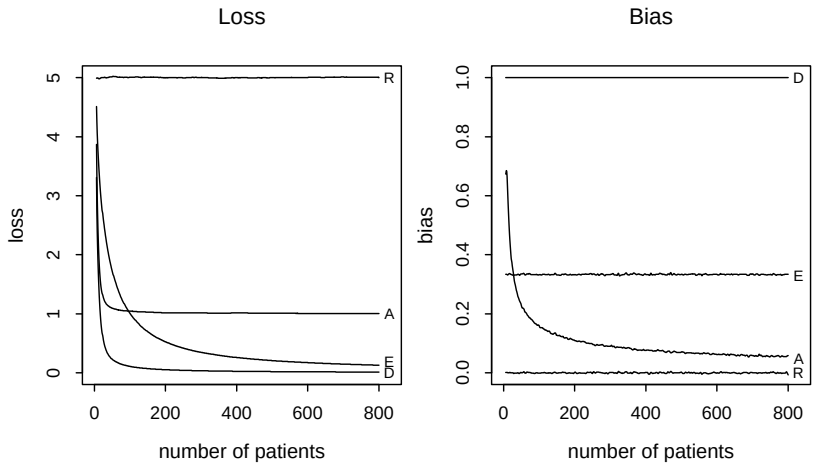

Figure 1: $\bar{L}_n$ and $\bar{B}_n$ for four non-adaptive rules, p=0.5: R, random; A; E, Efron's biased coin and D, deterministic. Results of 100,000 simulations, two treatments, $q = 5$, $n = 800$. Left-hand panel loss, right-hand panel smoothed bias

Large loss and small bias go together (can plot together)

MAXIMIZING UTILITY

- Ball, Smith, and Verdinelli (1993) suggest that the probabilities of treatment selection π_j be chosen to maximize a utility combining both the variance of parameter estimates and randomness.

$$\begin{aligned} U &= U_V - \gamma U_R \\ &= \sum_{j=1}^{t} \pi_j \phi_j - \gamma \sum_{j=1}^{t} \pi_j \left\{ \log(\pi_j / p_j) \right\}, \end{aligned}$$

- U_V provides estimates with low variance
- U_R contributes randomness
- The parameter γ provides a balance between these two.
- ϕ_j is a measure of information from applying treatment j

MAXIMIZING UTILITY 2

- Use of Lagrange multipliers shows

$$\pi_j = \frac{p_j\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma}}{\sum_{s=1}^{t} p_s\{1 + d_A(s, n, x_{n+1})\}^{1/\gamma}}.$$

- These designs give a balance between minimizing variance and parameter estimation that depends on n and γ.
- Near Rule D for small n and γ; tends to Rule R as $n \uparrow$

SIMULATIONS OF LOSS & BIAS

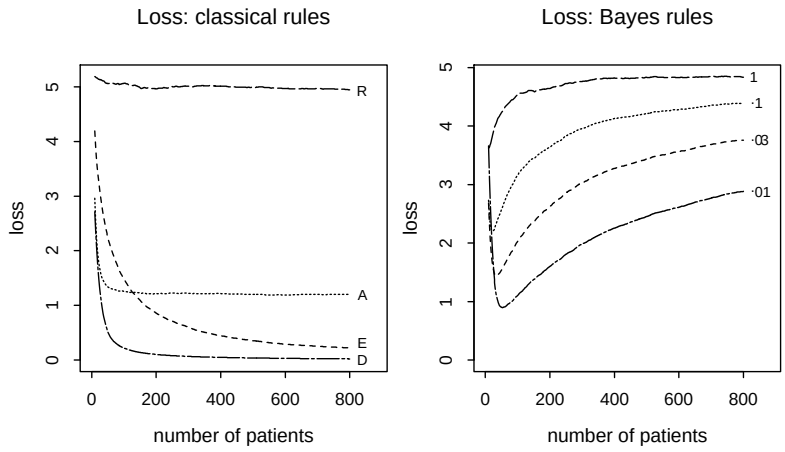

Figure 2: Designs for skewing proportion $p = 0.75$. Average losses L_n when $q = 5$ for eight allocation rules. Left-hand panel: A, D_A-optimality; D, deterministic; E, Efron's biased coin and R, random. Right-hand panel, Bayes rules: reading downward, $\gamma = 1$, 0.1, 0.03 and 0.01. Averages of 10,000 simulations

- Loss much higher for adaptive designs!

ADAPTIVE DESIGN: TARGET PROBABILITIES

- We choose the p_j adaptively to ensure greater allocation of the better treatments.
- Let $\hat{R}(j)$ be the estimated rank of treatment j (from ranking the $\hat{\alpha}_j$).
- Suppose $\hat{R}(j) = k$. Then $p_j = p_k^0$
- Use pre-assigned target probabilities

  $$p_1^0 \geq p_2^0 \geq \ldots \geq p_t^0,$$

  with at least one inequality.
- These reflect **ranks** but not **differences**.
- More stable than, for example, calculating p_k^0 from the $\hat{\alpha}_j$.

  $$\widehat{\Delta}_j = \hat{\alpha}_j - \bar{\alpha}.$$

  Then use normal cdf

  $$p'_j = \Phi(\widehat{\Delta}_j/T) \quad \text{and} \quad p_j = p'_j / \sum_{l=1}^{t} p'_l.$$

REGULARIZATION

- Some individual adaptive designs can be very unbalanced. This matters, since don't run several thousand trials (average properties not enough).
- Regularize by ensuring all n_j (numbers to treatments) $>$ some limit.
- 2 treatments. 5 & 5 for first 10 patients. Then if $n_j < \sqrt{n}$, that treatment is allocated when n is an integer squared. If $n = 800$, first regularize when $n = 36$, last $n = 784$.
- 3 treatments. 3, 3 & 3, then $\sqrt{n}$ rule, starting at $n = 16$.
- Effect arbitrary, but decreases with n.

EXAMPLE: THREE TREATMENT ADAPTIVE DESIGN

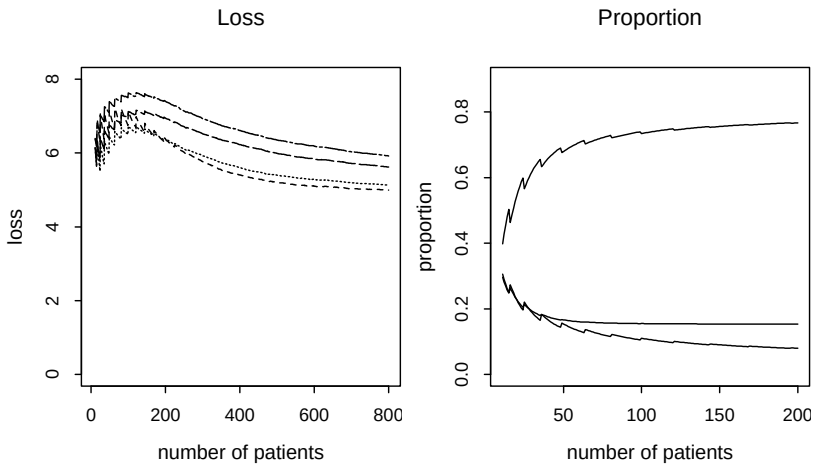

Figure 3: Rule G: regularised designs for three treatments. Left-hand panel: average losses $\bar{L}_n$ for four values of γ: reading downwards 1, 0.1, 0.03 and 0.01. Right-hand panel: average proportion $\bar{r}_{j,n}$ receiving each treatment when γ = 0.01. Averages of 10,000 simulations, α = (4.0, 0.65, 0)^T, p^* = (0.8, 0.15, 0.05)^T, $\sigma = 1.0$, $q = 5$.

The zig-zag pattern is caused by the operation of the regularisation rule: the decrease is caused by making the design slightly more balanced.

INDIVIDUAL TRIALS

- The purpose of regularisation is to avoid individual trials that become appreciably unbalanced.
- Now look at boxplots of properties of individual trials.

LOSS

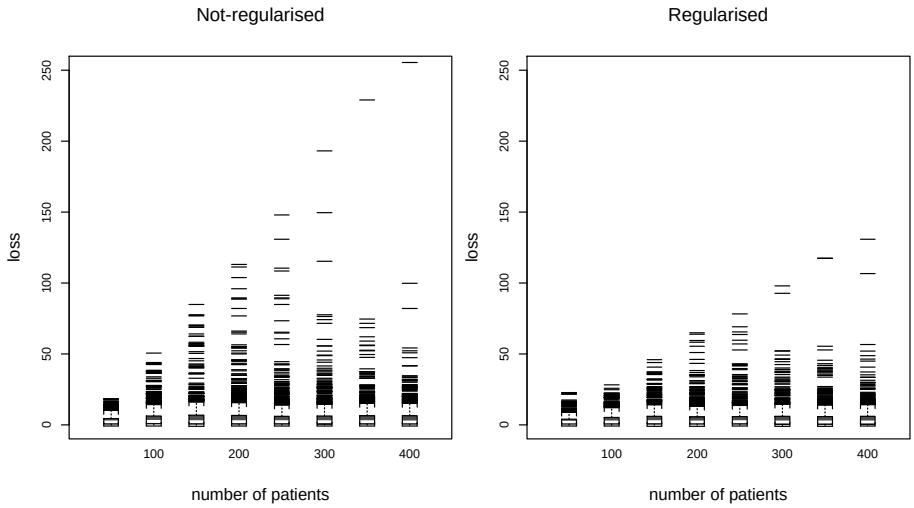

Figure 4: 1,000 individual adaptive designs: boxplots of loss L_n. Left-hand panel: not-regularised; right-hand panel, regularised. Rule A, $q = 5$, $\sigma = 1.1$

The effect of regularisation is striking, especially the comparatively larger number of high losses at lower values of n.

INFERENCE

- The parameter estimates are found assuming independent errors.
- But the allocation depends on the earlier responses and so the observations are not independent.
- With $t = 2$ the statistic of interest is the t-test for the hypothesis of no treatment effect, $\alpha_1 - \alpha_2 = 0$.
- We require the design to be skewing the allocation and so the null distribution of the statistic is not of interest.
- Instead we investigate the "pseudo-null" distribution by subtracting off the known value of $\alpha_1 - \alpha_2$.
- With $p_1 = 0.75$, at balance $3n/4$ patients will receive treatment 1.

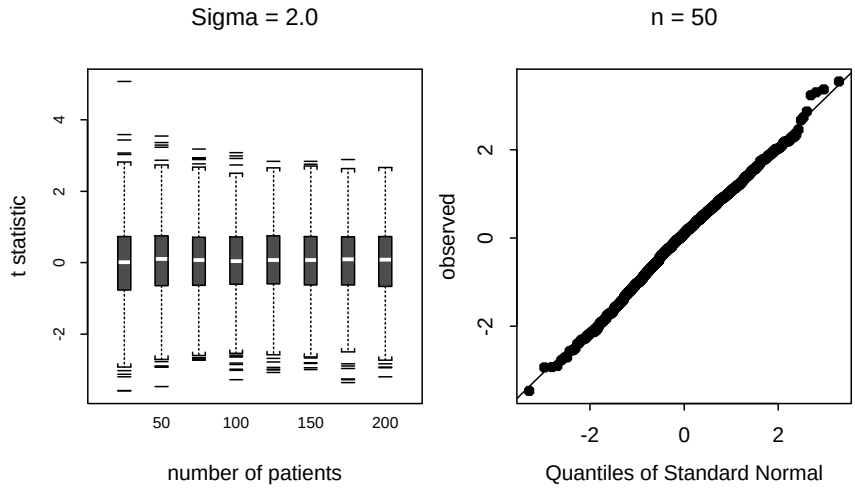

Figure 5: Pseudo-null distribution of the t-test for treatment equality: 1,000 individual regularized adaptive designs with two treatments when $q = 5$, μ = 0.6745, $\sigma = 2$ and $\gamma = 0.03$. Left-hand panel: boxplots of empirical distribution. Right-hand panel: normal QQ-plot when n = 50

- Figure 5 shows boxplots of 1,000 simulated values of the pseudo-null distribution of t. Even with $n = 25$, the statistics in the left-hand panel are centred close to zero with symmetrical distributions that are well behaved and seem close to normal. Standard tests can be used

Table 1: POWER AND SKEWING A Redesigned Trial: data on fluoxetine hydrochloride from Tamura et al. (1994). Average proportion of allocations to treatment 2 and average t-statistic from 1,000 simulations of 88 patient clinical trial.

Target p_2^*	Average proportion $\bar{r}_2$	Average statistic $\bar{t}$
0.5	0.500	2.563
0.55	0.546	2.549
0.6	0.592	2.512
0.65	0.637	2.450
0.7	0.681	2.371
0.75	0.722	2.266
0.8	0.760	2.140
0.85	0.796	1.970
0.9	0.820	1.810
0.95	0.833	1.712

CAKE AND THE EATING THEREOF

- The approach provides covariate-balanced adaptive designs with a controllable degree of randomness
- The skewed allocations provide greater allocation to better treatments (16 out of 88 in the redesigned trial)
- **BUT** there is a slight reduction in power of the test for treatment differences
- What kind of calculations are needed about present and future benefits to formalize this tradeoff?

THIS TALK

- Adjustment
- Sequential D-optimality
- Linear combinations of parameters
- An adaptive design
- Loss and Bias
- Regularization
- Average properties
- Individual trials
- Inference
- Power

References

Ball, F. G., A. F. M. Smith, and I. Verdinelli (1993). Biased coin designs with a Bayesian bias. *Journal of Statistical Planning and Inference 34*, 403–421.

Tamura, R. N., D. E. Faries, J. S. Andersen, and J. H. Heiligenstein (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association 89*, 768–776.