Optimal Few-Stage Designs for Clinical Trials

Janis Hardwick

University of Michigan

http://www.eecs.umich.edu/~jphard/

Research supported by National Science Foundation

Part of ongoing research program with Quentin Stout, Univ. Michigan

Basic Model

Imagine 2 populations of Bernoulli response data that represent patient responses to treatment arms 1 and 2, (T_1, T_2).

Sample N observations

From T_1 we get $X_{11}, X_{12}, \ldots \sim B(1, p_1)$ $\quad \searrow$

Independent with $(p_1, p_2) \in (0,1) \times (0,1)$

From T_2 we get $X_{21}, X_{22}, \ldots \sim B(1, p_2)$ $\quad \nearrow$

Sampling Procedures

Define an *allocation rule* or *design* as a sequence of indicators, δ_i, specifying the treatment for patient i; $i = 1, ..., N$.

$$\text{Thus,} \quad \delta(N) = (\delta_1, \delta_2, ..., \delta_N) \quad \text{such that} \quad \delta_i = \begin{cases} 1, & \text{if } T_1 \\ 0, & \text{if } T_2 \end{cases}$$

✓ $n_1 = \Sigma_1^N \, \delta_i$ and $n_2 = N - \Sigma_1^N \, \delta_i$

✓ For adaptive rules, δ_i may depend on data observed through time $i - 1$.

Part I: Loss + Cost

√ Let $\theta = (p_1 - p_2)$. Estimate $\hat{\theta}$ with MLEs or posterior means

√ Consider loss function incorporating normalized MSE$(\theta, \hat{\theta})$

along with a cost function reflecting failures during the study. ↙

$$L_N(\theta, \hat{\theta}_{n_1,n_2}) = N^2 \left[(\theta - \hat{\theta}_{n_1,n_2})^2\right] + \overbrace{n_1(1-p_1) + n_2(1-p_2)}$$

$$n_1 = \#\text{ from } T_1 \quad \text{and} \quad n_2 = \#\text{ from } T_2$$

$\Longrightarrow$ Goal: Find $\delta(N)$ to minimize the expected loss or "risk", $R_N(\theta, \hat{\theta}_{n_1,n_2})$

Risk Function

Analytically, given (p_1, p_2), one can find $n_1^* = n_1^*(N; p_1, p_2)$ to minimize

$$R_N(\theta, \hat{\theta}_{n_1,n_2}) = \mathbf{E}\left[N^2\left\{(\theta - \hat{\theta}_{n_1,n_2})^2\right\} + n_1(1-p_1) + n_2(1-p_2)\right]$$

where $\mathbf{E}$ is taken with respect to the binomial model, and $0 < n_1 < N$.

$\longrightarrow$ But this doesn't work unless we know (p_1, p_2).

$\surd$ We could use equal allocation, but lose on "ethical" cost

$\surd$ We could guess or estimate (p_1, p_2) and use $n_1^* = n_1^*(N; \hat{p_1}, \hat{p_2})$

$\Longrightarrow$ Use adaptive/sequential design.

Bayesian Design

- ⋆ Let p_1, p_2 have prior distribution, $\xi(p_1, p_2)$.
- ⋆ After m observations ξ is updated to get posterior $\xi_m(p_1, p_2 \mid \text{data})$
- ⋆ In our examples, ξ is the product of independent beta distributions.

Adjusted problem: *Locate $\delta(N)$ to minimize the Bayes risk :*

$$\mathcal{R}_N(\theta, \hat{\theta}_{n_1,n_2}) = \mathbf{E}^{\xi}\left[N^2\left\{(\theta - \hat{\theta}_{n_1,n_2})^2\right\} + n_1(1-p_1) + n_2(1-p_2)\right]$$

For $\mathbf{E}^{\xi}$ is expectation wrt prior, where the data follow the binomial model.

$\Rightarrow$ Seek designs that are *insensitive* to choice of prior distribution. $\Leftarrow$

Large Sample Approach

√ Optimal solutions exist theoretically via *dynamic programming*, but are infeasible to compute.

√ For normal rv's, W & H (1990) use a quasi-Bayesian approach to obtain a lower bound for the integrated risk (not quite Bayes since use MLEs).

√ This asymptotic lower bound is attainable to second order for a fairly large class of of allocation rules $\rightarrow$ call these Δ_A.

√ For a broad class of priors, the designs, $\delta(N)$, and estimators are independent of the Bayesian approach (in the limit).

Practical Sample Sizes

Hardwick (1991) handles binomial case. Compares asymptotic bound for minimum Bayes risk with exactly optimal design using dynamic programming.

N	MinBR/N	% Difference
10	1.153	2.75 %
50	1.134	1.15 %
160	1.127	0.05 %
320	1.125	0.03 %
∞	1.121	

$$\% \text{ Difference } = 1 - \frac{\text{Min Bayes Risk}}{N * \text{ Lower Bound}}$$

Note Hardwick (1991) cannot evaluate operating char. of optimal design.

Hardwick and Stout (1995) show that general evaluations can be carried out using new technique: *Path Induction*. However, despite this there are $\longrightarrow$

Objections to Fully Sequential Designs

⟶ They are complex to implement.

⟶ Responses assumed before next allocation.

⟶ Allocation of next patient deterministic.

⟶ While risk is minimized, trial time is maximized.

Preferable to Sample in Stages or Groups

⟶ Allows us to randomize within stages

⟶ Allows responses to be delayed somewhat

⟶ Concurrent patients reduces trial time

K-Stage Procedures

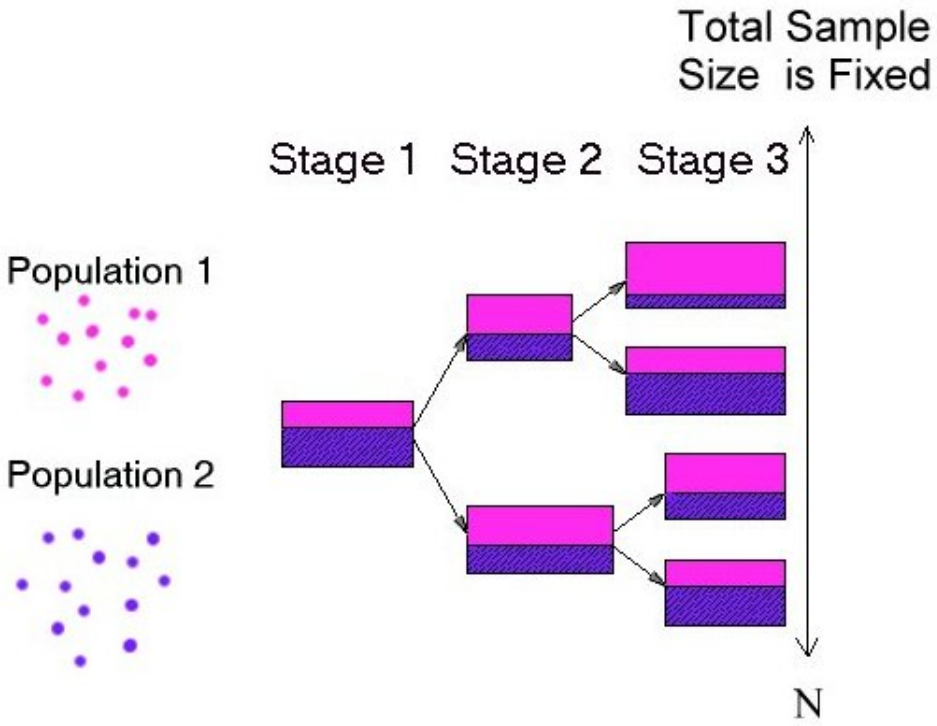
Let L_i = length of Stage i, and L_{1i} = prop. from T_1; $\; i = 1, \ldots, k$.

Then $\delta(\mathcal{L})$ describes a k-stage procedure where

$$\mathcal{L} = \begin{pmatrix} L_1 & L_{11} \\ \vdots & \vdots \\ L_k & L_{1k} \end{pmatrix}$$

with $\quad n - \sum_{j=1}^{i-1} L_j < L_i < n \quad$ and $\quad L_{i-1} \leq L_{1i} \leq L_i$

Flexible 3-Stage Procedure



Asymptotics Again

- √ There is cottage industry of literature on second order asymptotically optimal 3-stage designs.
- √ In fact, Δ_A, (from W & H) includes procedures that allocate patients in 3-stages.
- √ Nowhere are there good guidelines for choosing stage sizes in practice.
- √ *(As a rule, 2-stage designs achieve only first order asymptotic optimality.)*

Explicit W & H 3-Stage Procedure

Recall $n_1^*(N; p_1, p_2)$ is number on T_1 that minimizes $R_N(\theta, \hat{\theta}_{n_1,n_2})$

Stage 1: Sample $\frac{L_1}{2}$ from each treatment (so $L_{11} = L_{12} = \frac{L_1}{2}$).

Stage 2: Sample L_{21} more from T_1 and L_{22} more from T_2, where

$$L_{11} + L_{21} = \min\left\{N - L_{12} - L_3, \max\{L_{11},\ n_1^*(L_1 + L_2;\ \widehat{p_1}(L_{11}), \widehat{p_2}(L_{12})\}\right\}$$

$$\text{and } L_{22} = N - L_1 - L_3 - L_{21}.$$

Stage 3: Sample L_{31} more from T_1 and L_{32} more from T_2, where

$$\Sigma_1^3 L_{i1} = \min\left\{N - L_{12} + L_{22},\ \max\{L_{11} + L_{21}, n_1^*(N; \widehat{p_1}(L_{11} + L_{21}), \widehat{p_2}(L_{12} + L_{22})\}\right\}$$

$$\text{and } L_{32} = N - L_1 - L_2 - L_{31}.$$

How to choose stage sizes?

Well, that all looked very specific ... except that

$\Longrightarrow$ From W & H we get only

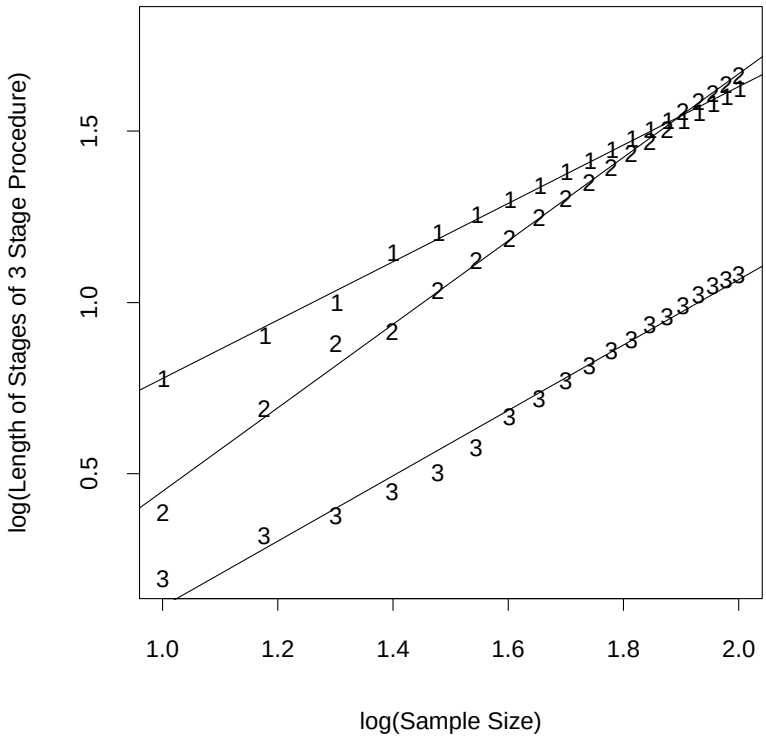
$$\lim_{N\to\infty} \frac{L_1 + L_3}{N} = 0 \qquad \text{and} \qquad \lim_{N\to\infty} \frac{N \log N}{L_3\sqrt{L_1}} = 0$$

We'll see that these guidelines are of little use for practical sample sizes. In fact, they're conceptually wrong for moderate N.

Optimal k-stage Procedures

- An *optimal* k-stage procedure is one that achieves $\min_{\delta(\mathcal{L})} \mathcal{R}(\xi, \delta(\mathcal{L}))$
- Use dynamic programming. Massive computing required.
- Determine best 1-stage procedure starting at any possible point in the experiment. (Can sometimes be done analytically)
- For all $1 < i < k$, determine optimal i-stage rule starting at any point in the experiment by evaluating all choices for sampling in this stage and then finishing with optimal (i-1)-stage continuation.
- Evaluate all possible parameter choices for the initial stage of the k-stage procedure using the optimal (k-1)-stage continuation.

Stage Lengths for Optimal Rule as N Increases



Efficiency Compared to Fully Sequential Design

$$0 \leq \text{ Efficiency}(\delta) = \frac{\text{Risk using any } \delta}{\text{Risk Optimal Procedure}} \leq 1$$

Design Type	L_1	E(L_2)	E(L_3)	Efficiency
Optimal 3-Stage	33	4	13	0.9994
Optimal WH	6	40	4	0.9990
Optimal 2-Stage	38	12	–	0.997
WH using guess	34	4	12	0.790

$$p_1 \sim Be(1, 10);\ p_2 \sim Be(10, 1) \text{ and } N = 50$$

Part II: Simple 2-Stage Problem

Goal: Maximize total successes among N observations, using 2 stages.

Note: Allocate all of 2nd stage to arm observed to be best during 1st stage.

Thus only need to determine

How many observations should be allocated to each arm for stage 1?

Answer will depend on N and priors.

Problem has an Extensive History

√ Colton (1965): 1^{st} stage equal allocation, optimal first stage size unknown.

√ Canner (1970): (Still EA for 1st stage) Analytically: Bayesian with uniform priors, optimal 1^{st} stage size $\approx \sqrt{2N+4} - 2$.

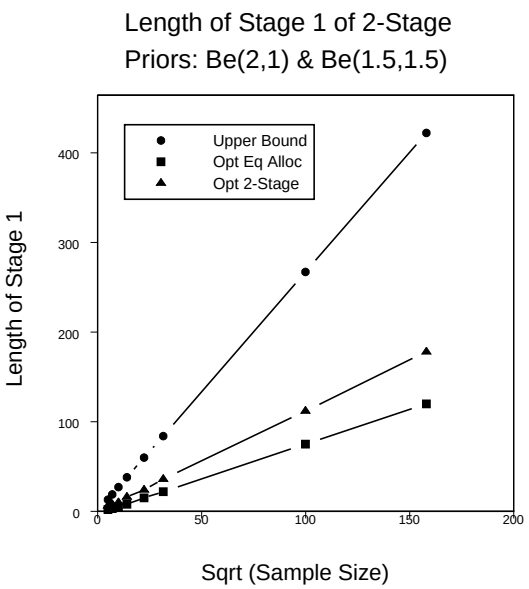
- ⋆ Conjecture: optimal 1st stage $\Theta(\sqrt{N})$ for arbitrary beta priors.

√ Cheng (1996): Analytically: $O(\sqrt{N})$ upper bound for optimal allocation to each arm on stage 1. No longer EA.

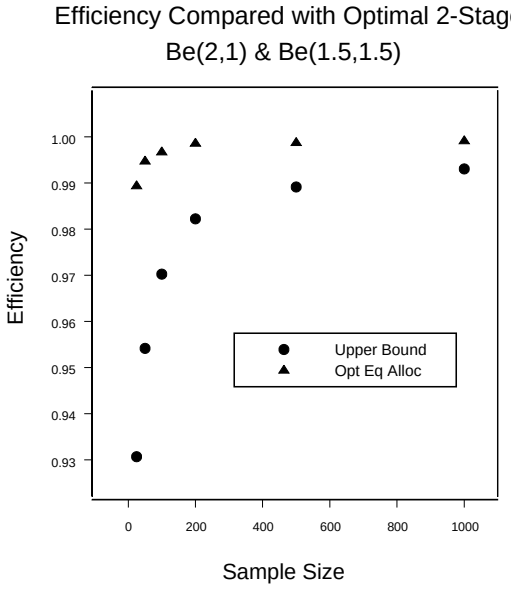
√ Hardwick & Stout (1995): Computationally: exact optimal allocation for arbitrary priors.

First Stage Size

"Upper Bound" is from Cheng, as are the priors



Stage 1 Length



Relative Efficiency

2-Stage 2-Arm Bandit Epilogue

Cheng, Su, and Berry (2002): *finally* obtain analytic asymptotically optimal 1st stage allocation for each arm for a general class of priors. Allocation is $\Theta(\sqrt{N})$.

This was after many years of work, by several people, on this simple objective function with a degenerate 2nd stage.

By changing one line in our program, we can optimize new objective functions, even very complex ones.

Results

- ⋆ May not always need optimal designs but they provide measuring stick for other designs
- ⋆ Asymptotic analytics often difficult to apply to useful sample size
- ⋆ Work applicable to arbitrary objective functions (must be additive, no minimax)
- ⋆ Optimal 2- and 3-stage designs can be highly efficient
- ⋆ New algorithms and good implementations allow one to optimize and analyze designs for practical sample sizes.

Some Future Work

- √ Applications to correct selection, hypothesis testing
- √ Incorporating decision and experimental costs
- √ Frequentist and Bayesian evaluations
- √ Extrapolating designs to sample sizes larger than can be optimized
- √ Visualizing designs to understand them better

Some Related Literature

- Canner, P.L. (1970), "Selecting one of two treatments when the responses are dichotomous", *J. Amer. Stat. Assoc.* **65**: 293–306.
- Cheng, Y. (1996), "Multistage bandit problems", *J. Stat. Plan. Infer.* **53**: 153–170.
- Colton T. (1965), "A two-stage model for selecting one of two treatments", *Biometrics* **21:** 169–180.
- Hardwick, J. (1991), "Computational problems associated with minimizing the risk in a simple clinical trial", *Contemporary Mathematics* **115:** 239–256.
- Hardwick, J. and Stout, Q.F. (1995), "Exact computational analyses for adaptive designs", *Adaptive Designs: Institute* **25** 223–37. Rosenberger & Flournoy, ed.'s.
- Hardwick, J., and Stout, Q.F. (2002), "Optimal few-stage designs", *J. Stat. Plan. Infer.* **104**: 121–145.
- Woodroofe, M. and Hardwick, J. (1991), "Sequential allocation for an estimation problem with ethical cost", *Ann. Stat.* **18**: 1358–1367.