Optimal Trial Design for Targeted Therapies

Sebastian Jobjörnsson

Chalmers University of Technology

EFSPI meeting on Biomarkers and Subgroups
Friday June 24, 2016, Leiden

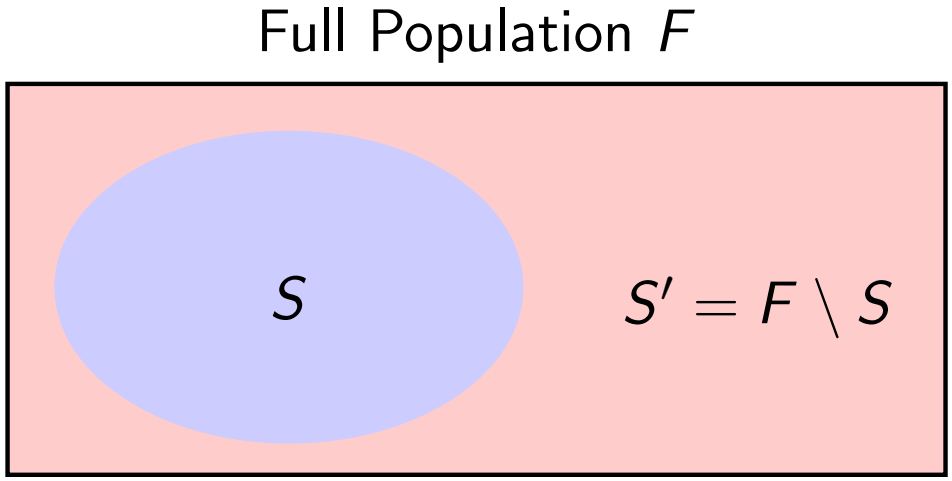
Joint work with Thomas Ondra, Robert Beckman, Carl-Fredrik Burman, Franz König, Nigel Stallard and Martin Posch

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under the IDEAL Grant Agreement no 602552, and the InSPiRe Grant Agreement no 602144.

Introduction

- An approach to the optimisation of confirmatory clinical trials will be presented.
- The efficacy of a treatment is tested in the full population and/or in a pre-specified, biomarker subpopulation.
- The optimisation is performed in the context of Bayesian decision theory. The optimal design is by definition the one maximising the expected utility.
- We compare two different perspectives:
  1. Commercial sponsor
  2. Public health decision maker

Framework and Basic Notation



- The full population F is partitioned into subgroups S and S'.
- λ_S denotes the prevalence of S in F.
- δ_S and $\delta_{S'}$ denote the true effects in the subgroups.
- This implies an overall effect $\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$.
- The hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$ are investigated.

The Three Design Types

Classical Design

Recruitment from the full population F.
No biomarker test is used.
Trial goal: reject $H_F : \delta_F \leq 0$.

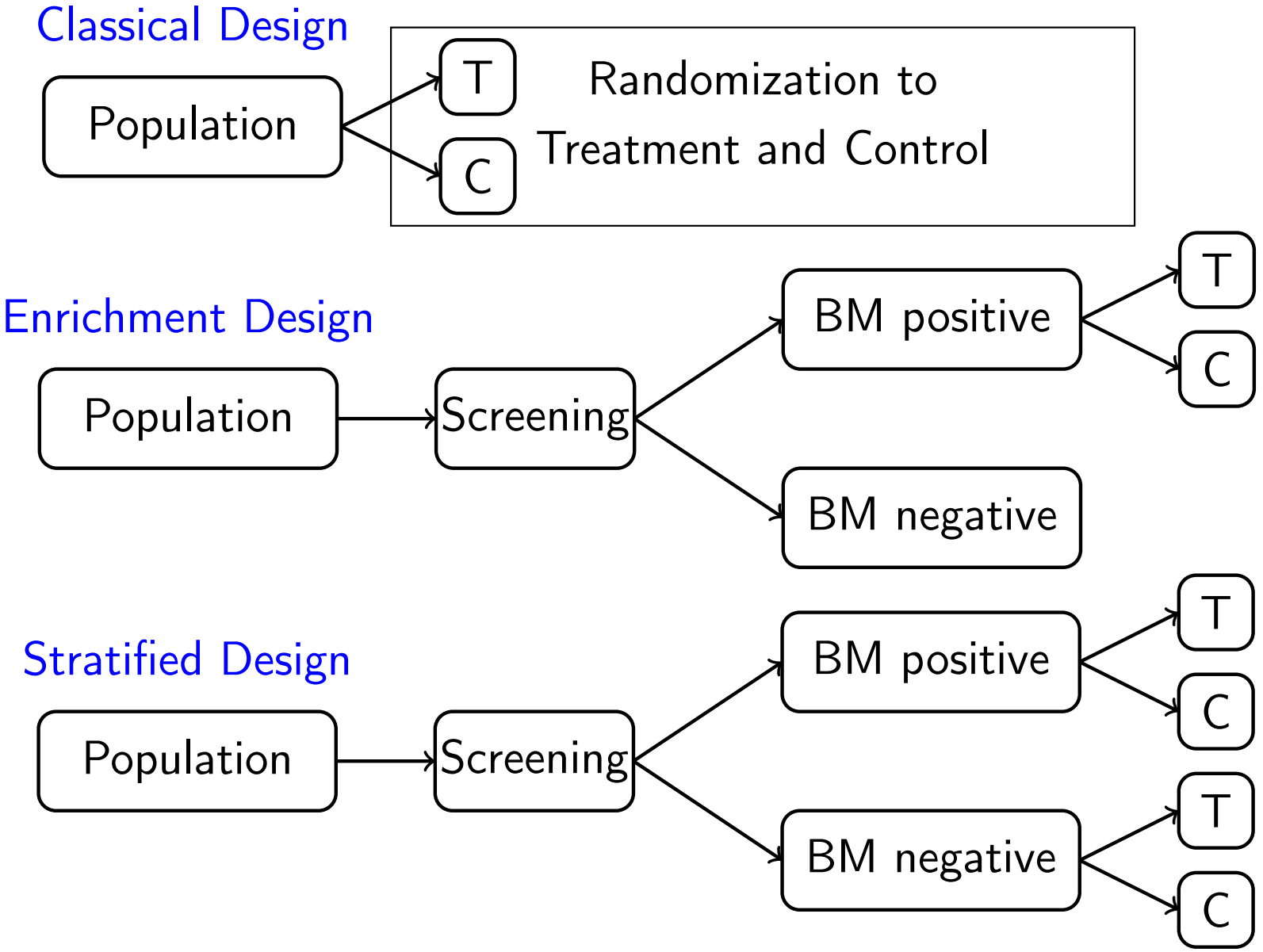
Enrichment Design

Recruitment from the subgroup S only.
Biomarker test is used to exclude patients in S'.
Trial goal: reject $H_S : \delta_S \leq 0$.

Stratified Design

Recruitment from the full population F.
Biomarker test is used to implement stratified randomization.
Trial goal: reject H_F or H_S.

Design Type Diagrams



Tests for Hypothesis Rejection

All tests are based on parallel group comparisons of sample means. Optimisation under the restriction that the type I error rate is $\leq \alpha$.

Classical Design:

H_F is tested using a z-test at level α.

Enrichment Design:

H_S is tested using a z-test at level α.

Stratified Design:

H_F and H_S are tested using a z-test at level α.

$H_F \cap H_S$ is tested at level α using the Spiessens-Debois procedure.

Closed testing principle implies strong control of the FWER[1].

To avoid (stratified) rejection of H_F driven by the effect in a single subgroup, the condition

$$p_S \leq \tau_S \text{ and } p_{S'} \leq \tau_{S'}$$

is used, where p_S and $p_{S'}$ are the unadjusted p-values for S and S'.

[1] Family-Wise Error Rate

The Spiessens-Debois Procedure

For adjusted significance levels α_F, α_S, reject $H_F \cap H_S$ if

$$p_F \leq \alpha_F \text{ or } p_S \leq \alpha_S,$$

where p_F, p_S are the p-values of the z-tests for H_F and H_S.

- For fixed α_F and α, α_S is chosen so that

$$\mathbb{P}_{H_F \cap H_S}\left(p_F < \alpha_F \text{ or } p_S < \alpha_S\right) = \alpha.$$

- Formulas well known from group sequential tests of normally distributed endpoints.

What Do We Optimise?

Design Type We choose either a classical, an enrichment or a stratified design.

Sample Size For each design type, we optimise the per-group sample size n.

Significance Levels For the stratified design, the signifiance levels α_F and α_S for the multiple testing procedure are optimised.

Hence, the design space may be written as

$$D = \{\mathtt{C}_n, \mathtt{E}_n, \mathtt{S}_{n,\alpha_S} \mid n \geq n_{\min}, \alpha_S \in [0, \alpha]\},$$

where $\mathtt{C}_n$ and $\mathtt{E}_n$ denote a classical and an enrichment design with sample size n, respectively, and $\mathtt{S}_{n,\alpha_S}$ denotes a stratified design with sample size n and a significance level for H_S equal to α_S.

The Form of the Utility Function

For both the sponsor and the public health perspective, we postulate a utility function of the form

$$U(d) = -C(d) + \begin{cases} \varphi_{F,d} & \text{if } \psi_{F,d} = 1 \\ \varphi_{S,d} & \text{if } \psi_{F,d} = 0 \text{ and } \psi_{S,d} = 1 \\ 0 & \text{if } \psi_{F,d} = 0 \text{ and } \psi_{S,d} = 0 \end{cases}$$

where

- $C(d) =$ Cost for the trial.
- $\varphi_{F,d} =$ Reward if the treatment is licenced in F.
- $\varphi_{S,d} =$ Reward if the treatment is licenced in S only.
- $\psi_{F,d} =$ Indicator function for the rejection of H_F.
- $\psi_{S,d} =$ Indicator function for the rejection of H_S.

The Reward Functions for the Two Perspectives

Sponsor view

$$\varphi_{F,d} = N \cdot r_F \cdot (\hat{\delta}_{F,d} - \mu_F)^+,$$
$$\varphi_{S,d} = \lambda_S \cdot N \cdot r_S \cdot (\hat{\delta}_{S,d} - \mu_S)^+,$$

Public health view

$$\varphi_{F,d} = N \cdot r_F \cdot (\delta_F - \mu_F),$$
$$\varphi_{S,d} = \lambda_S \cdot N \cdot r_S \cdot (\delta_S - \mu_S),$$

where

- N = Number of future patients.
- r_F, r_S = Revenue parameters.
- μ_F, μ_S = Clinically relevant thresholds.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ = Observed effect estimates.
- δ_S, δ_F = True effect sizes.

The Cost Function $C(d)$ for the Three Design Types

$$C(\mathtt{C}_n) = c_{\text{setup}} + 2nc_{\text{per-patient}},$$

$$C(\mathtt{E}_n) = c_{\text{setup}} + c_{\text{BMD}} + 2n\left(c_{\text{per-patient}} + \frac{c_{\text{BMS}}}{\lambda_S}\right),$$

$$C(\mathtt{S}_{n,\alpha_S}) = c_{\text{setup}} + c_{\text{BMD}} + 2n\left(c_{\text{per-patient}} + c_{\text{BMS}}\right),$$

where

- c_{setup} = Fixed setup cost for initiating the trial.
- $c_{\text{per-patient}}$ = Marginal cost per patient included in the trial.
- c_{BMD} = Development cost for the BM screening procedure.
- c_{BMS} = Marginal screening cost per patient.

Optimisation of the Expected Utility

The optimal design is given by

$$d^* = \text{argmax}_{d \in D} E_\pi \left[E_\Delta [U(d)] \right],$$

where the expectation is taken over a prior π over the effect sizes $\Delta = (\delta_S, \delta_{S'})$ and the sampling distribution.

The assumption of normally distributed sampling distributions implies

- $E_\Delta[U(d)]$ may be written in terms of the PDF and CDF of the standard normal distribution for the classical and enrichment design.
- $E_\Delta[U(d)]$ may be computed by a simple numerical integration for the stratified design.

Numerical Examples, Three Cases of Parameter Values

Case 1 (Large market, zero BM costs)

$N \cdot r_S = N \cdot r_F = 10{,}000$ MUSD[2] per unit of efficacy.

$c_{\text{BMD}} = c_{\text{BMS}} = 0.$

Case 2 (Small market, zero BM costs)

$N \cdot r_F = N \cdot r_S = 1000$ MUSD per unit of efficacy.

$c_{\text{BMD}} = c_{\text{BMS}} = 0.$

Case 3 (Small market, nonzero BM costs)

$N \cdot r_F = N \cdot r_S = 1000$ MUSD per unit of efficacy.

$c_{\text{BMD}} = 10$ MUSD and $c_{\text{BMS}} = 5000$ USD per patient.

The following parameters are the same for each case:

$$c_{\text{setup}} = 1 \text{ MUSD}, \quad c_{\text{per-patient}} = 0.05 \text{ MUSD},$$
$$\mu_F = \mu_S = 0.1,$$
$$\tau_S = \tau_{S'} = 0.3,$$
$$\sigma = 1, \quad n_{\min} = 50, \quad \alpha = 0.025.$$

[2] Million US Dollars

Numerical Examples, Two Different Priors for $\Delta = (\delta_S, \delta_{S'})$

Weak BM Prior

Weak prior evidence that the biomarker is predictive.

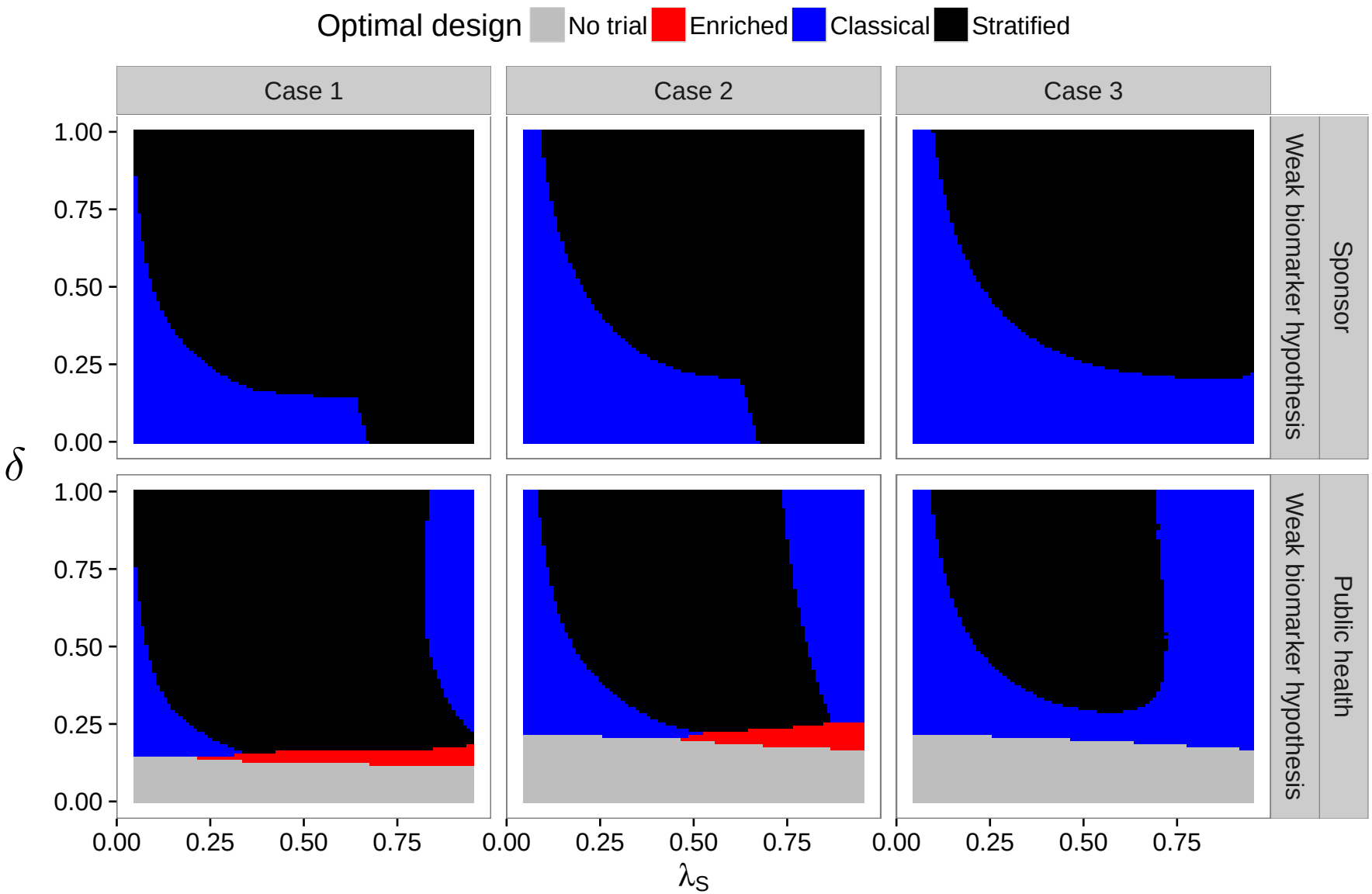
Strong BM Prior

Strong prior evidence that the biomarker is predictive.

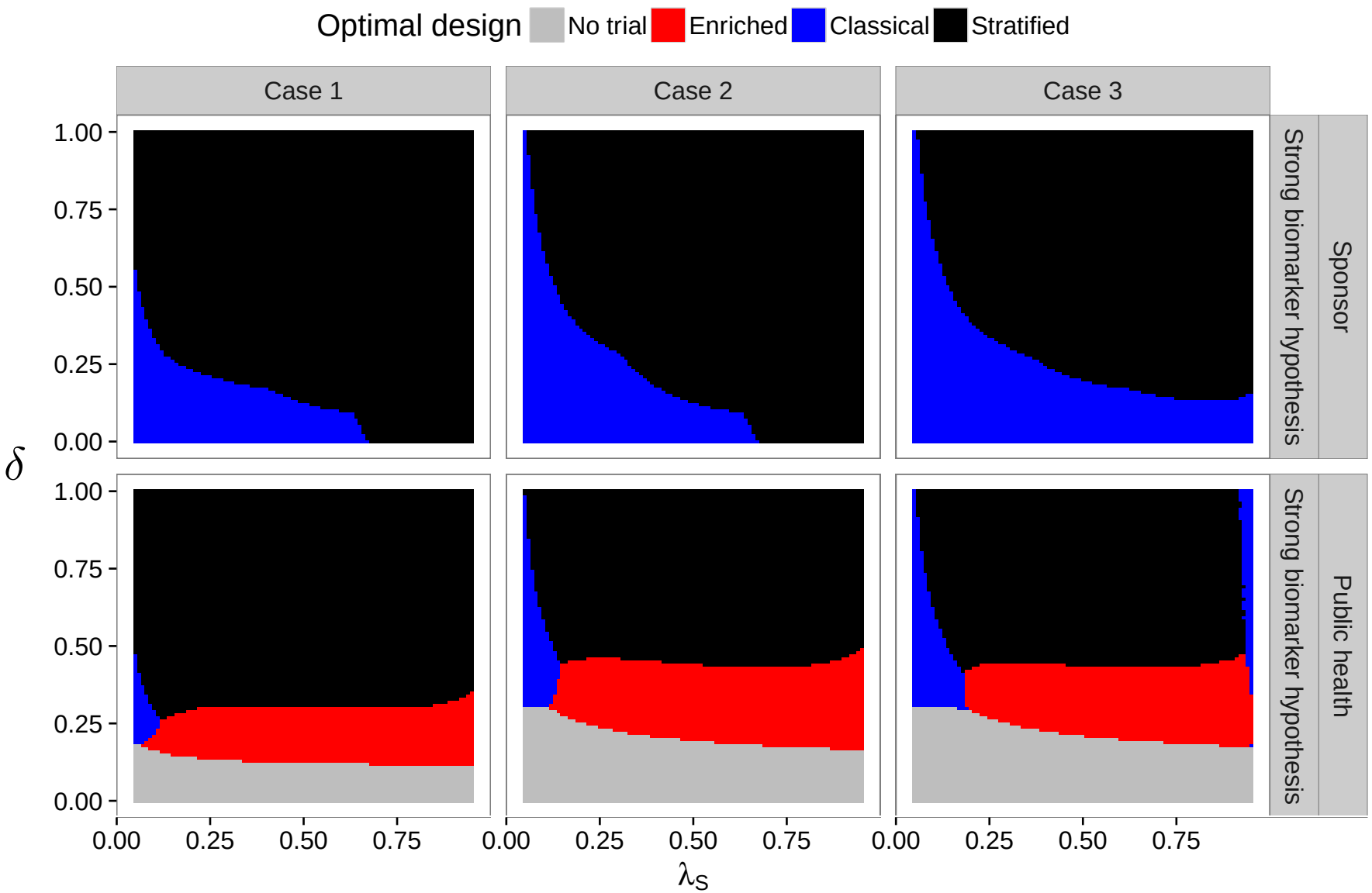
	δ_S	0	δ	δ	δ
	$\delta_{S'}$	0	0	$\delta/2$	δ
Weak BM Prior		0.2	0.2	0.3	0.3
Strong BM Prior		0.2	0.6	0.1	0.1

The constant $\delta > 0$ parametrizes the effect sizes in the prior.

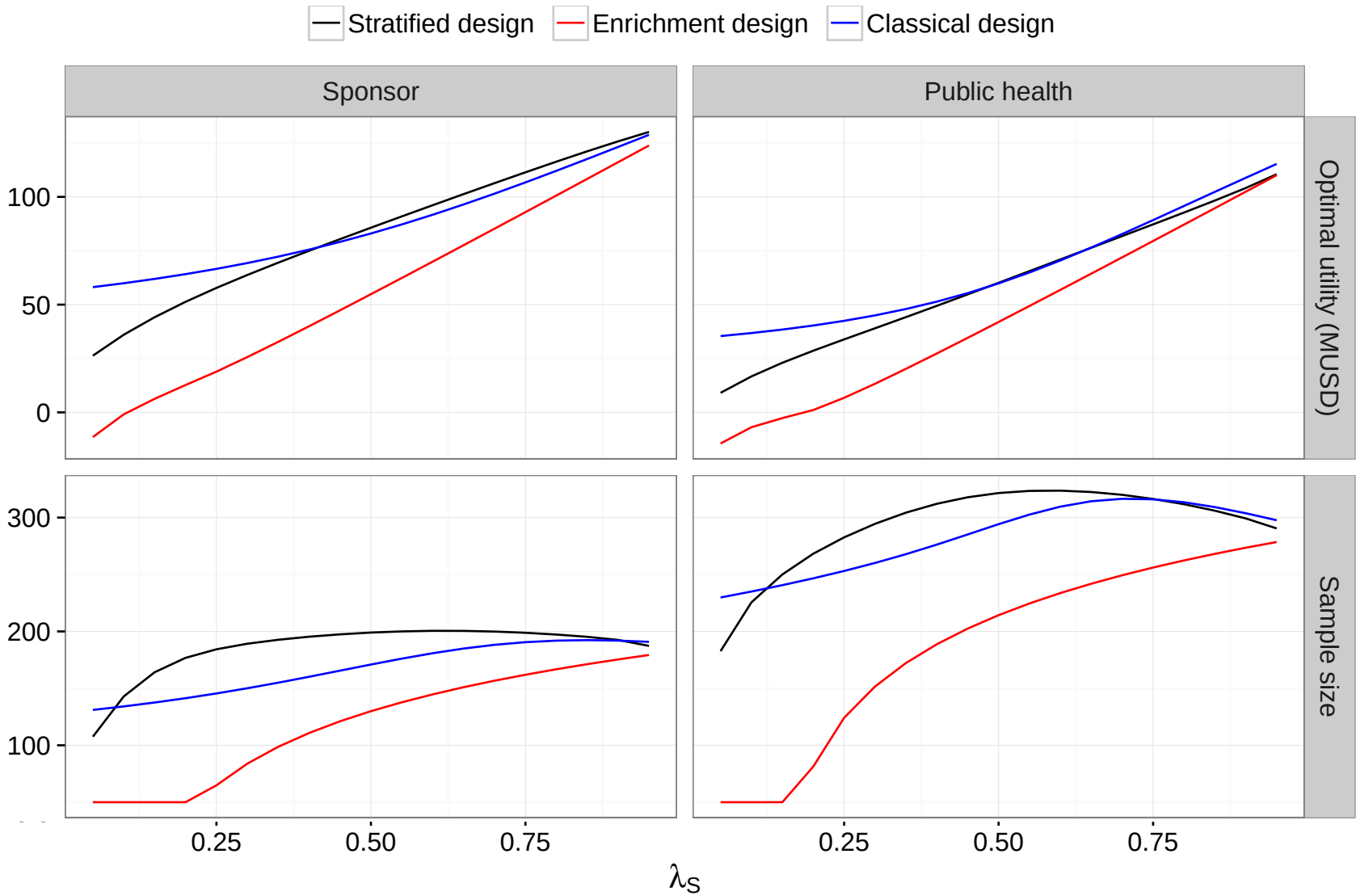
Optimal Trial Design Type (Weak BM Prior)



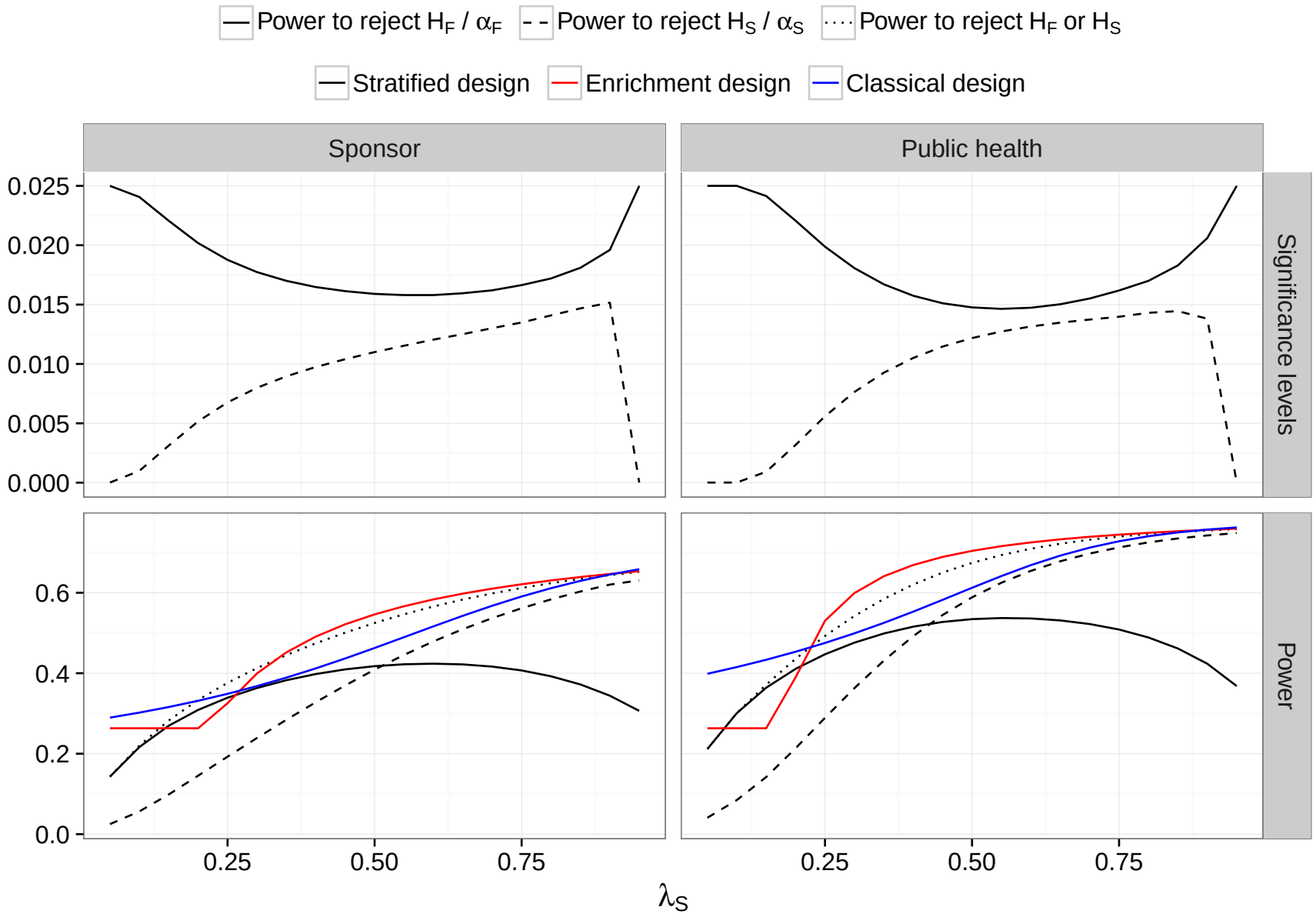
Optimal Trial Design Type (Strong BM Prior)



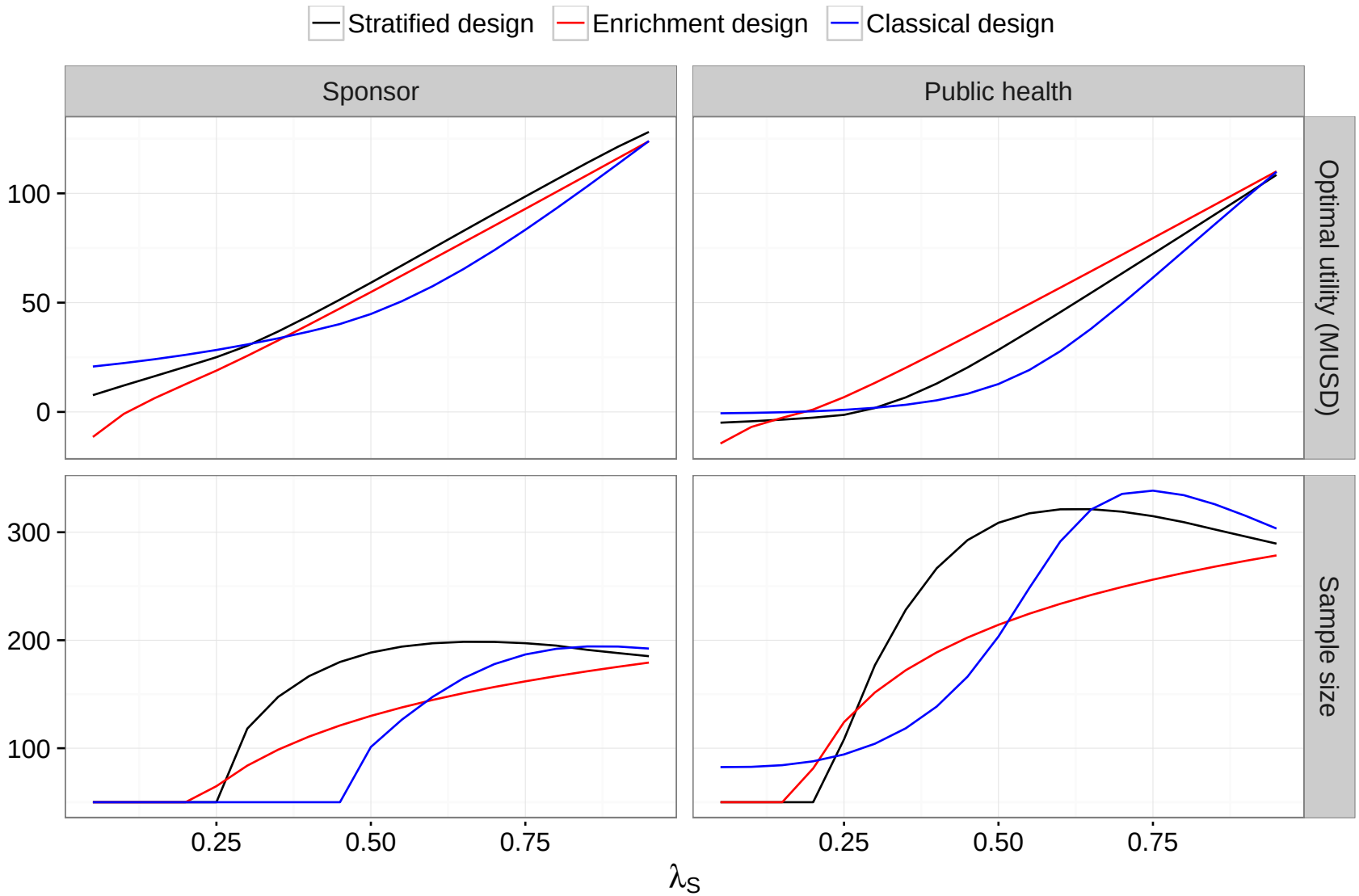
Opt. U and n (Case 3, Weak BM Prior, $\delta = 0.3$)



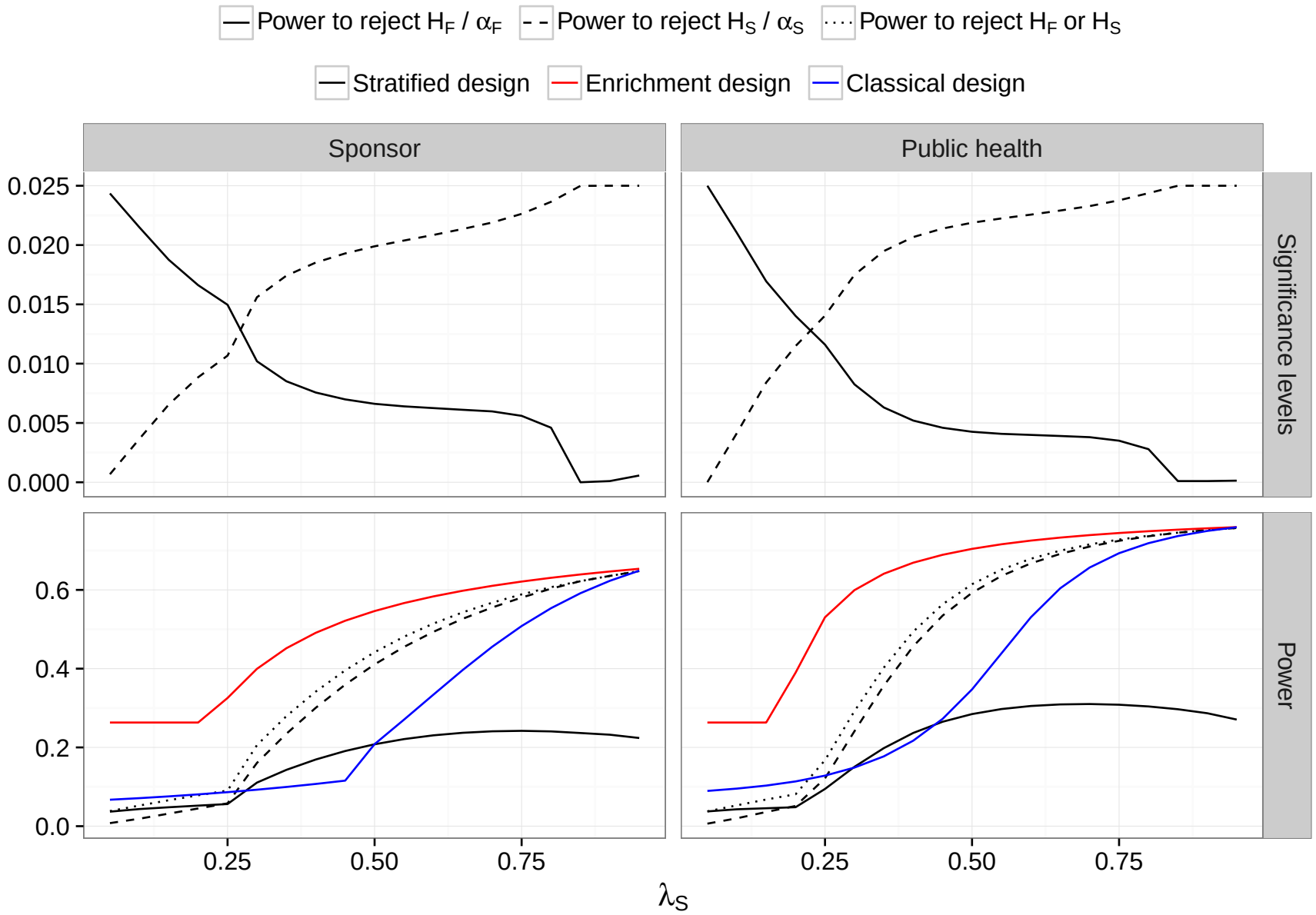
Opt. α_F, α_S and Power (Case 3, Weak BM Prior, $\delta = 0.3$)



Opt. U and n (Case 3, Strong BM Prior, $\delta = 0.3$)



Opt. α_F, α_S and Power (Case 3, Strong BM Prior, $\delta = 0.3$)



Conclusions

- In principle, a decision theoretic approach can be used to select appropriate values for all trial design variables.
- In practice, the complexity of the resulting optimisation problem sets the limits of the approach.
- The examples presented show that the optimal design depends strongly on the particulars of the situation:
  - Subgroup prevalence.
  - Trial costs.
  - Initial beliefs (the prior).
- For the numerical examples considered, we observed that:
  - Optimal sample sizes larger for public health than for the sponsor.
  - For low prevalences, the classical design outperforms the designs that are based on the biomarker.
  - For the sponsor, the enrichment design never maximizes the expected utility.

Possible Extensions

The model can be extended in several directions:

- Allow for partial enrichment.
- Include adaptive enrichment designs, leading to sequential optimisation.
- Make the reward functions more realistic.
- Allow N to depend on the trial type.
- Include a discount rate for the sponsor.

Key References

Beckman, R. A., J. Clark, and C. Chen (2011). Integrating predictive biomarkers and classifiers into oncology clinical development programmes. *Nature Reviews Drug Discovery 10*(10), 735–748.

Graf, A. C., M. Posch, and F. Koenig (2015, January). Adaptive designs for subpopulation analysis optimizing utility functions. *Biometrical Journal 57*, 76–89.

Ondra, T., S. Jobjörnsson, R. A. Beckman, C.-F. Burman, F. König, N. Stallard, and M. Posch (2016). Optimizing trial designs for targeted therapies. *arXiv:1606.03987*.

Song, Y. and G. Y. H. Chi (2007, August). A method for testing a prespecified subgroup in clinical trials. *Statistics in Medicine 26*(19), 3535–3549.

Spiessens, B. and M. Debois (2010, November). Adjusted significance levels for subgroup analyses in clinical trials. *Contemporary Clinical Trials 31*(6), 647–656.