Optimizing Confirmatory Clinical Trial Designs for Targeted Therapies

A Decision Theoretic Approach

Martin Posch

Section for Medical Statistics, Medical University of Vienna

Institut für Angewandte Statistik | Johannes Kepler Universität Linz

June 2016

Joint work with Thomas Ondra, Sebastian Jobjörnsson, Robert Beckman, Carl-Fredrik Burman, Franz König, Nigel Stallard

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.

Clinical Trials for "Precision Medicine"

For the development of targeted therapies, clinical trials with complex objectives, confirming treatment effects in sub-populations and/or in the overall populations are required.

- Knowledge on the genetic basis of many diseases is rapidly increasing and therapies that target underlying molecular mechanisms are developed.
- Patients' responses are predicted to targeted treatments based on genetic features or other biomarkers.

Objective: Identify subgroups (based on biomarkers) where the treatment has a positive benefit risk balance.

Patient Populations

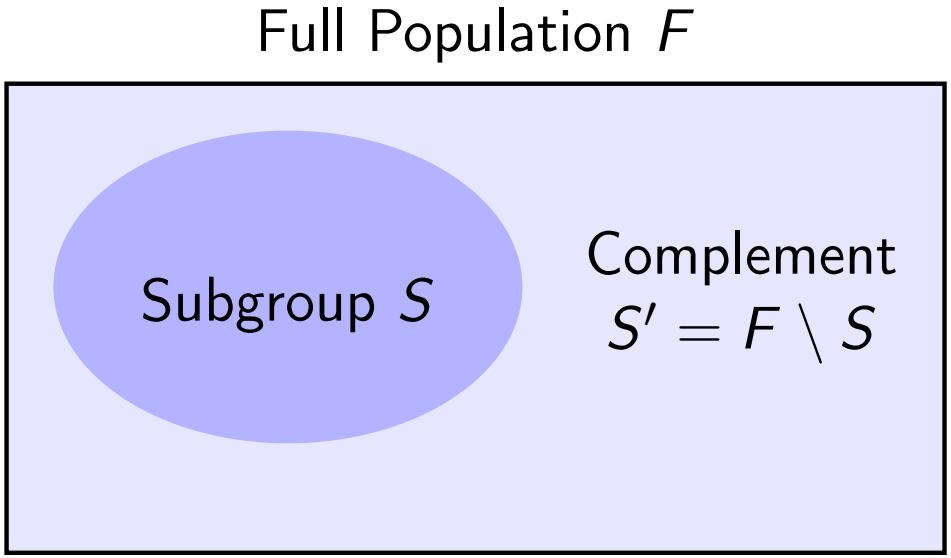
Full Population F

- The overall treatment effect is

$$\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$$

  where λ_S is the prevalence of subgroup S.
- Assume $\delta_{S'} \leq \delta_S$.
- Test of hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$.

Patient Populations

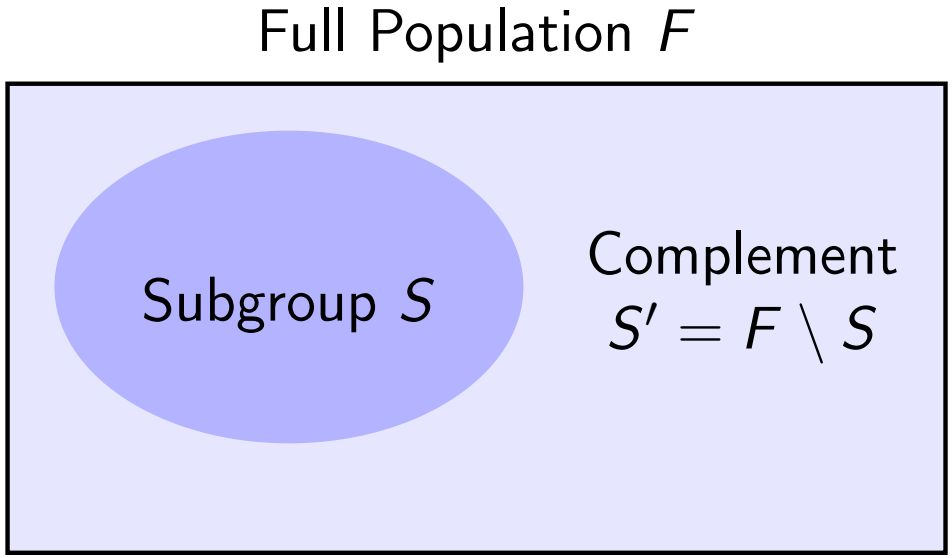


- The overall treatment effect is

$$\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$$

  where λ_S is the prevalence of subgroup S.
- Assume $\delta_{S'} \leq \delta_S$.
- Test of hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$.

Patient Populations

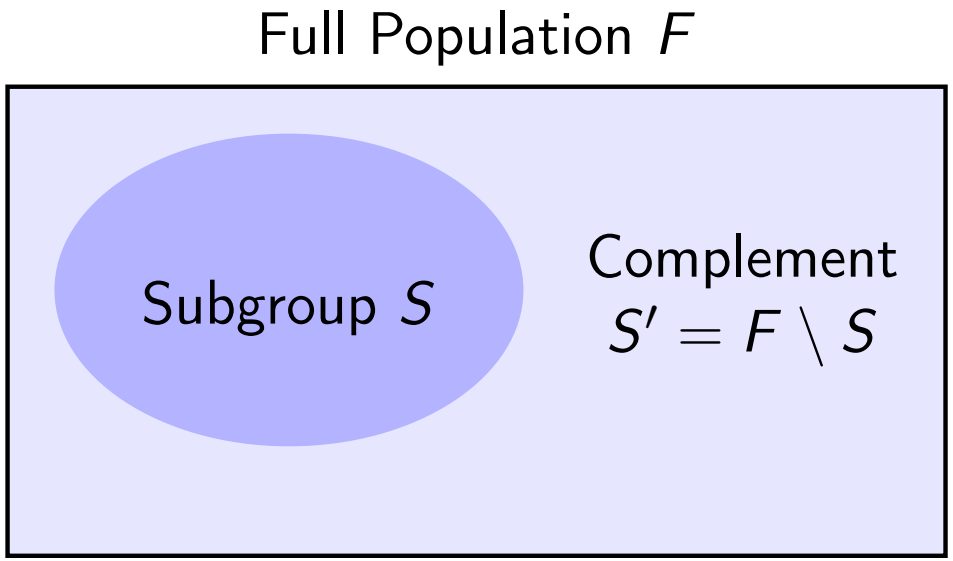


- The overall treatment effect is

$$\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$$

  where λ_S is the prevalence of subgroup S.
- Assume $\delta_{S'} \leq \delta_S$.
- Test of hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$.

Patient Populations

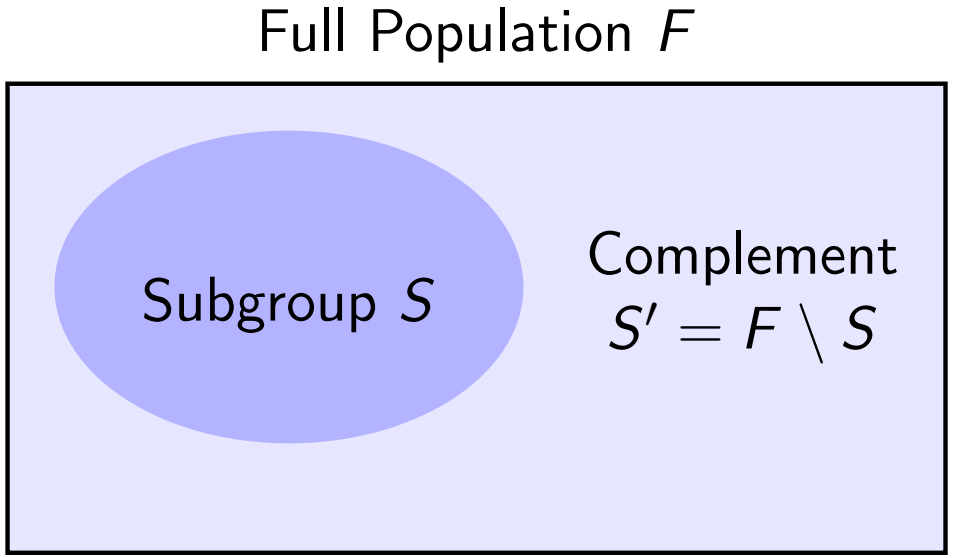


- The overall treatment effect is

$$\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$$

  where λ_S is the prevalence of subgroup S.
- Assume $\delta_{S'} \leq \delta_S$.
- Test of hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$.

Patient Populations



- The overall treatment effect is

$$\delta_F = \lambda_S \delta_S + (1 - \lambda_S)\delta_{S'}$$

  where λ_S is the prevalence of subgroup S.
- Assume $\delta_{S'} \leq \delta_S$.
- Test of hypotheses $H_F : \delta_F \leq 0$ and $H_S : \delta_S \leq 0$.

Which clinical trial design to choose?

Classical Design:

Recruitment from population F.
No Biomarker is determined.
Test of H_F.

Stratification Design:

Recruitment from population F.
Stratified randomization by Biomarker.
Test of H_F and H_S.

Enrichment Design:

Recruitment only from population S.
Test of H_S

Biomarker Designs

Which clinical trial design to choose?

Classical Design:

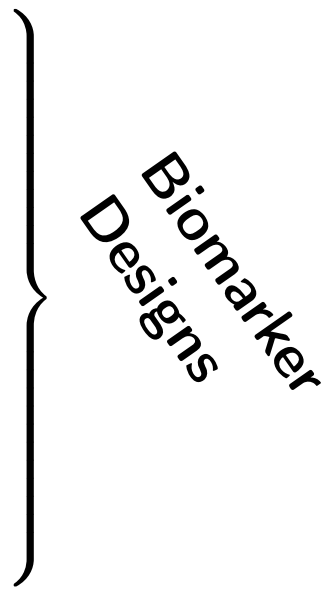
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

Stratification Design:

Recruitment from population F.
Stratified randomization by Biomarker.
Test of H_F and H_S.

Enrichment Design:

Recruitment only from population S.
Test of H_S



Which clinical trial design to choose?

Classical Design:

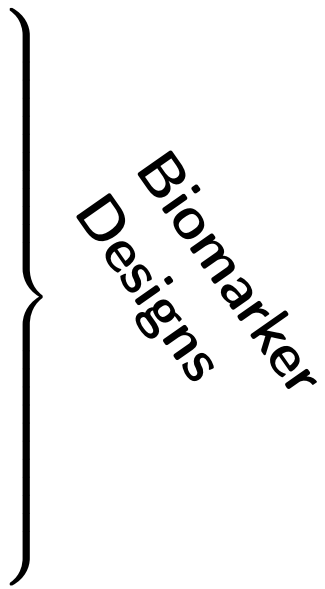
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

Stratification Design:

Recruitment from population F.
Stratified randomization by Biomarker.
Test of H_F and H_S.

Enrichment Design:

Recruitment only from population S.
Test of H_S



Which clinical trial design to choose?

Classical Design:

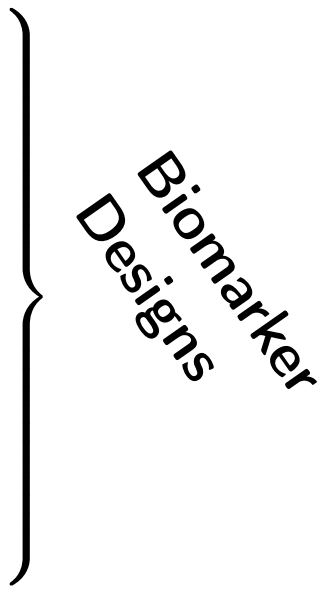
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

Stratification Design:

Recruitment from population F.
Stratified randomization by Biomarker.
Test of H_F and H_S.

Enrichment Design:

Recruitment only from population S.
Test of H_S



Optimizing Clinical trial designs

- When is a biomarker (BM) design beneficial compared to a classical design?
- When to choose stratified, when an enrichment design?
- Which sample size?
- Which multiple test for the stratified design is optimal?

Perform a Trial in the Full or Only the Subpopulation?

- The power to reject at least one hypothesis depends on the effect sizes $\delta_S, \delta_{S'}$ and the prevalences
- Assume we suspect that $\delta_{S'} \leq \delta_S$ but believe that $\delta_{S'} > \delta_S$ is not plausible.
- Then, the design in the subpopulation (recruiting only patients in S) always leads to the highest power:
  - If $\delta_{S'} = \delta_S$ the enrichment design has a larger power than the stratification design (which requires multiple testing) and equal power as the classical design.
  - If $\delta_{S'} < \delta_S$ the enrichment design has larger power than stratification & classical design.

Thus, is enrichment always preferable?

Perform a Trial in the Full or Only the Subpopulation?

- The power to reject at least one hypothesis depends on the effect sizes $\delta_S, \delta_{S'}$ and the prevalences
- Assume we suspect that $\delta_{S'} \leq \delta_S$ but believe that $\delta_{S'} > \delta_S$ is not plausible.
- Then, the design in the subpopulation (recruiting only patients in S) always leads to the highest power:
  - If $\delta_{S'} = \delta_S$ the enrichment design has a larger power than the stratification design (which requires multiple testing) and equal power as the classical design.
  - If $\delta_{S'} < \delta_S$ the enrichment design has larger power than stratification & classical design.

Thus, is enrichment always preferable?

Different Gains, different Costs...

The argument results from an oversimplification. The Power to reject any null hypotheses is not the only criteria.

- The stratification design tests the full population $H_F : \delta_F \leq 0$, demonstrating that the treatment works "on average".
- The enrichment design tests a limited null hypothesis $H_S : \delta_S \leq 0$ leading to a limited indication.
- Ethical problem if patients that may benefit are excluded.
- Enrichment maybe costly (e.g. due to longer patient recruitment, ...).

We apply a utility based approach, (cf. Beckman et al., 2011; Graf et al., 2015), to model the expected utilities of a particular trial design from a sponsor's and a public health view.

The Utility Function

$$U(d) = \underbrace{-\,C(d)}_{\text{Cost}} + \underbrace{\begin{cases} \varphi_{F,d} & \text{if } H_F \text{ is rejected} \\ \varphi_{S,d} & \text{if only } H_S \text{ is rejected} \\ 0 & \text{if no hypothesis is rejected} \end{cases}}_{\text{Reward}}.$$

where, $d \in \mathcal{D}$ denotes a trial design (Classical, Stratification, Enrichment) together with a specific sample size, and testing procedure.

The Hypothesis Test for the Stratification Design

Controlling the Familywise Error Rate

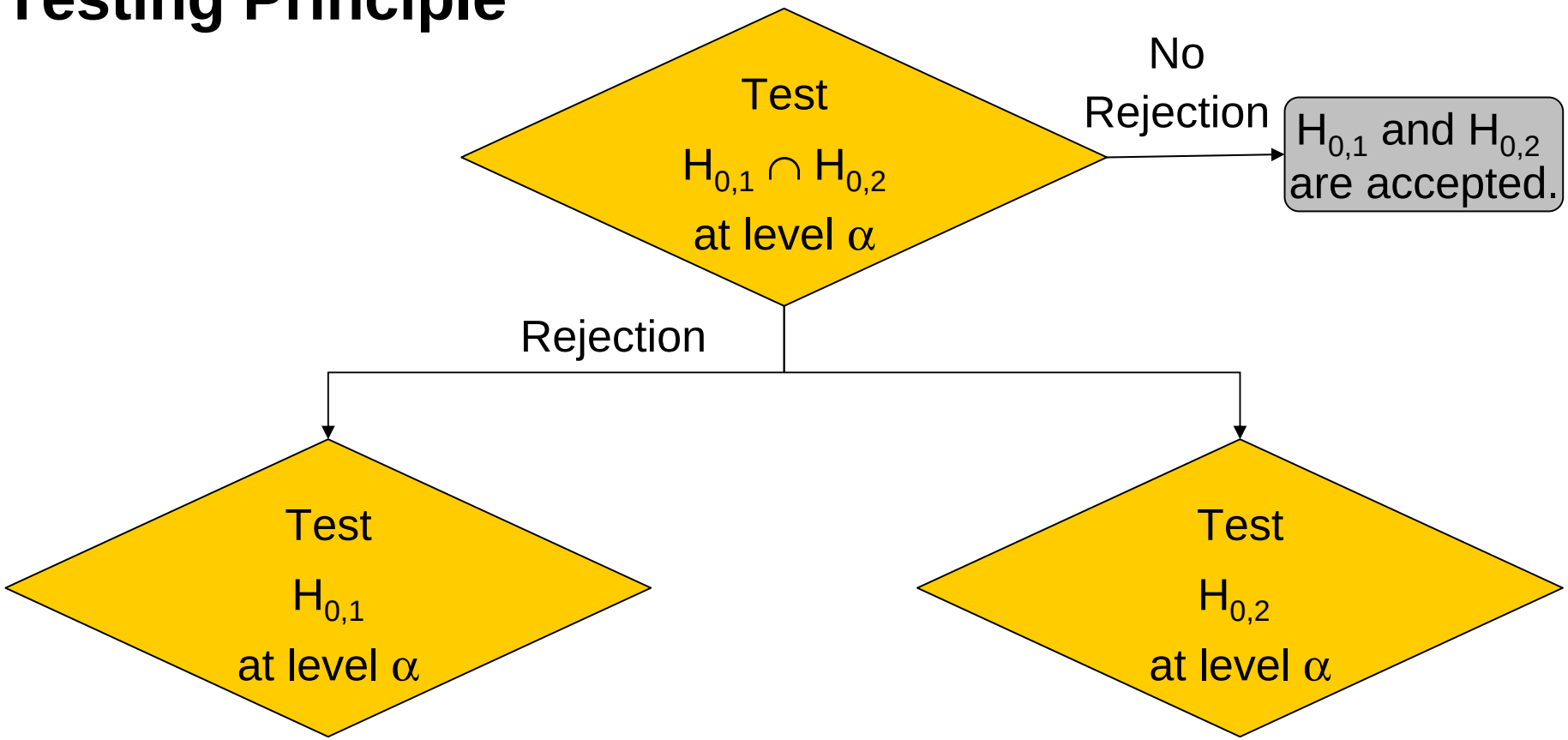
- In the stratification design two hypotheses H_F, H_S are tested.
- Regulatory authorities require the control of the Familywise Error Rate defined as

  The probability to reject at least one true null hypothesis.

- A general, powerful approach is the Closed Testing Principle.

Marcus, R; Peritz, E; Gabriel, KR (1976)

The Closed Testing Principle



The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

The Spiessens-Debois Test for the Intersection Hypothesis

- Let p_S, p_F denote p-values for z-tests of H_F and H_S
- The Spiessens and Debois test rejects the intersection hypothesis $H = H_S \cap H_F$ if

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

- If the intersection hypothesis is rejected, the elementary hypotheses are tested at level α.
- α_F, α_S are chosen such that

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

- The correlation of p_S, p_F and therefore also the levels α_S, α_F depend on λ.
- For fixed α, λ_S and α_S, the level α_F is determined by the above equation.
- If the prevalence increases, the correlation increases and for α_S fixed, α_F increases.

Testing Procedures for Parallel Group Comparison of Means

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F. If a hypothesis is rejected, the other is tested at level α.
- To reject H_F, also the consistency condition

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

for parameters $\tau_S, \tau_{S'}$, must be satisfied.

Classical Design:

H_F is tested with a z-test.

Enrichment Design:

H_S is tested with a z-test.

Testing Procedures for Parallel Group Comparison of Means

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F. If a hypothesis is rejected, the other is tested at level α.
- To reject H_F, also the consistency condition

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

for parameters $\tau_S, \tau_{S'}$, must be satisfied.

Classical Design:

H_F is tested with a z-test.

Enrichment Design:

H_S is tested with a z-test.

Testing Procedures for Parallel Group Comparison of Means

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F. If a hypothesis is rejected, the other is tested at level α.
- To reject H_F, also the consistency condition

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

for parameters $\tau_S, \tau_{S'}$, must be satisfied.

Classical Design:

H_F is tested with a z-test.

Enrichment Design:

H_S is tested with a z-test.

Testing Procedures for Parallel Group Comparison of Means

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F. If a hypothesis is rejected, the other is tested at level α.
- To reject H_F, also the consistency condition

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

for parameters $\tau_S, \tau_{S'}$, must be satisfied.

Classical Design:

H_F is tested with a z-test.

Enrichment Design:

H_S is tested with a z-test.

Testing Procedures for Parallel Group Comparison of Means

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F. If a hypothesis is rejected, the other is tested at level α.
- To reject H_F, also the consistency condition

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

for parameters $\tau_S, \tau_{S'}$, must be satisfied.

Classical Design:

H_F is tested with a z-test.

Enrichment Design:

H_S is tested with a z-test.

The Utility Function

$$U(d) = \underbrace{-\,C(d)}_{\text{Cost}} + \underbrace{\begin{cases} \varphi_{F,d} & \text{if } H_F \text{ is rejected} \\ \varphi_{S,d} & \text{if only } H_S \text{ is rejected} \\ 0 & \text{if no hypothesis is rejected} \end{cases}}_{\text{Reward}}.$$

where, $d \in \mathcal{D}$ denotes a trial design (Classical, Stratification, Enrichment) together with a specific sample size, and testing procedure.

The Rewards

Sponsor view

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

- N ... number of future patients (market size).
- r_F, r_S ... revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ ... efficacy estimates.
- μ_F, μ_S ... clinically relevant thresholds.

Public health view

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

- δ_S, δ_F ... true effect sizes.

The Rewards

Sponsor view

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

- N . . . number of future patients (market size).
- r_F, r_S . . . revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ . . . efficacy estimates.
- μ_F, μ_S . . . clinically relevant thresholds.

Public health view

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

- δ_S, δ_F . . . true effect sizes.

The Rewards

Sponsor view

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

- N ... number of future patients (market size).
- r_F, r_S ... revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ ... efficacy estimates.
- μ_F, μ_S ... clinically relevant thresholds.

Public health view

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

- δ_S, δ_F ... true effect sizes.

The Rewards

Sponsor view

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

- N ... number of future patients (market size).
- r_F, r_S ... revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ ... efficacy estimates.
- μ_F, μ_S ... clinically relevant thresholds.

Public health view

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

- δ_S, δ_F ... true effect sizes.

The Rewards

Sponsor view

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

- N ... number of future patients (market size).
- r_F, r_S ... revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ ... efficacy estimates.
- μ_F, μ_S ... clinically relevant thresholds.

Public health view

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

- δ_S, δ_F ... true effect sizes.

The Rewards

Sponsor view

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

- N ... number of future patients (market size).
- r_F, r_S ... revenue parameters.
- $\hat{\delta}_{F,d}, \hat{\delta}_{S,d}$ ... efficacy estimates.
- μ_F, μ_S ... clinically relevant thresholds.

Public health view

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

- δ_S, δ_F ... true effect sizes.

Trial Costs $C(d)$

- Classical Design

$$c_{\text{setup}} + 2n\, c_{\text{per-patient}}.$$

- Stratified Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + c_{\text{BM determination}}).$$

- Enrichment Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + \frac{c_{\text{BM determination}}}{\lambda_S}).$$

Trial Costs $C(d)$

- Classical Design

$$c_{\text{setup}} + 2n\, c_{\text{per-patient}}.$$

- Stratified Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + c_{\text{BM determination}}).$$

- Enrichment Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + \frac{c_{\text{BM determination}}}{\lambda_S}).$$

Trial Costs $C(d)$

- Classical Design

$$c_{\text{setup}} + 2n\, c_{\text{per-patient}}.$$

- Stratified Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + c_{\text{BM determination}}).$$

- Enrichment Design

$$c_{\text{setup}} + c_{\text{BM development}} + 2n(c_{\text{per-patient}} + \frac{c_{\text{BM determination}}}{\lambda_S}).$$

The Expected Utility of Trial Designs in the Planning Stage

Identifying optimized trial designs

Expected Utility:

$$E_\pi \{E_\Delta [U(d)]\}$$

The expectation is taken over

- the prior π on the effect sizes $\Delta = (\delta_S, \delta_{S'})$ and
- the sampling distribution

Optimal design: Choose the design with maximal expected utility with

- Optimal type of design (classical/stratified/enrichment)
- Optimal sample size
- Optimal α allocation (for the stratified design)

Prior Distributions π on the Effect Sizes $\delta_S, \delta_{S'}$

δ_S	0	θ	θ	θ
$\delta_{S'}$	0	0	$\theta/2$	θ
Weak Biomarker Prior	0.2	0.2	0.3	0.3
Strong Biomarker Prior	0.2	0.6	0.1	0.1

where $\theta > 0$ is an effect size parameter.

Scenario

- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$
- **Consistency parameters** $\tau_S = \tau_{S'} = 0.3.$

Scenario

- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$
- **Consistency parameters** $\tau_S = \tau_{S'} = 0.3.$

Scenario

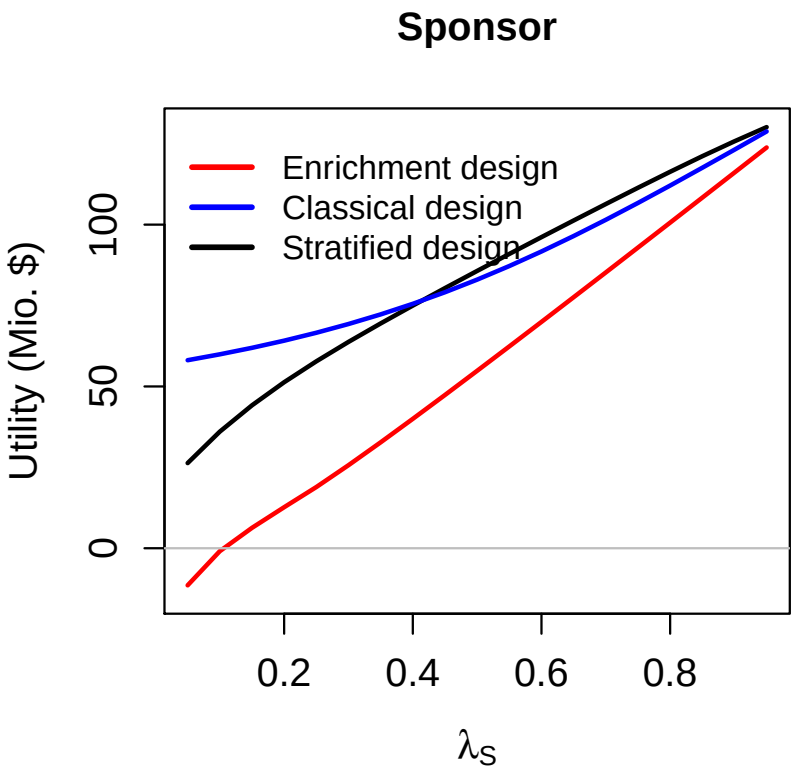
- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$
- **Consistency parameters** $\tau_S = \tau_{S'} = 0.3.$

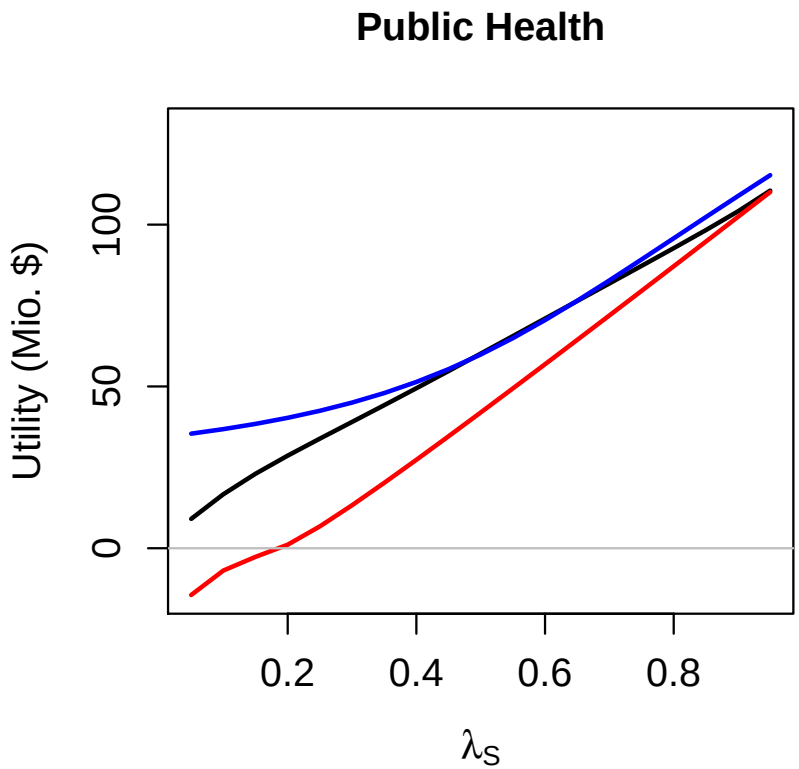
Scenario

- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$
- **Consistency parameters** $\tau_S = \tau_{S'} = 0.3.$

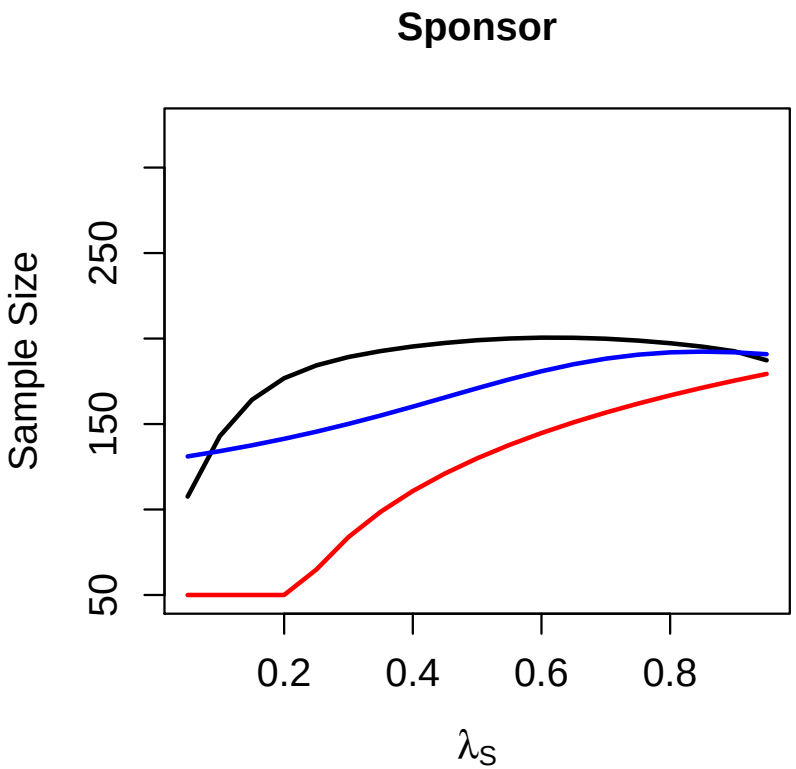
Optimized Expected Utilities

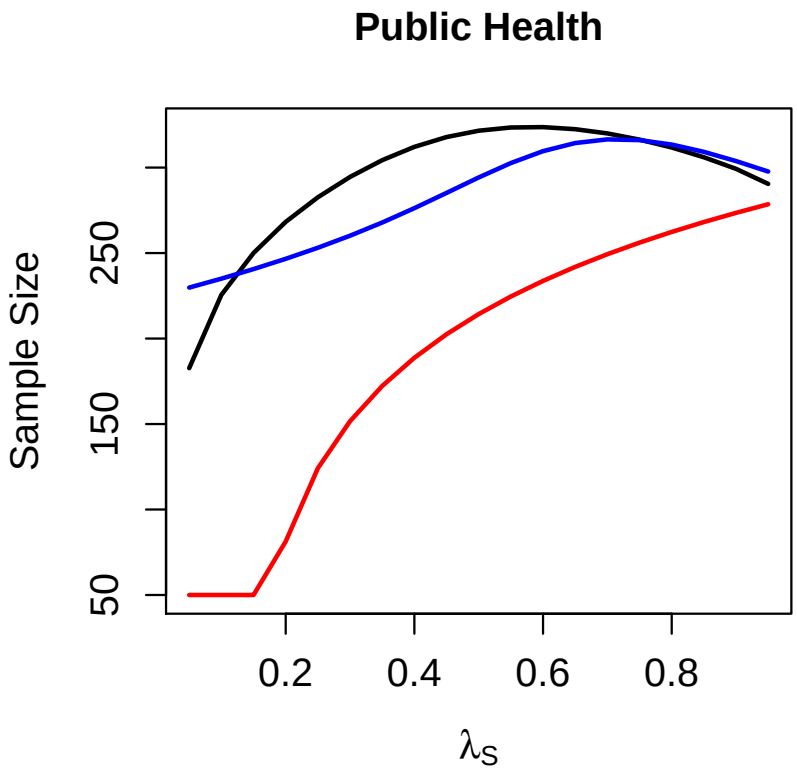
Weak Biomarker Prior





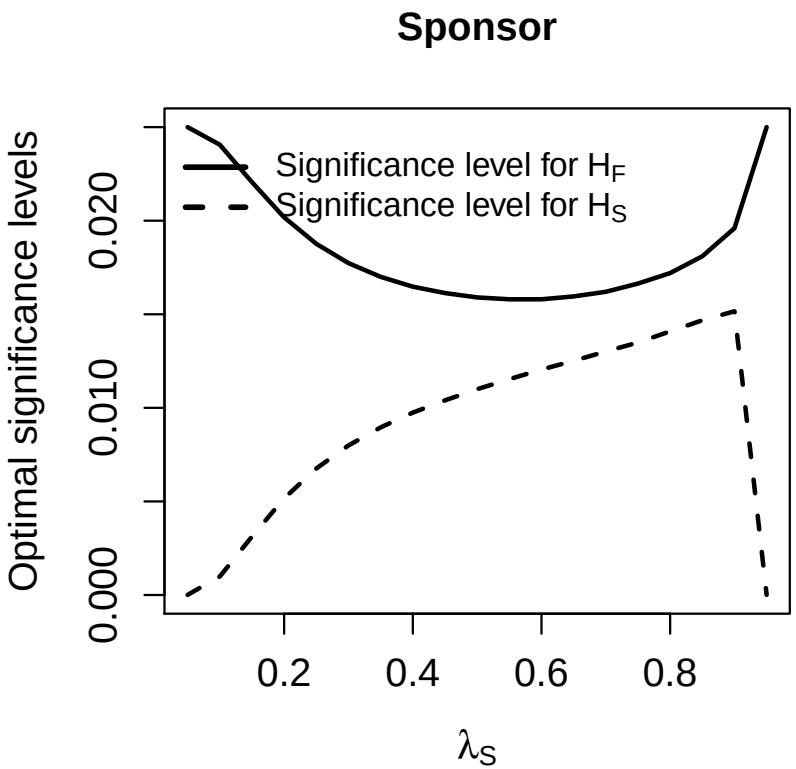
Optimized Sample Size
Weak Biomarker Prior

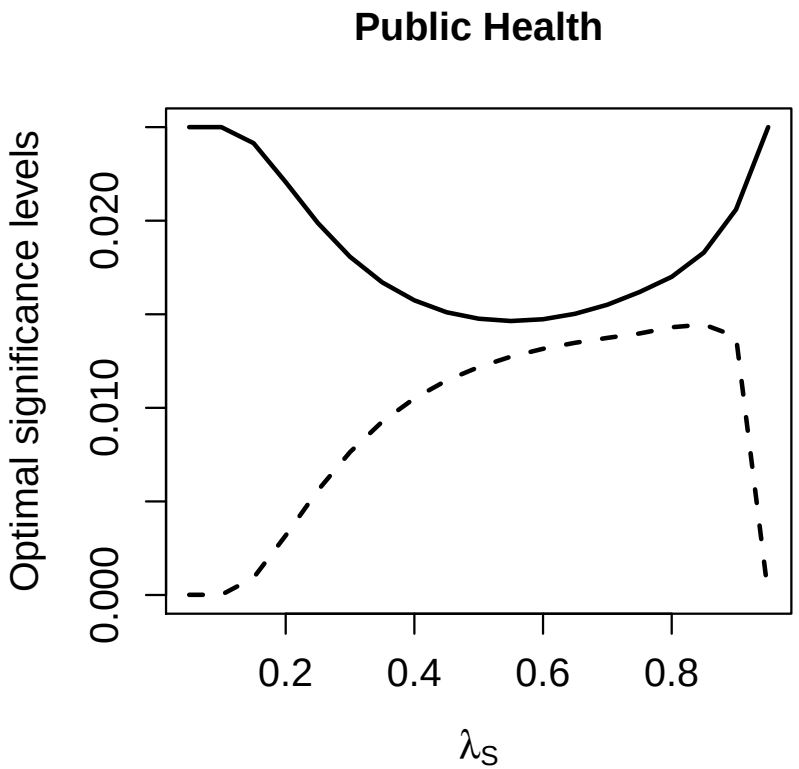




Optimized Alpha Allocation

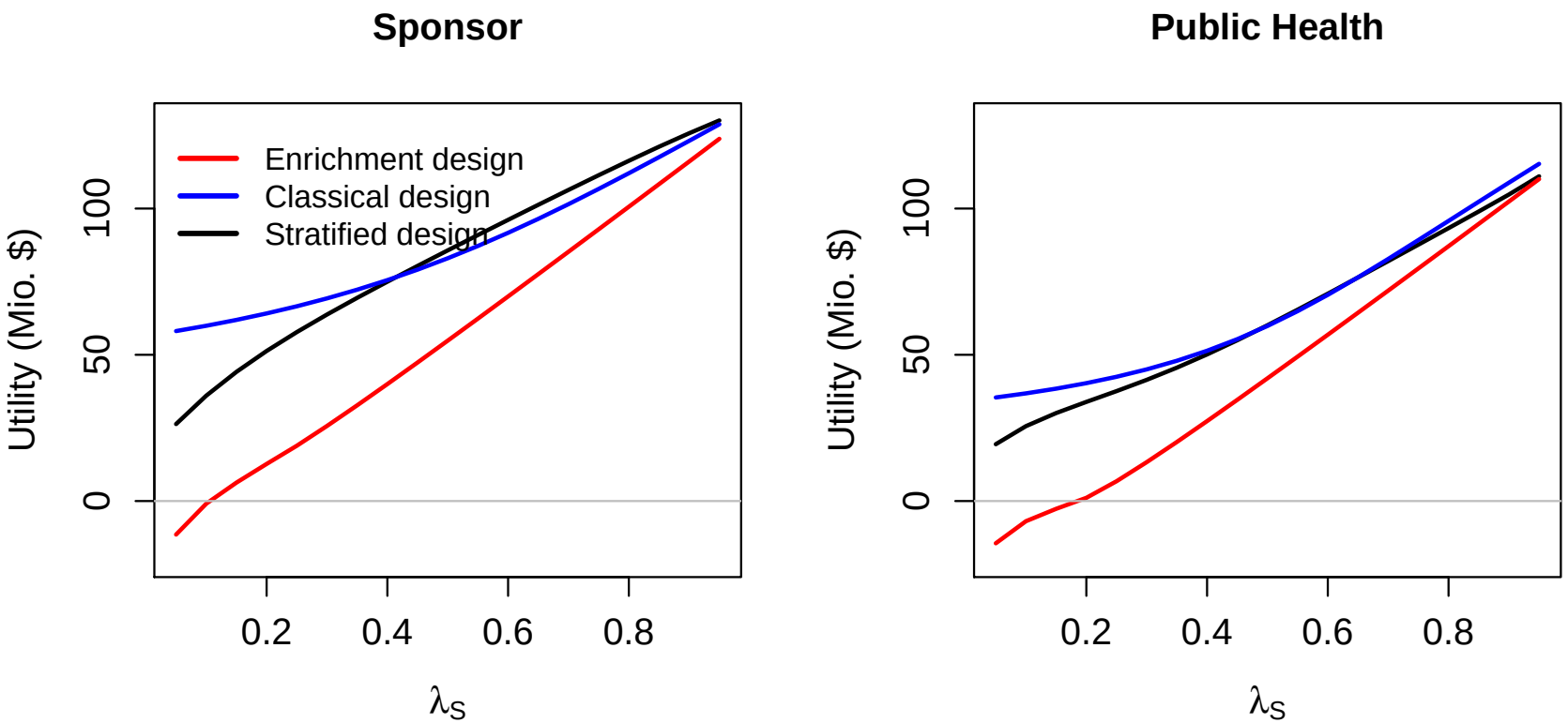
Weak Biomarker Prior





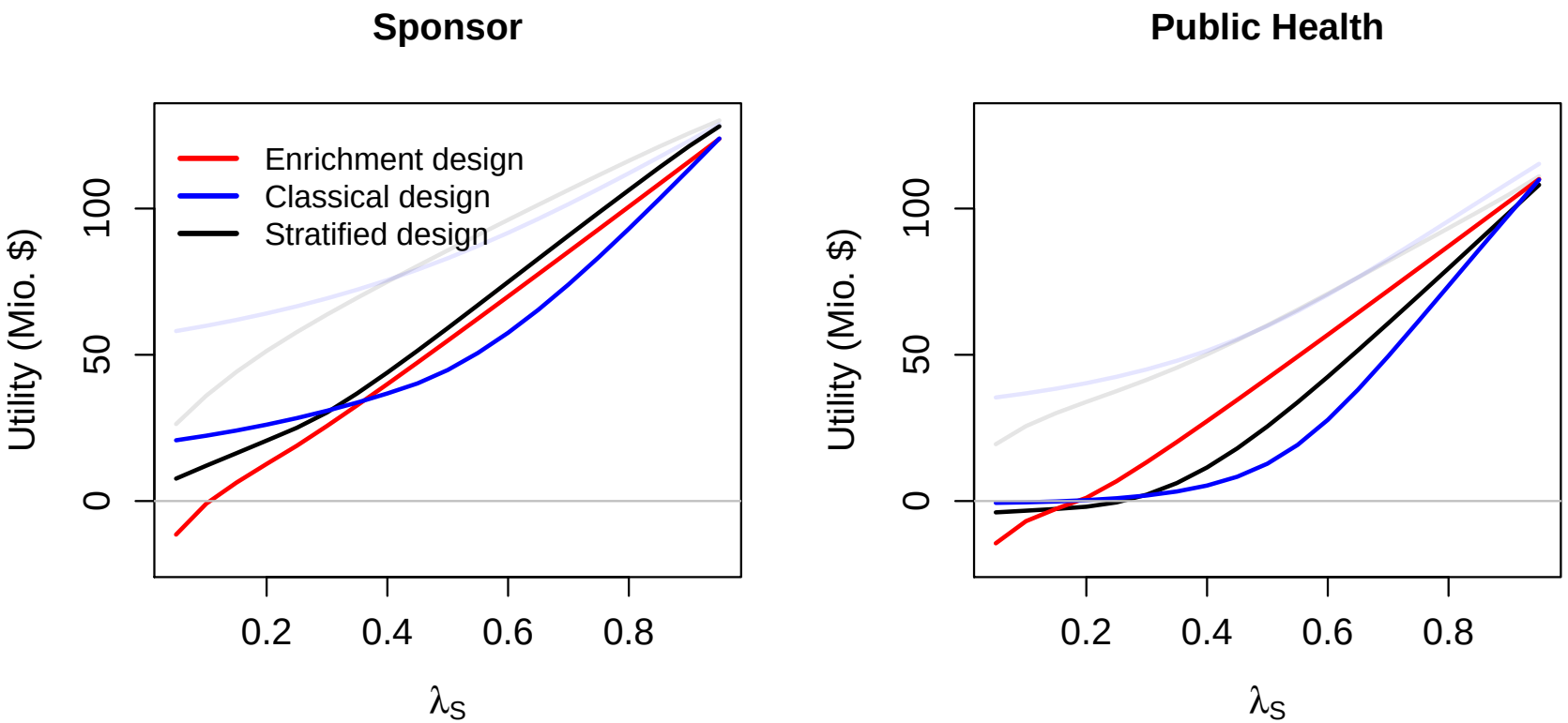
Optimized Expected Utilities – Impact of the Prior

Weak Biomarker Prior

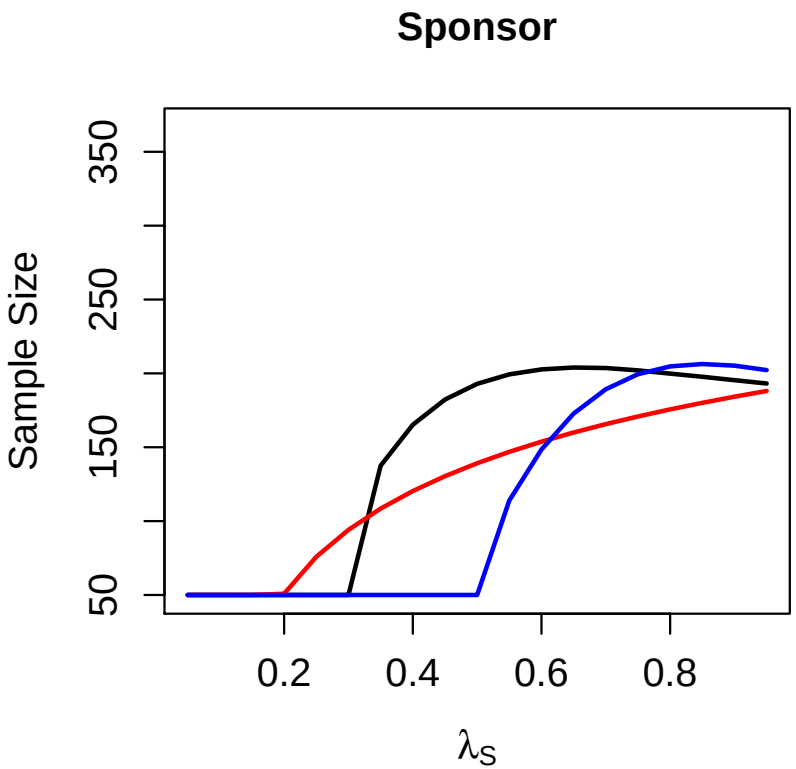


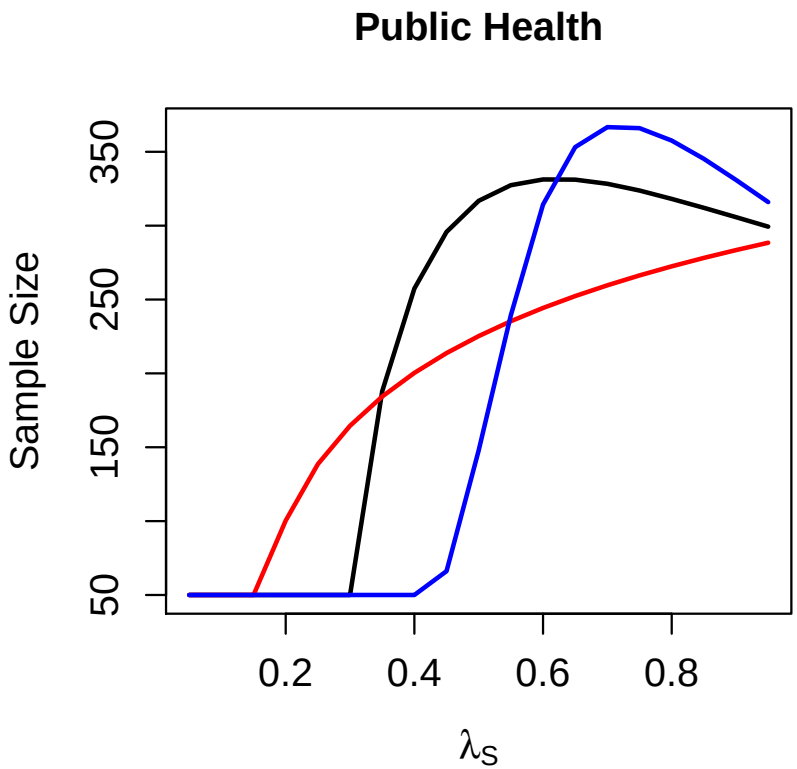
Optimized Expected Utilities – Impact of the Prior

Strong Biomarker Prior



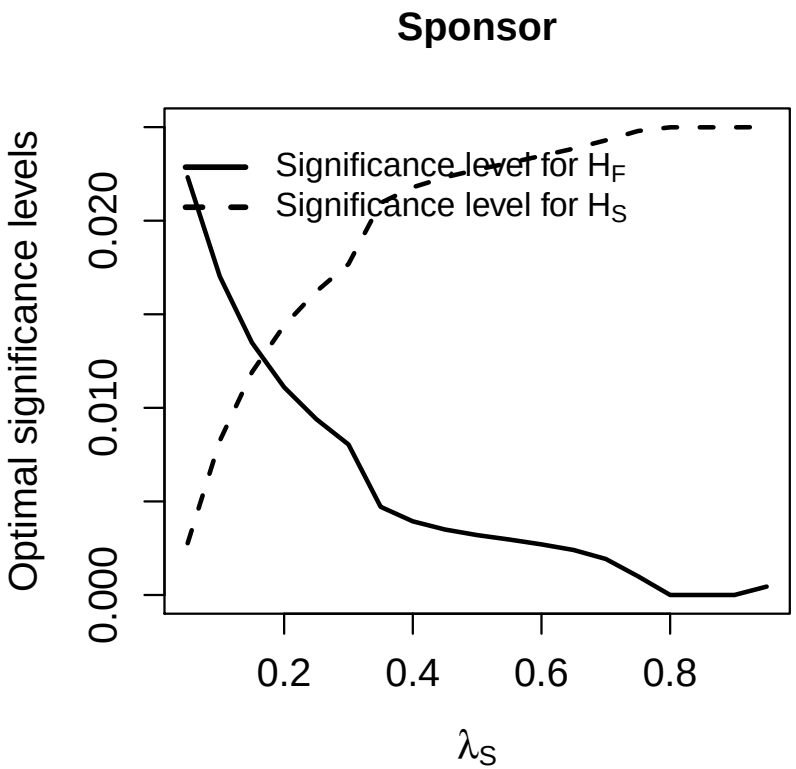
Optimized Sample Size
Strong Biomarker Prior

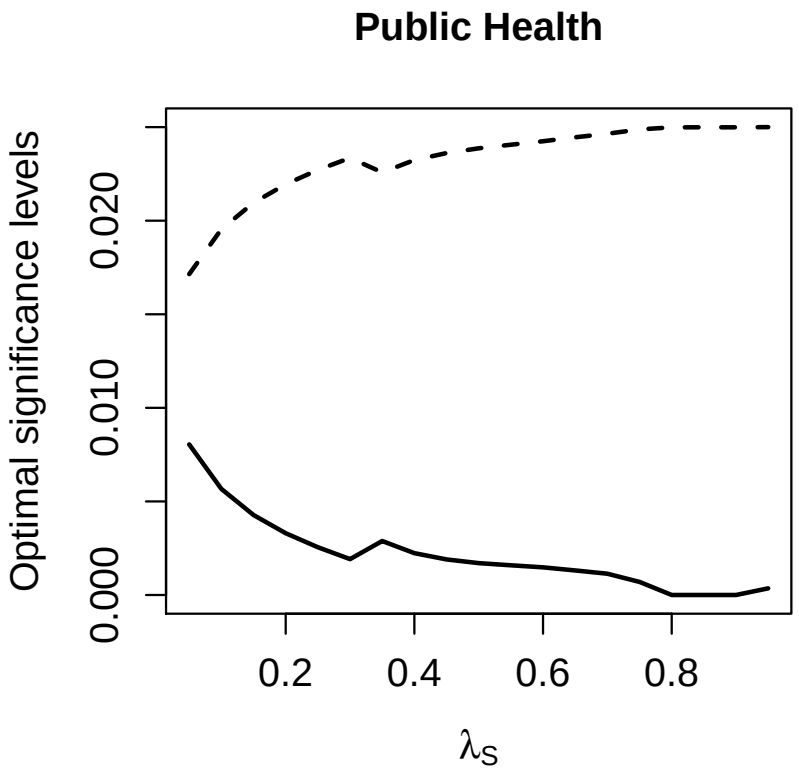




Optimized Alpha Allocation

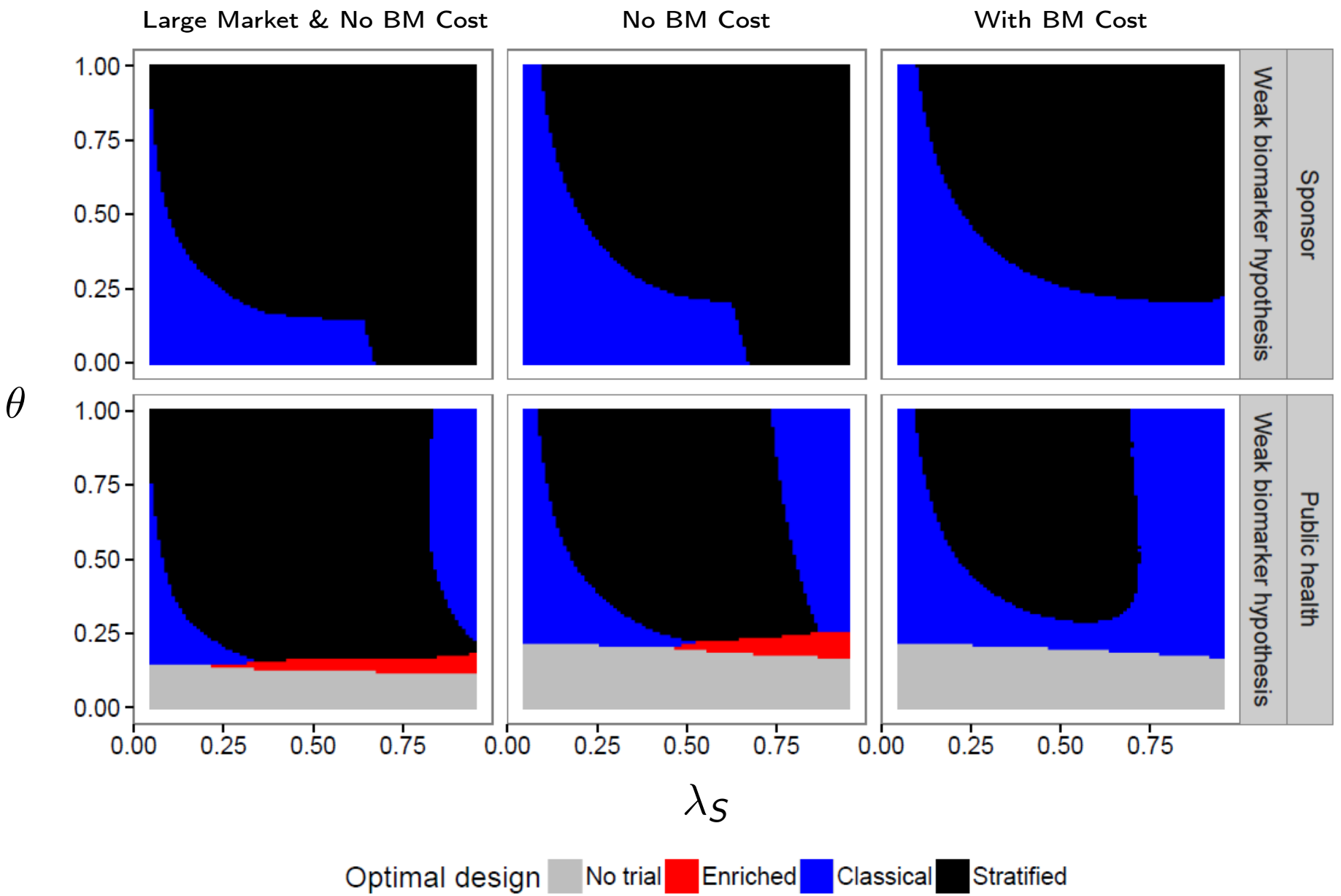
Strong Biomarker Prior





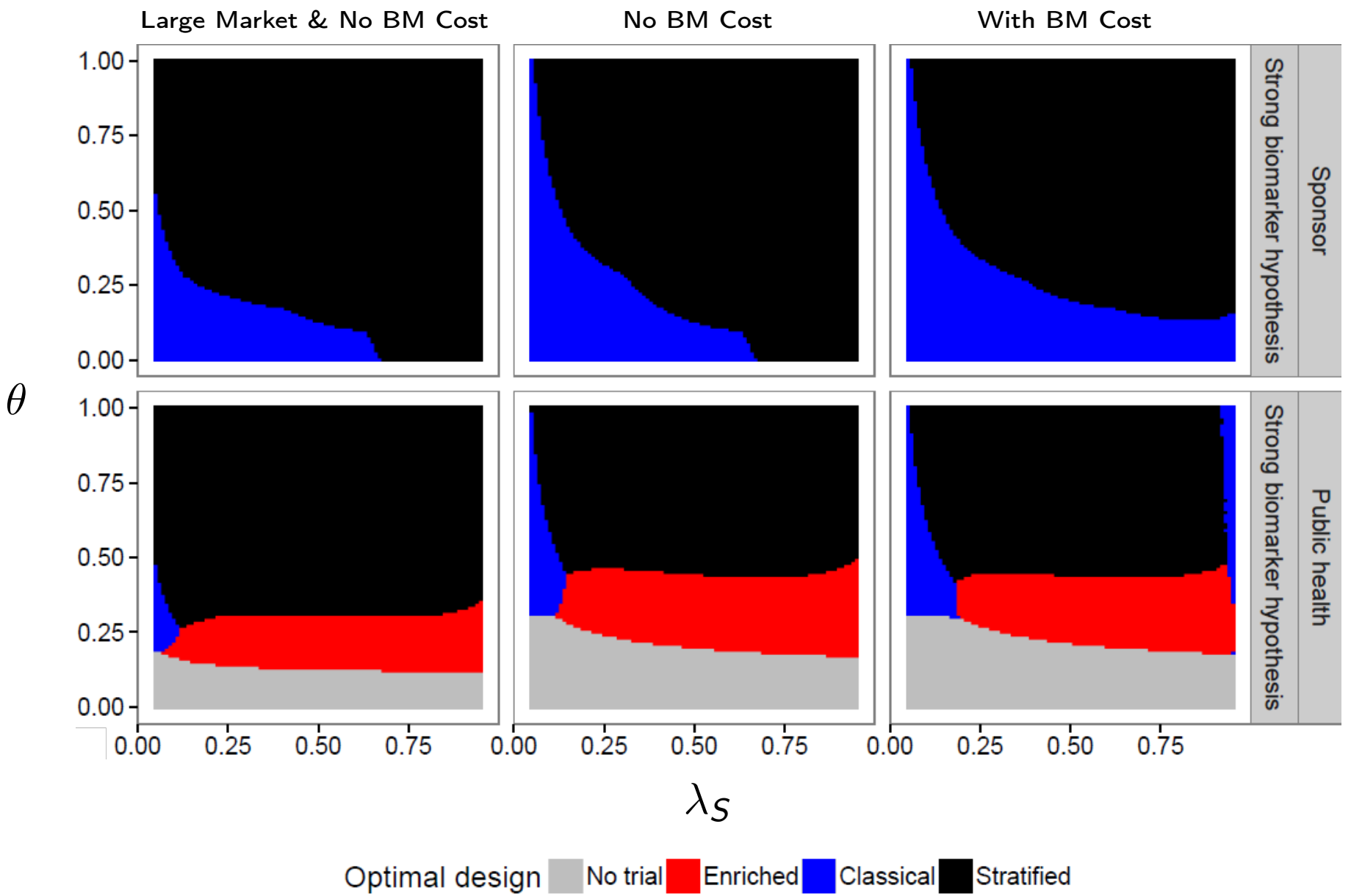
Optimal Trial Designs

Weak Biomarker Prior



Optimal Trial Designs

Strong Biomarker Prior



Some General Observations and Conclusion

- The decision theoretic model can inform the choice of
  - the type of trial design,
  - the sample size and the weights in the multiple test.
- The optimal sample size under the public health view is typically larger than in the sponsor view.
- The enrichment design is never optimal for the sponsor view
- For some scenarios the optimized designs differ, but the expected utilities of different design options are often small.
- The optimal design depends strongly on the particulars of the situation: Subgroup prevalence, trial costs and initial beliefs.

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. Burman, F. König, S. Stallard, and P. M (2016). Optimizing trial designs for targeted therapies.

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.