Optimizing Confirmatory Clinical Trial Designs for Targeted Therapies

A Decision Theoretic Approach

Martin Posch

Section for Medical Statistics, Medical University of Vienna

Institut für Mathematik | Universität Wien
Oktober 2016

Acknowledgement

Optimized Enrichment Designs

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

Two Player Game

Michael Pearce, Nigel Stallard, Simon Day, Frank Miller, Jason Madan, Sahra Zohar,....

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.

Drug Development

Phase 1- Phase IV Clinical Trials:

- **Phase I** Safety, dosage range determination, side effects are tested in a small group.
- **Phase II** Proof of principle to show the drug is effective and to further evaluate its safety.
- **Phase III** Confirmation of the efficacy of a drug, monitoring of side effects.
- **Phase IV** Postmarketing studies (long term use, safety,...)

Drug Development

Phase 1- Phase IV Clinical Trials:

- Phase I: Safety, dosage range determination, side effects are tested in a small group.
- Phase II: Proof of principle to show the drug is effective and to further evaluate its safety.
- Phase III: Confirmation of the efficacy of a drug, monitoring of side effects.
- Phase IV: Postmarketing studies (long term use, safety,...)

The standard Phase III Trial to Demonstrate Efficacy

- Randomize n patients to treatment and control group
- Compare the two groups in a pre-defined primary endpoint
- Perform a hypothesis test with significance level $\alpha = 0.025$ (one-sided)
- The sample size n is chosen to achieve a power of 80% or 90% for a *minimal clinical relevant effect size*.

Example z-Test

- Consider an approximately normally distributed endpoint X (larger values beneficial).
- $$X_{i,C} \sim N(\nu_C, \sigma^2), \quad X_{i,T} \sim N(\nu_T, \sigma^2), i = 1, \ldots, n$$
  observations in the treatment and control group ($2n$ patients in total).
- Test the hypotheses
$$H_0 : \nu_T = \nu_C \text{ versus } H_1 : \nu_T > \nu_C .$$
- Assuming that σ known H_0 is rejected if
$$Z_n = \frac{\bar{X}_C - \bar{X}_T}{\sigma} \sqrt{n/2} > z,$$
  where z is a critical value and $\bar{X}_k = \frac{1}{n} \sum_{i=1}^{n} X_{i,k}$, $k = C, T$.
- If $z = z_{1-\alpha}$, the $1 - \alpha$ quantile of the standard normal distribution, the type I error rate is α.

Hypothesis Testing in Phase III Clinical Trials

Regulator Sets the required critical value z.
A standard requirement is a one-sided significance level of $\alpha = 0.025$ which corresponds to $z = 1.96$.

Payer Sets the price p (given the treatment is approved)

Sponsor Chooses the sample size $2n$ of the trial.
E.g.,

$$n = 2\sigma^2 \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\delta^2}$$

for a power of $1 - \beta = 0.8$ or 0.9, where δ is a minimal clinically relevant treatment effect.

However, the values for α, β are ad-hoc choices.

Payoffs for Regulators and Payers

Regulator/Payer payoff

$$P_R = \begin{cases} N(\delta - p - r) & \text{if } Z_n \geq z \\ 0 & \textit{else} \end{cases}$$

Sponsor payoff

$$P_S = \begin{cases} Np - 2n\,c & \text{if } Z_n \geq z \\ -2n\,c & \textit{else} \end{cases},$$

where

- N The size of the future population
- $2n$ Trial Sample size
- p The price of the drug to treat one patient
- δ The true treatment effect ($\delta = \nu_T - \nu_C$).
- c the cost to recruit one patient into the trial
- r Additional cost to the regulator.

Expected Utilities

At the planning stage

- the distribution of Z given δ is known;
- we assume that both players have the same prior belief on the effect size δ, given by a prior distribution $N(\mu_0, \sigma_0)$.

Therefore, given p, z, n we can compute the expected utilities

$$U_R(p,z,n) \quad := \quad E_\delta E_Z(P_R) = N\left[(\mu_0 - r - p)\Phi(-\xi) + (\sigma_0^2/\sigma_x)\phi(\xi)\right]$$

$$U_S(p,z,n) \quad := \quad E_\delta E_Z(P_S) = N\, p\, \Phi(-\xi) - 2cn$$

where $\sigma_x^2 = \sigma_0^2 + 2\sigma^2/n$ and $\xi = (2z\sigma/n - \mu_0)/\sigma_x$ and Φ, ϕ are the standard normal CDF and density.

Optimized Strategies

The optimal n^* for the Sponsor, given z and p is

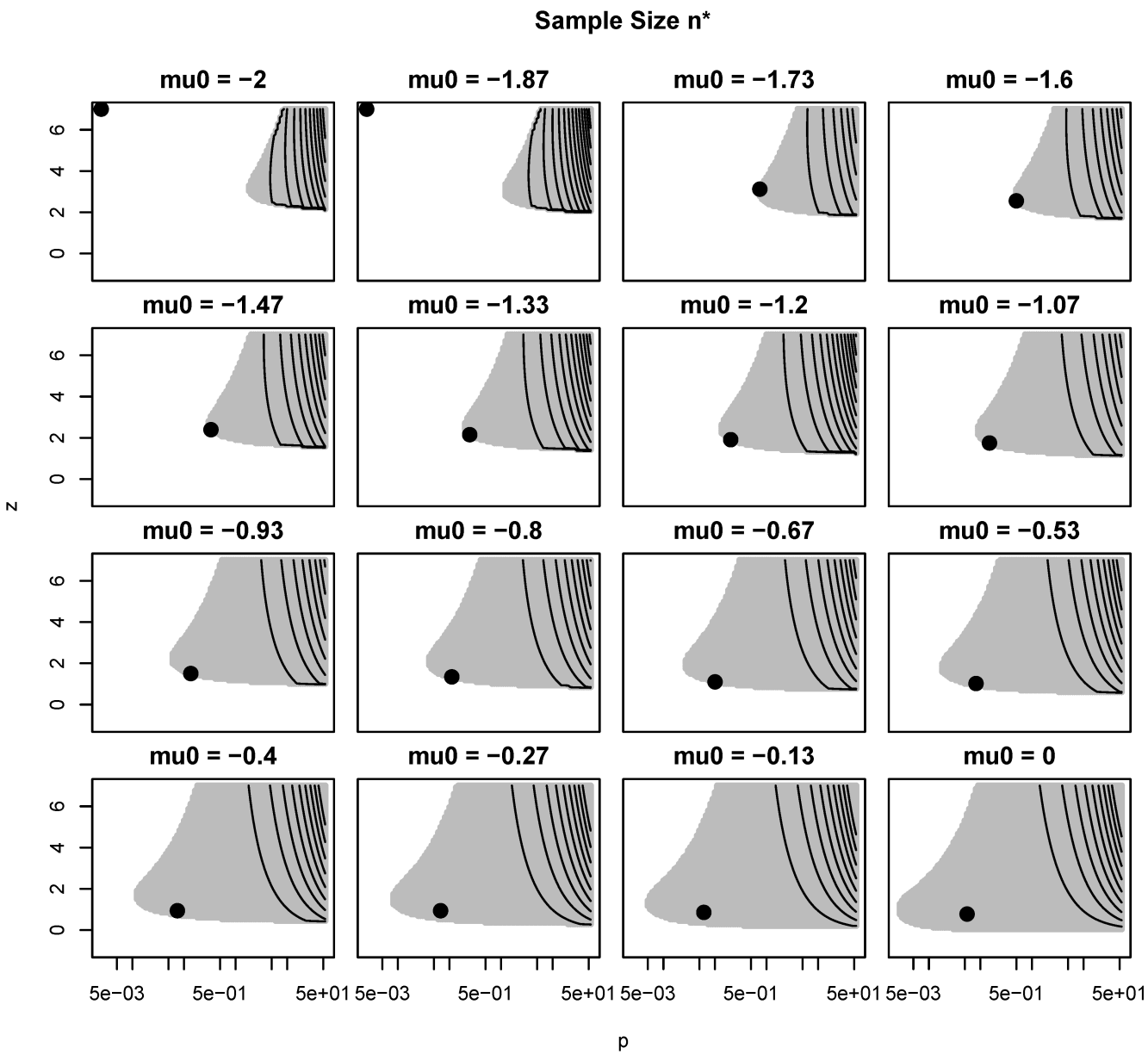
$$n^*(p, z) = argmax_n U_S(p, z, n)$$

The optimal z and p is given by

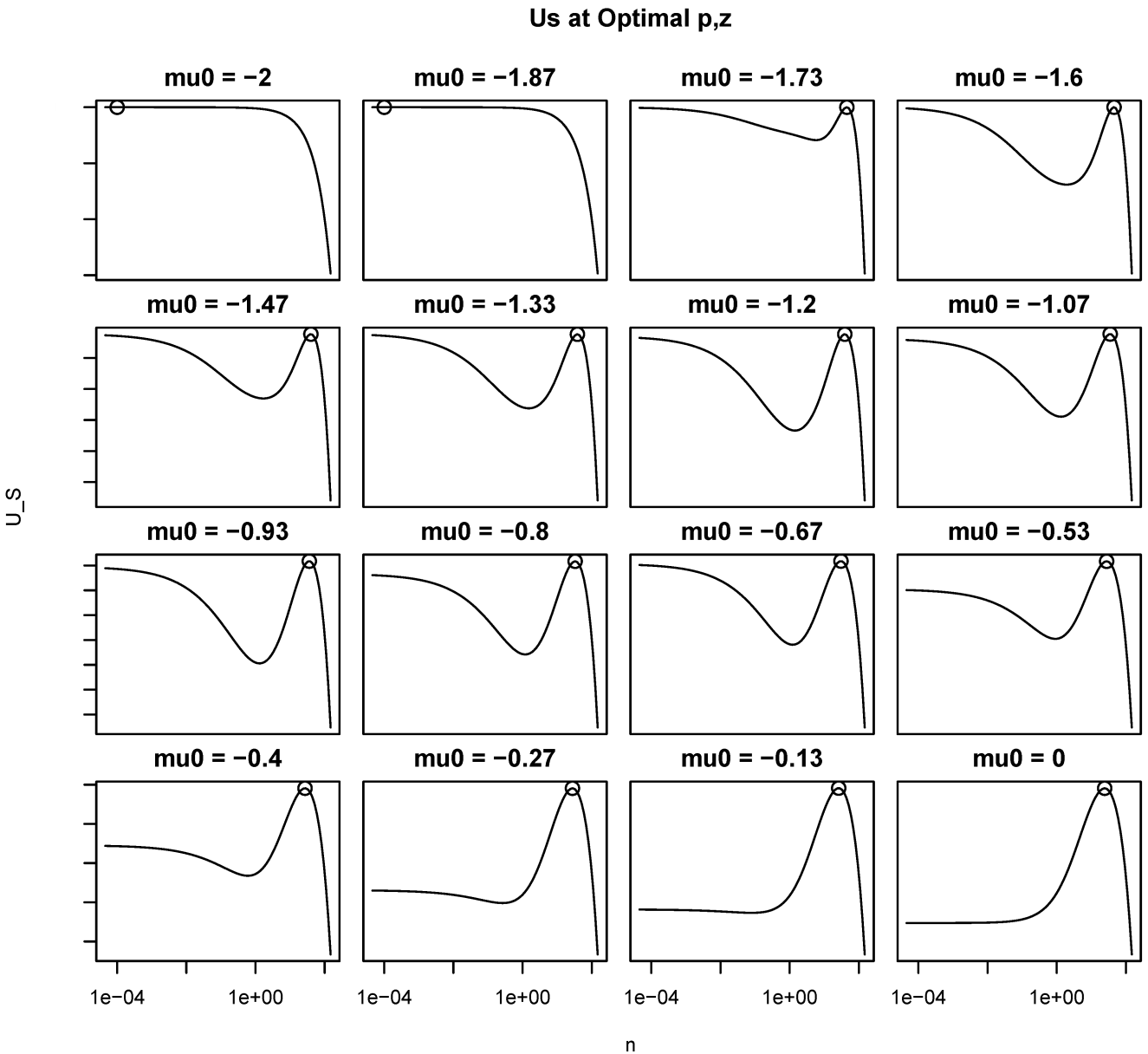
$$(p^*, z^*) = argmax_{p,z} U_R(p, z, n^*(z, p))$$

Numerical example: $N = 10000$ $\sigma = 2.12$, $\sigma_0 = 1$, $c = 0.5$.

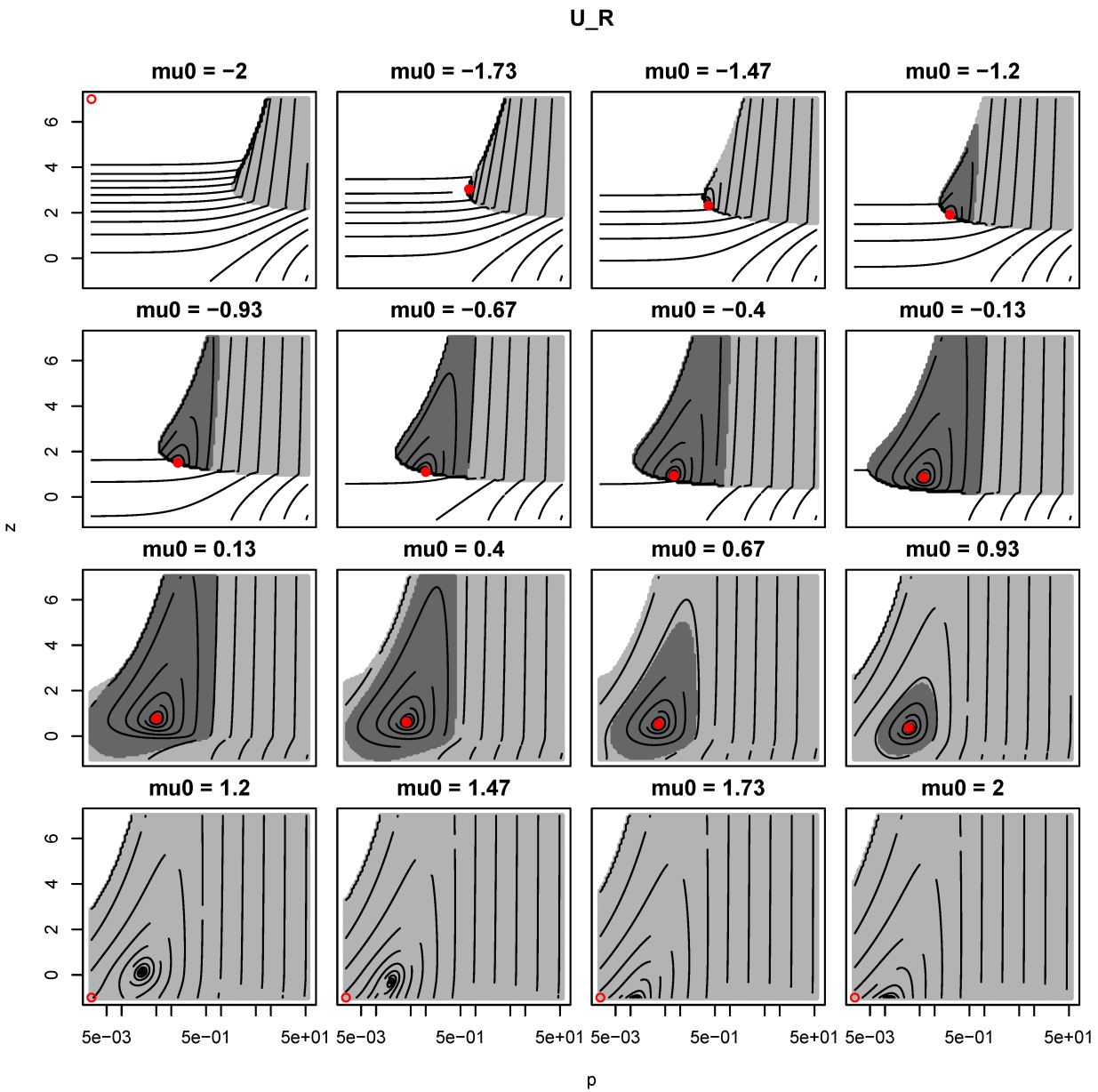
Optimal Sample Size Chosen by the Sponsor $n^*(z, p)$



Sponsor's Utility $U_S(p^*, z^*)$ as Function of n



Regulator's Utility U_R if Sponsor Chooses $n^*(p, z)$



□ $n* = 0$, ■ $n^* > 0$ and $U_R < u$, ■ $n^* > 0$ and $U_R > u$, where $u = \max(U_R(\infty, 0), U_R(-\infty, 0))$

Conclusion

The optimal rule for the Regulator/Payer depends on the prior:

Very Sceptical $z^* = \infty$, $p^* = 0 \Rightarrow$ No trial is performed.

Sceptical Intermediate z^* and p^*: Local optima of U_S at $n = 0$ and some $n > 0$.

Optimistic Intermediate z^* unique local optimum > 0.

Very Optimistic $z = -\infty$ and $p = 0$: Sponsor makes no profit.

Clinical Trials for "Precision Medicine"

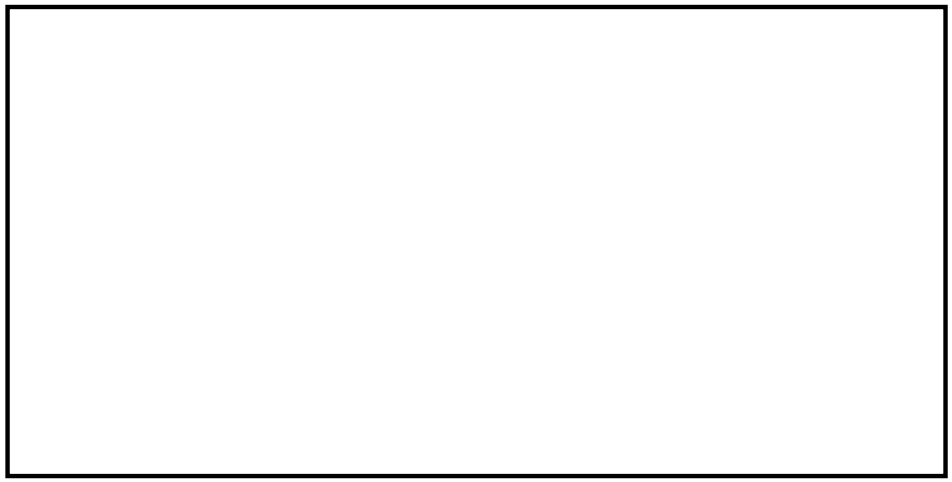
For the development of targeted therapies, clinical trials with complex objectives, confirming treatment effects in sub-populations and/or in the overall populations are required.

- 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

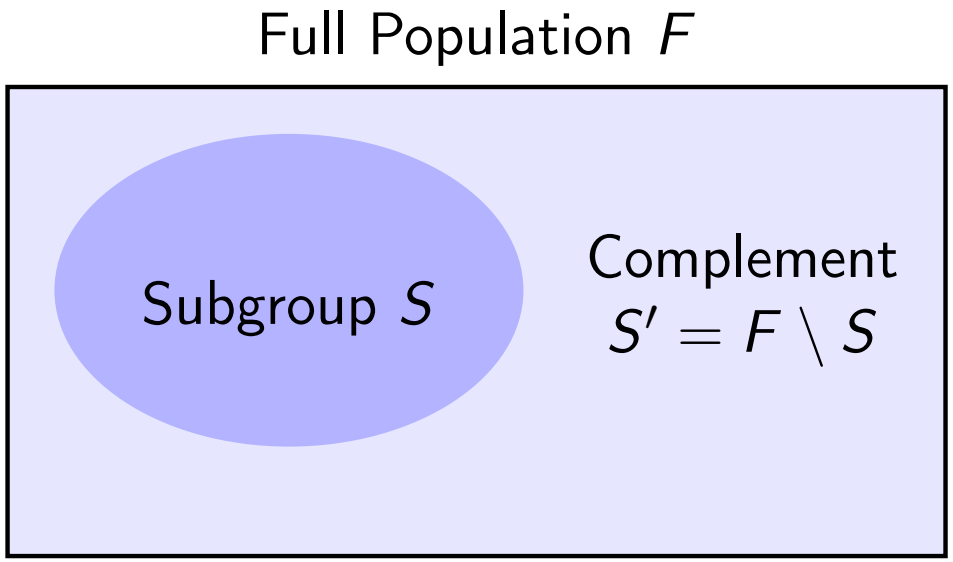


- Assume the treatment effects in the subgroups satisfy $\delta_S \geq \delta_{S'}$.
- The treatment effect in F is

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

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

Patient Populations

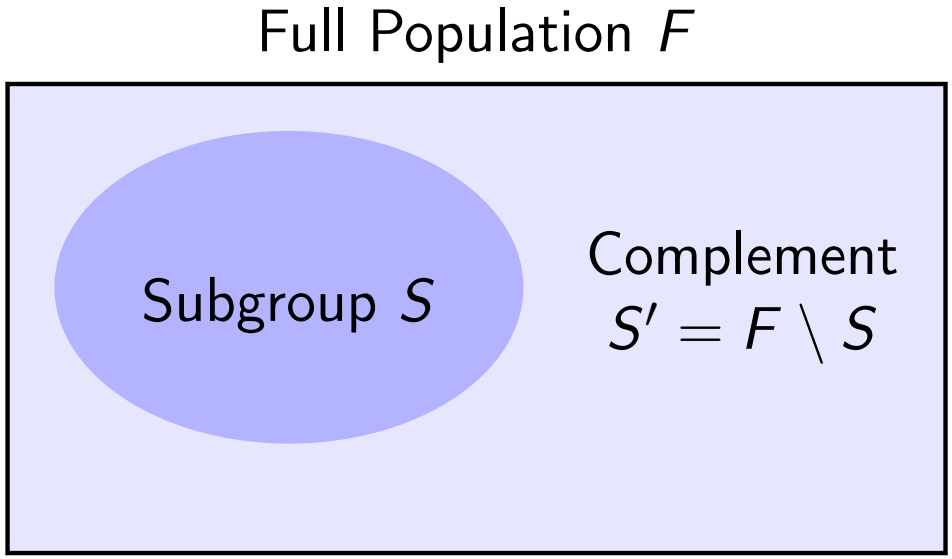


- Assume the treatment effects in the subgroups satisfy $\delta_S \geq \delta_{S'}$.
- The treatment effect in F is

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

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

Patient Populations

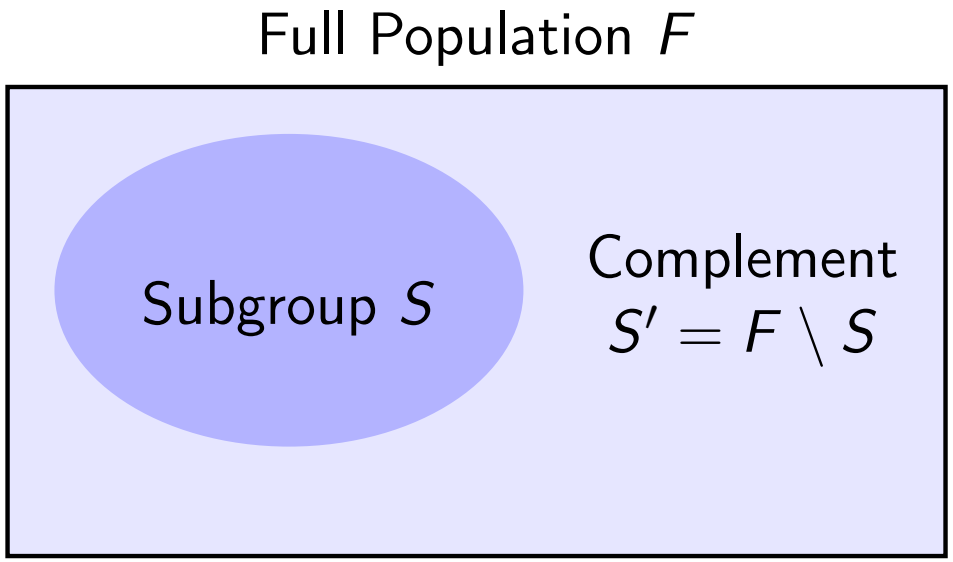


- Assume the treatment effects in the subgroups satisfy $\delta_S \geq \delta_{S'}$.
- The treatment effect in F is

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

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

Patient Populations

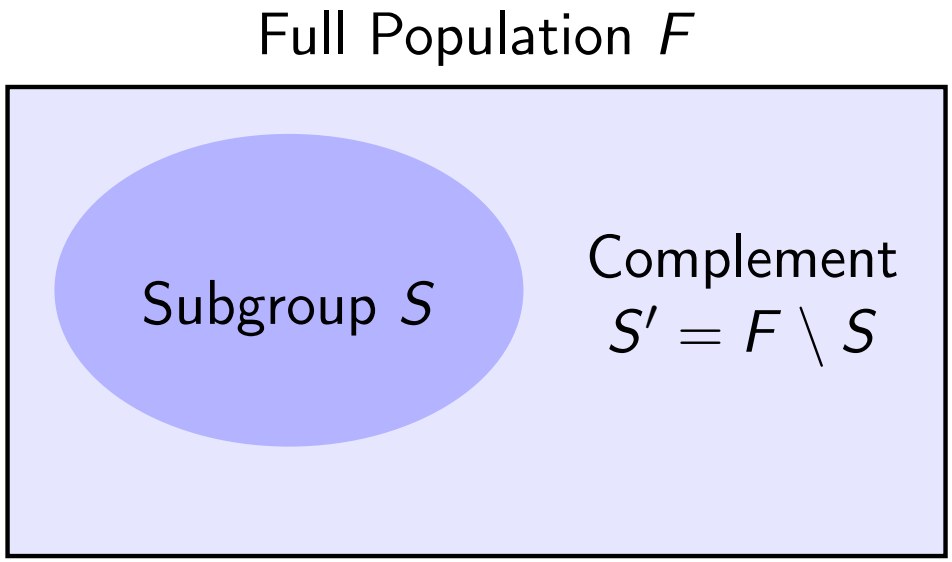


- Assume the treatment effects in the subgroups satisfy $\delta_S \geq \delta_{S'}$.
- The treatment effect in F is

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

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

Patient Populations



- Assume the treatment effects in the subgroups satisfy $\delta_S \geq \delta_{S'}$.
- The treatment effect in F is

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

  where λ_S is the prevalence of subgroup 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:

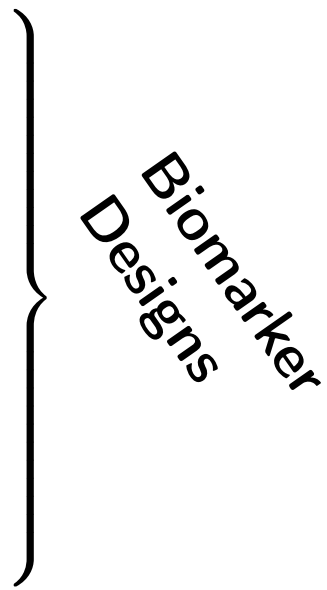
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

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:

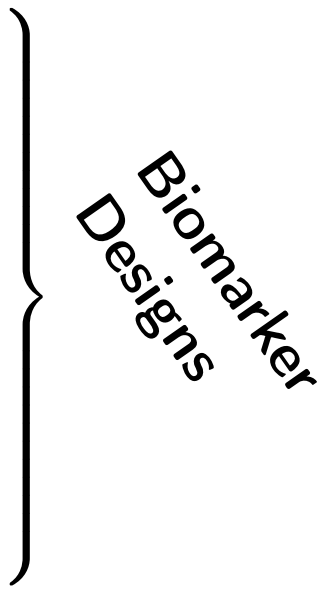
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

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:

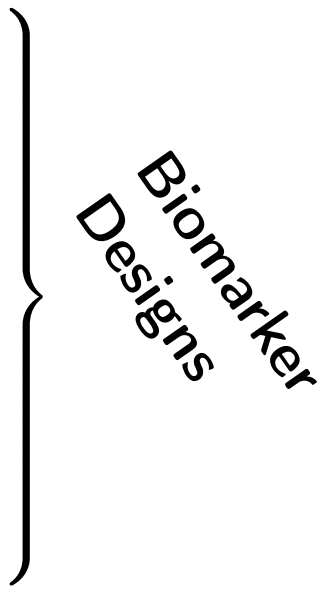
Recruitment from population F.
No Biomarker is determined.
Test of H_F.

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



Testing Procedures for Parallel Group Comparison of Means

Classical Design:

H_F is tested with a z-test at level $\alpha = 0.025$.

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F such that the overall type I error rate is $\alpha = 0.025$.
- 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.

Enrichment Design:

H_S is tested with a z-test at level $\alpha = 0.025$.

Testing Procedures for Parallel Group Comparison of Means

Classical Design:

H_F is tested with a z-test at level $\alpha = 0.025$.

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F such that the overall type I error rate is $\alpha = 0.025$.
- 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.

Enrichment Design:

H_S is tested with a z-test at level $\alpha = 0.025$.

Testing Procedures for Parallel Group Comparison of Means

Classical Design:

H_F is tested with a z-test at level $\alpha = 0.025$.

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F such that the overall type I error rate is $\alpha = 0.025$.
- 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.

Enrichment Design:

H_S is tested with a z-test at level $\alpha = 0.025$.

Testing Procedures for Parallel Group Comparison of Means

Classical Design:

H_F is tested with a z-test at level $\alpha = 0.025$.

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F such that the overall type I error rate is $\alpha = 0.025$.
- 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.

Enrichment Design:

H_S is tested with a z-test at level $\alpha = 0.025$.

Testing Procedures for Parallel Group Comparison of Means

Classical Design:

H_F is tested with a z-test at level $\alpha = 0.025$.

Stratification Design:

- H_S and H_F are tested with a closed Spiessens-Debois (2010) test at levels α_S, α_F such that the overall type I error rate is $\alpha = 0.025$.
- 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.

Enrichment Design:

H_S is tested with a z-test at level $\alpha = 0.025$.

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 significance levels α_F and α_S for H_F and H_S in the weighted multiple test for the stratified design are optimal?

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 for a design d

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

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 optimizing over

- Type of design (classical/stratified/enrichment)
- Sample size
- α allocation (for the stratified design)

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 optimizing over

- Type of design (classical/stratified/enrichment)
- Sample size
- α allocation (for the stratified design)

Prior Distributions π On the Effects $\delta_S, \delta_{S'}$

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

$\theta > 0 \ldots$ 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

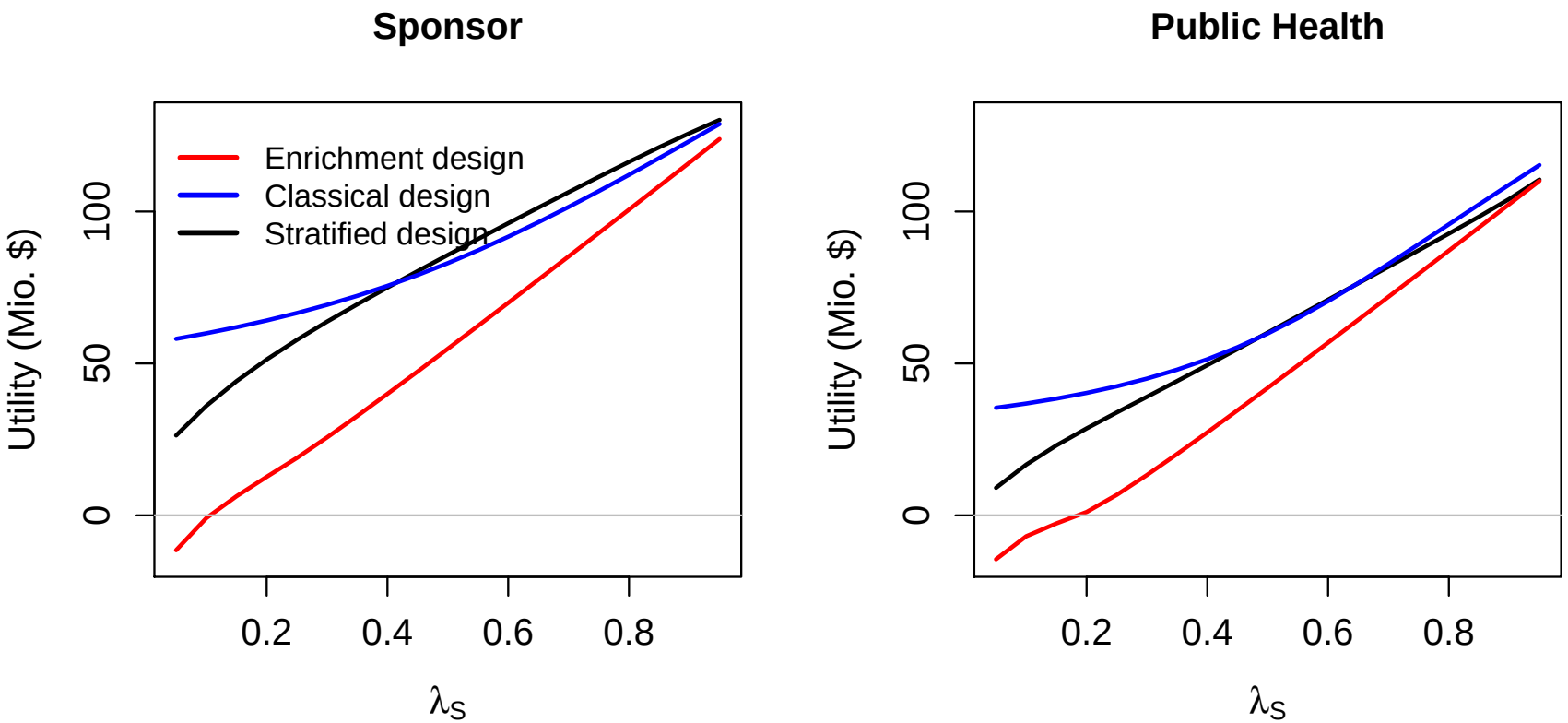
- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$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

Scenario

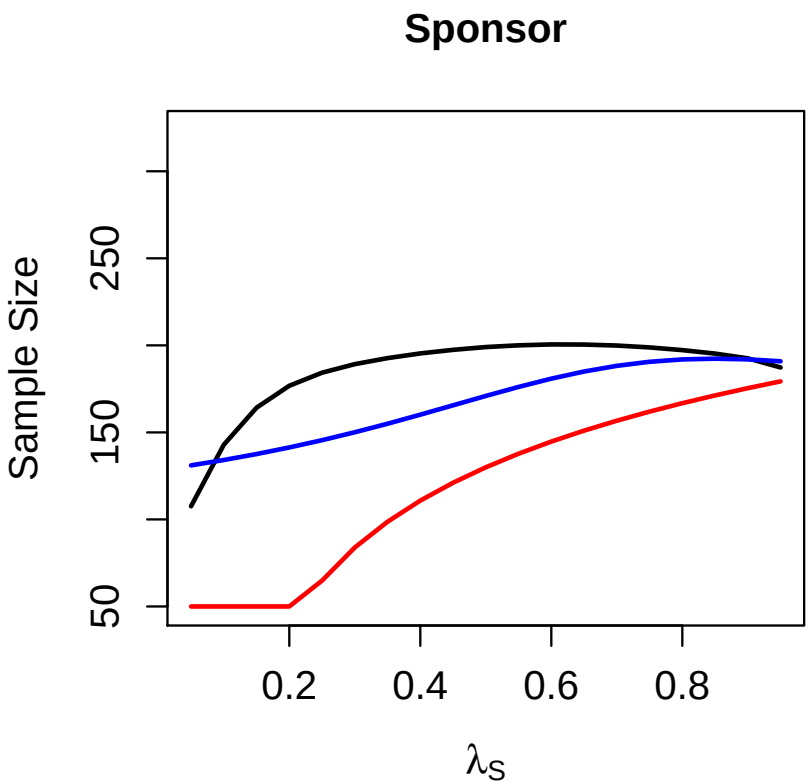
- **Effect size parameter in the prior**
  $\theta = 0.3$
- **Reward parameters**
  $Nr_F = Nr_S = 1000\text{MUSD}$
  $\mu_F = \mu_S = 0.1.$
- **Cost Parameters in (MUSD)**
  $c_{\text{setup}} = 1$
  $c_{\text{per-patient}} = 0.05$
  $c_{\text{BM development}} = 1$
  $c_{\text{BM determination}} = 0.005.$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

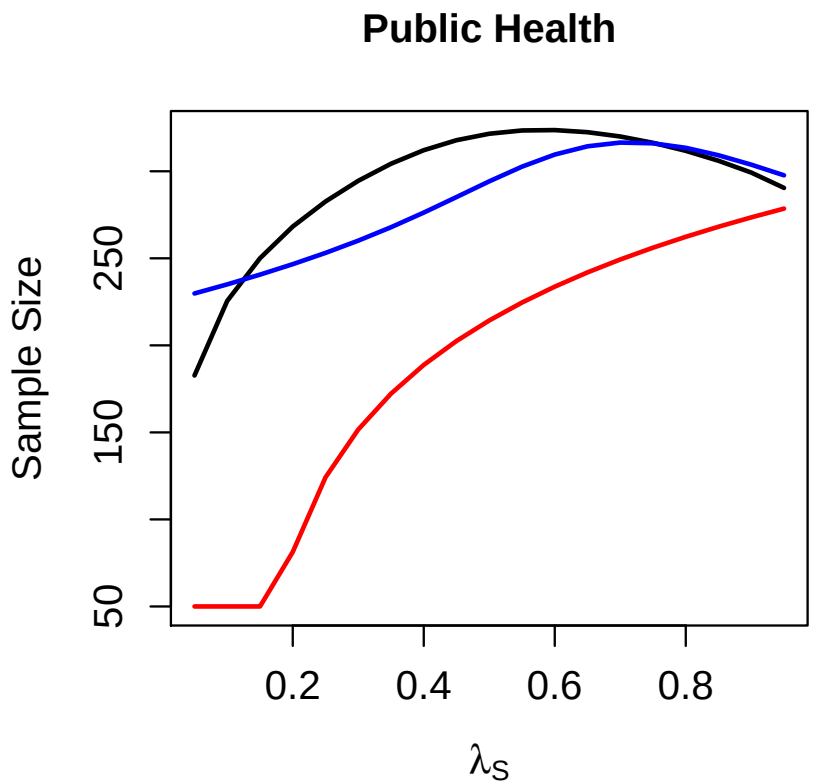
Optimized Expected Utilities

Weak Biomarker Prior



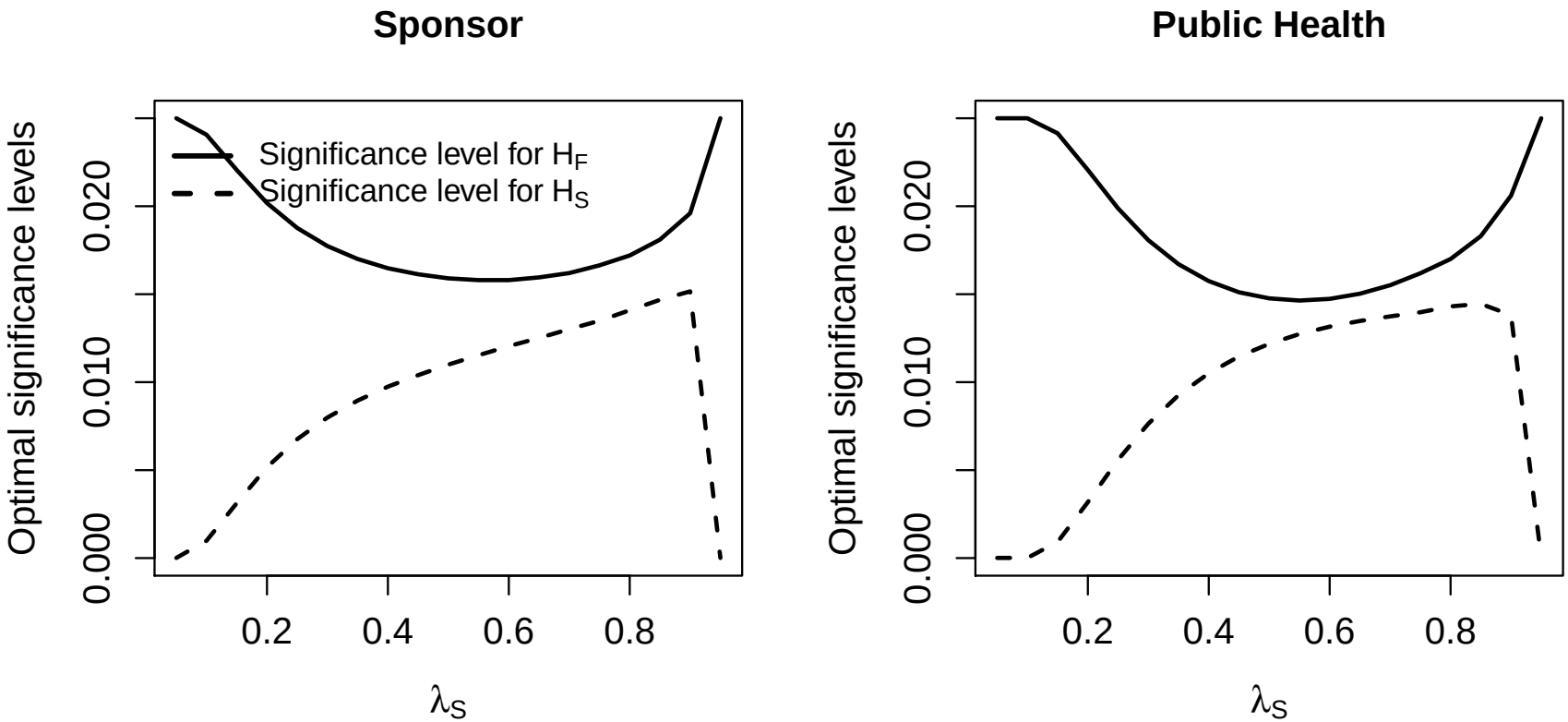
Optimized Sample Size
Weak Biomarker Prior



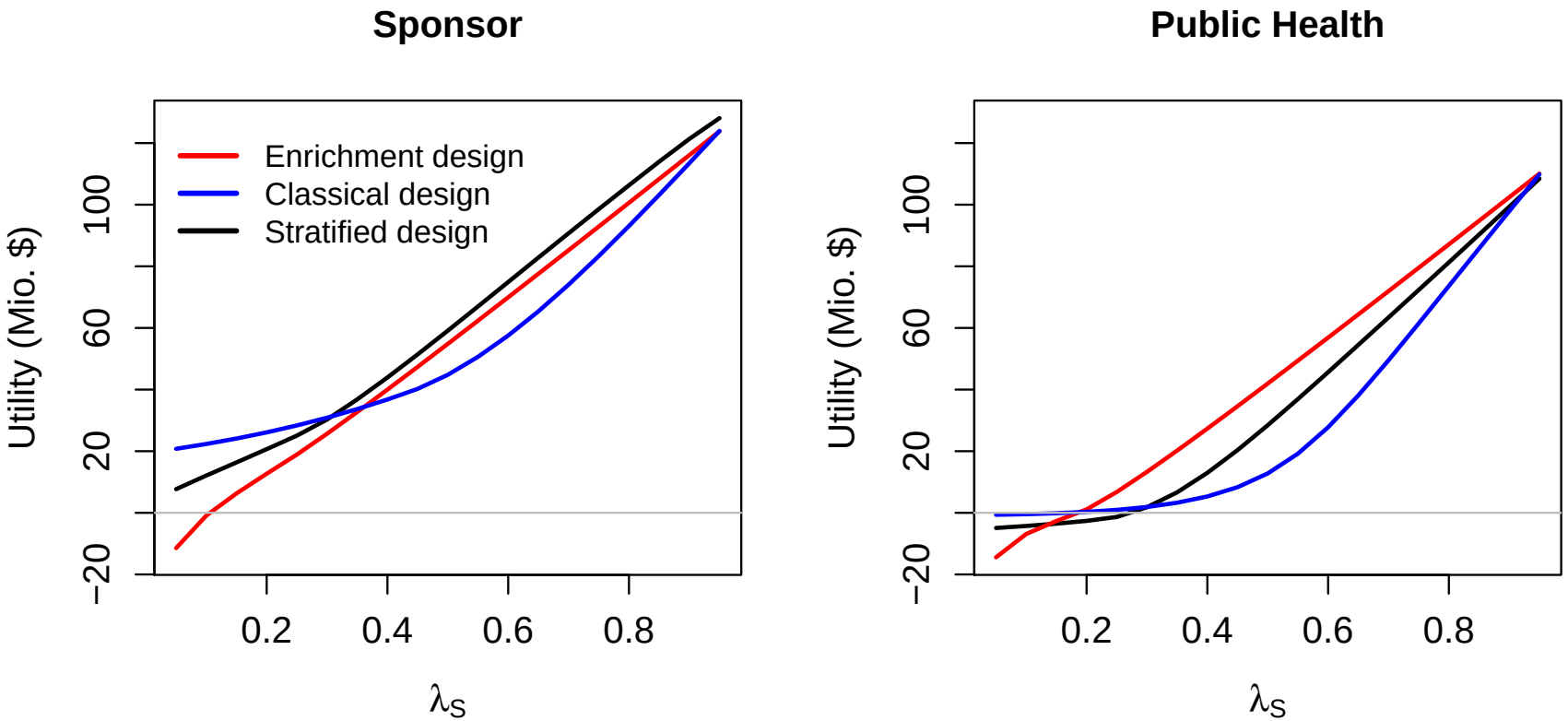


Optimized Alpha Allocation

Weak Biomarker Prior

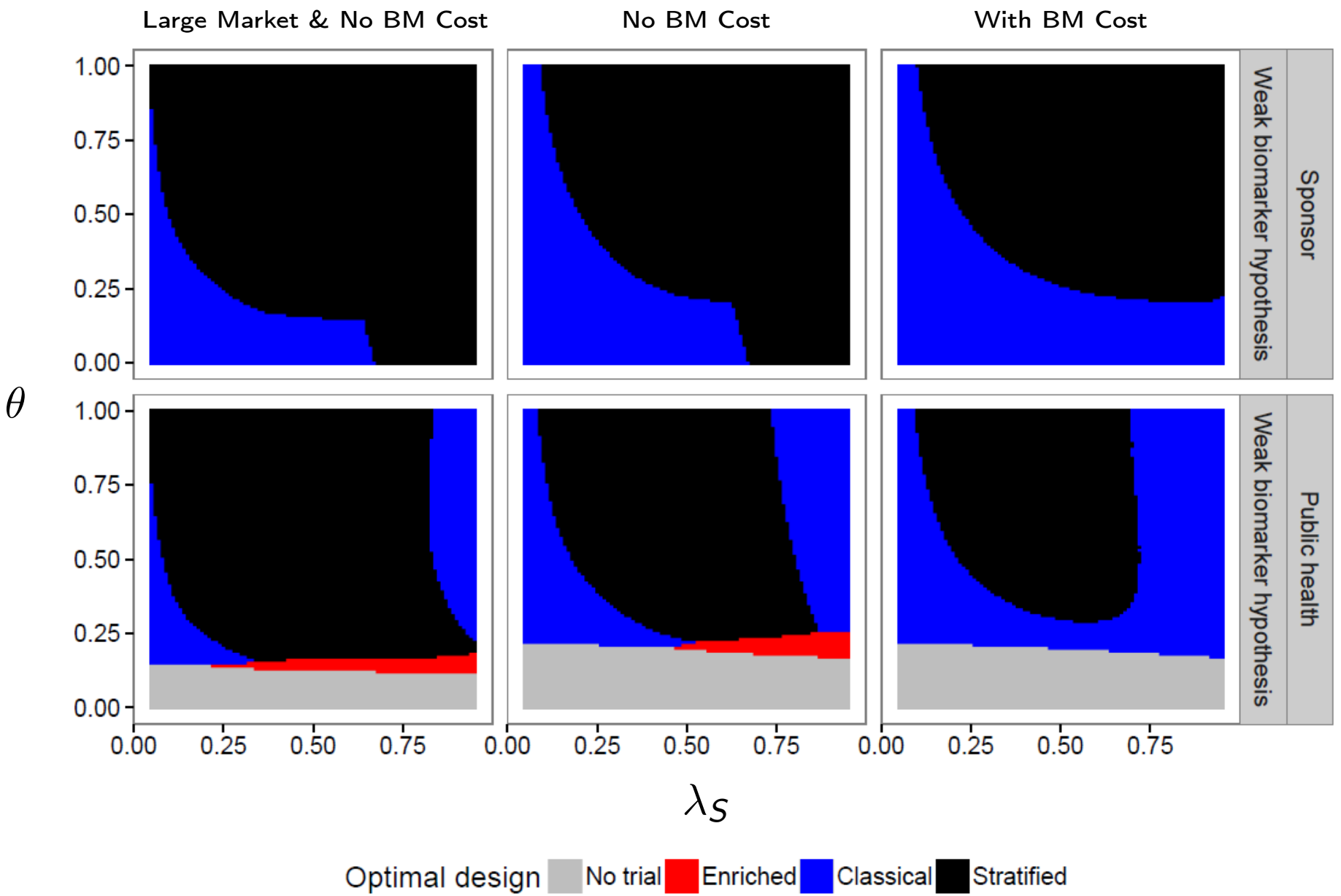


Optimal Designs for the Strong Biomarker Prior



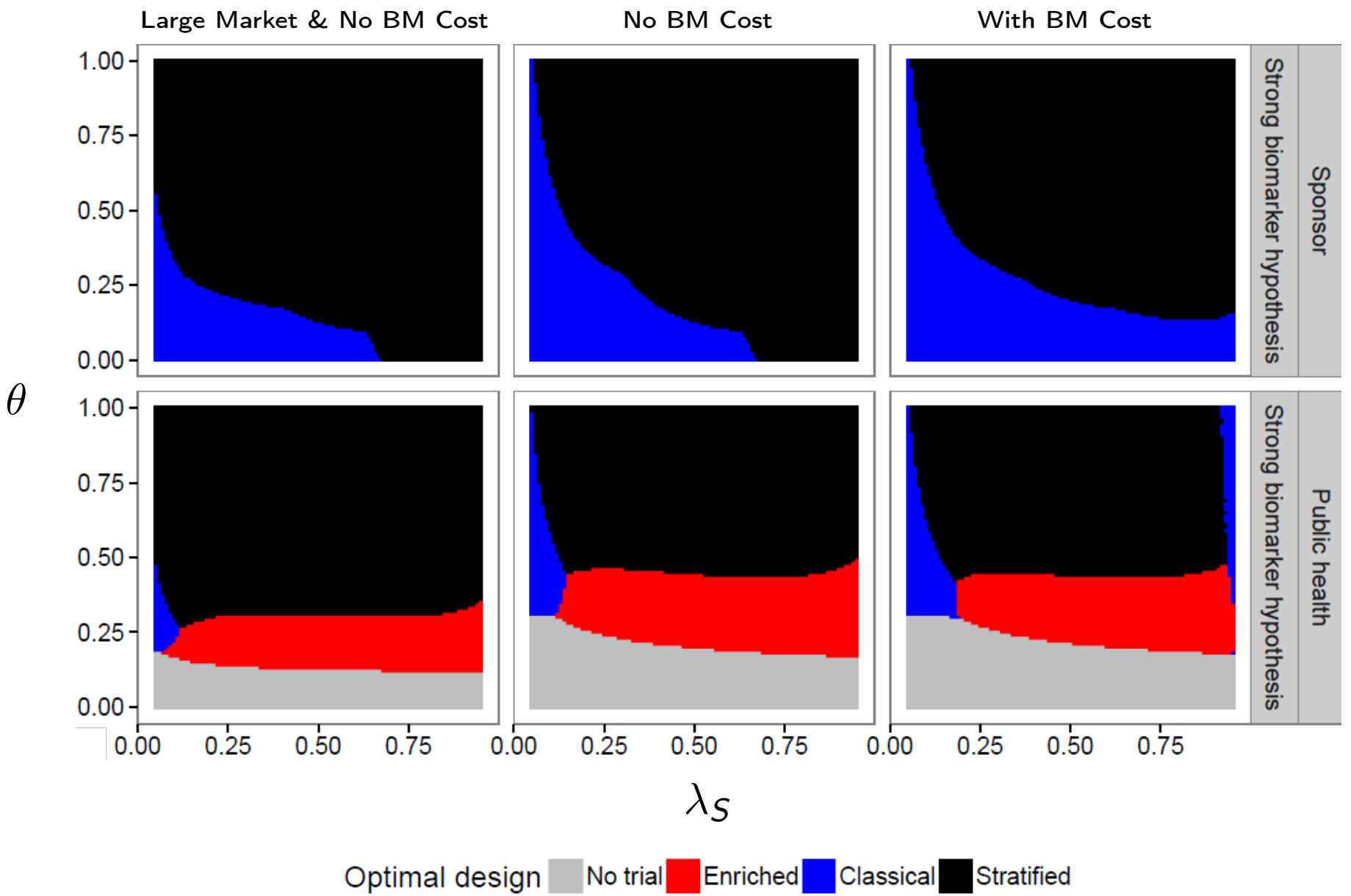
Optimal Trial Designs

Weak Biomarker Prior



Optimal Trial Designs

Strong Biomarker Prior



Discussion

- The optimal sample size under the public health view is typically larger than in the sponsor view.
- For the considered priors, the enrichment design is never optimal for the sponsor view
- The optimal design depends strongly on the particulars of the situation: Subgroup prevalence, trial costs and initial beliefs.

Two Extensions of the Trial Designs

Partial Enrichment Design

The prevalence of the subgroup in the trial λ_T is a design parameter and may differ from λ_S, the prevalence in the population.
Special cases are the stratified design ($\lambda_T = \lambda_S$) and the (full) enrichment design ($\lambda_T = 1$).

E.g., Zhao et al. (2010)

Adaptive Enrichment Designs

Two stage design, where the second stage sample size and second stage trial prevalence may depend on the first stage outcome.

E.g., Brannath et al. (2009); Beckman et al. (2011); Jenkins et al. (2011); Friede et al. (2012); Simon and Simon (2013); Graf et al. (2015)

For simplicity we use as multiple testing procedure a single step unweighted Bonferroni test.

Partial Enrichment Design: Hypothesis test of H_F

Because trial and population prevalence do not match, the standard z-test is not a valid test for H_F.

- H_F is tested with a reweighted z-statistics

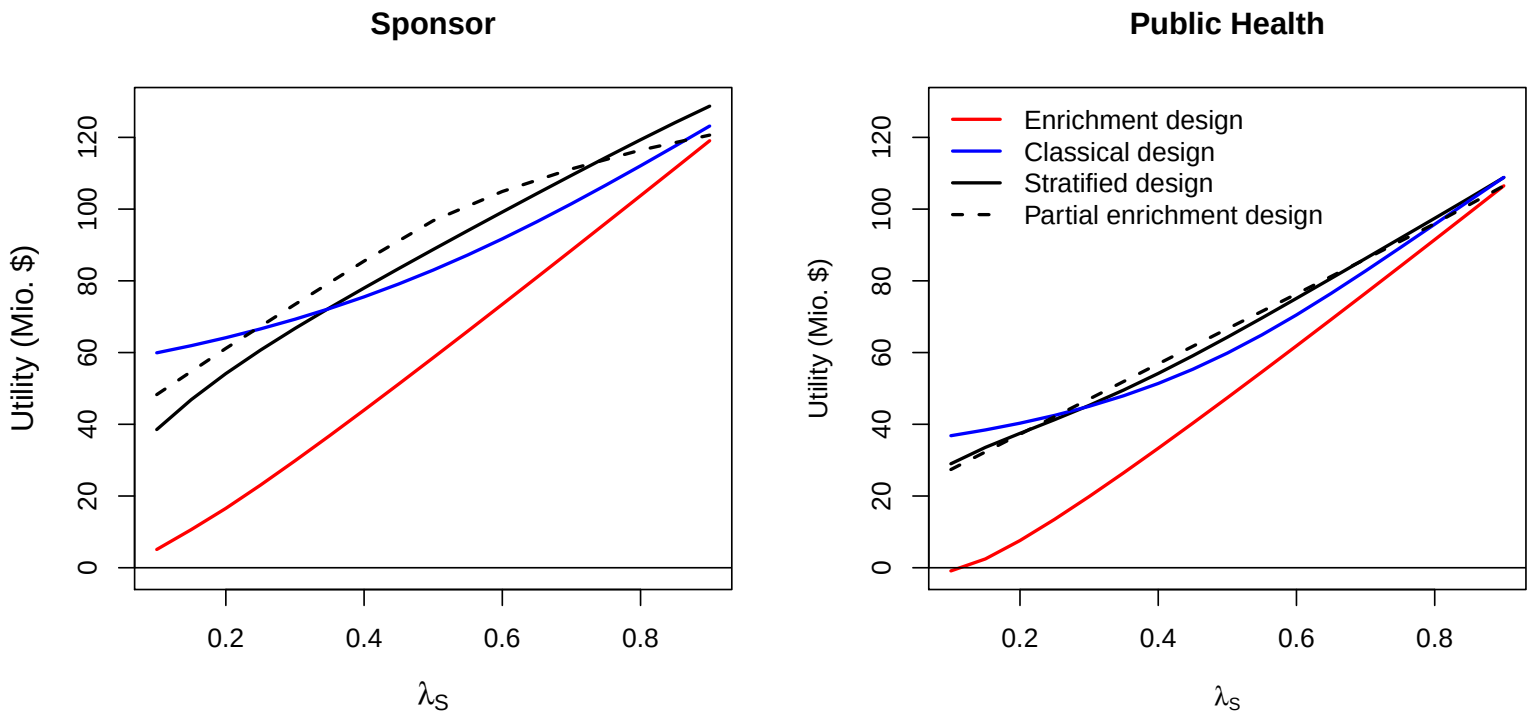
$$\tilde{Z}_F = \xi \left(\frac{\lambda_S}{\sqrt{\lambda_T}} Z_S + \frac{1-\lambda_S}{\sqrt{1-\lambda_T}} Z_{S'} \right),$$

  where $Z_S, Z_{S'}$ denote the z-statistics for the subgroups S, S' and ξ is a normalizing constant.

  e.g., Zhao et.al. 2010

- As above, to reject H_F, in addition the consistency condition needs to be fulfilled.

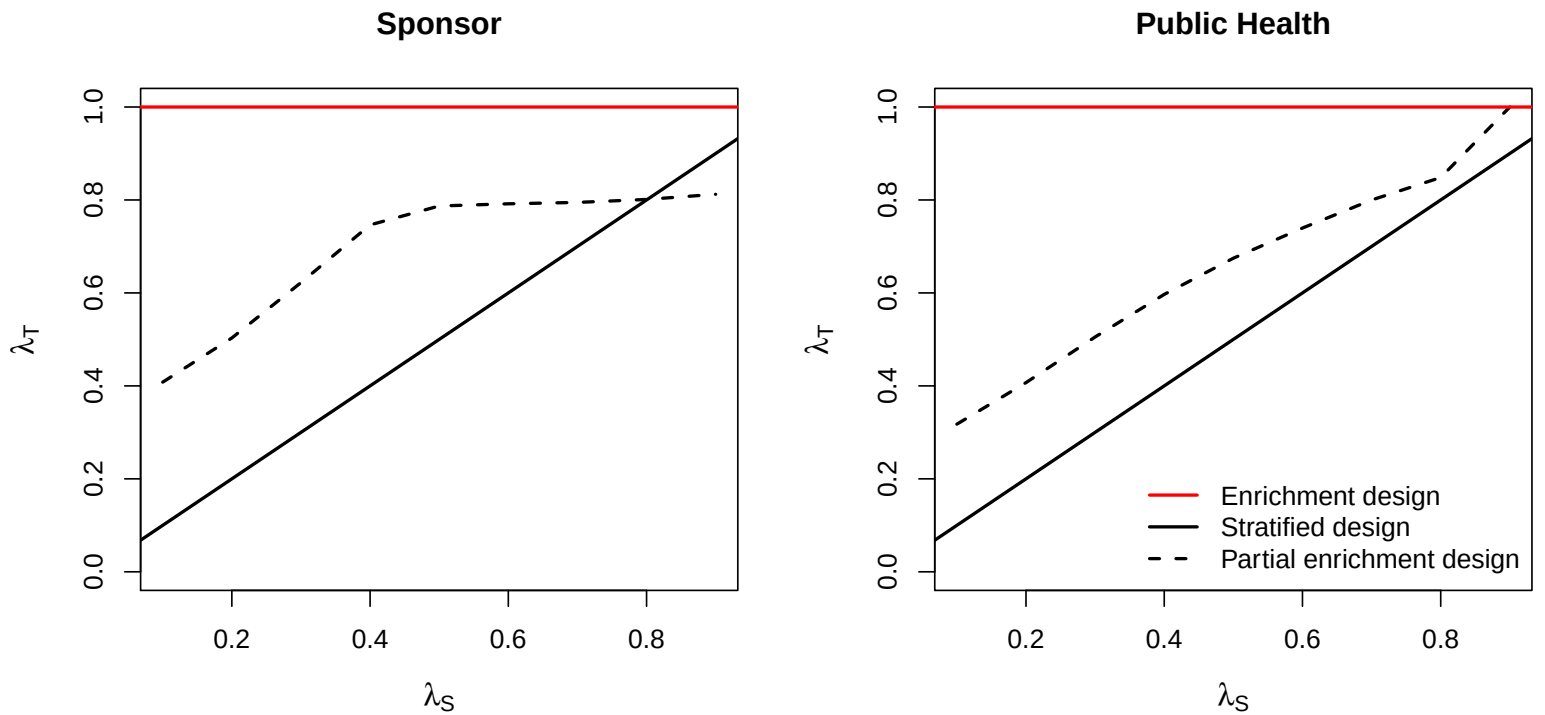
Optimized Utilities
Weak Biomarker Prior



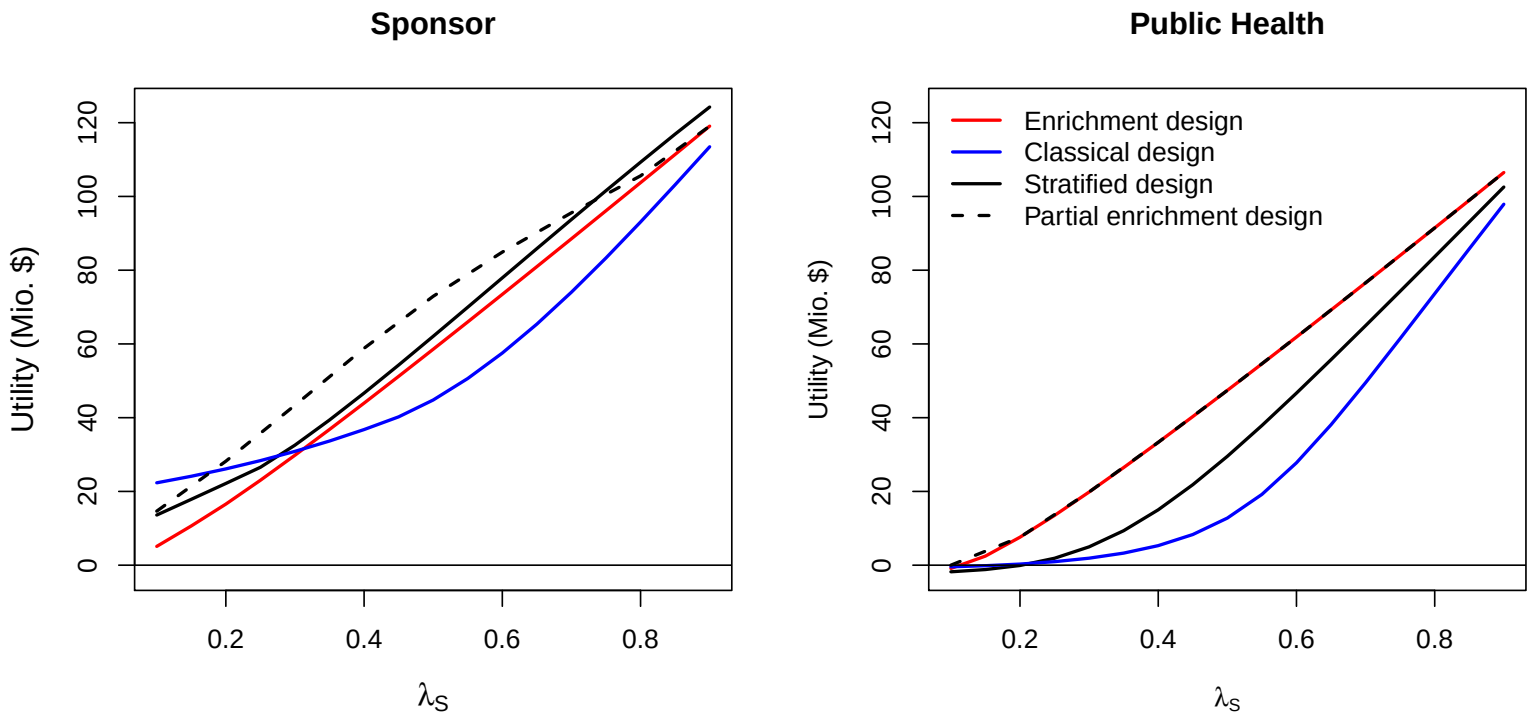
$(Nr_F = Nr_S = 1000, c_{\text{setup}} = c_{\text{BM development}} = 1, c_{\text{per-patient}} = 0.05, c_{\text{BM determination}} = 0)$

Optimal Subgroup Prevalence λ_T

Weak Biomarker Prior



Optimized Utilities
Strong Biomarker Prior



Adaptive Design

First stage

Sample size n_1, subgroup trial prevalence λ_T^1

Second Stage

Second stage sample size n_2 and subgroup trial prevalence λ_T^2
are chosen based on first stage data.

Testing procedure

- Overall test statistics computed with combination function:

$$\begin{aligned} Z_S &= \sqrt{\tfrac{1}{2}}\, Z_S^1 + \sqrt{\tfrac{1}{2}}\, Z_S^2 \\ \tilde{Z}_F &= \sqrt{\tfrac{1}{2}}\, \tilde{Z}_F^1 + \sqrt{\tfrac{1}{2}}\, \tilde{Z}_F^2 \end{aligned}$$

  where Z_S^1, Z_S^2 and $\tilde{Z}_F^1, \tilde{Z}_F^2$ are stage wise z-statistics.
- Unweighted Bonferroni test boundaries applied to Z_S and $\tilde{Z}_F$ (if $\lambda_T^2 < 1$).

Optimal Adaptive Design
Based on Backwards Induction

Optimal Adaptation Rule

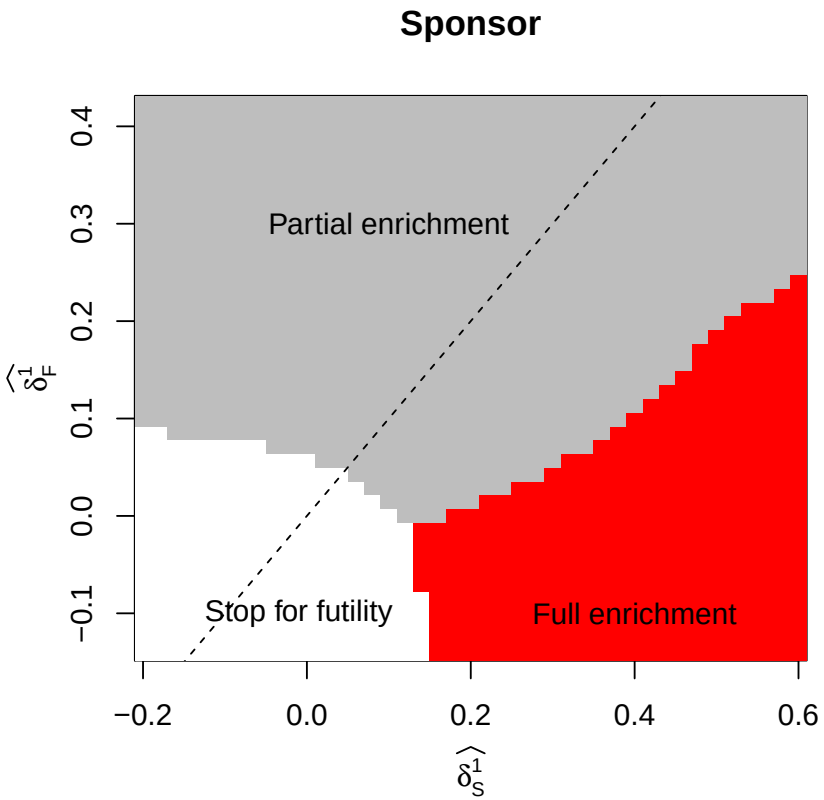
- In the interim analysis choose n_2, λ_T^2 such that the expected utility conditional on the first stage data is maximized.
- Especially, $n_2 = 0$ corresponds to stopping for futility, $\lambda_T = 1$ to a second stage enrichment design.

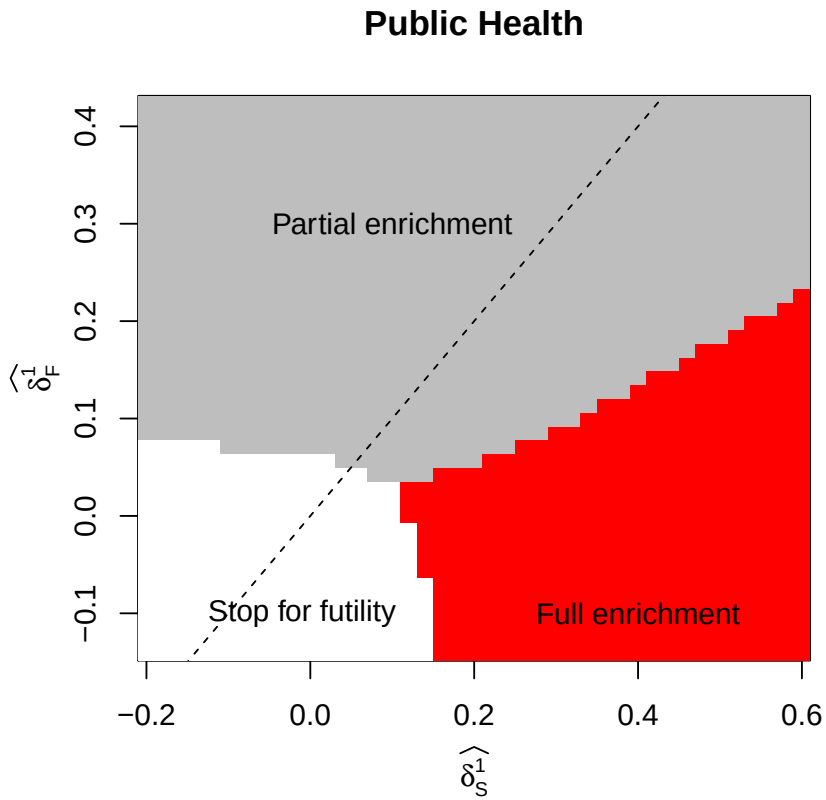
Optimizing first stage parameters

- Choose n_1, λ_T^1 such that the expected utility (given the optimal adaptation rule is applied at interim) is maximized.

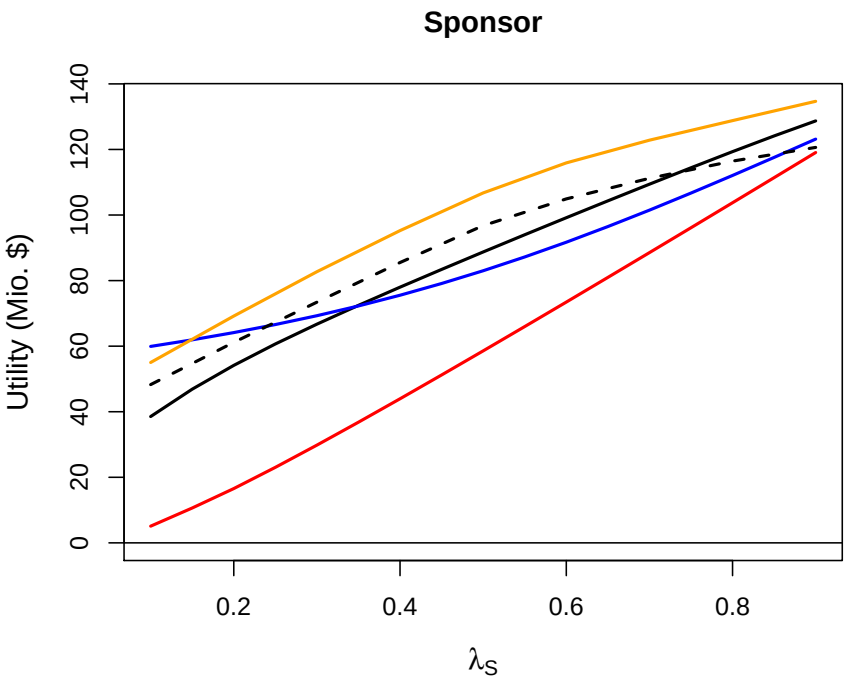
Example for the Optimal Decision Rule

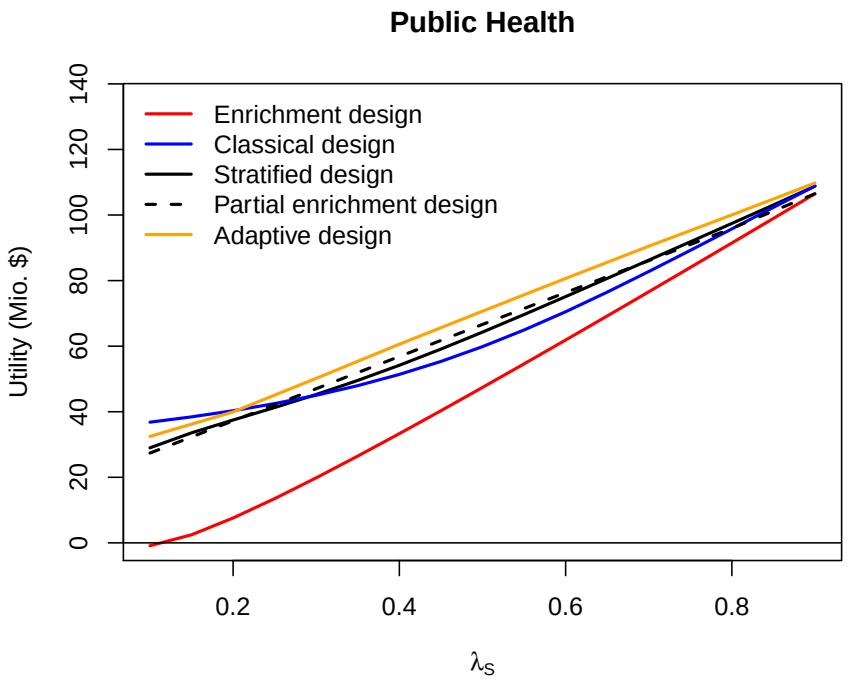
Weak Biomarker Prior ($n_1 = 100, \lambda_S = \lambda_T^1 = 0.5$)



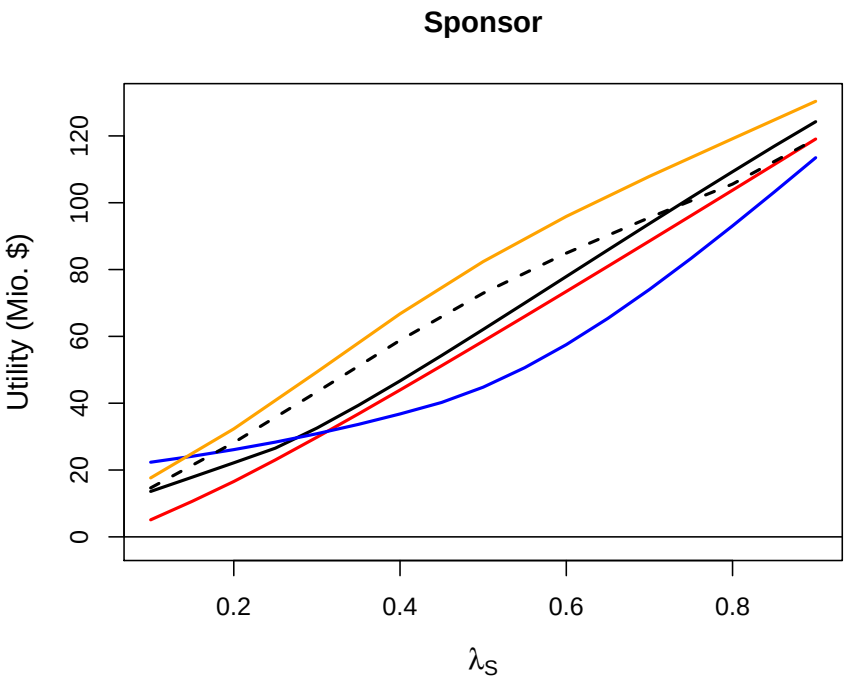


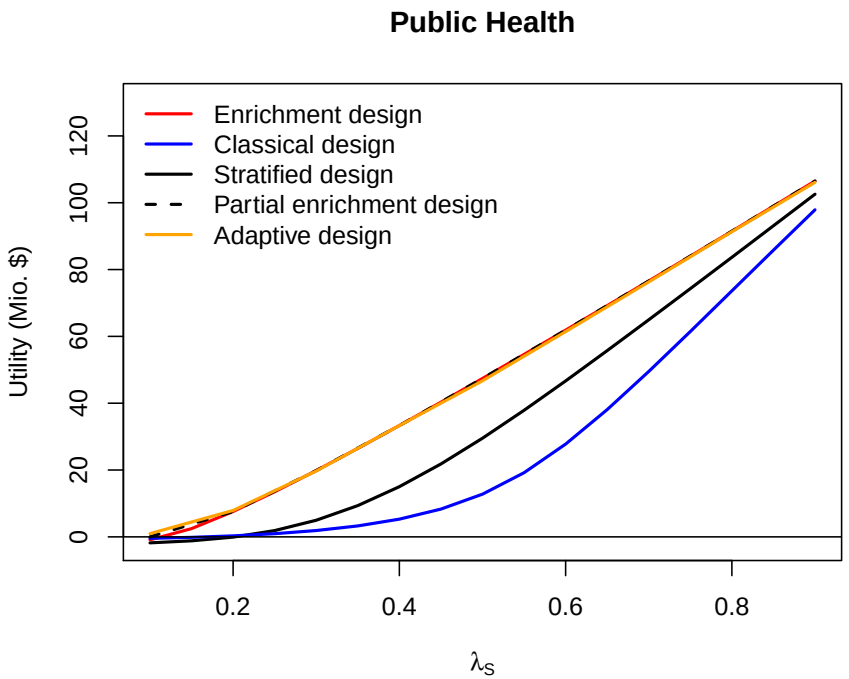
Optimized Utilities
Weak Biomarker Prior





Optimized Utilities
Strong Biomarker Prior





Summary

- Partial Enrichment Designs can increase the utility mainly for the sponsor utility function.
- Adaptive Enrichment Designs further increase the expected utility, also for the public health utility function.

References I

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.

Brannath, W., E. Zuber, M. Branson, F. Bretz, P. Gallo, M. Posch, and A. Racine-Poon (2009, May). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine 28*(10), 1445–1463.

Friede, T., N. Parsons, and N. Stallard (2012, December). A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine 31*(30), 4309–4320.

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

Jenkins, M., A. Stone, and C. Jennison (2011, July). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints†. *Pharmaceutical Statistics 10*(4), 347–356.

Ondra, T., S. Jobjörnsson, R. A. Beckman, C. Burman, F. König, S. Stallard, and P. M (2016). Optimizing trial designs for targeted therapies. *PlosOne*.

References II

Simon, N. and R. Simon (2013). Adaptive enrichment designs for clinical trials. *Biostatistics 14*(4), 613–625.

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.

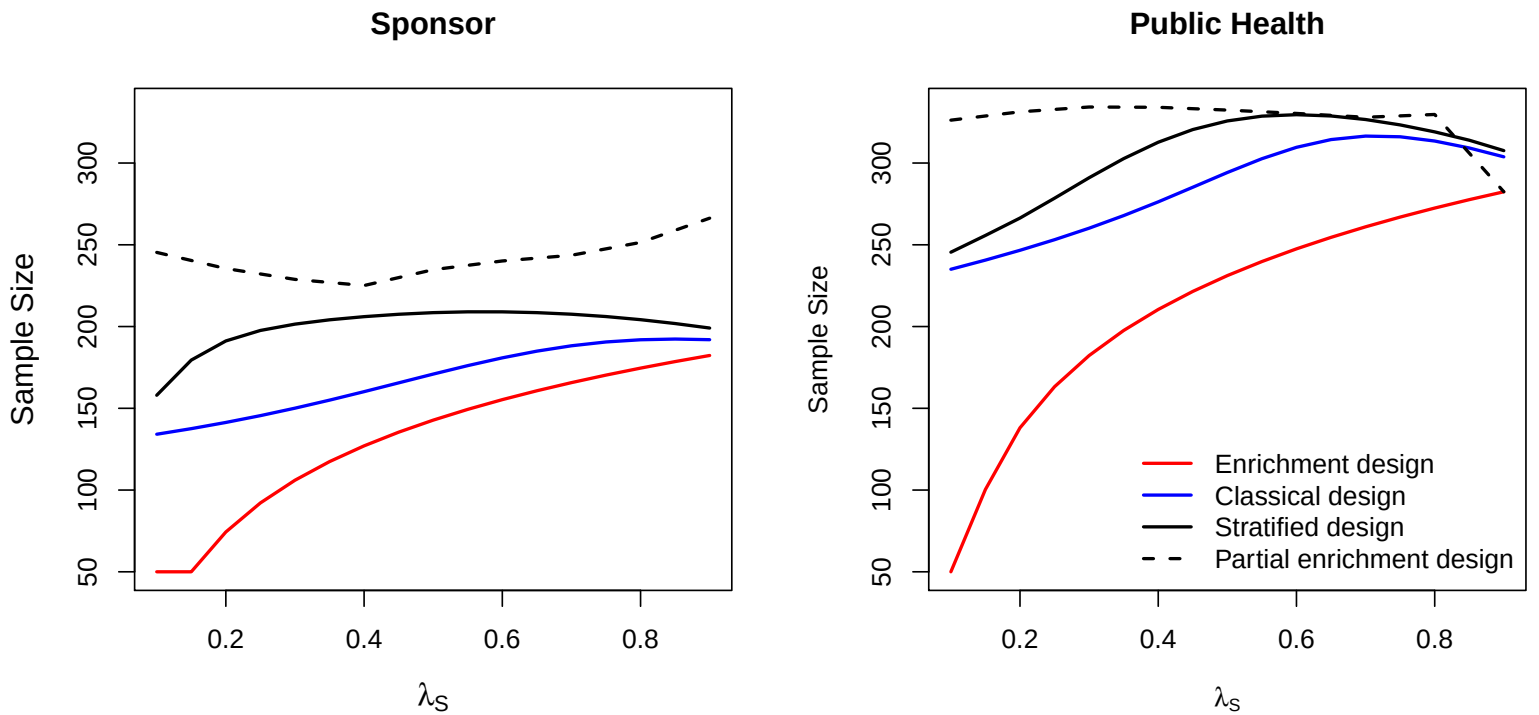
Spiessens, B. and M. Debois (2010, November). Adjusted significance levels for subgroup analyses in clinical trials. *Contemporary Clinical Trials 31*(6), 647–656.

Zhao, Y. D., A. Dmitrienko, and R. Tamura (2010). Design and analysis considerations in clinical trials with a sensitive subpopulation. *Statistics in Biopharmaceutical Research 2*(1), 72–83.

Backup Slides

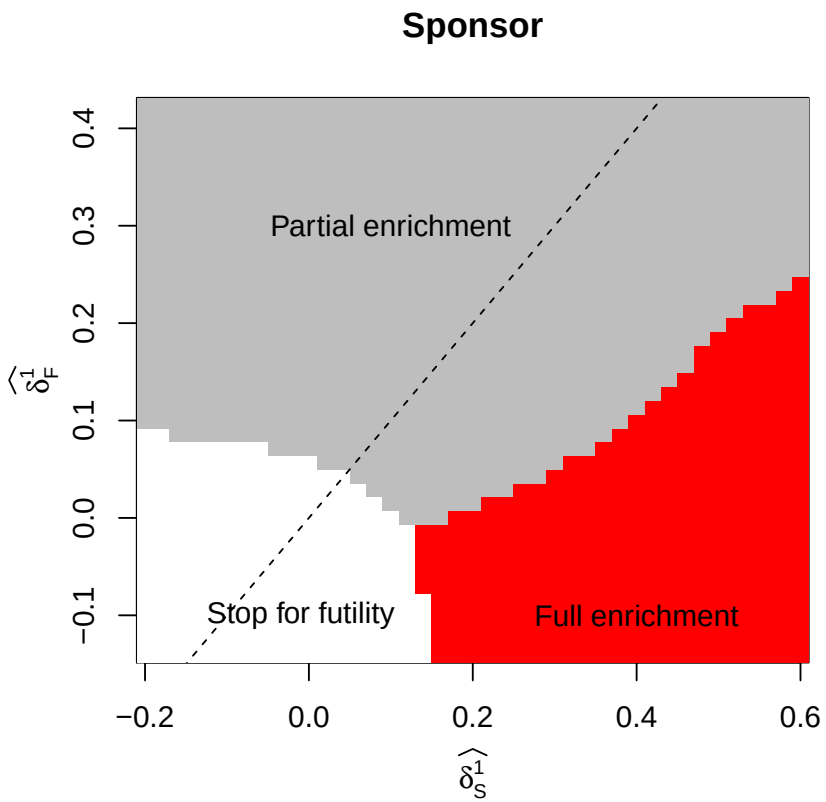
Optimal Sample Size of the Partial Enrichment Design

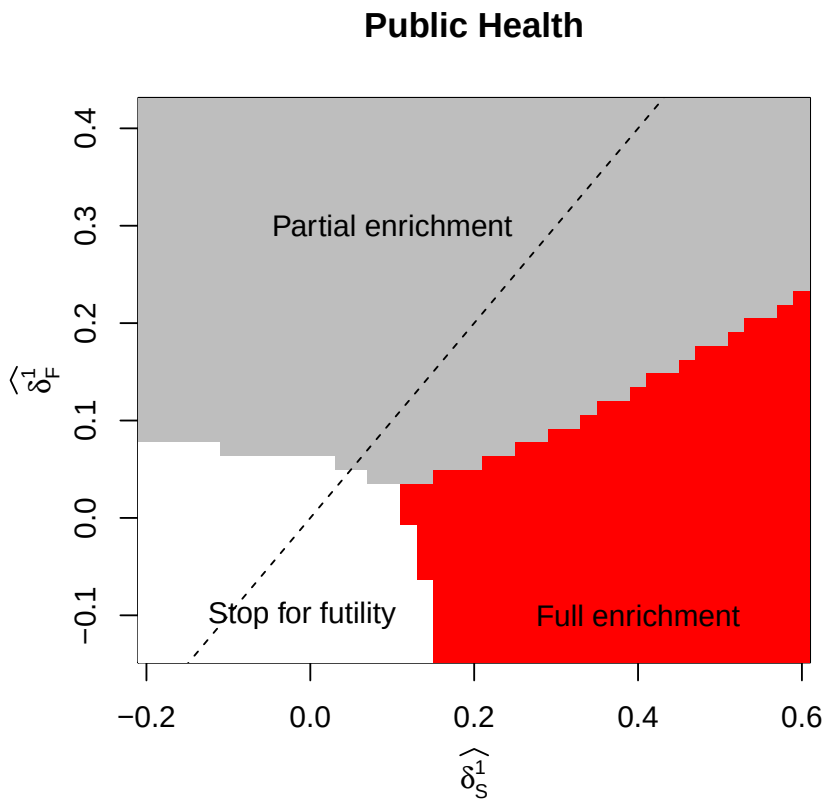
Weak Biomarker Prior



Example for the Optimal Decision Rule

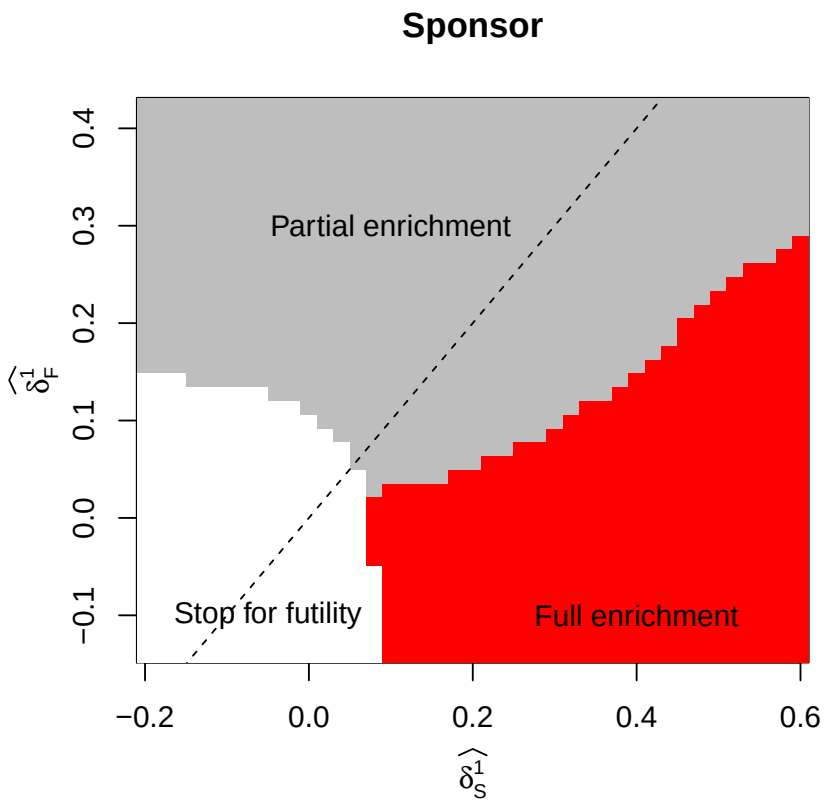
Weak Biomarker Prior ($n_1 = 100, \lambda_S = \lambda_T^1 = 0.5$)

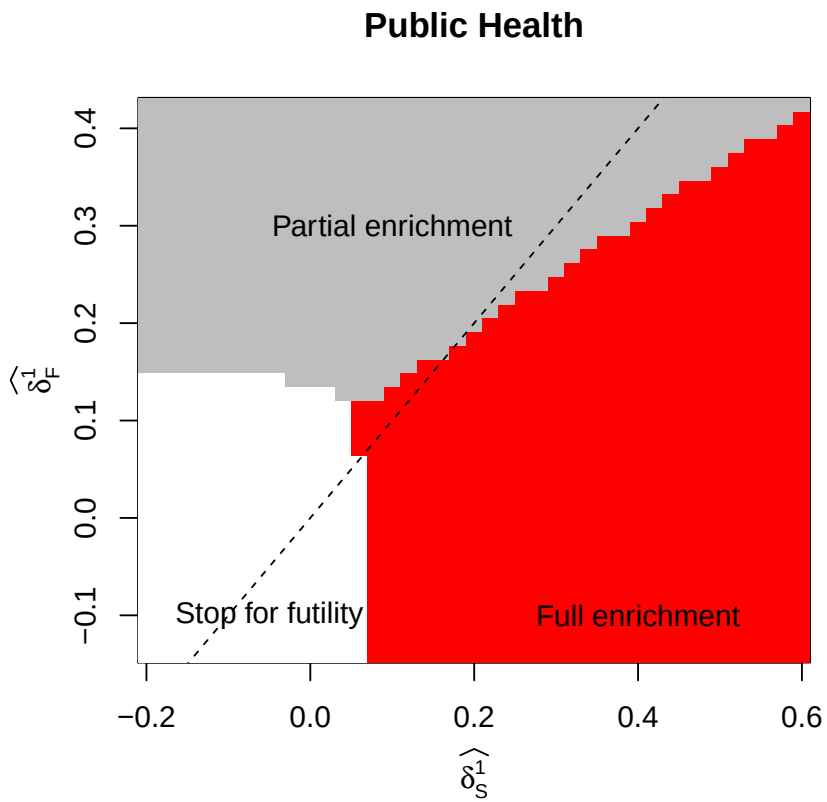




Example for the Optimal Decision Rule

Strong Biomarker Prior ($n_1 = 100, \lambda_S = \lambda_T^1 = 0.5$)





mu0 = −2

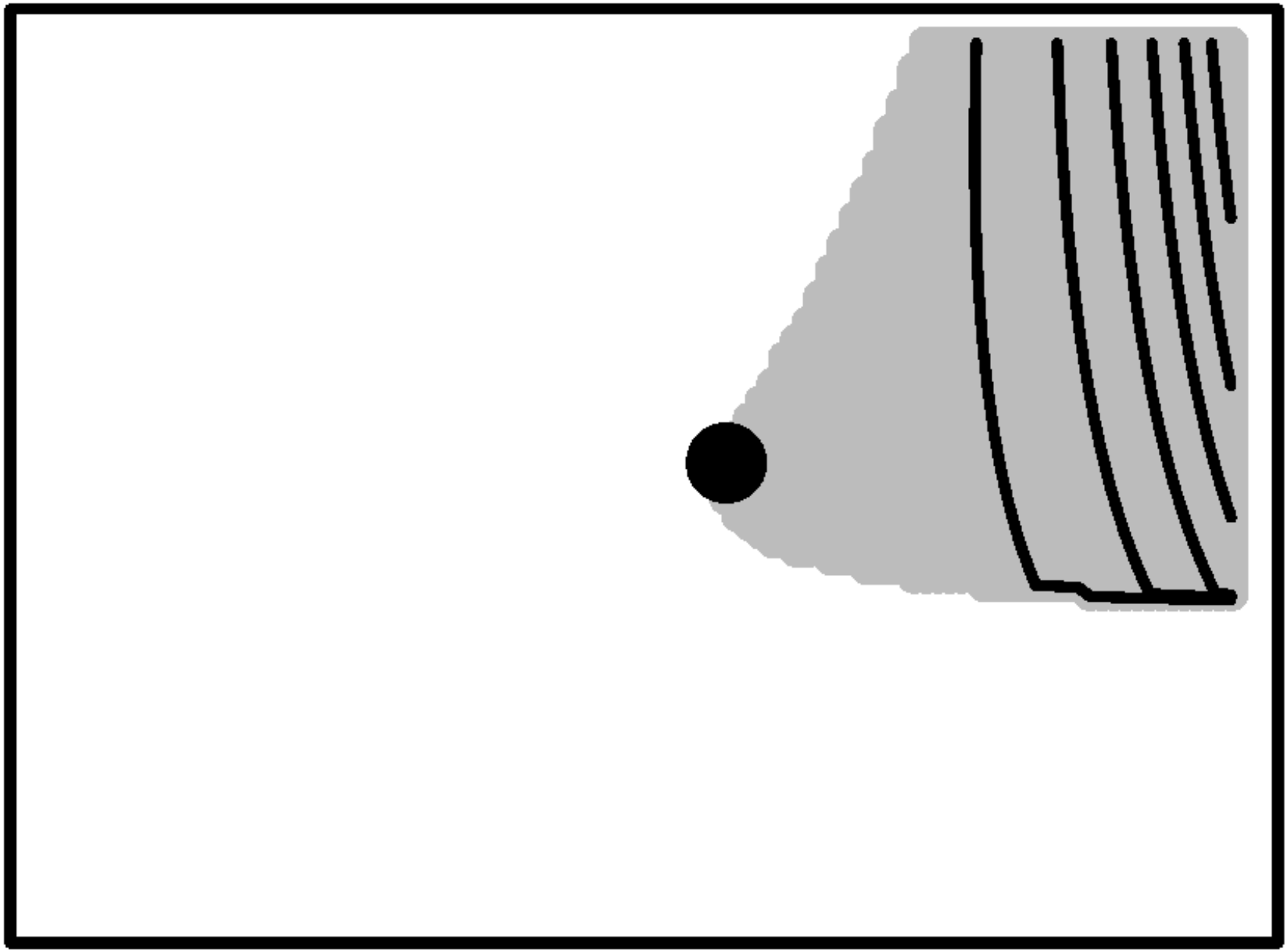
0

2

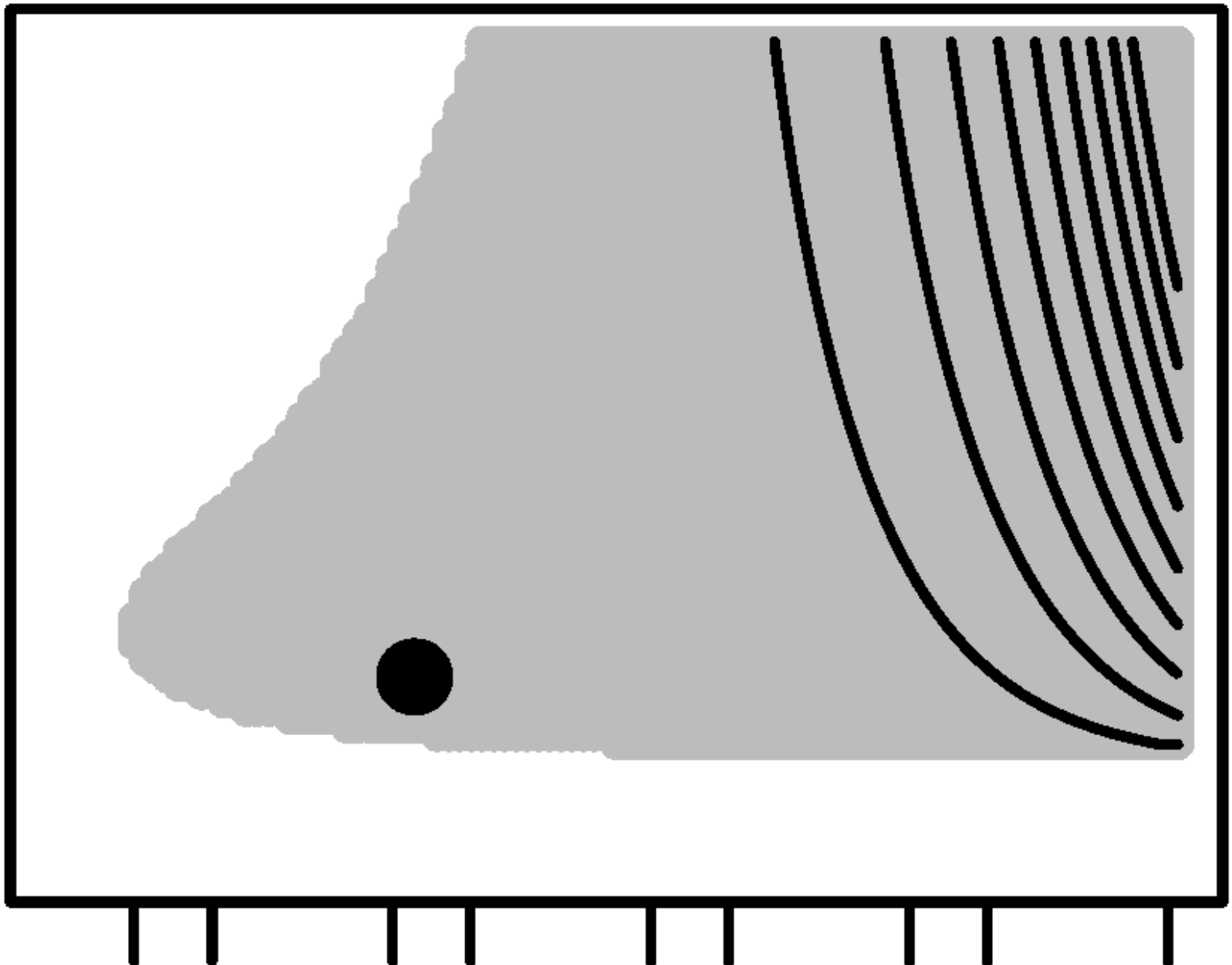
4

6

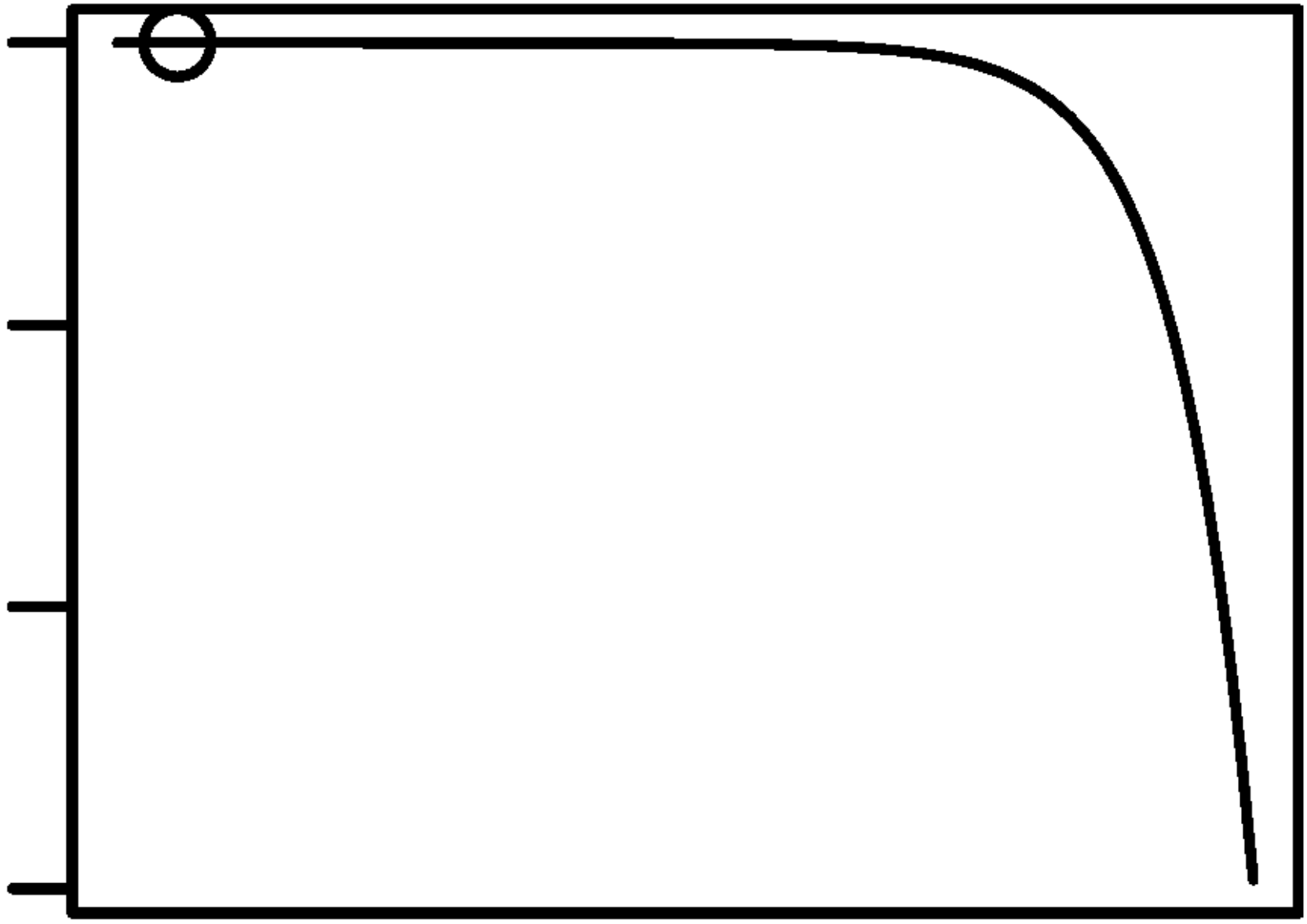
mu0 = −1.73



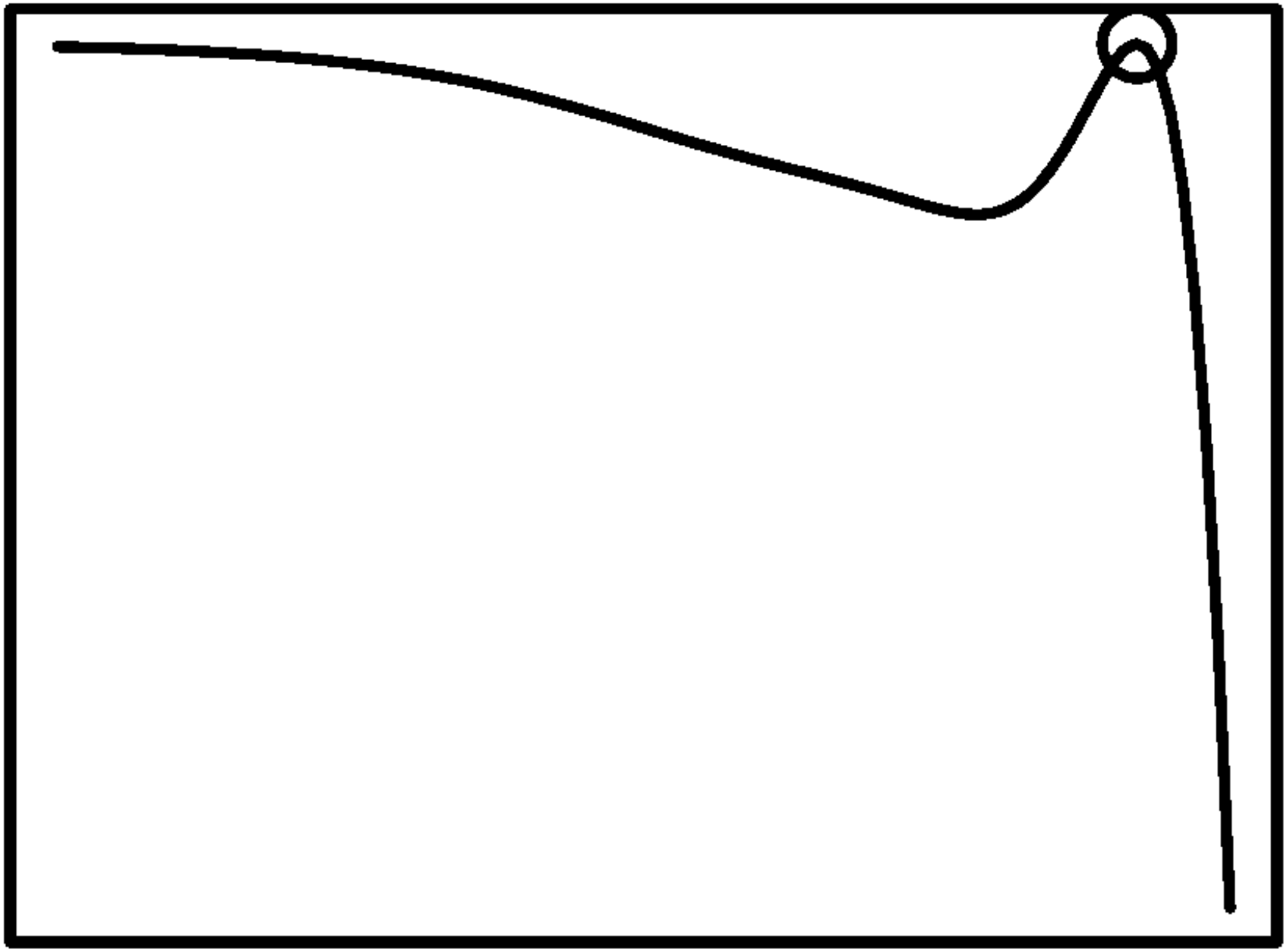
mu0 = −0.13

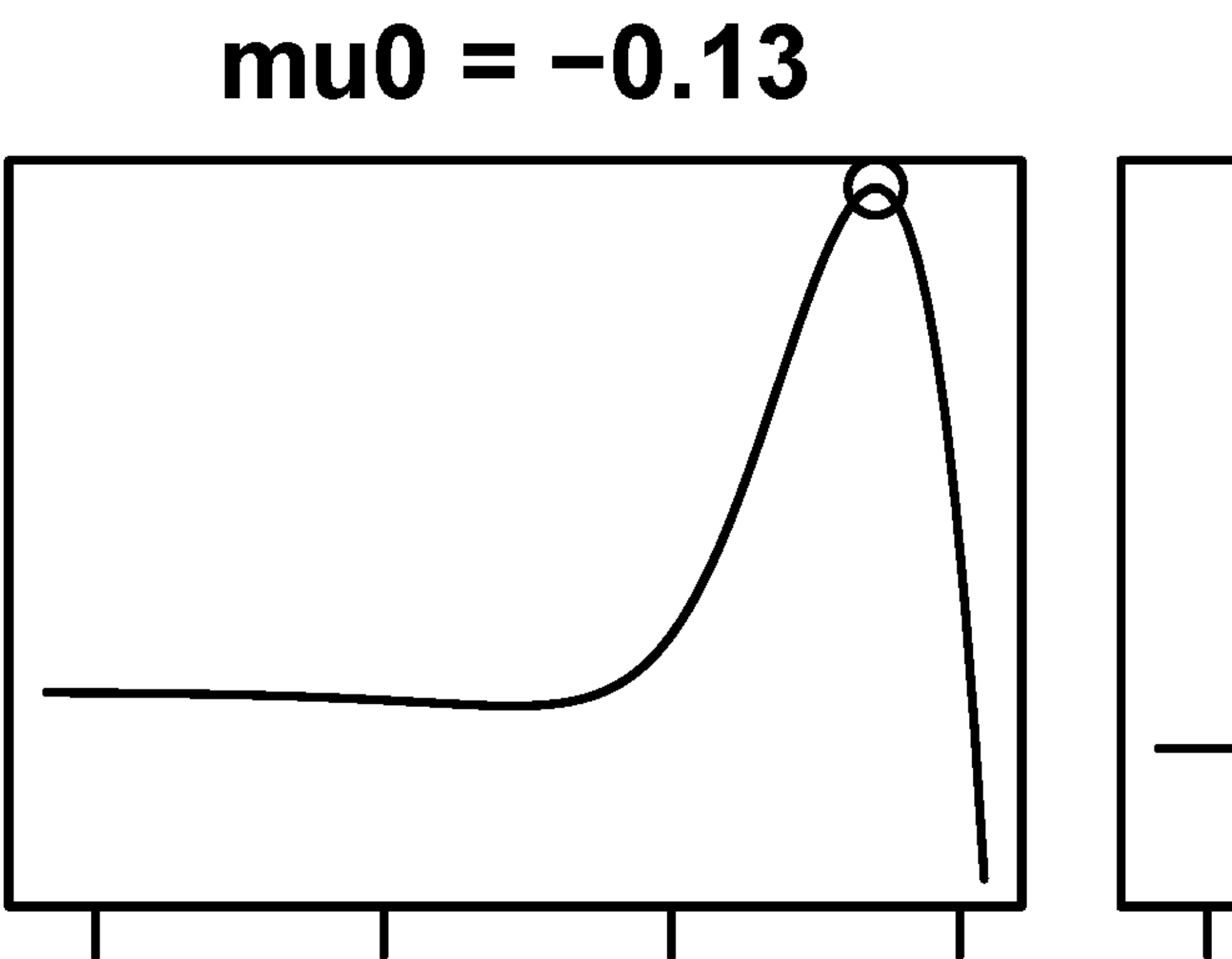


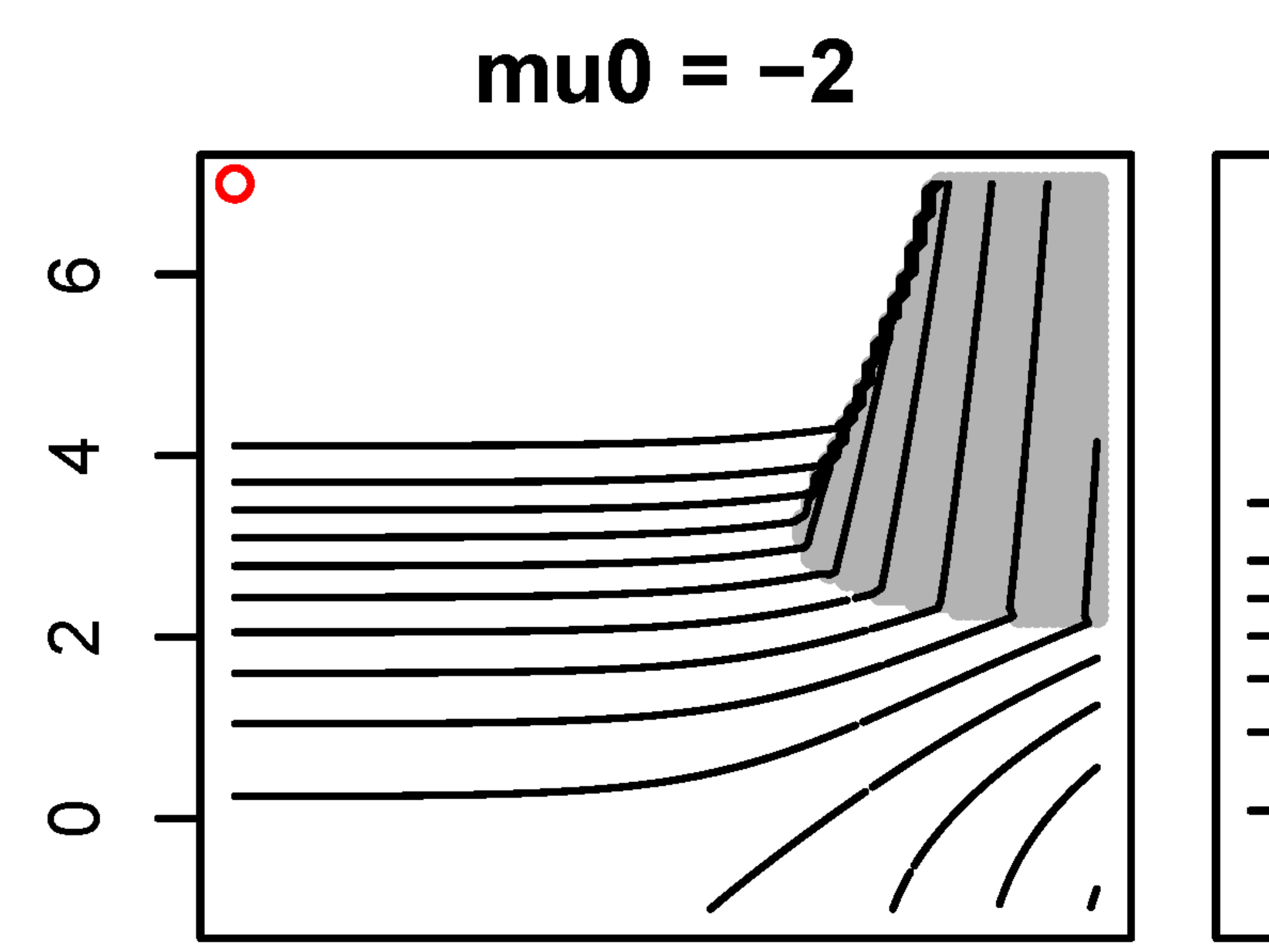
mu0 = −2



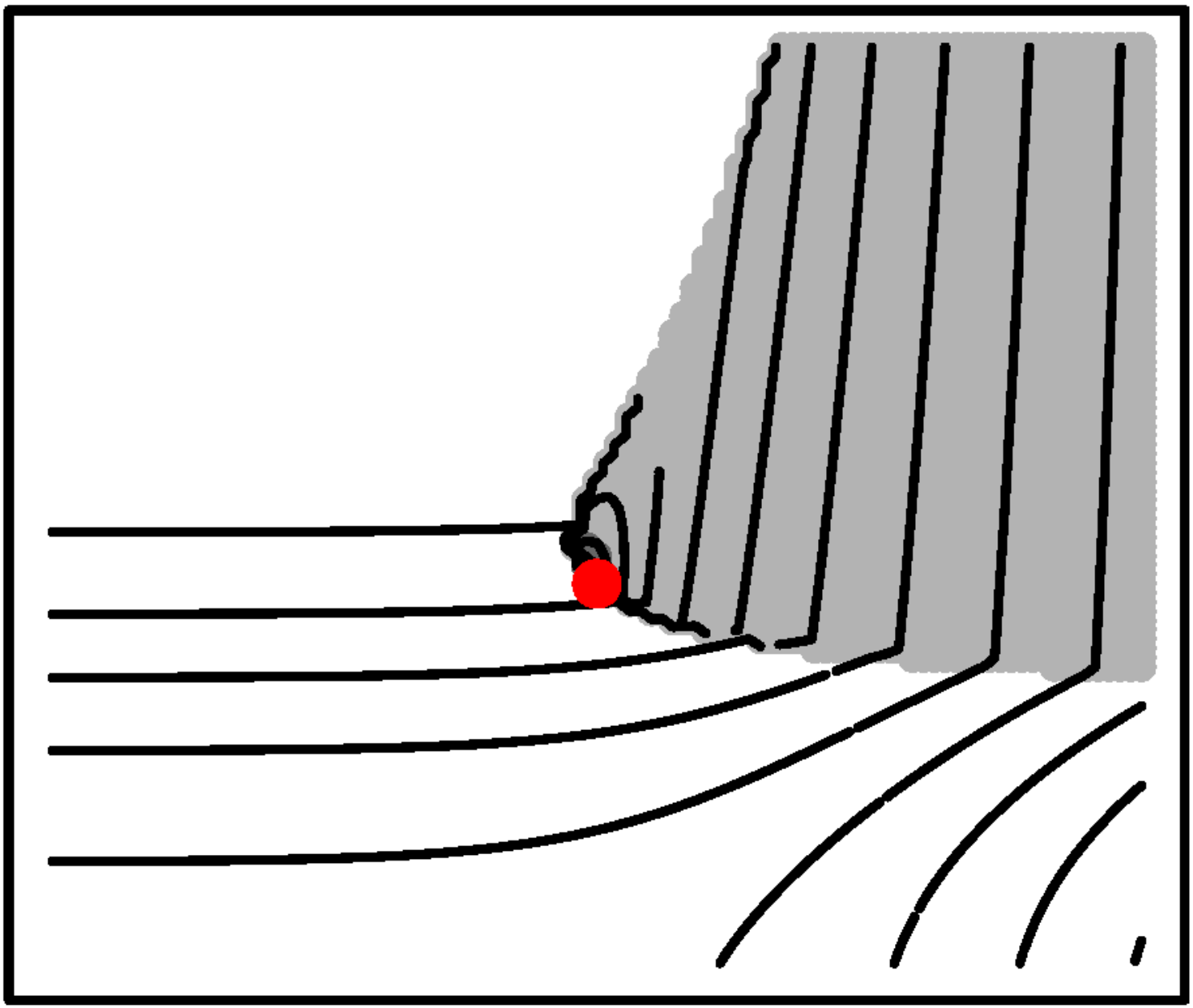
mu0 = −1.73



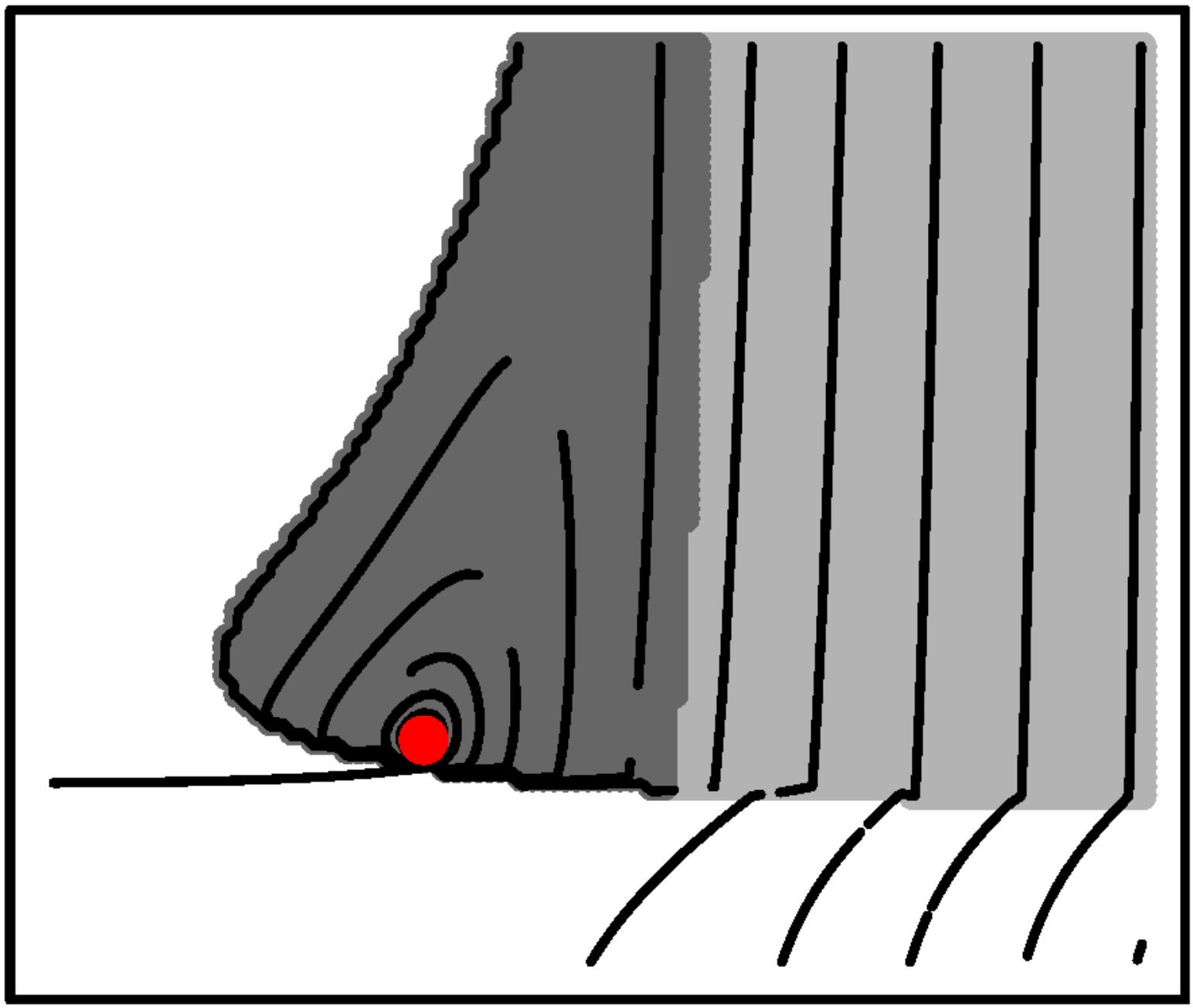

mu0 = −0.13


mu0 = −2
0
2
4
6

mu0 = −1.47



mu0 = −0.4



mu0 = 1.73