Two-Stage Designs for proteomic and gene expression studies applying methods differing in costs

Alexandra Goll   Peter Bauer

Section of Medical Statistics
Medical University of Vienna

supported by FWF-Fund Nr. P18698-n15

ROeS Seminar - Bern 2007

Motivation

In genomic or proteomic studies we generally search for genes or proteins which explain the occurrence of a disease, the curse of a disease, ...

Limitations

- very large number of candidate genes (proteins)
- only a very small fraction of genes (proteins) show noticeable effects

Constraints

- Limit of resources: total costs C are fixed

Test Problem:

Consider

m_1 hypotheses for the mean of independent normally distributed observations μ_i with known variance σ^2

$$H_{0i}: \ \mu_i = 0 \quad \text{versus} \quad H_{1i}: \ \mu_i > 0 \quad i = 1, \ldots, m_1$$

assuming independence of observations across hypotheses

Motivation: Example

Consider an experiment with

- $m_1 = 1000$ ... number of hypotheses tests
- $C = 20000$ ... fixed total costs

Given

- $\pi_0 = 0.99$ ... proportion of true null hypotheses among all m_1 hypotheses
- $\Delta = \mu/\sigma = 0.75$ ... effect size

Conventional Single-Stage Design

- The costs per observation are set to $c_1 = 1$.
- Distribute total costs $C = 20000$ equally among the hypotheses

$$\Rightarrow 20000/1000 = 20 \text{ observations per hypothesis test}$$

Motivation: Example

Consider an experiment with

- $m_1 = 1000$ ... number of hypotheses tests
- $C = 20000$ ... fixed total costs

Given

- $\pi_0 = 0.99$ ... proportion of true null hypotheses among all m_1 hypotheses
- $\Delta = \mu/\sigma = 0.75$ ... effect size

Conventional Single-Stage Design

- The costs per observation are set to $c_1 = 1$.
- Distribute total costs $C = 20000$ equally among the hypotheses

$$\Rightarrow 20000/1000 = 20 \text{ observations per hypothesis test}$$

Error Rates

Notation

R ... number of rejected hypotheses
FP ... number of false positive decisions

Family Wise Type I Error Rate (FWE)

is defined as the probability of at least one Type I error:

$$FWER = P(FP \geq 1)$$

False Discovery Rate (FDR)

is the expected proportion of Type I errors among the rejected hypotheses (Benjamini and Hochberg, 1995):

$$FDR = E\left(\frac{FP}{\max\{R, 1\}}\right)$$

Error Rates

Notation

R ... number of rejected hypotheses
FP ... number of false positive decisions

Family Wise Type I Error Rate (FWE)

is defined as the probability of at least one Type I error:

$$FWER = P(FP \geq 1)$$

False Discovery Rate (FDR)

is the expected proportion of Type I errors among the rejected hypotheses (Benjamini and Hochberg, 1995):

$$FDR = E\left(\frac{FP}{\max\{R, 1\}}\right)$$

Controlling the Family Wise Type I Error Rate

Bonferroni adjusted level

- Reject all hypotheses with a p-value $p_i \leq \gamma$
- $\gamma = \alpha / m_1$

Example: Single-Stage Design with 20 observations

$\alpha = 0.05$: $\gamma = 0.05/1000 = 0.00005$

Assuming an effect size of $\Delta = 0.75$ the Power (expected fraction of correctly identified effective genes/proteins) is 0.296

Because we have to adjust for multiplicity, the power of the single-stage design is very low as compared to a single hypothesis test with the same sample size.

Controlling the Family Wise Type I Error Rate

Bonferroni adjusted level

- Reject all hypotheses with a p-value $p_i \leq \gamma$
- $\gamma = \alpha / m_1$

Example: Single-Stage Design with 20 observations

$\alpha = 0.05$: $\gamma = 0.05/1000 = 0.00005$

Assuming an effect size of $\Delta = 0.75$ the Power (expected fraction of correctly identified effective genes/proteins) is 0.296

Because we have to adjust for multiplicity, the power of the single-stage design is very low as compared to a single hypothesis test with the same sample size.

Controlling the False Discovery Rate (Storey 2002)

Storey's procedure to control the FDR at level α

- m_1 hypothesis tests
- Reject all hypotheses with a p-value $p_i \leq \gamma$
- To control the FDR at level α choose the maximal γ so that

$$\widehat{FDR}_\gamma(p_1, \ldots, p_{m_1}) = \frac{\gamma\,\widehat{\pi}_0 m_1}{\max(\sharp\{p_i \leq \gamma\}, 1)} \leq \alpha$$

where $\widehat{\pi}_0 = \sharp\{p_i > \lambda\}/\{(1-\lambda)m_1\}$ and $0 < \lambda < 1$

Example: Single-Stage Design with 20 observations

For $\Delta = 0.75$: $Power = 0.443$ (FWE: $Power = 0.296$)

Controlling the False Discovery Rate (Storey 2002)

Storey's procedure to control the FDR at level α

- m_1 hypothesis tests
- Reject all hypotheses with a p-value $p_i \leq \gamma$
- To control the FDR at level α choose the maximal γ so that

$$\widehat{FDR}_\gamma(p_1, \ldots, p_{m_1}) = \frac{\gamma\, \widehat{\pi}_0 m_1}{\max(\sharp\{p_i \leq \gamma\}, 1)} \leq \alpha$$

where $\widehat{\pi}_0 = \sharp\{p_i > \lambda\}/\{(1-\lambda)m_1\}$ and $0 < \lambda < 1$

Example: Single-Stage Design with 20 observations

For $\Delta = 0.75$: $Power = 0.443$ (FWE: $Power = 0.296$)

Single-stage designs

Single-stage designs have low power to detect existing effects.

Two-stage designs

Screen for promising hypotheses in the first stage which are further investigated in the second stage:

- Limit of resources: total costs C are fixed
- A fraction r of the resources C are used in the first stage for screening
- The remaining resources $(1 - r)C$ are used for second stage
- Costs per observation are set to 1

Single-stage designs

Single-stage designs have low power to detect existing effects.

Two-stage designs

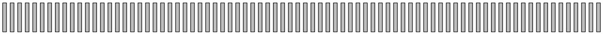
Screen for promising hypotheses in the first stage which are further investigated in the second stage:

- Limit of resources: total costs C are fixed
- A fraction r of the resources C are used in the first stage for screening
- The remaining resources $(1 - r)C$ are used for second stage
- Costs per observation are set to 1

The Two-Stage Design

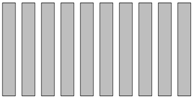
m_1 hypotheses

$$n_1 = \frac{rC}{m_1}$$



Select H_i: $p_{i1} < \gamma_1$

m_2 (random)



$$n_2 = \frac{(1-r)C}{m_2}$$

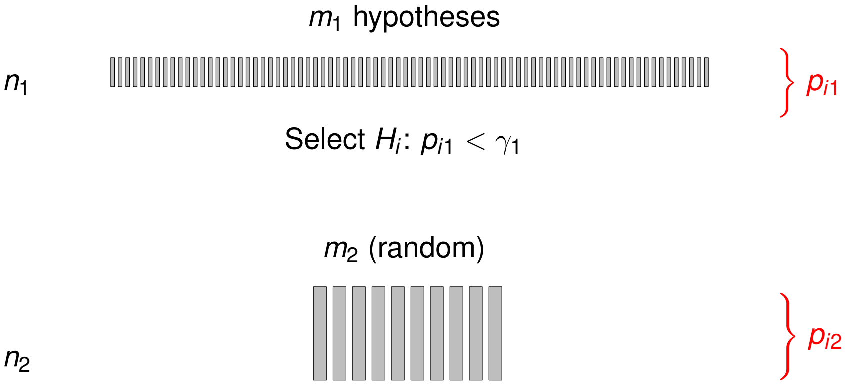
Test decision: Reject H_i: $p_{i2} < \gamma_2$

The computation of the p-value after the second stage depends on the ...

...design of the two-stage experiment

- Pilot Design
- Integrated Design

The Pilot Design



Test decision: Reject H_i: $p_{i2} < \gamma_2$

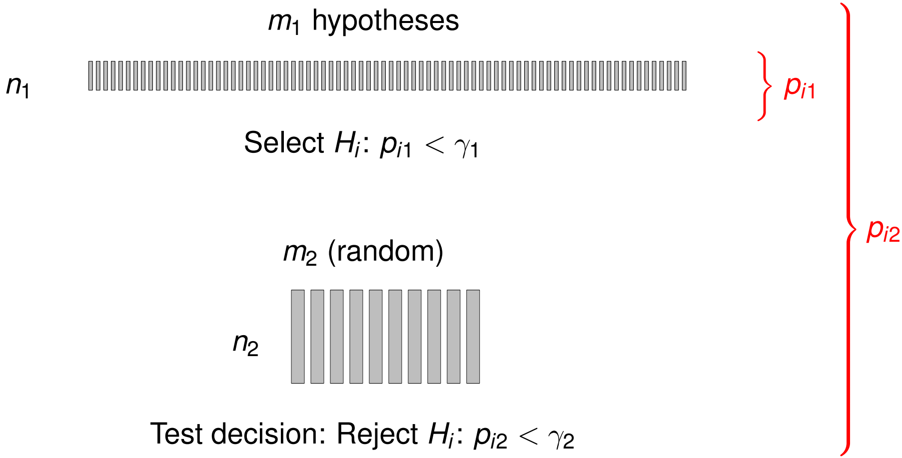
The Pilot Design

The final test decision after the second stage is only based on the second stage data:

$\rightarrow$ Sample size for the m_2 final hypothesis tests is n_2.

The correction for multiplicity is performed for the m_2 selected hypotheses using Bonferroni's correction (FWE) or Storey's procedure (FDR).

The Integrated Design



The Integrated Design

The final test decision is based on the pooled observations over both stages:

→ Sample size for m_2 final hypothesis tests is $n_1 + n_2$.
→ Price to be payed: the correction has to be performed for all initial m_1 hypotheses.

Error Rate Control

- early acceptance decision in a sequential test procedure
- define sequential p-values for all m_1 hypotheses on the Tsiatis-Rosner-Metha (1984) ordering in the sample space:
  For hypotheses early accepted we use the first stage p-value, for the final analysis we integrate over all sample path more extreme than the observed one.
- Despite than m_2 and n_2 are random the sequential p-values are independent and uniformly distributed under the null hypothesis. (see Zehetmayer et al, 2005)

The Integrated Design

The final test decision is based on the pooled observations over both stages:

→ Sample size for m_2 final hypothesis tests is $n_1 + n_2$.
→ Price to be payed: the correction has to be performed for all initial m_1 hypotheses.

Error Rate Control

- early acceptance decision in a sequential test procedure
- define sequential p-values for all m_1 hypotheses on the Tsiatis-Rosner-Metha (1984) ordering in the sample space:
  For hypotheses early accepted we use the first stage p-value, for the final analysis we integrate over all sample path more extreme than the observed one.
- Despite than m_2 and n_2 are random the sequential p-values are independent and uniformly distributed under the null hypothesis. (see Zehetmayer et al, 2005)

The Integrated Design

The final test decision is based on the pooled observations over both stages:

→ Sample size for m_2 final hypothesis tests is $n_1 + n_2$.
→ Price to be payed: the correction has to be performed for all initial m_1 hypotheses.

Error Rate Control

- early acceptance decision in a sequential test procedure
- define sequential p-values for all m_1 hypotheses on the Tsiatis-Rosner-Metha (1984) ordering in the sample space:
  For hypotheses early accepted we use the first stage p-value, for the final analysis we integrate over all sample path more extreme than the observed one.
- Despite than m_2 and n_2 are random the sequential p-values are independent and uniformly distributed under the null hypothesis. (see Zehetmayer et al, 2005)

Asymptotic optimal designs

Two-stage designs

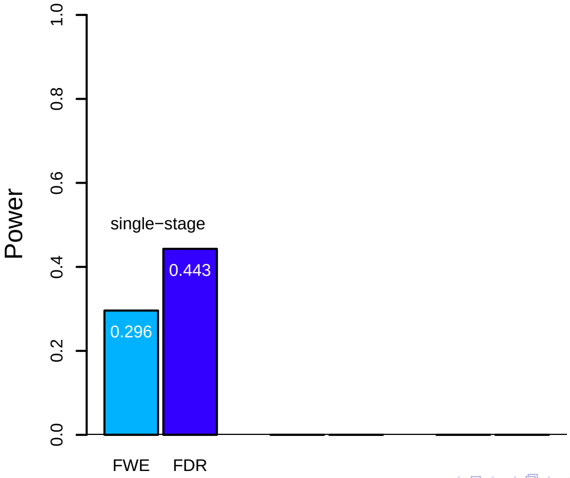
For given overall costs C, proportion of true null hypotheses π_0 and effect size $\Delta = \mu/\sigma$ the power of a two-stage design can be optimized with respect to:

- r ..fraction of total costs used in the first stage
- γ_1 ..selection boundary after first stage

Asymptotic: for $m_1 \to \infty$ and $C = d * m_1$ for $d > 0$

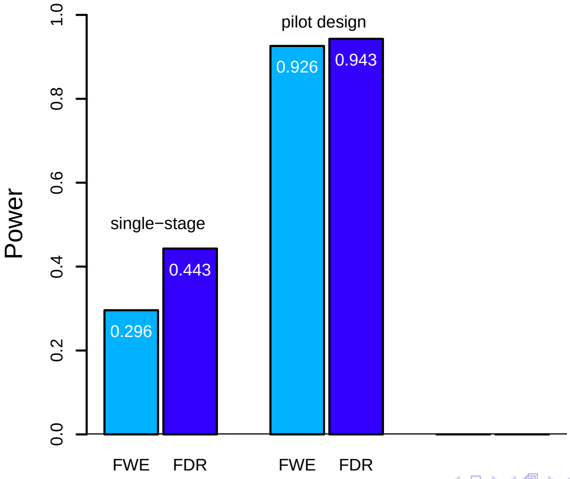
Asymptotic optimal designs

$C = 20000$, $m_1 = 1000$, $\Delta = 0.75$, $\alpha = 0.05$, $\pi_0 = 0.99$



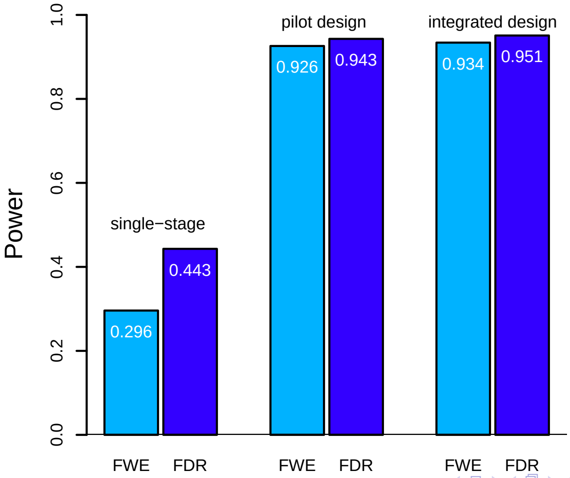
Asymptotic optimal designs

$C = 20000,\ m_1 = 1000,\ \Delta = 0.75,\ \alpha = 0.05,\ \pi_0 = 0.99$



Asymptotic optimal designs

$C = 20000$, $m_1 = 1000$, $\Delta = 0.75$, $\alpha = 0.05$, $\pi_0 = 0.99$



Example:

Pilot Design controlling the FWE

- 71.8 % of C are used in the first stage
- $n_1 = 20000 * 0.718/1000 \approx 14$
- $\gamma_1 = 0.092$
- $m_2 \approx$ 100 hypotheses are selected for the second stage
- $\gamma_2 \approx 0.05/100 = 0.0005$
- $n_2 = (1 - 0.718) * 20000/100 \approx 56$

In Genomic or Proteomic studies ...

Scenario 1: Different costs

Different costs per observation may arise at both stages:

- costs per observation in the first stage set to $c_1 = 1$
- cost ratio between stages $c_2 > 1$

Scenario 2: Different costs and effect sizes

There is an increasing focus on using a less accurate assay in early stages and a more accurate one in later stages:

- cost ratio $c_2 > 1$
- effect size ratio between stages $k > 1$

In Genomic or Proteomic studies ...

Scenario 1: Different costs

Different costs per observation may arise at both stages:

- costs per observation in the first stage set to $c_1 = 1$
- cost ratio between stages $c_2 > 1$

Scenario 2: Different costs and effect sizes

There is an increasing focus on using a less accurate assay in early stages and a more accurate one in later stages:

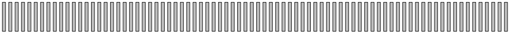
- cost ratio $c_2 > 1$
- effect size ratio between stages $k > 1$

The Two-Stage Design: Scenario 1

m_1 hypotheses

effect size Δ

$$n_1 = \frac{rC}{m_1}$$

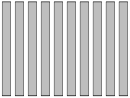


Select H_i: $p_{i1} < \gamma_1$

m_2 (random)

effect size Δ

$$n_2 = \frac{(1-r)C}{c_2 m_2}$$



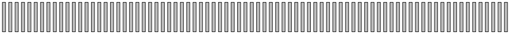
Test decision: Reject H_i: $p_{i2} < \gamma_2$

The Two-Stage Design: Scenario 2

m_1 hypotheses

effect size Δ

$n_1 = \frac{rC}{m_1}$

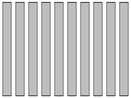


Select H_i: $p_{i1} < \gamma_1$

m_2 (random)

effect size $k\Delta$

$n_2 = \frac{(1-r)C}{c_2 m_2}$



Test decision: Reject H_i: $p_{i2} < \gamma_2$

Scenario 1: Example

Consider again the experiment with

- $m_1 = 1000$ ... number of hypotheses tests
- $C = 20000$ ... fixed total costs (limit of resources)

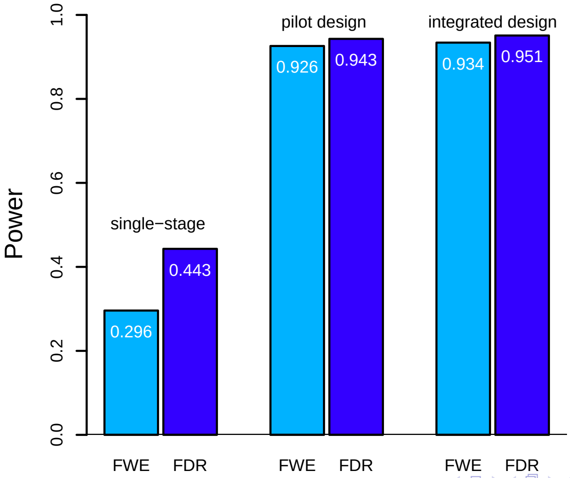
Given

- $\pi_0 = 0.99$ ... proportion of true null hypotheses among all m_1
- $\Delta = 0.75$ ... effect size

We now assume that there a cost ratio between stages $c_2 > 1$.

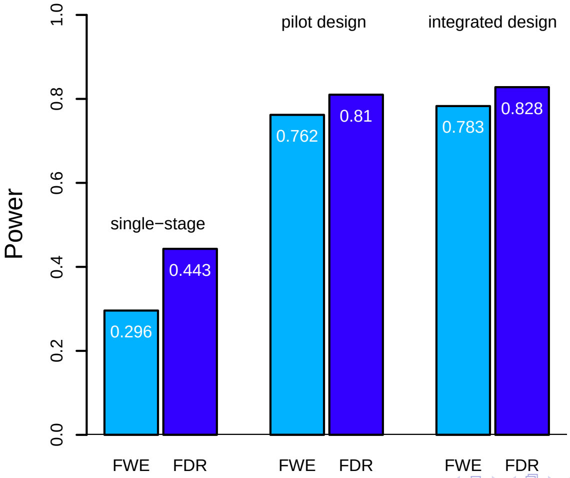
Asymptotic optimal designs

$C = 20000$, $m_1 = 1000$, $\Delta = 0.75$, $\alpha = 0.05$, $\pi_0 = 0.99$, $c_2 = 1$



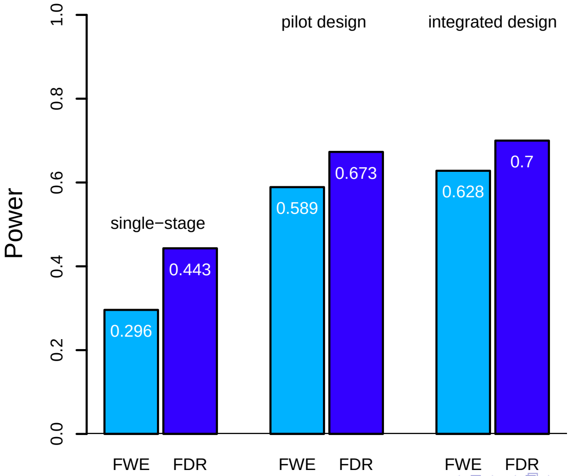
Asymptotic optimal designs

$C = 20000$, $m_1 = 1000$, $\Delta = 0.75$, $\alpha = 0.05$, $\pi_0 = 0.99$, $c_2 = 5$



Asymptotic optimal designs

$C = 20000$, $m_1 = 1000$, $\Delta = 0.75$, $\alpha = 0.05$, $\pi_0 = 0.99$, $c_2 = 15$



Break Even Point in Cost Ratio

If the cost ratio c_2 increases, the power of the two-stage design decreases.

Question

Is there a cost ratio c_2^*, where it does not make sense to apply a two-stage design as compared to the single-stage design?

Integrated design

c_2^* does not exist:
The power of the asymptotic optimal integrated design

- is always larger than of the corresponding single-stage design.
- converges to the power of the single-stage design ($\lim_{c_2 \to \infty} r = 1$)

Break Even Point in Cost Ratio

If the cost ratio c_2 increases, the power of the two-stage design decreases.

Question

Is there a cost ratio c_2^*, where it does not make sense to apply a two-stage design as compared to the single-stage design?

Integrated design

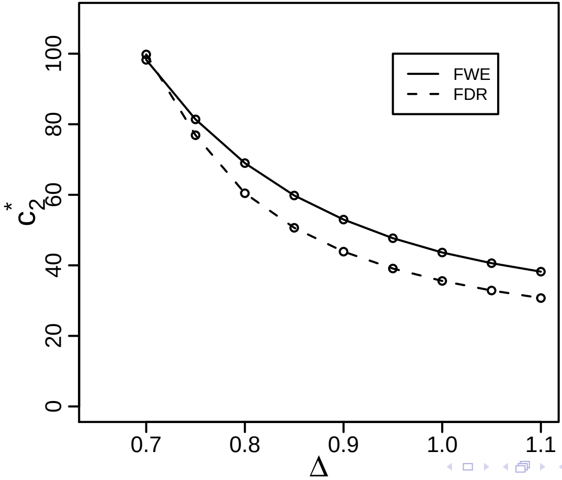
c_2^* does not exist:

The power of the asymptotic optimal integrated design

- is always larger than of the corresponding single-stage design.
- converges to the power of the single-stage design ($\lim_{c_2 \to \infty} r = 1$)

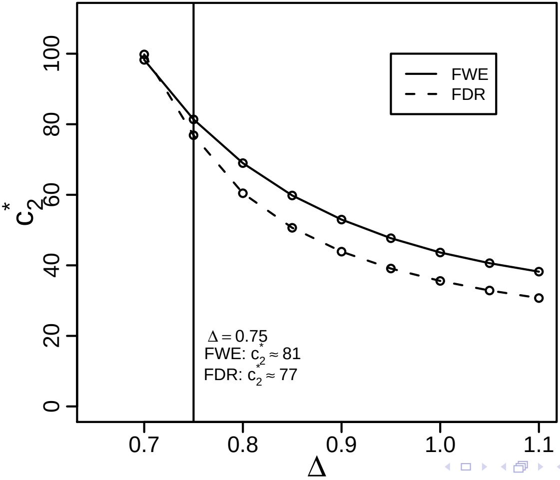
Break Even Point: Pilot Design

$C = 20000$, $m_1 = 1000$, $\pi_0 = 0.99$, $\alpha = 0.05$



Break Even Point: Pilot Design

$C = 20000$, $m_1 = 1000$, $\pi_0 = 0.99$, $\alpha = 0.05$



Break Even Point in Cost Ratio

- For the integrated design no such Break Even Point exists.
- For the pilot design such point exist but for small Δ it is large.
- Two-stage designs are a good option to increase the power even if the cost ratio per observation between stages is fairly high.

Impact of Design Misspecifications

Whereas costs are usually known a priori the optimal designs depend on the unknown parameters π_0 and Δ.

Is there an amount of misspecification where it would have been better to use a single-stage design?

Example

- $C = 20000$
- $m_1 = 1000$
- $c_2 = 15$

r and γ_1 planned for the situation:

- $\pi_0 = 0.99$
- $\Delta = 0.75$

Impact of Design Misspecifications

Whereas costs are usually known a priori the optimal designs depend on the unknown parameters π_0 and Δ.

Is there an amount of misspecification where it would have been better to use a single-stage design?

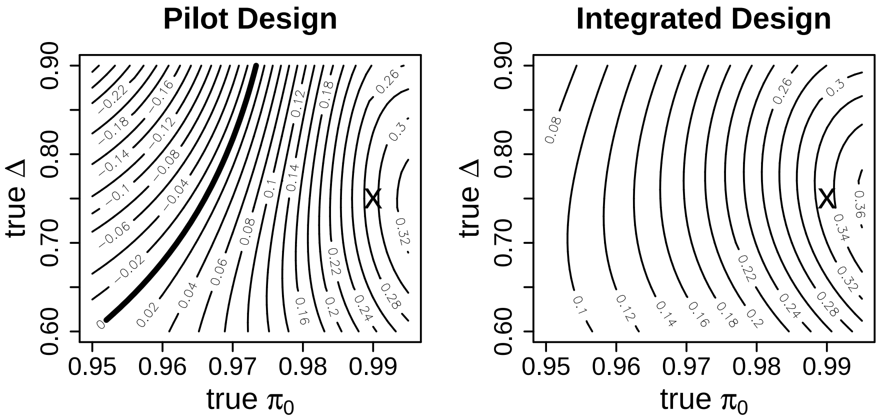
Example

- $C = 20000$
- $m_1 = 1000$
- $c_2 = 15$

r and γ_1 planned for the situation:

- $\pi_0 = 0.99$
- $\Delta = 0.75$

Difference of power values between the two-stage designs and the corresponding single-stage designs (Control of FWE):



X ... point for which two-stage design parameters are optimized

Misspecifications

- The integrated design is more robust against design misspecifications
- if the planned π_0 is larger than the true one:
  loss of power as compared to the single-stage design
- if the planned π_0 is smaller than the true one:
  increase of power as compared to the single-stage design

Scenario 2

The experimenter has two different candidate methods for the measurements from the very beginning:

- low-cost standard method: effect size $= \Delta$
  costs per observation = 1
- high-cost improved method: effect size $= k\Delta$
  costs per observation > 1
- cost ratio between methods $c_2 > 1$
- effect size ratio between methods $k > 1$

Two-Stage Procedures

- first stage: low second stage: low
- first stage: low second stage: high
- first stage: high second stage: high

Scenario 2

The experimenter has two different candidate methods for the measurements from the very beginning:

- low-cost standard method: effect size$= \Delta$
  costs per observation = 1
- high-cost improved method: effect size$= k\Delta$
  costs per observation > 1
- cost ratio between methods $c_2 > 1$
- effect size ratio between methods $k > 1$

Two-Stage Procedures

- first stage: low second stage: low
- first stage: low second stage: high
- first stage: high second stage: high

Pilot Design controlling the FWE

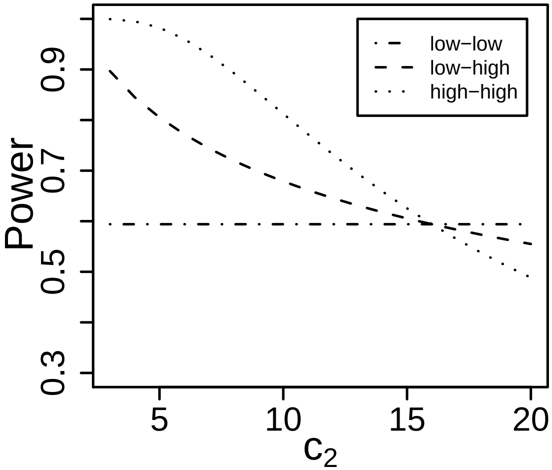
Consider an experiment with

- $\alpha = 0.05$
- $C = 20000$, $m_1 = 1000$
- low-cost method: effect size $\Delta = 0.5$
- high-cost method: $k = 4$ effect size: $0.5 * 4 = 2$

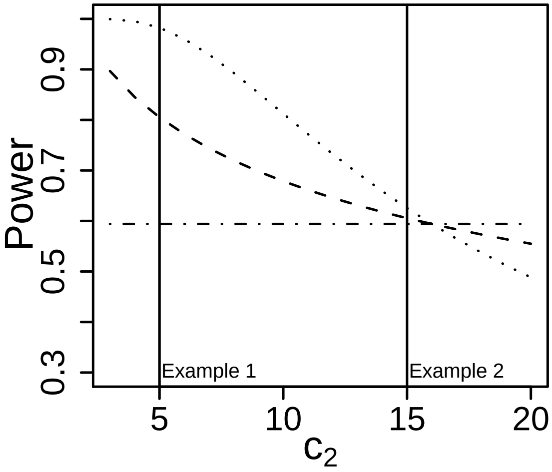
Question

Looking at different values for c_2:
Which of the 3 procedures has the maximal Power?

Common crossing point at $c_2 = k^2 = 16$



Common crossing point at $c_2 = k^2 = 16$



Examples:

Procedure		Example 1: $c_2 = 5$	Example 2: $c_2 = 15$
low-low	Power	0.594	0.594
	n_1	≈ 13	≈ 13
low-high	Power	0.805	0.605
	n_1	≈ 10	≈ 13
high-high	Power	0.983	0.625
	n_1	≈ 3	< 1

Scenario 2

- The low-high procedure may only be preferable if the high-cost method is too expensive to achieve the necessary repetitions at the first stage.
- In our example the high-cost method is not to expensive but highly effective. If we have only double effects for the expensive method than the low-low procedure would be already preferable if the costs are only 4 times larger.
- If the improved method is much more expensive it has to be much more effective to apply a high-high or a low-high procedure.

Conclusions

Different costs in both stages

Two-stage designs are a good option to improve the power even if the cost ratio between stages c_2 is fairly high.

Misspecification

- The integrated design is more robust against misspecification than the pilot design.
- Optimism in the planning phase with regard to the number of true alternatives may help to avoid loss of power.

If two different methods are available

- Depending on c_2 and k it is preferable to run two-stage designs which apply either the low-cost or the high-cost method at both stages.
- Switching from the low-cost to the high-cost method may only be advisable if there is lack of finance so that n_1 for the high-cost method is to small.

Selected References

- Benjamini,Y, Hochberg,Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing.
  J. R. Statist. Soc. B, 57, 289-300.
- Goll,A, Bauer,P (2007) Two-stage designs applying methods differing in costs.
  Bioinformatics, 23, 1519-1526.
- Satagopan,JM et al. (2004) Two-stage designs for gene-disease association studies. *Biometrics*, 58, 163-170.
- Storey,JD (2002) A direct approach to false discovery rates.
  J. R. Statist. Soc. B, 64, 479-498.
- Storey,JD, Taylor,JE, Siegmund,D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc. B*, 66, 187-205.
- Wang,H et al. (2006) Optimal two-stage genotyping designs for genome-wide association scans, *Genetic Epidemiology*, 30, 356-368.
- Zehetmayer,S, Bauer,P, Posch,M (2005) Two-stage designs for experiments with a large number of hypotheses.
  Bioinformatics, 21, 3771-3777.