Multi-Stage Designs controlling the False Discovery or the Family Wise Error Rate

Sonja Zehetmayer Peter Bauer Martin Posch

Section of Medical Statistics
Medical University of Vienna

Supported by FWF-Fund Nr. P18698-n15

Multiple Comparison Procedures
Vienna 2007

Motivation

- In gene disease association studies genetic markers (DNA sequences or genes) are searched which predict the effect of a therapy, the occurrence of a disease, ...

Limitations

- Very large number of candidate markers
- Only a very small number of markers have an influence

CONSTRAINT

- Not the number of patients is limited but the total costs or the total number of gene evaluations

Motivation

Consider a gene-disease association study with

- $m_1 = 5000$ hypotheses tests (candidate genes)
- limited resources of $N = 40000$ overall gene evaluations (total costs)

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

Single-stage design - Example

Distribute 40000 observations equally among the hypotheses

$\Rightarrow 40000/5000 = 8$ observations per hypothesis test

Control of the Family Wise Error Rate (FWE)

Definition

The probability of at least one Type I error among all m_1 hypotheses of an experiment.

Bonferroni adjustment - Example

- $N = 40000$, $m_1 = 5000$, 8 observations per hypothesis test
- For $\alpha = 0.05$: $\alpha/m_1 = 0.00001$
- Effect size $\Delta/\sigma = 0.5$
- *Power* $= 0.0022$

Control of the False Discovery Rate (FDR)

Definition

$$FDR = E(\frac{V}{max\{R, 1\}})$$

V . . . number of erroneously rejected hypotheses

R . . . number of rejected hypotheses

e.g., Benjamini and Hochberg (1995)

Storey's procedure to control the FDR at level α

- m_1 hypothesis tests
- Reject all hypotheses whose p-value p_i satisfies

$$p_i \leq \gamma,$$

- where γ is the largest real number satisfying

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

and

$\#\{p_i \leq \gamma\}$ . . . number of p-values not exceeding γ

$\hat{\pi}_0$ . . . estimate of proportion of true H_0 based on the p_i

Controlling the False Discovery Rate (Storey 2002)

Storey, Taylor and Siegmund (2004)

This procedure controls the FDR if the p-values corresponding to the true null hypotheses are independent and uniformly distributed.

FDR - Example

- $N = 40000$, $m_1 = 5000$, 8 observations per hypothesis test
- $\alpha = 0.05$
- Effect size $\Delta/\sigma = 0.5$

  FWE: *Power* $= 0.0022$

  FDR: *Power* $= 0.0025$

Multi-stage design

Design Constraint for gene-disease association studies

- Not the number of patients is limited but the total costs represented by the total number of gene evaluations.

E.g., Satagopan et al. (2002), Zehetmayer et al. (2005): Two-stage designs

We investigate

- increasing the number of stages (general framework)
- FWE versus FDR
- asymptotic optimal designs

Multi-stage design

m_1 hypotheses

$n_1 = \frac{r_1 N}{m_1}$

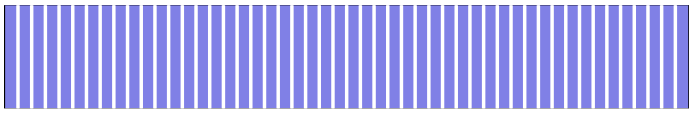
sample size

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

m_2 (random)

$n_2 = \frac{r_2 N}{m_2}$

(random)



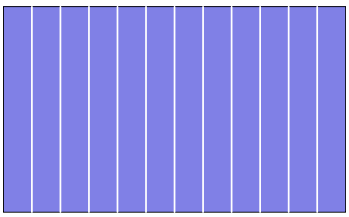
Select H_i: $p_{i2} < \gamma_2$

$\vdots$ $\qquad$ $\vdots$

m_k (random)

$n_k = \frac{r_k N}{m_k}$

(random)



Test decision: Reject H_i: $p_{ik} < \gamma_k$

The integrated design - procedure

m_1 hypotheses

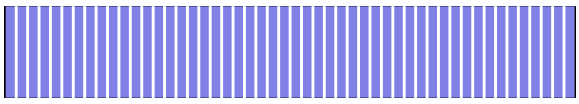
n_1

$\} p_{i1}$

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

m_2

n_2



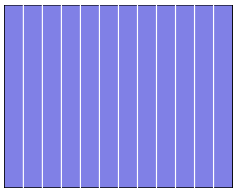
$\} p_{i2}$

Select H_i: $p_{i2} < \gamma_2$

$\vdots$

m_k

n_k



$\} p_{ik}$

Test decision: Reject H_i: $p_{ik} < \gamma_k$

The integrated design

The sequential p-value

At the end of the trial - either after

- early acceptance
- or reaching the final stage

an overall sequential p-value based on a monotonic ordering can be calculated for each hypothesis (Tsiatis et al, 1984).

Properties of the sequential p-values

Property for random sample sizes

The sequential p-values p_i^s of the integrated multi-stage design with random stage-wise per-hypothesis sample sizes are

- independent and
- uniformly distributed

under the null hypothesis (when the observations are independent).

REMARK

For the proof the γ_i are specified via a futility spending function.

Asymptotic optimal designs

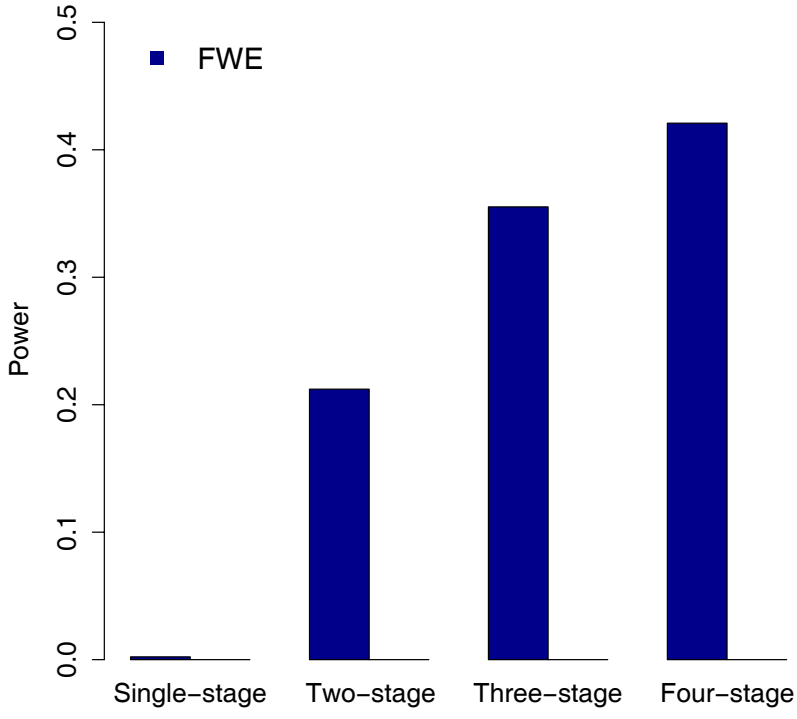
Optimization

- For given N, π_0, Δ/σ and α we optimize the probability to reject an alternative with respect to the parameters
  - $r_1, \ldots, r_k$ (fraction of total number of observation N for stage $1, \ldots, k$)
  - $\gamma_1, \ldots, \gamma_{k-1}$ (selection boundaries)
- letting $m_1 \to \infty$ and assuming $N = dm_1$ for $d > 0$,
- The optimal power and optimal parameters depend on N, m_1, and Δ/σ only via $\sqrt{N/m_1}\Delta/\sigma$.

Asymptotic optimal designs - FWE

Reject all H_{0i} with $p_i^s < \frac{\alpha}{m_1}$

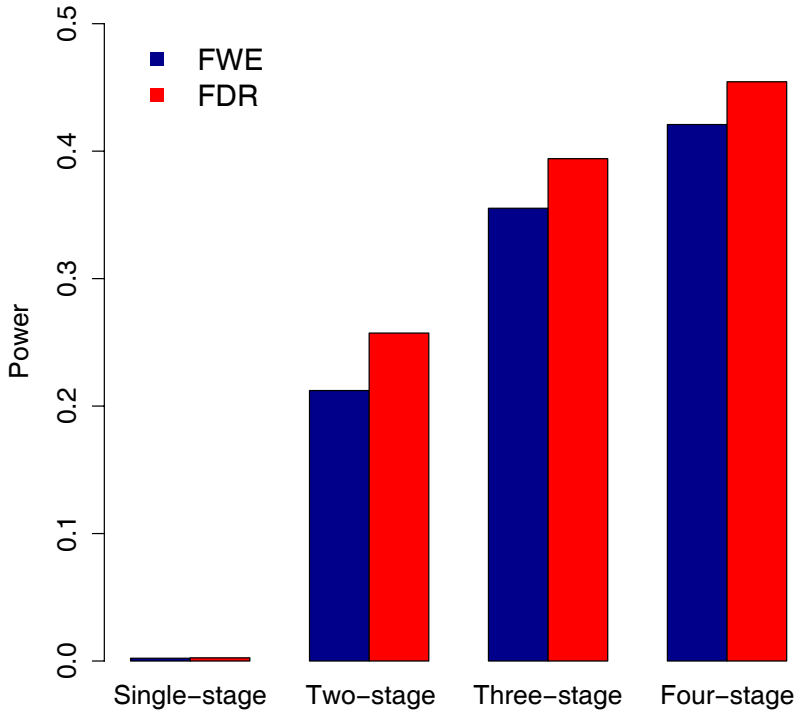
$N = 40000$
$m_1 = 5000$
$\pi_0 = 0.99$
$\Delta/\sigma = 0.5$



Asymptotic optimal designs - FDR

$\widehat{FDR}_{g_k}(p^s_1, \ldots, p^s_{m_1}) \leq \alpha$

$N = 40000$
$m_1 = 5000$
$\pi_0 = 0.99$
$\Delta/\sigma = 0.5$



Three-stage design - Asymptotic optimal parameters

- $N = 40000$ overall observations
- $m_1 = 5000$ hypotheses
- $\alpha = 0.05$

Consider we are planning an experiment with three stages for $\Delta/\sigma = 0.5$ and $\pi_0 = 0.99$.

FWE

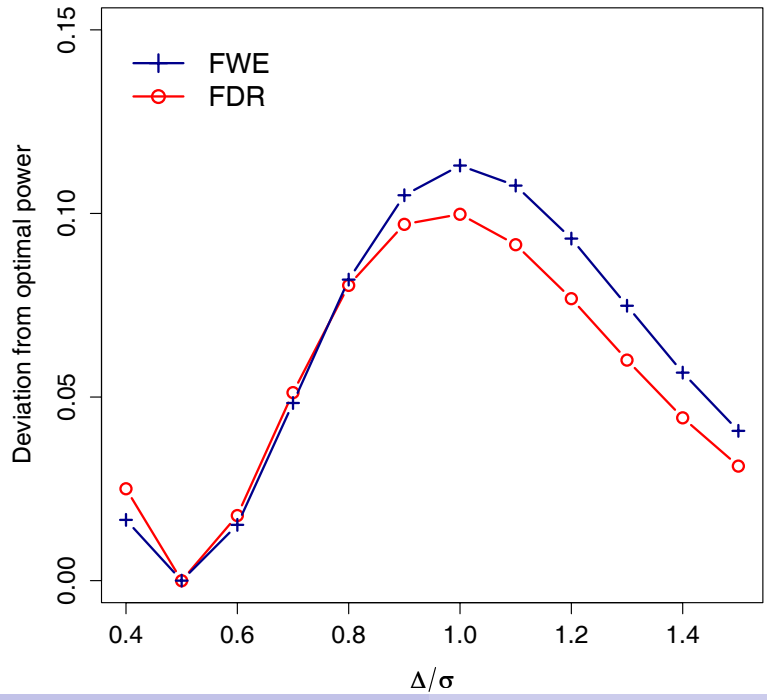
- $Power^{\infty} = 0.355$
- $m_2^{\infty} = 1413.1$
- $m_3^{\infty} = 191.9$
- $n_1^{\infty} = 3.2$
- $n_2^{\infty} = 16.8$
- $n_3^{\infty} = 90.5$

FDR

- $Power^{\infty} = 0.394$
- $m_2^{\infty} = 1683.4$
- $m_3^{\infty} = 316.9$
- $n_1^{\infty} = 3.2$
- $n_2^{\infty} = 14.1$
- $n_3^{\infty} = 59.2$

Asymptotic optimal designs - Miss-specification

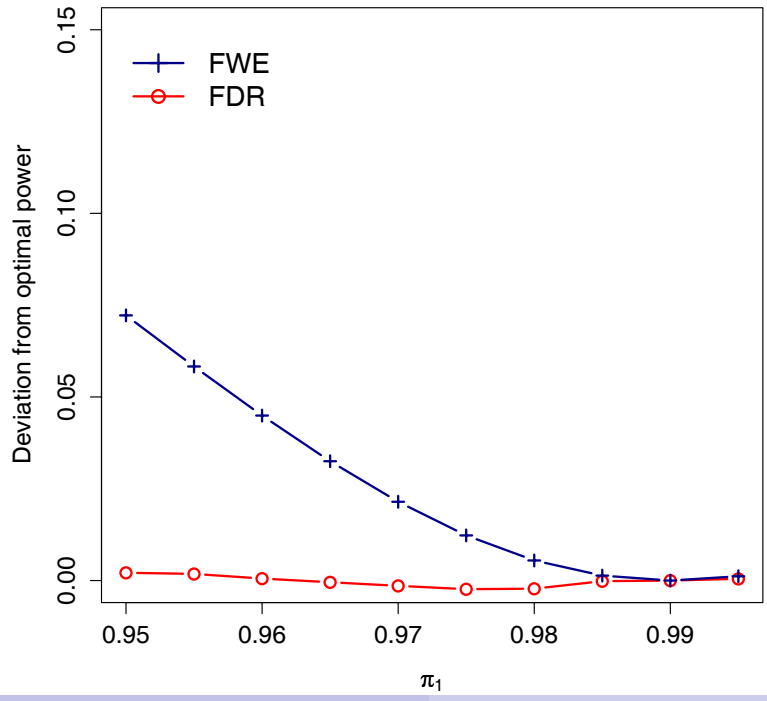
Difference between optimal power (if trial was planned based on actual values) and non-optimal power if Δ/σ was miss-specified?



Planned for:
$N = 40000$
$m_1 = 5000$
$\pi_0 = 0.99$
$\Delta/\sigma = 0.5$

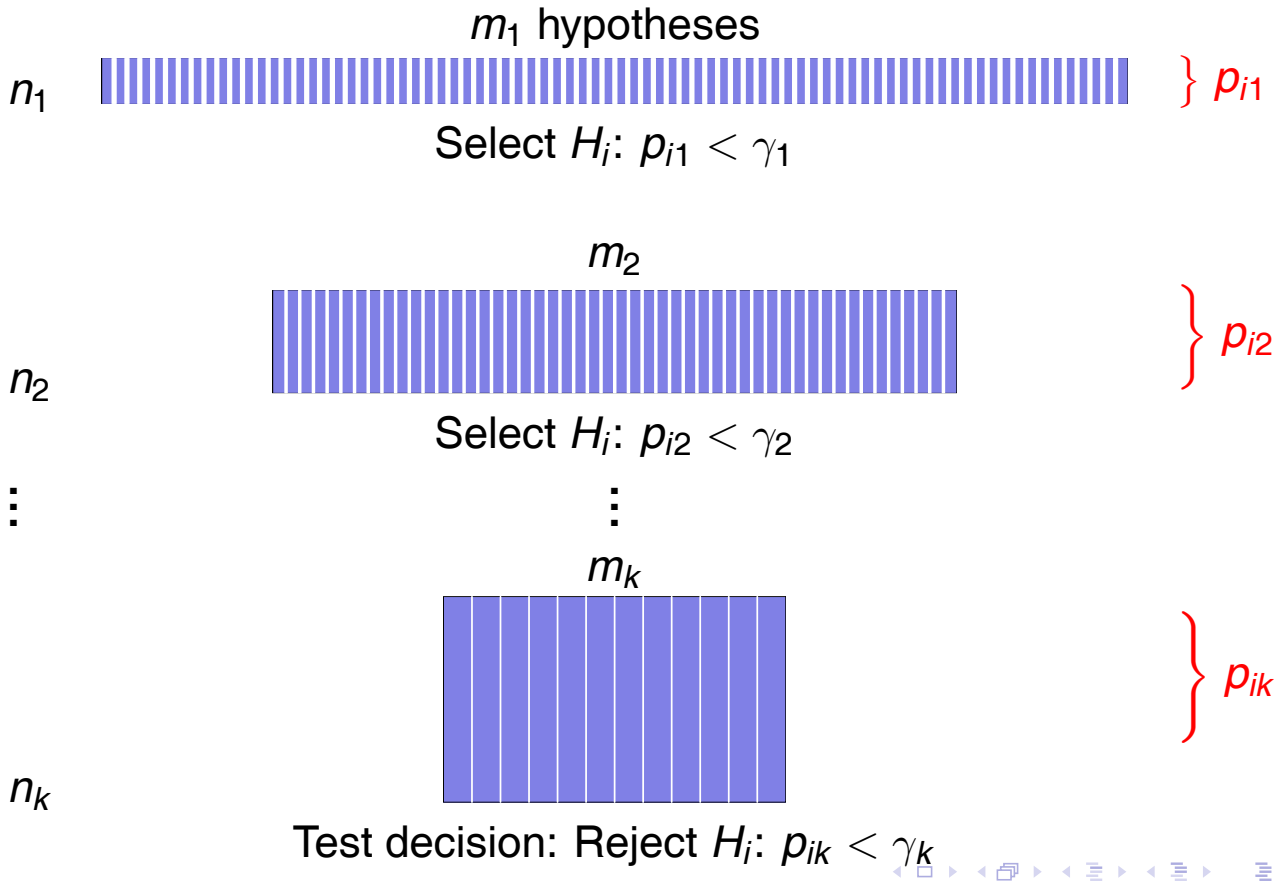
Asymptotic optimal designs - Miss-specification

Difference between optimal power (if trial was planned based on actual values) and non-optimal power if π_0 was miss-specified?



Planned for:
$N = 40000$
$m_1 = 5000$
$\pi_0 = 0.99$
$\Delta/\sigma = 0.5$

Remark: The pilot design



The pilot design

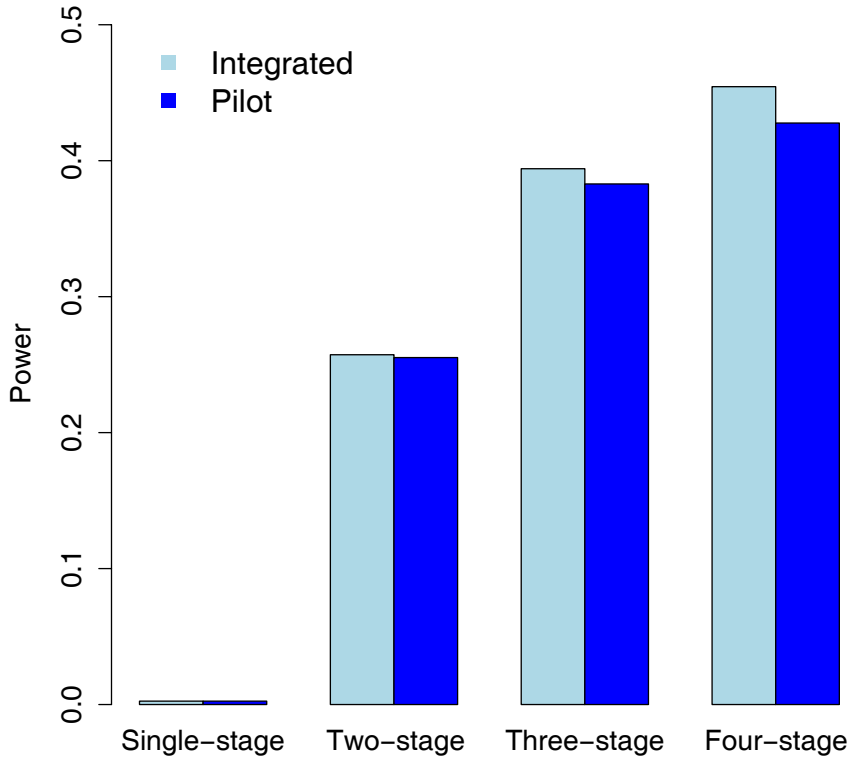
Procedure

- at each interim analysis, the p-value is computed using the observations from the preceding stage only
- The hypothesis tests are performed with the last stage data only

Error rates

- FWE: $\gamma_k = \frac{\alpha}{m_k}$
- FDR:
  - the stages can be considered as single-stage designs
  - the procedure by Storey for independent p-values can be applied
  - $\widehat{FDR}_{\gamma_k}(p_{1k}, \ldots, p_{m_k k})$

Pilot design controlling the FDR



$N = 40000$
$m_1 = 5000$
$\pi_0 = 0.99$
$\Delta/\sigma = 0.5$

Conclusions

- Multi-stage designs are strikingly superior to single-stage designs
- even when in the simple pilot design only the last stage data is used for the final test decision

The crucial point is not

- the choice of the error rate - FWE versus FDR
- the type of design - integrated or pilot

but skipping non-promising hypotheses in early phases of the experiment!

Under miss-specification:
integrated design controlling the FDR seems to be the best choice

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.
- Satagopan,JM, et al. (2002) Two-stage designs for gene-disease association studies. *Biometrics*, **58**, 163-170.
- 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.
- Tsiatis,AA, Rosner,GL, Mehta,CR (1984) Exact confidence intervals following a group sequential test, *Biometrics*, 40, 797-804.
- Zehetmayer,S, Bauer,P, Posch,M (2005) Two-stage designs for experiments with a large number of hypotheses, *Bioinformatics*, 21, 3771-3777.
- Zehetmayer,S, Bauer,P, Posch,M (2007) Optimized multi-stage designs controlling the FDR or the FWE. Submitted.