Nuisance-parameter based sample size reestimation in adaptive enrichment designs with an application in major depression

Marius Placzek, Tim Friede

Department of Medical Statistics, University Medical Center Göttingen, Germany

BMBF project (BundesMinisterium für Bildung und Forschung) „**BIOSTATISTISCHE METHODEN ZUR EFFIZIENTEN EVALUATION VON INDIVIDUALISIERTEN THERAPIEN (BIMIT)“.**

Motivating examples in major depression

Combination of baseline variables to predict treatment response

- Frank et. al (2011) compared a pharmacotherapy and a psychotherapy in a randomized controlled study (n=318)
- 17-item Hamilton Depression Rating Scale (HDRS-17) over 12 weeks as efficacy endpoint
- A number of baseline variables which are predictive of treatment outcome were identified and combined to an *optimal moderator* of treatment effect (Kraemer HC, 2013)
- For patients above (below) a certain threshold of the optimal moderator psychotherapy was superior (inferior) to pharmacotherapy

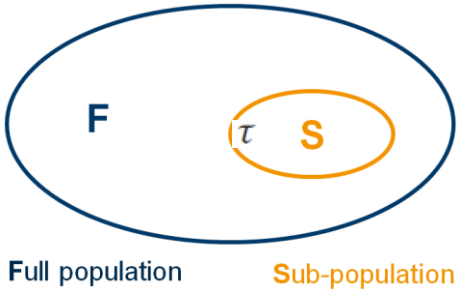
Serum BDNF levels as predictor of treatment response

- A number of small uncontrolled studies identified baseline serum levels of brain derived neurotrophic factors (sBDNF) to predict treatment response to various pharmacological treatments (including duloxetine)
- HDRS-17 over 6 weeks (Mikoteit et al, 2014) or 8 weeks (Wolkowitz et. al, 2011)

In both examples the identified subgroups need to be confirmed in a RCT!

Adaptive Enrichment Design

Basic concept:



- **Stage 1**: Recruit patients from full population (F)
- **Interim analysis**: make the decisions on …
  - whether trial is stopped for futility
  - if trial is continued, decide whether recruitment in **Stage 2** is from full population (F) or subpopulation (S) (enrichment)
    - e.g. epsilon-decision rule (Kelly et al 2005)
  - testing strategy in final analysis
- **Final analysis**: test for an effect in F and/ or S

Hypotheses and Test Statistics

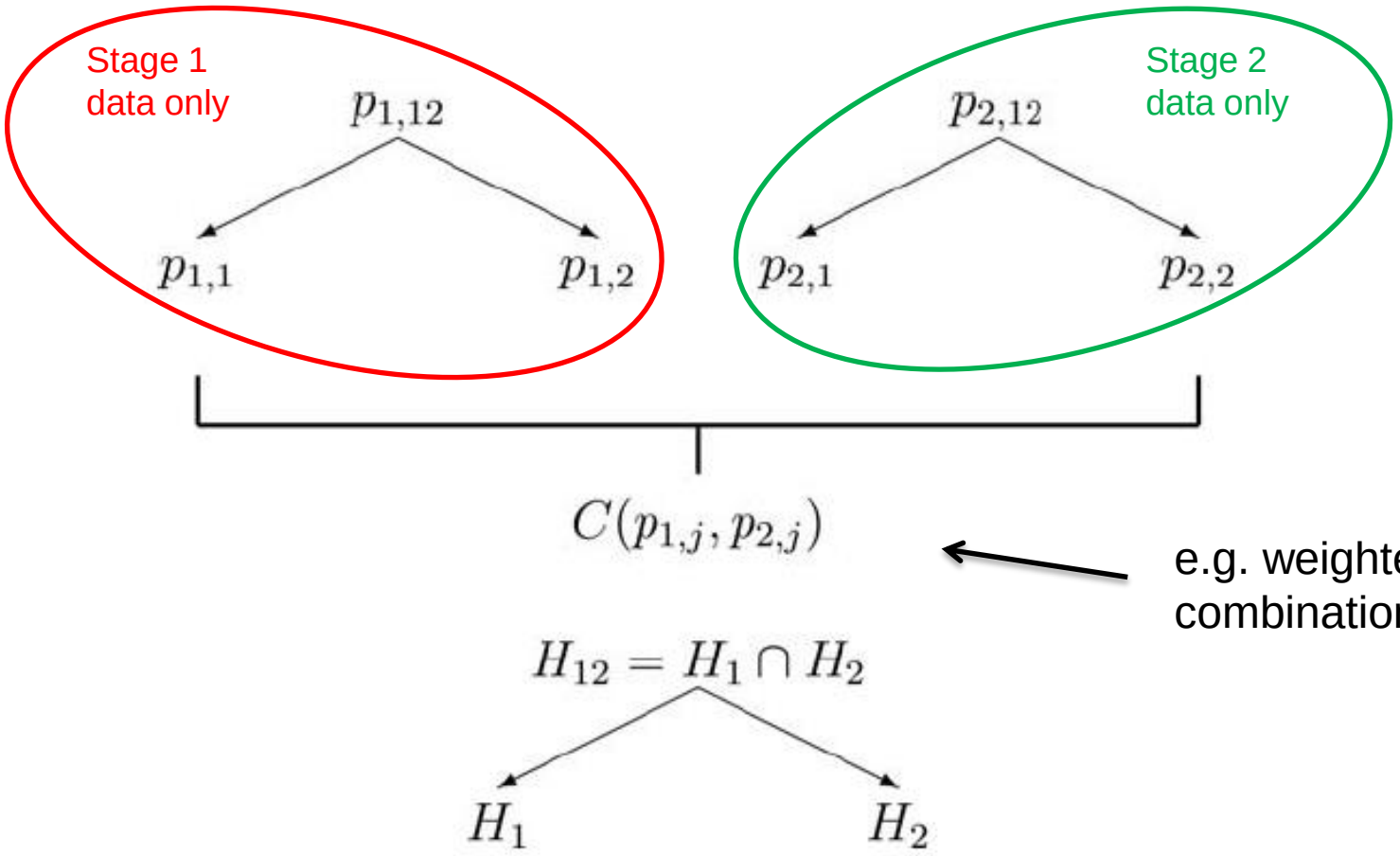
UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

- Normal distributed endpoints
- individual hypotheses $H_0^{\{F\}}$ (no effect in full population)

  $H_0^{\{S\}}$ (no effect in subpopulation)
- intersection hypothesis $H_0^{\{F,S\}}$ (no effect in full and subpopulation)
- standardized test statistics $Z^{\{F\}} = \sqrt{\frac{n}{2}} \frac{\bar{X}_F^T - \bar{X}_F^P}{\hat{\sigma}_F}$, $Z^{\{S\}} = \sqrt{\frac{n\hat{\tau}}{2}} \frac{\bar{X}_S^T - \bar{X}_S^P}{\hat{\sigma}_S}$
  - depend on estimates of nuisance parameters σ_F^2 , σ_S^2 , τ
- under $H_0^{\{F,S\}}$

$$\begin{pmatrix} Z^{\{F\}} \\ Z^{\{S\}} \end{pmatrix} \dot{\sim} MN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\tau} \\ \sqrt{\tau} & 1 \end{pmatrix} \right)$$

Combination Test and Closure Principle

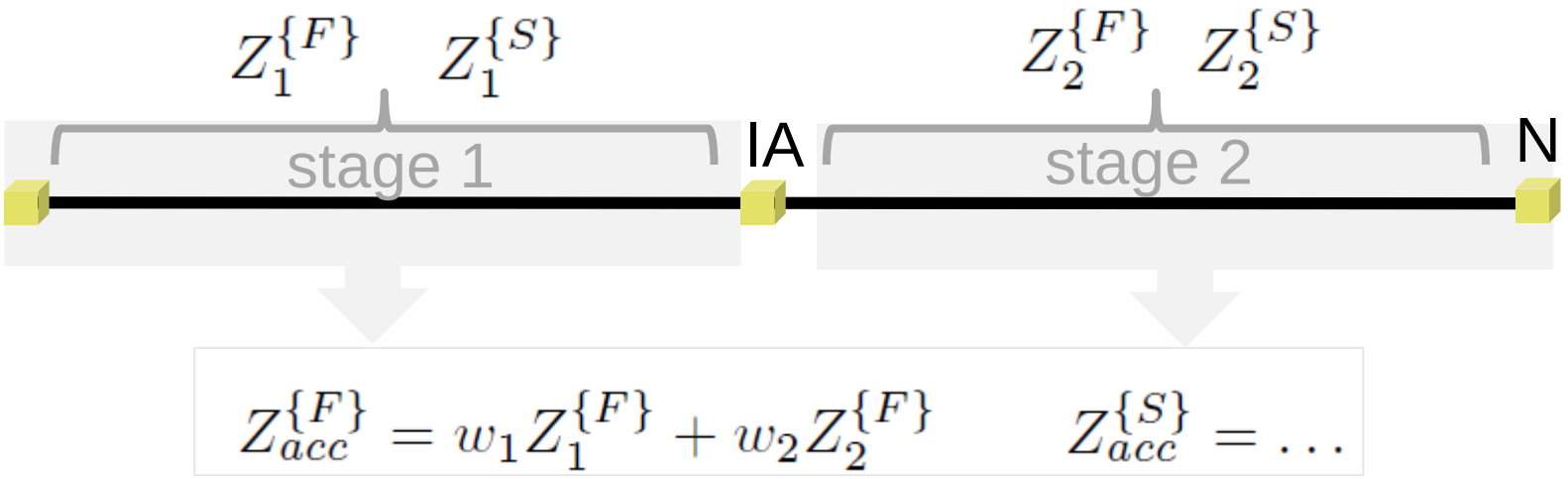


e.g. weighted inverse normal combination function

Figure 4 Closure principle for testing adaptively $n = 2$ null hypotheses H_1 and H_2.

Figure taken from Bretz el al (2006) Biometrical Journal

Conditional Error Function Approach



- use **Stage 1** data to calculate the conditional error

$$CE = \mathbf{P}\left(\text{reject } H_0(Z_{acc}^{\{F\}}, Z_{acc}^{\{S\}}) | z_1^{\{F\}}, z_1^{\{S\}}\right)$$

- after **Stage 2** test with the accumulated data to the level of the conditional error
- For each individual hypothesis → apply closed testing procedure

Sample Size Calculation

- under the alternative

$$\begin{pmatrix} Z^{\{F\}} \\ Z^{\{S\}} \end{pmatrix} \dot{\sim} \mathcal{N}\left(\underbrace{\begin{pmatrix} \sqrt{\frac{n}{2}} \frac{\Delta_F}{\sigma_F} \\ \sqrt{\frac{n\tau}{2}} \frac{\Delta_S}{\sigma_S} \end{pmatrix}}_{\boldsymbol{\delta}}, \underbrace{\begin{pmatrix} 1 & \sqrt{\tau} \\ \sqrt{\tau} & 1 \end{pmatrix}}_{\boldsymbol{V}} \right)$$

- let $G_{\mathcal{N}(\boldsymbol{\delta},\boldsymbol{V})}$ denote the distribution function of $\mathcal{N}(\boldsymbol{\delta},\boldsymbol{V})$ and $z_{\mathcal{N}(\mathbf{0},\boldsymbol{V}),1-\alpha}$ the $(1-\alpha)$ - equicoordinate quantile of $\mathcal{N}(\mathbf{0},\boldsymbol{V})$
- use estimates of nuisance parameters and effect sizes, e.g. based on previous studies, to calculate the initial sample size via

$$N_{init} = \min\ n,\ \text{s.t. } 1 - G_{\mathcal{N}(\boldsymbol{\delta},\boldsymbol{V})}(z_{\mathcal{N}(\mathbf{0},\boldsymbol{V}),1-\alpha}) \geq 1-\beta$$

Problems?

- misspecifications of nuisance parameters
- example: variance of 17-HDRS outcome
- Cipriani et al. (2012)

Duloxetine versus other anti-depressive agents for depression (Review)

The Cochrane Library

Baseline HDRS score	22.3±5.1
(Mikoteit 2014)	22.0±4.1

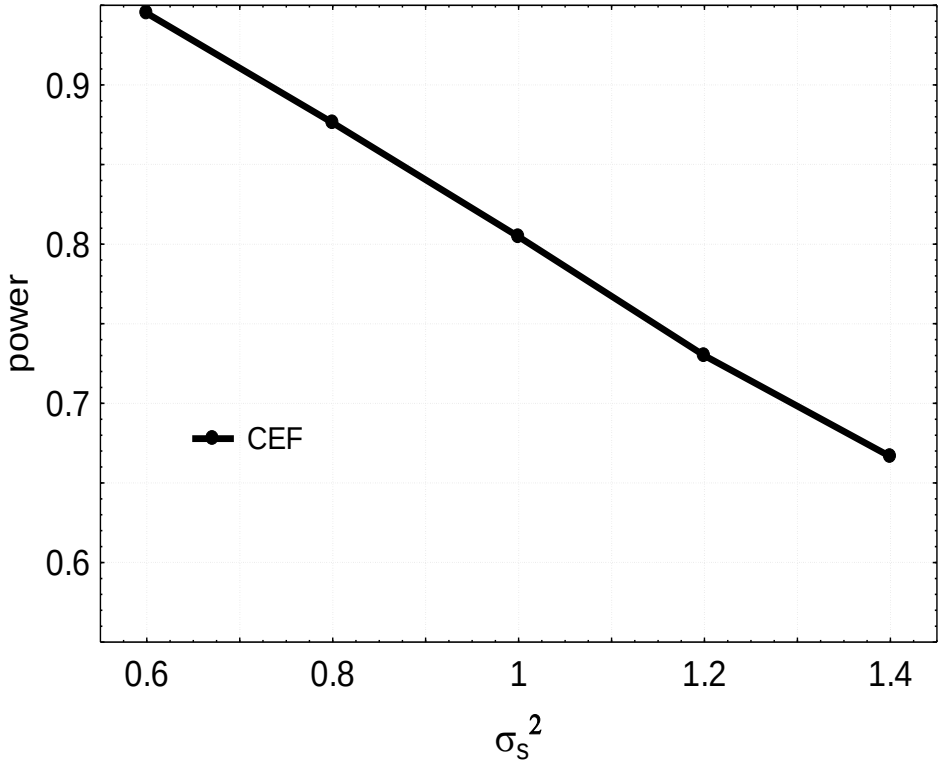
(Wolkowitz 2011)	Escitalopram	
	Baseline	(Week 8)
HDRS-17	26.1 ± 8.3	13.2 ± 8.9

Study or subgroup	Duloxetine N	Mean(SD)	SSRIs N	Mean(SD)	Differer IV,Random,9
1 versus Escitalopram					
Khan 2007 (SCT-MD-23)	126	-9.6 (7.6)	136	-11.1 (6.9)	
Nierenberg 2007 (HMCR)	273	-7.61 (6.93)	274	-7.22 (6.62)	
Wade 2007	146	-11.51 (6.3)	141	-12.77 (6.3)	
Subtotal (95% CI)	**545**		**551**		
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 4.63$, df = 2 (P = 0.10); I^2 =57%					
Test for overall effect: Z = 1.04 (P = 0.30)					
2 versus Fluoxetine					
Goldstein 2002 (HMAQ A)	68	-7.87 (6.98)	33	-6.09 (7.55)	
ID#3327 (HMAQ B)	81	-6.25 (7.12)	37	-7.05 (6.64)	
Subtotal (95% CI)	**149**		**70**		
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 1.53$, df = 1 (P = 0.22); I^2 =35%					
Test for overall effect: Z = 0.32 (P = 0.75)					
3 versus Paroxetine					
Detke 2004 (HMAY A)	93	-11.32 (6.29)	85	-11.06 (5.93)	
Goldstein 2004 (HMAT B)	86	-7.72 (7.67)	84	-6.06 (8.12)	
Higuchi 2009a	66	9.8 (5.8)	128	9.6 (6.3)	
ID#4091 (HMAT A)	81	-5.54 (6.68)	87	-6.2 (6.97)	
Lee 2007 (HMCV)	238	11.73 (4.57)	240	11.94 (4.59)	
Perahia 2006 (HMAY B)	102	-11.72 (5.78)	97	-10.84 (7)	
Subtotal (95% CI)	**666**		**721**		

Problems?

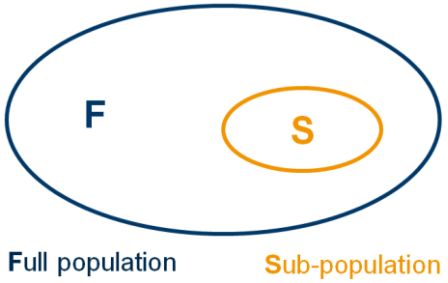
- For example here: misspecification of σ_S^2

  - Adaptive design: CEF approach
  - nsim=10,000, N=128
  - prevalence $\tau = 0.5$
  - $\alpha = 0.025,\ 1 - \beta = 0.8$
  - $\Delta_{F \setminus S} = 0,\ \Delta_S = 0.5$
  - $\sigma_F^{2*} = \sigma_F^2 = 1,\ \ \sigma_S^{2*} = 1$



- motivation for sample size recalculation procedure
  - Internal Pilot Study Design (Wittes & Brittain, 1990)

IPS design with Blinded Review



- here: nuisance parameters σ_F^2, σ_S^2 and τ
- after $n_1 = p^* N_0$ subjects per group (treatment/control):
  - blinded reestimation via „lumped variance“

$$\widehat{\sigma}^2_{F,OS}\ , \qquad \widehat{\sigma}^2_{S,OS} = \frac{1}{2n_{1S}-1} \sum_{i\in\{T,C\}} \sum_{j=1}^{n_{1S}} (X_{ij} - \overline{X}_S)^2\ , \qquad \widehat{\tau} = n_{1S}/n_1$$

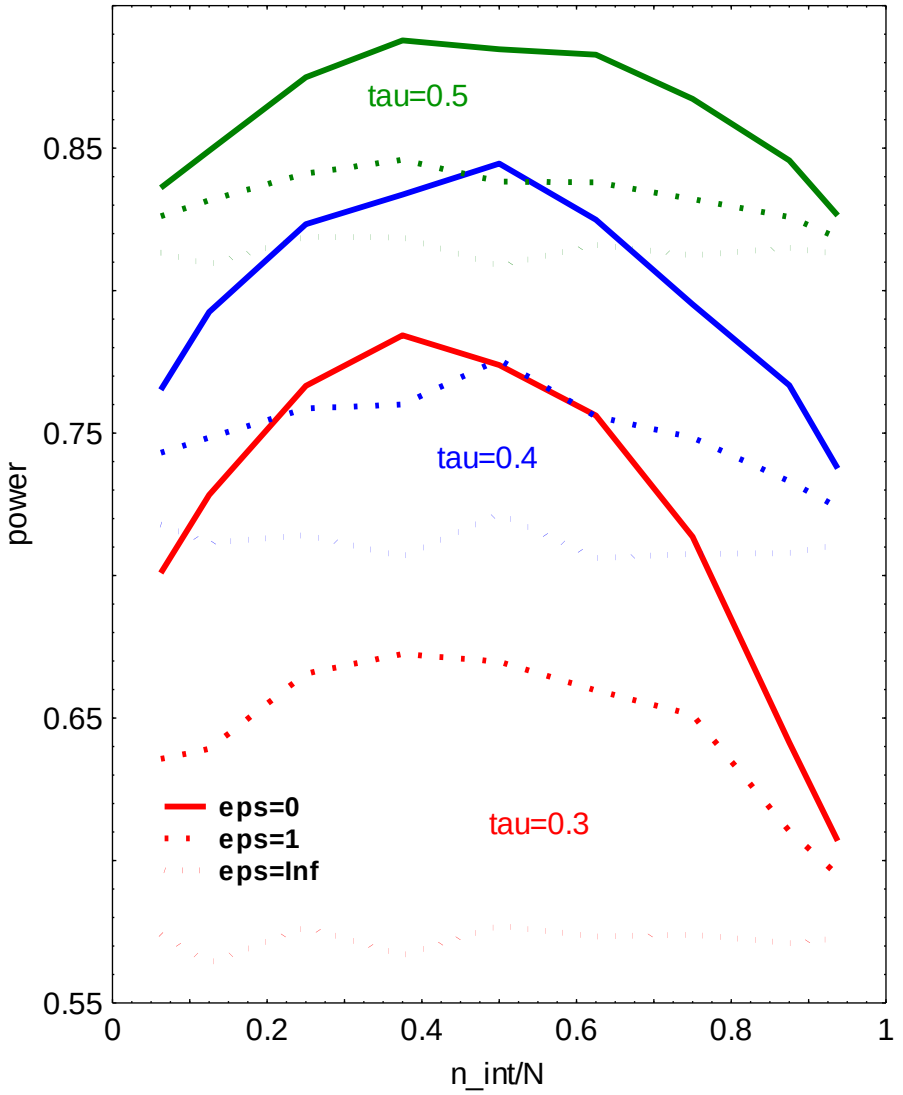
  - here OS=OneSample means no unblinding of treatment/control group
  - plug in new estimates and recalculate sample size N= n_1+n_2 for final analysis

Optimal Timepoint for Interim Analysis?

- Adaptive design: CEF approach
- Simulation results for nsim=10,000
- N=400 subjects per group (treatment/placebo)
- under the alternative

$$\Delta_{F\setminus S} = 0,\ \Delta_S = 0.3$$

- maximum in power after 40-50% of the subjects



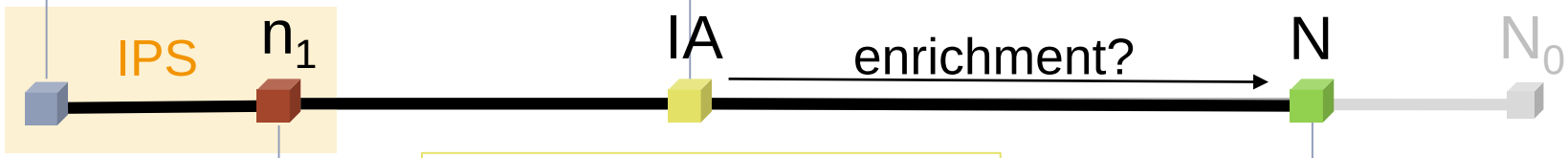
Combine BSSR and Adaptive Enrichment Methods

use model assumptions to calculate initial sample size N_0

Stage 1 data: $n_1 = 0.3 \cdot N_0$

Stage 2 data: 70 % of these N

stop for interim analysis



IA: decision on

→ population

→ testing strategy

final analysis

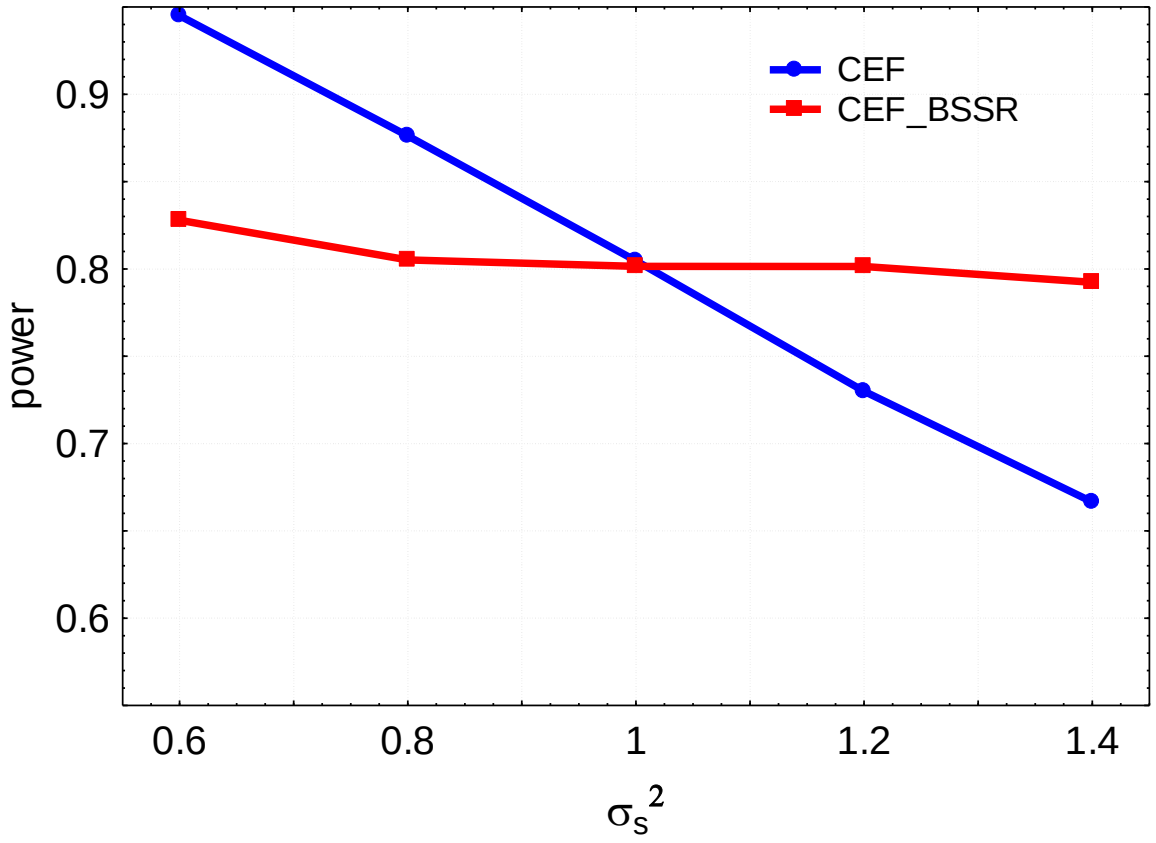
at prespecified percentage of subjects, e.g. 30 %, ($n_1 = 0.3 \cdot N_0$) perform BSSR and calculate final sample size N

i) only in full cohort

ii) both populations

Combine BSSR and Adaptive Enrichment Methods

- nsim=10,000, $\tau = 0.5$
- $\alpha = 0.025,\ 1-\beta = 0.8$
- $\Delta_{F\setminus S} = 0,\ \Delta_S = 0.5$
- $\sigma_F^{2*} = \sigma_F^2 = 1,\ \ \sigma_S^{2*} = 1$
- BSSR at 30% of N_0
- Interim Analysis at 50% of N ($\varepsilon = 1$)



Combine BSSR and Adaptive Enrichment Methods

type-I-error rates

- nsim=100,000, $\tau^* = 0.4$
- $\sigma_F^2 = \sigma_S^2 = \sigma_F^{2*} = \sigma_S^{2*} = 1$
- $\alpha = 0.025$
- $\Delta_{F \setminus S} = \Delta_S = 0$
- BSSR at 30% of N_0
- Interim Analysis at 50% of N ($\varepsilon = 1$)

tau	N_0	N	CTSD	CEF
0.3	662	843	0.02435	0.02509
0.4	662	664	0.02491	0.02581
0.5	662	537	0.02426	0.02489
0.3	371	475	0.02487	0.02528
0.4	371	375	0.02426	0.02528
0.5	371	303	0.02506	0.02516
0.3	237	305	0.02561	0.02594
0.4	237	241	0.02484	0.02558
0.5	237	194	0.02462	0.02500
0.3	169	213	0.02531	0.02564
0.4	169	168	0.02566	0.02578
0.5	169	136	0.02553	0.02583

Conclusions & Discussion

- Combination of BSSR and Adaptive Enrichment Methods leads to robust and flexible design
- increasing computational time due to computational complexity with increasing number of subgroups (simulations in planning stage)
- extension to nonnormal endpoints, e.g. count data
- include modeling of drop-outs

References

- Friede T, Parsons N, Stallard N (2012) *A conditional error function approach for subgroup selection in adaptive clinical trials*. Statistics in Medicine 31: 4309-4320
- Wittes J, Brittain E (1990). *The role of internal pilot studies in increasing the efficiency of clinical trials*. Statistics in Medicine 9: 65–72.
- Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A (2009) *Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology*. Statistics in Medicine 28:1445–1463
- Bauer P, Köhne K (1994) *Evaluation of experiments with adaptive interim analyses*. Biometrics 50: 1029-1041
- Bretz F, Schmidli H, König F, Racine A, Maurer W (2006) *Confirmatory seamless phase II/III clinical trials with hypothesis selection at interim: general concepts*. Biometrical Journal 48: 623-634

References

- Frank E et al. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. Psychological Medicine 41(1): 151–162.
- Kraemer HC (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. Statistics in Medicine 32: 1964-1973
- Mikoteit T et al. (2014). High baseline BDNF serum levels and early psychopathological improvement are predictive of treatment outcome in major depression. Psychopharmacology 231: 2955-2965
- Wolkowitz OM (2011). Serum BDNF levels before treatment predict SSRI response in depression. Progress in Neuro-Psychopharmacology & Biological Psychiatry 35: 1623-1630
- Cipriani A et al. (2012). Duloxetine versus other anti-depressive agents for depression (Review). The Cochrane Library 10