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

- 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
  
  \[
  Z^F = \sqrt{\frac{n}{2}} \frac{\bar{X}_F - \bar{X}_P}{\hat{\sigma}_F}, \quad Z^S = \sqrt{\frac{n\hat{\tau}}{2}} \frac{\bar{X}_S - \bar{X}_P}{\hat{\sigma}_S}
  \]

  - depend on estimates of nuisance parameters \( \sigma_F^2, \sigma_S^2, \tau \)

- under \( H_0^{F,S} \)

\[
\begin{pmatrix}
  Z^F \\
  Z^S
\end{pmatrix} \sim MN \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\tau} \\ \sqrt{\tau} & 1 \end{pmatrix} \right)
\]
Combination Test and Closure Principle

Stage 1
data only

Stage 2
data only

$C(p_{1,j}, p_{2,j})$

$H_{12} = H_1 \cap H_2$

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

Figure taken from Bretz et al (2006) Biometrical Journal

e.g. weighted inverse normal combination function
Conditional Error Function Approach

\[ Z_{\text{acc}}^{\{F\}} = w_1 Z_1^{\{F\}} + w_2 Z_2^{\{F\}} \quad Z_{\text{acc}}^{\{S\}} = \ldots \]

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

\[ CE = \mathbb{P}\left( \text{reject } H_0(Z_{\text{acc}}^{\{F\}}, Z_{\text{acc}}^{\{S\}}) \mid z_1^{\{F\}}, z_1^{\{S\}} \right) \]

- after **Stage 2** test with the accumulated data to the level of the conditional error

- For each individual hypothesis \( \rightarrow \) apply closed testing procedure
Sample Size Calculation

- under the alternative

\[
\left( \begin{array}{c}
Z\{F\} \\
Z\{S\}
\end{array} \right) \sim \mathcal{N} \left( \left( \begin{array}{c}
\sqrt{\frac{n}{2}} \frac{\Delta_F}{\sigma_F} \\
\sqrt{\frac{n}{2}} \frac{\Delta_S}{\sigma_S}
\end{array} \right), \left( \begin{array}{cc}
1 & \sqrt{\tau} \\
\sqrt{\tau} & 1
\end{array} \right) \right)
\]

- let \( G_{\mathcal{N}(\delta, V)} \) denote the distribution function of \( \mathcal{N}(\delta, V) \) and the \( (1 - \alpha) \)-equicoordinate quantile of \( \mathcal{N}(0, V) \)

- use estimates of nuisance parameters and effect sizes, e.g. based on previous studies, to calculate the initial sample size via

\[
N_{\text{init}} = \min n, \text{ s.t. } 1 - G_{\mathcal{N}(\delta, V)} \left( z_{\mathcal{N}(0, V), 1-\alpha} \right) \geq 1 - \beta
\]
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)

| 17 | 26.1±8.3 | 13.2±8.9 |

---

**Subtotal (95% CI)**

- 666
- 721
Problems?

- For example here: misspecification of $\sigma^2_S$.

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

- motivation for sample size recalculation procedure
  - Internal Pilot Study Design (Wittes & Brittain, 1990)
IPS design with Blinded Review

- here: nuisance parameters $\sigma_F^2$, $\sigma_S^2$ and $\tau$

- after $n_1=p^* \ N_0$ subjects per group (treatment/control):
  - blinded reestimation via „lumped variance“

\[
\hat{\sigma}^2_{F,\text{OS}}, \quad \hat{\sigma}^2_{S,\text{OS}} = \frac{1}{2n_{1S}-1} \sum_{i \in \{T,C\}} \sum_{j=1}^{n_{1S}} (X_{ij} - \bar{X}_S)^2, \quad \hat{\tau} = n_{1S}/n_1
\]

- 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 50% of these

Stage 1 data 40-50% of these

stop for interim analysis

Stage 2 data

IA

enrichment?

IA: decision on at prespecified percentage of subjects, e.g. 30%, \( n_1 = 0.3 \times N_0 \), perform BSSR and calculate final sample size \( N \)

\[ N = N_0 \]

\[ n_1 \]

\[ \text{testing strategy} \]

Nuisance-parameter based sample size reestimation in adaptive enrichment designs, Marius Placzek, 24.06.2015 © UMG
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 $N$ \( (\varepsilon = 1) \)

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


References


- 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


- Cipriani A et al. (2012). Duloxetine versus other anti-depressive agents for depression (Review). The Cochrane Library 10