Group sequential designs with negative binomial data

Ekkehard Glimm$^1$  Tobias Mütze$^{2,3}$

$^1$Statistical Methodology, Novartis, Basel, Switzerland
$^2$Department of Medical Statistics, University Medical Center Göttingen
$^3$DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany

Thanks to Tim Friede and Heinz Schmidli

PSI/BBS One-day meeting Basel

September 13, 2016
Motivating examples

Fixed design

Group sequential designs

Assessing operating characteristics

Discussion and outlook
Example 1: Clinical trials in heart failure

- Heart failure (HF) with preserved ejection fraction (HFpEF)
- Primary endpoint: Number of heart failure hospitalizations (HFH)
- HFH can be modeled with negative binomial distribution (Rogers et al., 2014)
- Example: the CHARM-Preserved trial (Yusuf et al., 2003)

**Table: Heart failure hospitalizations in CHARM-preserved**

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Candesartan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>1509</td>
<td>1514</td>
</tr>
<tr>
<td>Total follow-up years</td>
<td>4374.03</td>
<td>4424.62</td>
</tr>
<tr>
<td>Patients with $\geq 1$ admission</td>
<td>278</td>
<td>230</td>
</tr>
<tr>
<td>Total admissions</td>
<td>547</td>
<td>392</td>
</tr>
</tbody>
</table>

- Rate ratio for recurrent heart failure hospitalizations according to negative binomial model $\theta = 0.71$
Example 2: Clinical trials in relapsing-remitting multiple sclerosis

- Primary endpoint: number of combined unique active lesions (CULAs)
- CULAs are modeled using the negative binomial distribution
- Example: Phase II study of Siponimod (Selmaj et al., 2013)
  - Placebo and five doses of Siponimod
  - Equal follow-up times (in general either 3 or 6 months)

**Table:** Monthly number of lesions (at 3 months)

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Siponimod 0.25 mg</th>
<th>Siponimod 0.5 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>61</td>
<td>51</td>
<td>43</td>
</tr>
<tr>
<td>Monthly CULAS</td>
<td>1.39</td>
<td>0.78</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Statistical model

- Number of counts for patient $i = 1, \ldots, n_j$ receiving treatment $j = 1, 2$
  \[ Y_{ij} | \lambda_{ij} \sim \text{Pois}(t_{ij} \lambda_{ij}) \]

- Follow-up per patient: $t_{ij}$

- Gamma-mixture for the rates
  \[ \lambda_{ij} \sim \Gamma \left( \frac{1}{\phi}, \frac{1}{\phi \mu_j} \right) \]

- Marginal distribution of counts
  \[ Y_{ij} \sim \text{NB} \left( t_{ij} \mu_j, \phi \right) \]

- Expected value and variance
  \[ \mathbb{E} [Y_{ij}] = t_{ij} \mu_j \]
  \[ \text{Var} [Y_{ij}] = t_{ij} \mu_j (1 + \phi t_{ij} \mu_j) \]
Hypothesis testing I

- Statistical hypothesis

\[ H_0 : \frac{\mu_1}{\mu_2} \geq 1 \quad \text{vs.} \quad H_1 : \frac{\mu_1}{\mu_2} < 1. \]

- Hypothesis is tested using a Wald-type test of the maximum-likelihood estimators \( \hat{\beta}_j \) of the log-rates \( \beta_j = \log(\mu_j) \)

- Wald-type test statistic

\[
T = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\frac{1}{\hat{I}_{\beta_1}} + \frac{1}{\hat{I}_{\beta_2}}}} \quad H_0 \overset{\text{asymp.}}{\sim} \mathcal{N}(0, 1)
\]
Hypothesis testing II

- Fisher information of log-rates $\beta_j$

$$I_{\beta_j} = \sum_{i=1}^{n_j} \frac{t_{ij} \exp(\beta_j)}{1 + \phi t_{ij} \exp(\beta_j)} = \sum_{i=1}^{n_j} \frac{t_{ij} \mu_j}{1 + \phi t_{ij} \mu_j}.$$  

(Reminder: $I_{\beta_j}^{-1}$ is the asymptotic variance of the MLE $\hat{\beta}_j$.)

- Information level $I_{fix}$ describes "knowledge" about unknown treatment effect

$$I_{fix} = \frac{1}{I_{\beta_1}} + \frac{1}{I_{\beta_2}} = \frac{I_{\beta_1} \cdot I_{\beta_2}}{I_{\beta_1} + I_{\beta_2}}$$

- Sample size planning by solving equation

$$I_{fix} = \frac{(q_{1-\beta} - q_{1-\alpha})^2}{(\beta_1 - \beta_2)^2}$$
Group sequential designs: Overview

- Test the hypothesis $H_0$ at several interim analyses and stop the trial if $H_0$ can be rejected (stop for efficacy)

- The interim analyses are performed with the Wald-type test using all data available up to that point in time

- Counts of patient $i$ in treatment $j$ at analysis $k$: $Y_{ijk} \sim NB(t_{ijk}\mu_j, \phi)$

- $t_{ijk}$ is the follow-up time until analysis $k$

- The final analysis is performed when a prespecified information level $I_{max}$ is attained (maximum information trial)
Critical values of the individual tests $c_k$ must be chosen such that global type I error $\alpha$, i.e.

$$\alpha \leq \mathbb{P}_{H_0} \left( T_k < c_k \text{ for at least one } k = 1, \ldots, K \right).$$

Allocate global type I error $\alpha = \sum_{k=1}^K \pi_k$

Type I error rate $\pi_k$ for analysis $k$

$$\mathbb{P}_{H_0} \left( T_1 \geq c_1, \ldots, T_{k-1} \geq c_{k-1}, T_k < c_k \right) = \pi_k$$

Choose $\pi_k$ through error spending function $f : [0, \infty) \rightarrow [0, \alpha]$ with $f(0) = 0$ and $f(t) = \alpha$, $t \geq 1$:

$$\pi_1 = f \left( \mathcal{I}_1 / \mathcal{I}_{\text{max}} \right),$$

$$\pi_k = f \left( \mathcal{I}_k / \mathcal{I}_{\text{max}} \right) - f \left( \mathcal{I}_{k-1} / \mathcal{I}_{\text{max}} \right) \quad k = 2, 3, \ldots.$$
Critical values

- First critical value is the normal quantile \( c_1 = q_{\pi_1} \)

- Joint distribution \((T_1, \ldots, T_k)\) required to calculate critical value \( c_k \)

- Asymptotic normality of joint distribution has canonical form
  [Scharfstein et al., 1997]

\[
(T_1, \ldots, T_k)' \to \mathcal{N}(0, \Sigma_k)
\]

with

\[
(\Sigma_k)_{(k_1,k_2)} = (\Sigma_k)_{(k_2,k_1)} = \sqrt{\frac{I_{k_1}}{I_{k_2}}}, \quad 1 \leq k_1 \leq k_2 \leq k.
\]
Practical considerations

- Information level depends on rates $\mu_j$, shape parameter $\phi$, follow-up times $t_{ijk}$, and sample size $n_j$.

- At analysis $k$, $\mathcal{I}_k$ not known and is estimated by plugging in the rate and shape maximum-likelihood estimators.

- Critical value $c_k$ is not determined prior to the trial but at the time of analysis $k$.

- $\hat{\mathcal{I}}_k$ is the estimated information level of stage $k$ obtained with the data available at interim $k$. 
Practical considerations continued

- In practice the following estimators are considered

\[
\hat{\pi}_1 = f \left( \frac{\hat{I}_1}{\hat{I}_{\text{max}}} \right)
\]

\[
\hat{\pi}_k = f \left( \frac{\hat{I}_k}{\hat{I}_{\text{max}}} \right) - f \left( \frac{\hat{I}_{k-1}}{\hat{I}_{\text{max}}} \right) \quad k = 2, 3, \ldots
\]

\[
\left( \hat{\Sigma}_k \right)_{(k_1, k_2)} = \sqrt{\frac{\hat{I}_{k_1}}{\hat{I}_{k_2}}}
\]

- Estimated information might decrease if sample sizes or time between analyses is small, i.e. \( \hat{I}_k < \hat{I}_{k-1} \)
  - then analysis is skipped \( \leftrightarrow \) critical value \( c_k = \infty \)

- ”Locally” allocated type I error preserves the global type I error

\[
\sum_{i=1}^{K} \hat{\pi}_k = \alpha
\]
Planning of group sequential trials

- Power for given set of critical values \( c_1, \ldots, c_K \)

\[
\text{Power} = 1 - \Pr_{H_1} ( T_1 \geq c_1, \ldots, T_K \geq c_K )
\]

- For rate ratio \( \theta^\ast \) in alternative, joint distribution \( ( T_1, \ldots, T_K ) \)
  approximately normal with mean vector \( \log(\theta^\ast)(\sqrt{I_1}, \ldots, \sqrt{I_K})' \)

- For planning purposes, we write

\[
I_k = w_k I_{\max}, \quad k = 1, \ldots, K, \quad w_k \in (0, 1]
\]

- Calculate maximum information \( I_{\max} \) required to obtain power of \( 1 - \beta \) by solving

\[
1 - \Pr_{\theta^\ast} ( T_1 \geq c_1, \ldots, T_K \geq c_K ) = \beta
\]

- Sample size, study duration, etc must be selected such that the maximum information is obtained
Simulation study - preface

- In the simulation, interim analysis time points are determined by theoretical information levels $\mathcal{I}_k$. The actual estimated information levels $\hat{\mathcal{I}}_k$ differ from this.

- Use of spending functions which imitate critical values of Pocock’s test and O’Brien & Fleming’s test

- Recruitment times uniform in fixed accrual period
Simulation scenarios - type I error

- Simulation scenarios motivated by the number of hospitalizations from Example 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error rate $\alpha$</td>
<td>0.025</td>
</tr>
<tr>
<td>Annual rates $\mu_1 = \mu_2$</td>
<td>0.08, 0.1, 0.12</td>
</tr>
<tr>
<td>Shape parameter $\phi$</td>
<td>2, 3, 4, 5</td>
</tr>
<tr>
<td>Group sample size $n_1 = n_2$</td>
<td>600, 1000, 1400</td>
</tr>
<tr>
<td>Stages $K$</td>
<td>2, 5</td>
</tr>
<tr>
<td>Study duration</td>
<td>3.5 (years)</td>
</tr>
<tr>
<td>Recruitment period</td>
<td>1.25 (years)</td>
</tr>
</tbody>
</table>

- 25,000 Monte Carlo replications per scenario
Results - type I error

<table>
<thead>
<tr>
<th>Fix</th>
<th>O'Brien &amp; Fleming</th>
<th>Pocock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal group specific sample size</td>
<td>Type I error rate</td>
<td>Type I error rate</td>
</tr>
<tr>
<td>600</td>
<td>0.022</td>
<td>0.026</td>
</tr>
<tr>
<td>1000</td>
<td>0.023</td>
<td>0.025</td>
</tr>
<tr>
<td>1400</td>
<td>0.024</td>
<td>0.024</td>
</tr>
</tbody>
</table>

You can see the graph which shows the type I error rates for different sample sizes with Fix, O'Brien & Fleming, and Pocock methods.
Simulation scenarios - power

- Parameters for the Monte Carlo simulation study of the power

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error rate $\alpha$</td>
<td>0.025</td>
</tr>
<tr>
<td>Annual rate $\mu_1$</td>
<td>0.0875</td>
</tr>
<tr>
<td>Annual rate $\mu_2$</td>
<td>0.125</td>
</tr>
<tr>
<td>Rate ratio $\mu_1/\mu_2$</td>
<td>0.7</td>
</tr>
<tr>
<td>Group sample size $n_1 = n_2$</td>
<td>600, 650, ... , 1500</td>
</tr>
<tr>
<td>Shape parameter $\phi$</td>
<td>5</td>
</tr>
<tr>
<td>Stages $K$</td>
<td>2, 5</td>
</tr>
<tr>
<td>Study duration</td>
<td>3.5 (years)</td>
</tr>
<tr>
<td>Recruitment period</td>
<td>1.25 (years)</td>
</tr>
</tbody>
</table>

- 25 000 Monte Carlo replications per scenario
Results - power

- Maximum number of stages: 2
- Maximum number of stages: 5

Design
- Fix
- O'Brien & Fleming
- Pocock

Maximum group specific sample size $n_1$
Results - stopping times

- Rejections by stage (O’Brien-Fleming, total power of 80%)
Results - gains from stopping early

- Study times at which in theory 25%, 50%, 75%, and 100% of the maximum information level $I_{max}$ is attained
Simulation scenarios - type I error

- Simulation scenarios motivated by the CULAs from Example 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error rate $\alpha$</td>
<td>0.025</td>
</tr>
<tr>
<td>6-month rates $\mu_1 = \mu_2$</td>
<td>6, 8, 10</td>
</tr>
<tr>
<td>Shape parameter $\phi$</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>Group sample size $n_1 = n_2$</td>
<td>50, 70, \ldots, 150</td>
</tr>
<tr>
<td>Stages $K$</td>
<td>2, 3</td>
</tr>
<tr>
<td>Individual follow-up</td>
<td>0.5 (years)</td>
</tr>
<tr>
<td>Recruitment period</td>
<td>1.5 (years)</td>
</tr>
</tbody>
</table>

- 18 scenarios per group sample size for group sequential designs and 9 scenarios for the fixed design
- 25,000 Monte Carlo replications per scenario
Results - type I error

<table>
<thead>
<tr>
<th>Method</th>
<th>Type I Error Rate</th>
<th>Maximum Group Specific Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix</td>
<td></td>
<td>50, 75, 100, 125, 150</td>
</tr>
<tr>
<td>O'Brien &amp; Fleming</td>
<td></td>
<td>50, 75, 100, 125, 150</td>
</tr>
<tr>
<td>Pocock</td>
<td></td>
<td>50, 75, 100, 125, 150</td>
</tr>
</tbody>
</table>

The diagram illustrates the type I error rate for different maximum group specific sample sizes for the Fix, O'Brien & Fleming, and Pocock methods.
Simulation scenarios - power

- Parameters for the Monte Carlo simulation study of the power

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error rate $\alpha$</td>
<td>0.025</td>
</tr>
<tr>
<td>6-month rate $\mu_1$</td>
<td>4.2</td>
</tr>
<tr>
<td>6-month rate $\mu_2$</td>
<td>8.4</td>
</tr>
<tr>
<td>Group sample size $n_1 = n_2$</td>
<td>70, 75, ..., 140</td>
</tr>
<tr>
<td>Shape parameter $\phi$</td>
<td>3</td>
</tr>
<tr>
<td>Stages $K$</td>
<td>2</td>
</tr>
<tr>
<td>Individual follow-up</td>
<td>0.5 (years)</td>
</tr>
<tr>
<td>Recruitment period</td>
<td>1.5 (years)</td>
</tr>
</tbody>
</table>

- 25,000 Monte Carlo replications per scenario
Results - power

Maximum number of stages: 2

Maximum number of stages: 3

Design
- Fix
- O'Brien & Fleming
- Pocock
Results - analysis specific rejection rate

- Rate of stopping at a specific analysis at a power of 80%

![Graph showing rejection rates for different maximum stages](image-url)
Results - gains from stopping early

- Study times at which in theory 25%, 50%, 75%, and 100% of the maximum information level $I_{max}$ is attained
Discussion and outlook

- Maximum-likelihood theory for negative binomial data results asymptotically in canonical form of joint distribution of test statistic

- Information level depends on rates, shape parameter, follow-up times, and sample size

- Future research on group sequential with negative binomial endpoints
  - Blinded information monitoring
  - Adaptive group sequential designs
  - Optimal designs

- Extend approach to quasi-Poisson models in the future
Bibliography

Semiparametric efficiency and its implication on the design and analysis of group-sequential studies.

S. Yusuf et al. (2003)
Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial.

Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved.

Siponimod for patients with relapsing-remitting multiple sclerosis (BOLD): an adaptive, dose-ranging, randomised, phase 2 study.