Inference for Multi-Arm Multi-Stage Adaptive Clinical Trials

Cyrus Mehta, Ph.D.
Cytel Inc.
Acknowledgements

• This is joint work with
  • Ping Gao, The Medicines Company
  • Lingyun Liu, Cytel Inc
  • Pranab Ghosh, Cytel Inc
1. Quick introduction to Adaptive Group Sequential Designs

2. Preservation of $\alpha$ and parameter estimation in 2-arm classical GSD

3. Preservation of $\alpha$ and parameter estimation in 2-arm adaptive GSD

4. Extension to multi-arm adaptive GSD

5. Future directions
Adaptive Group Sequential Designs

• An independent data monitoring committee monitors the accruing data:
  • Monitoring is unblinded
  • Monitoring times are pre-specified but flexible
• Possible decisions at each monitoring time
  • Stop early due to overwhelming efficacy
  • Stop early for futility
  • Make changes to future course of trial: (increase sample size, change number of looks, drop arms)
Classical and Adaptive GSDs

Design (1): Non-Adaptive

Design (2): Adaptive

Sample size, number and spacing of future looks were adapted at look 2 of GSD
MDCO-216 is a macromolecular complex of a recombinant protein and phospholipid.

MCDO-216 is delivered by IV infusion.

It reduces plaque burden (as measured by ultrasound).

This clinical trial will compare four doses (4mg/kg, 8 mg/kg, 16 mg/kg, 32 mg/kg) of MDCO-216 to a standard of care control arm.

Three analyses with possible dose selection and sample size re-estimation.
Current State of Research

• Methods for $\alpha$ error control in adaptive trials are well developed

• Methods for p-values, confidence intervals and parameter estimates not well developed

• Available Methods for two-arm multi stage:
  • Extension of RCI (Mehta et al., 2008): do not exhaust the $\alpha$, hence conservative
  • Extension of SWCI (Brannath et al. 2009, Gao et al. 2013): exhaust the $\alpha$, hence provide exact coverage

• Nothing available so far for multi-arm multi-stage adaptive group sequential designs
For a K-look group sequential design define:

\[
\begin{align*}
\delta & = \mu_1 - \mu_0 \text{ is the treatment effect} \\
\hat{\delta}_j & = \text{the mle of } \delta \text{ at look } j, \text{ for } j = 1, 2, \ldots, K \\
I_j & = \text{the Fisher information for } \delta \text{ at look } j \\
W_j & = \hat{\delta}_j \sqrt{I_j} \text{ is the score statistic at look } j \\
W_j & \sim N(\hat{\delta}I_j, I_j) \text{ with independent increments}
\end{align*}
\]

- Independent increments means that \(W_j\) and \((W_k - W_j)\) are independent for any \(k > j\)
- This property enables us to create the design efficiently for any value of \(K\)
Three-look Group Sequential Design

Boundaries \((u_1, l_1), (u_2, l_2), (u_3, l_3)\) and \(I_{\text{max}}\) obtained from:

\[
P_0(W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq u_3) = \alpha
\]

\[
P_\delta \sqrt{I_{\text{max}}}(W_1 \leq l_1 \text{ or } W_2 \leq l_2 \text{ or } W_3 \leq u_3) = \beta
\]
Look 1 of Classical Three-Look Design

Stopping Boundaries for Integrated Design

Sample Size

Stopping Boundary: Z Scale

Test Statistic
Look 2 of Classical Three-Look Design

Stopping Boundaries for Integrated Design

Sample Size

Stopping Boundary: Z Scale

Test Statistic
Null hypothesis that $\delta=0$ is rejected. But what about P-value and CI?
• Stage wise ordering of sample space
  • Stopping at the same look with larger value of test statistic is more extreme
  • Stopping at an earlier look is more extreme than stopping at a later look
• Compute the probability of the more-extreme region based on this ordering

Classical SWCI Method: Tsiatis, Rosner, Mehta (Biometrics, 1984)
Classical SWCI: P-Value Computation

\[ P_{\text{value}} = P_0 \left( W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq w_3 \right) \]
Find $\delta_{\text{low}}$ such that: $P_{\delta_{\text{low}}} (W_{1} \geq u_{1} \text{ or } W_{2} \geq u_{2} \text{ or } W_{3} \geq w_{3}) = 0.975$

Find $\delta_{\text{up}}$ such that: $P_{\delta_{\text{up}}} (W_{1} \geq u_{1} \text{ or } W_{2} \geq u_{2} \text{ or } W_{3} \geq w_{3}) = 0.025$
What if we adapt at look 2?
Created two additional looks, changed the spending function and increase sample size
How Type-1 Error is Preserved

Design (1): Non-Adaptive

\[ P_0(W_3^{(1)} \geq u_3^{(1)} | w_2^{(1)}) = \varepsilon_0^{(1)} \]

Design (2): Adaptive

\[ P_0 \cup (W_3^{(2)} \geq u_3^{(2)}, W_4^{(2)} \geq u_4^{(2)}, W_5^{(2)} \geq u_5^{(2)} | w_2^{(1)}) = \varepsilon_0^{(2)} \]

Identical Conditional Errors: \( \varepsilon_0^{(1)} = \varepsilon_0^{(2)} \)
Monitor the Adaptive Trial: Look 3

Stopping Boundaries for Integrated Design

Test Statistic

Sample Size

IISA, Poona. 21 Dec 2015
Null hypothesis that $\delta=0$ is rejected
What about P-value and confidence interval?
1. Find the backward image, in the classical sample space, of the statistic $w_4^{(2)}$ obtained in the adaptive sample space
2. Use the classical SWCI method to compute P-value and CI for the backward image
Backward Image, \( w^*(0) \), for Obtaining the P-Value

- Find \( w^*(0) \) such that:
  
  \[
  P_0 (W_3^{(1)} \geq w^*(0) \mid w_2^{(1)}) = P_0 (W_3^{(2)} \geq u_3^{(2)} \text{ or } W_4^{(2)} \geq w_4^{(2)} \mid w_2^{(1)})
  \]

- Then find the p-value corresponding to \( w^*(0) \) by classical SWCI method
1. \( P_\delta (W_3^{(1)} \geq w^*(\delta) | w_2^{(1)}) = P_\delta (W_3^{(2)} \geq u_3^{(2)} , W_4^{(3)} \geq w_4^{(3)} | w_2^{(1)}) \)

2. \( P_{\delta_{low}} (W_1^{(1)} \geq u_1^{(1)} \text{ or } W_2^{(1)} \geq u_2^{(1)} \text{ or } W_3^{(1)} \geq w^*(\delta_{low})) = 0.975 \)
Simulation of a 3-look LD(OF) GSD with adaption at look 2 to a 3-look LD(PK) GSD (100,000 simulations)

<table>
<thead>
<tr>
<th>True Value of $\delta$</th>
<th>Median of 100,000 estimates of $\delta$</th>
<th>Prop. of 90% CI’s containing $\delta$</th>
<th>Proportion of 90% CI’s excluding $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>From below From above</td>
</tr>
<tr>
<td>-0.15</td>
<td>-0.14972</td>
<td>0.90007</td>
<td>0.05022 0.04971</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00027</td>
<td>0.90073</td>
<td>0.04920 0.05007</td>
</tr>
<tr>
<td>0.15</td>
<td>0.14986</td>
<td>0.89866</td>
<td>0.04955 0.05179</td>
</tr>
<tr>
<td>0.30</td>
<td>0.29999</td>
<td>0.90087</td>
<td>0.04940 0.04973</td>
</tr>
<tr>
<td>0.45</td>
<td>0.44963</td>
<td>0.89929</td>
<td>0.05083 0.04988</td>
</tr>
</tbody>
</table>

Results for Backward Image (BWCI) Method

IISA, Poona. 21 Dec 2015
Simulation of a 3-look LD(OF) GSD with adaption at look 1 to a 2-look LD(OF) GSD (100,000 simulations)

Comparison with other adaptive methods

<table>
<thead>
<tr>
<th>True value of $\delta$</th>
<th>Probability of lower 95% CL $&gt; \delta$</th>
<th>Probability of upper 95% CL $&lt; \delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BWCI</td>
<td>SWCI</td>
</tr>
<tr>
<td>-0.15</td>
<td>0.02505</td>
<td>0.0256</td>
</tr>
<tr>
<td>0.0</td>
<td>0.02462</td>
<td>0.0251</td>
</tr>
<tr>
<td>0.15</td>
<td>0.02473</td>
<td>0.0256</td>
</tr>
<tr>
<td>0.3</td>
<td>0.02411</td>
<td>0.0253</td>
</tr>
<tr>
<td>0.45</td>
<td>0.02470</td>
<td>0.0259</td>
</tr>
</tbody>
</table>

• SWCI Method: Brannnath, Mehta, Posch (*Biometrics*, 2009)
The Two-Arm Problem was Published

Next we consider extending the approach to Multi-Arm Multi-Stage (MAMS) Designs
Generalization to Multi-arm trials

- Suppose there are D dose comparisons with respect to a common control arm
- Select the best dose at end of look 1
- Future looks are then based on a two-arm trial

\[ \bar{\delta} = (\delta_1, \delta_2, \ldots, \delta_D) \] where effect of i-th dose is
\[ \delta_i = \mu_1 - \mu_0 \] for \( i = 1, 2, \ldots, D \)
\[ \bar{W}_1 = (\hat{\delta}_{11}\sqrt{I_{11}}, \hat{\delta}_{21}\sqrt{I_{21}}, \ldots, \hat{\delta}_{D1}\sqrt{I_{D1}}) \]

Score statistics for the D dose groups at look 1
\[ W_{\text{max}, 1} = \max(\hat{\delta}_{11}\sqrt{I_{11}}, \hat{\delta}_{21}\sqrt{I_{21}}, \ldots, \hat{\delta}_{D1}\sqrt{I_{D1}}) \]

Boundaries \((u_1, l_1), (u_2, l_2), (u_3, l_3)\) and \(I_{\text{max}}\) obtained from:
\[ P_0(W_{\text{max}, 1} \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq u_3) = \alpha \]
\[ P_{\delta\sqrt{I_{\text{max}}}}(W_{\text{max}, 1} \leq l_1 \text{ or } W_2 \leq l_2 \text{ or } W_3 \leq u_3) = \beta \]
Boundary Comparisons

Two arms, 3 looks, $\gamma(-5)$ error spending functions, $\alpha=0.025$

<table>
<thead>
<tr>
<th>Cum. $\alpha$ Spent</th>
<th>Efficacy Boundary</th>
<th>Cum. $\beta$ Spent</th>
<th>Futility Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0007</td>
<td>3.18319</td>
<td>0.0028</td>
<td>-0.86843</td>
</tr>
<tr>
<td>0.0046</td>
<td>2.64034</td>
<td>0.0177</td>
<td>0.55167</td>
</tr>
<tr>
<td>0.0250</td>
<td>1.98335</td>
<td>0.0967</td>
<td>1.98335</td>
</tr>
</tbody>
</table>

Five arms, 3 looks, $\gamma(-5)$ error spending functions, $\alpha=0.025$

<table>
<thead>
<tr>
<th>Cum. $\alpha$ Spent</th>
<th>Efficacy Boundary</th>
<th>Cum. $\beta$ Spent</th>
<th>Futility Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0007</td>
<td>3.55470</td>
<td>0.0037</td>
<td>-0.05339</td>
</tr>
<tr>
<td>0.0046</td>
<td>3.05130</td>
<td>0.0233</td>
<td>1.17190</td>
</tr>
<tr>
<td>0.0250</td>
<td>2.46592</td>
<td>0.1270</td>
<td>2.46592</td>
</tr>
</tbody>
</table>
Superimpose the two boundaries

![Graph showing stopping boundaries vs information fraction for Two Arm and Four Arm trials.](image)
Three-look design; two active doses vs control; select best dose at end of look 1; increase sample size at end of look 2 if conditional power is between 30% and 90%

<table>
<thead>
<tr>
<th>True Parameter $(\delta_1, \delta_2)$</th>
<th>97.5% lower CI for $\delta_s$</th>
<th>Probability that $(50% \text{ CI is } \leq \delta_s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.5, 0)</td>
<td>0.97581</td>
<td>0.50043</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>0.97535</td>
<td>0.50092</td>
</tr>
<tr>
<td>(2.5, 0)</td>
<td>0.97535</td>
<td>0.50092</td>
</tr>
<tr>
<td>(1.5, 2.5)</td>
<td>0.97540</td>
<td>0.49750</td>
</tr>
<tr>
<td>(2, 2.5)</td>
<td>0.97538</td>
<td>0.49765</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>0.97520</td>
<td>0.49878</td>
</tr>
</tbody>
</table>

These and other results have been submitted to a technical journal
Summary of Results

• For Two Arm Trials
  • Backward image method is the only published method with exact two-sided coverage
  • Applicable to SSR, alteration of number and spacing of interim looks and spending function
  • SWCI method (Brannath et. al. 2009) provides comparable results but only for one-sided case

• For Multi-Arm Trials
  • Special Case (select best dose at look 1): results shown here have been submitted for publication
  • Results hold conservatively if best dose not selected
  • General Case (select any number of doses and drop losers along the way): still being investigated. Hypothesis testing problem published (Gao et al. 2015) and available in East.
Stage wise adjusted confidence intervals for group sequential designs provide exact coverage and median-unbiased point estimates.

These methods extend in a natural way to adaptive group sequential design.

Also extend to MAMS trials in which ONE dose is selected for further testing versus control.

Fully general case in which any number of doses can be selected at each stage, still under investigation.