Sequential Monitoring of Multiple Endpoints in Clinical Trials

Ronald Gangnon, Ellen Roecker and Thomas Cook

Statistical Data Analysis Center
Department of Biostatistics and Medical Informatics
University of Wisconsin
Madison, Wisconsin

May 3, 1999

e-mail: ronald@biostat.wisc.edu
Internet: http://www.biostat.wisc.edu/~ronald
Outline

1. Example
2. Proposed Methodology
3. (Partial) Analysis of Example
4. Conclusions
Example

• Randomized, multicenter clinical trial

• Three treatment arms: Placebo, Low dose, High dose
  – Comparisons of interest:
    Placebo v. Low dose and Placebo v. High dose

• Overall $\alpha=0.05$ (0.025 for each comparison)

• Primary endpoint: Time to a Composite Event
  – Death; MI; Urgent Revascularization

• Test statistic: (Weighted) log-rank test
Original Interim Monitoring Plan

- Independent Data Safety Monitoring Board (DSMB)

- Lan-DeMets $\alpha$-spending function implementation of O’Brien-Fleming boundary

- Information fraction = number of events (in all 3 arms)/total expected (for $\alpha$-spending)

- Two planned interim analyses at approximately 70% and 90% information fractions
FDA Comment

• “Trial should not be stopped early unless there is also a benefit in mortality”

• Eventual response: Introduce 2nd primary endpoint:
  – Time to Death or MI
  – Included MI due to low death rate

• Questions:
  – How to divide overall $\alpha$ between endpoints?
  – How to divide $\alpha$ at interim analyses?
  – How to account for correlation between endpoints?
Working \( \alpha \)-Spending Functions

- Define working \( \alpha \)-spending function for each endpoint

- For 1\(^{st}\) primary endpoint, allocate overall \( \alpha \) of 0.02
  - Use an extremely conservative O’Brien-Fleming boundary until final analysis
  - Compute early stopping boundary as if overall \( \alpha \) is 0.002

- For 2\(^{nd}\) primary endpoint, allocate overall \( \alpha \) of 0.005
  - O’Brien-Fleming boundary truncated at 3.50

- Information fraction based on 1\(^{st}\) primary endpoint
Proposed $\alpha$-Spending Functions

Original Alpha-Spending Function

Working Alpha-Spending Function for 1st Primary Endpoint

Working Alpha-Spending Function for 2nd Primary Endpoint

Overall Alpha-Spending Function (Both Endpoints)
Calculating Appropriate Boundaries

- Simple approach: Independent calculation of boundaries
  - Equivalent to Bonferroni-type correction
  - Conservative approach with correlated outcomes

- Alternative approach: Account for correlation in endpoints through randomization test

- 1\textsuperscript{st} primary endpoint:
  - Calculate boundary using standard methods

- 2\textsuperscript{nd} primary endpoint
  - Calculate boundary via simulation to ensure desired cumulative alpha is spent
Simulation Technique

1. Calculate boundary for 1st primary endpoint ignoring 2nd primary endpoint
2. Reassign treatment groups according to randomization scheme
3. For each endpoint at each interim, calculate the test statistics based on the new randomization
4. For a range of possible boundary values for 2nd primary endpoint, count the number of boundary crossings (on either test at any interim)
5. Repeat Steps 2-4 as needed
6. Select boundary value for 2nd primary endpoint that achieves the desired cumulative alpha at current interim
Analysis of Example

• Only consider Placebo v. Low dose comparison

• Total of 3 analyses (2 interim analyses)

• First interim analysis at 52% information
  – Boundary for 1\textsuperscript{st} primary: $Z_c = \pm 4.28$
  – Boundary for 2\textsuperscript{nd} primary: $Z_c = \pm 3.50$ (truncated)

• Second interim analysis at 80% information
  – Boundary for 1\textsuperscript{st} primary: $Z_c = \pm 3.46$
  – Nominal boundary for 2\textsuperscript{nd} primary: $Z_c = \pm 3.19$
  – 500,000 simulations used to estimate actual boundary
Simulation Results for the 2\textsuperscript{nd} Interim Analysis

![Graph showing simulation results for the 2\textsuperscript{nd} Interim Analysis. The graph plots tail probability (2-sided) against test statistic. The x-axis represents test statistics ranging from 3.1181 to 3.1754, and the y-axis represents tail probabilities from 0.0030 to 0.0015. The graph includes lines indicating different tail probabilities.]
Importance Sampling

- Large number of simulations required to estimate boundary
- Sampling method is inefficient
- Under null hypothesis, observations near boundary are very rare

- Alternative approach: Importance sampling
  - Sample values near boundary disproportionately often (don’t sample under null hypothesis)
  - Reweight this sample to reflect desired distribution
  - Weight = ratio of probability sample selected under null to probability of sample under importance sampling
Importance Sampling

• How to produce samples with large value of test statistic?
• Sample more events in group A

• Randomization distribution of the # of events in group A is the hypergeometric distribution
  – Equivalent to drawing balls from an urn with $n_A$ black balls and $n_B$ white balls
• For importance samples, increase the # of black balls in urn
  – Draw balls from an urn with $K \cdot n_A$ black balls and $n_B$ white balls

• 100,000 simulations with $K=1.25$
Importance Sampling Results for the 2\textsuperscript{nd} Interim Analysis

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Tail Probability (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0749</td>
<td>0.0015</td>
</tr>
<tr>
<td>3.1319</td>
<td>0.0020</td>
</tr>
<tr>
<td>3.1958</td>
<td>0.0022</td>
</tr>
</tbody>
</table>
Final Analysis

• Final boundary for 1\textsuperscript{st} primary endpoint: $Z_c = \pm 2.33$
  – Equivalent to final boundary under original scheme

• Boundary for 2\textsuperscript{nd} primary endpoint
  – Nominal boundary: $Z_c = \pm 2.85$
  – 200,000 simulations under null hypothesis
  – Estimated boundary (95% CI): $Z_c = \pm 2.58 \ (2.53, 2.65)$

  – Marginal boundary crossing probability: 0.010 (v. 0.005)
  – All benefit occurred at final analysis
Interim Monitoring Boundary for 1st Primary Endpoint

- Z-score:
  - 0.0 0.52 0.80 1.00
  - -4 -2 0 2 4
  - 1.25

- Graph:
  - Z-score vs. Information Fraction
  - Continue Trial
  - Reject Null Hypothesis
  - Crosses at -0.17 and -0.26
Interim Monitoring Boundary for 2\textsuperscript{nd} Primary Endpoint

![Graph showing the interim monitoring boundary for the 2\textsuperscript{nd} primary endpoint. The graph includes a line indicating the decision boundaries at Z-scores of -0.2 and -0.21, with a dotted line at Z-score of 1.66. The graph also marks points at Z-scores of 0.0, 0.52, 0.80, and 1.00 that correspond to different information fractions. The boundary lines and points are labeled as 'Reject Null Hypothesis' or 'Continue Trial.'
Extensions to Three (or More) Endpoints

• Simple adjustment
  – Adjust only one endpoint (e.g., mortality)
  – Use ordinary (nominal) boundary for other endpoints
  – Simulate boundary for selected endpoint

• Hierarchical adjustment
  – Select an ordering of the endpoints (1\textsuperscript{st}, 2\textsuperscript{nd}, …)
  – Use ordinary (nominal) boundary for 1\textsuperscript{st} endpoint
  – Simulate boundary for 2\textsuperscript{nd} endpoint accounting for correlation with the 1\textsuperscript{st} endpoint
  – Simulate boundary for 3\textsuperscript{rd} endpoint accounting for correlation with the 1\textsuperscript{st} and 2\textsuperscript{nd} endpoints, etc.
Conclusions

- Simulation-based method for monitoring correlated endpoints
- Need to select an endpoint for which boundary is adjusted
- Will account for other inaccuracies in monitoring boundary (to ensure all alpha is spent)
- Easily adapted to other situations (non-survival outcomes)
- Natural extensions to three or more endpoints