Adaptive designs for time-to-event trials

Dominic Magirr$^1$, Thomas Jaki$^1$, Franz König$^2$ and Martin Posch$^2$

$^1$Department of Mathematics and Statistics, Lancaster University
$^2$Institut für Medizinische Statistik, Medical University of Vienna

January 2014
Outline

Motivating example

Adaptive methods

Ignored data
Lung cancer trial (Schäfer and Müller, 2001)

- Patients randomized to “Radiotherapy + Chemotherapy” (E) or “Chemotherapy” (C)
- Median survival on C ≈ 14 months
- Anticipated survival on E ≈ 20 months
- Sample size: 255 deaths ($\alpha = 0.025$, $\beta = 0.2$)
- Exponential model ... this could be achieved with 40 months recruitment and 20 months min follow-up.
40 months into the trial...

(a) patient recruitment was much slower than expected
   – only 136 patients had been randomized
(b) the hazard rate had been over-estimated in the planning
   – only 56 deaths had been observed

Recommendation of Schäfer and Müller:
“abandon the trial because there [is] no chance of achieving the planned sample size within a reasonable time”
Counterproposal of study group

- Look at the data to see if there is a larger treatment effect than originally anticipate.
- If so, reduce the initially planned sample size (required number of events).
- Larger the observed treatment effect → earlier the study ends.
Reverse scenario (Irle & Schäfer, 2012)

- Look at data to see if there is a smaller than anticipated treatment effect.
- If so, increase the sample size (required number of events) to give a better chance of achieving a statistically significant result.

Standard analysis will not control the type I error rate...
Adaptive design with immediate responses

E.g., under $H_0$,

$$\frac{1}{\sqrt{2}} \Phi^{-1} \left\{ 1 - p_1(X_1^{\text{int}}) \right\} + \frac{1}{\sqrt{2}} \Phi^{-1} \left\{ 1 - p_2(Y) \right\} \sim \mathcal{N}(0, 1)$$
Adaptive design with delayed responses

\[
\frac{1}{\sqrt{2}} \Phi^{-1} \left\{ 1 - p_1(X_1^{\text{int}}) \right\} + \frac{1}{\sqrt{2}} \Phi^{-1} \left\{ 1 - p_2(Y) \right\} \sim \mathcal{N}(0, 1)
\]
When is this valid?

✓ Interim decision strategy based solely on (primary endpoint) treatment effect estimate.

✗ Interim decisions are based on partial information from patients who are yet to provide full primary endpoint response e.g. second-stage sample size is chosen on basis of progression-free survival when primary endpoint is overall survival.
Potential solution

\[
\frac{1}{\sqrt{2}} \Phi^{-1} \{1 - p_1(X_1)\} + \frac{1}{\sqrt{2}} \Phi^{-1} \{1 - p_2(Y)\} \sim \mathcal{N}(0, 1)
\]

E.g., Liu & Pledger (2005) – Gaussian responses
Schmidli, Bretz & Racine-Poon (2007) – Binary responses
Extra problem with time-to-event endpoint?
Jenkins, Stone & Jennison (2011); Irle & Schäfer (2012)

• Must pre-specify end of follow-up of first-stage patients, $T_1$, in definition of $p_1$.

• Otherwise, $p_1(X_1) \sim U[0, 1]$ under $H_0$, and type I error may be inflated.
Data is “thrown away”

• Final test decision only depends on a subset of the recorded survival times; part of the observed data is ignored.
• Particularly damaging if long-term survival is of most concern (it is the survival times of earliest recruited patients that is ignored).
• Therefore, we investigated the effect of naïvely incorporating this illegitimate data into the final test statistic...
Adaptive log-rank test

“Correct” adaptive test statistic

\[ Z^{\text{CORRECT}} = w_1 L_1(T_1) + w_2 \Phi^{-1}(1 - p_2) \]

“Naïve” adaptive test statistic

\[ Z^{\text{NAIVE}} = w_1 L_1(T^*) + w_2 \Phi^{-1}(1 - p_2) \]

- \( L_1(t) \) is the log-rank statistic based on Stage 1 patients, followed up until calendar time \( t \).
- \( w_i \) are explicitly (Jenkins et al.) or implicitly (Irle & Schäfer) fixed weights with \( w_1^2 + w_2^2 = 1 \).
- \( T_1 \) is the (implicitly) fixed end of first-stage follow up.
- \( T^* \) is the time of final analysis (dependent on interim decisions).
Worst-case assumption

- The null distribution of $Z^{\text{CORRECT}}$ is $\mathcal{N}(0, 1)$.
- The null distribution of $Z^{\text{NAIVE}}$ is completely unknown.
- However, we can look at the stochastic process

$$Z(t) = w_1 L_1(t) + w_2 \Phi^{-1}(1 - p_2), \quad t \in [T_1, T^{\text{max}}].$$

**Worst-case:** the interim data (PFS, early endpoints, etc) can be used to predict exactly when $L_1(t)$ reaches its maximum.
Upper bound on type I error

An upper bound can be found assuming second-stage design is engineered such that $T^*$ coincides with $\arg \max L_1(t)$:

$$\max \alpha = P_{H_0} \left\{ \max_{t \geq T_1} w_1 L_1(t) + w_2 \Phi^{-1}(1 - p_2) > 1.96 \right\}$$

$$= \ldots$$

$$\approx \int_0^1 P_{H_0} \left[ \max_{u=u_1} B(u) > \sqrt{u} \frac{1.96 + w_2 \Phi^{-1}(x)}{w_1} \right] \, dx,$$

with $u_1 = \{\# \text{ stage 1 deaths at } T_1\} / \{\# \text{ stage 1 deaths at } T^{\max}\}$
Figure: Worst case type I error for various choices of weights and information fractions.
Irle & Schäfer example revisited

- Original trial design:
  - 248 deaths; 40 months recruitment; 20 months min follow-up.

- Interim analysis after 23 months:
  - 190 patients recruited; 60 deaths.
  - Treatment effect in terms of PFS less impressive than expected.
  - Decision made to increase required number of deaths from 248 to 400.

- At calendar time $T_1 \approx 60$ months:
  - 170 first-stage patient deaths; 78 second-stage patient deaths.
  - $u_1 = 170/190; w_1 = \sqrt{170/248}; \text{max } \alpha = 0.040.$
Guaranteed level-$\alpha$ test

Simply increase cut-off such that $P_{H_0} \{ \max_{t \geq T_1} Z(t) \geq k^* \} = \alpha$.

In assessing the effect on power, four relevant probabilities are:

A. $P_{\theta = \theta_R} \left( Z^{\text{CORRECT}} > 1.96 \right)$
B. $P_{\theta = \theta_R} \left( Z^{\text{CORRECT}} > k^* \right)$
C. $P_{\theta = \theta_R} \left\{ Z(T^{\text{max}}) > k^* \right\}$
D. $P_{\theta = \theta_R} \left\{ \max_{t \geq T_1} Z(t) > k^* \right\}$
Figure: Conditional power as defined by A (thin line), B (medium line), C (thick line) and D (dashed line) given $p_2$, under two scenarios.

(a): $u_1 = 170/190$, $w_1 = (170/248)^{1/2}$ and $\theta_R = 0.36$.
(b): $u_1 = 147/288$, $w_1 = (147/248)^{1/2}$ and $\theta_R = 0.36$. 
**Conclusion: methods trade-off**

<table>
<thead>
<tr>
<th></th>
<th>Type I control</th>
<th>Informed interim decisions</th>
<th>All survival times in test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Independent increments&quot;</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>&quot;Correct&quot; adaptive</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>&quot;Naïve&quot; adaptive</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>&quot;Naïve&quot; + k*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>

Other research avenue: use joint distribution of, e.g., PFS and overall survival.
References


