Sample size considerations for clinical trials with two primary time-to-event outcomes

Toshimitsu Hamasaki, PhD, pstat®
National Cerebral and Cardiovascular Center

Tomoyuki Sugimoto, PhD
Hirosaki University School of Sciences and Technology

Koko Asakura, PhD
National Cerebral and Cardiovascular Center

Scott R. Evans, PhD
Harvard School of Public Health

Mayumi Fukuda, MD
National Cerebral and Cardiovascular Center

Haruko Yamamoto, MD, PhD
National Cerebral and Cardiovascular Center
To discuss sample size determination when using two time-to-event outcomes for comparing two interventions in superiority clinical trials as co-primary or primary contrasts

- A use of two primary or co-primary time-to-event endpoints has become common in clinical trials evaluating interventions in many disease areas such as infectious disease, oncology, or cardiovascular disease.

Structure of presentation
1. Design issues in clinical trials with two time-to-event outcomes
2. Study designs, hypothesis testing and powers
3. Time dependency association and censoring scheme
4. Behavior of sample size
5. Summary
Examples: clinical trials with two time-to-event outcomes

<table>
<thead>
<tr>
<th>Disease area</th>
<th>Outcomes</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV (Kaposi’s sarcoma in HIV-infected)</td>
<td>◆ Kaposi’s sarcoma (KS) progression ◆ HIV virologic failure</td>
<td>◆ Both events are not fatal and each event is not censored by other event. ◆ Subjects who do not experience both events yet are censored at the same time (e.g., by the end of the study or patient drop-out) in the end of follow-up period.</td>
</tr>
<tr>
<td>Oncology</td>
<td>◆ Overall survival (OS) ◆ Time to progression (TTP) or Progression-free survival (PFS)</td>
<td>◆ OS requires long follow-up periods after disease progressio, which leads to quite long and also expensive studies. ◆ PFS is often included as a short-term primary endpoint, defined as the time from randomization until tumor progression or death from any cause, whichever may occurs earlier than OS.</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>◆ Major cardiac adverse event (MACE) ◆ Death</td>
<td>◆ MACE is a composite endpoints, including multiple types of clinical events of varying degrees of relatedness. ◆ Death is included as a component of MACE and it is the most important event.</td>
</tr>
</tbody>
</table>
Design Issues: sample size

(i) Inferential goal for multiple outcomes
◆ To evaluate a joint statistical significance on BOTH outcomes
  “Multiple Co-Primary Endpoints”
◆ To evaluate a statistical significance on AT LEAST ONE outcome
  “Multiple Primary Endpoints”

Sample sizing for two time-to-event outcomes could be more complex compared with other scale outcomes such as continuous or binary outcomes- the following aspects should carefully be considered in sample size determination in clinical trials with two time-to-event outcomes

(ii) Censoring scheme between two outcomes
◆ Whether an event of interest is FATAL or NON-FATAL
◆ A non-fatal event could be censored by the fatal event (DEPENDENT censoring)

(iii) Time dependent association between two outcomes
◆ Whether the association between the two time-to-event outcomes could be changed with the time
Superiority clinical trials with two time-to-event outcomes

Endpoint (EP1): \((T_{i1}^*, C_{i1})\)  
“Positively” correlated  
Endpoint (EP2): \((T_{i2}^*, C_{i2})\)

Underlying continuous survival time and potential censoring time for the \(k\)th outcome for the \(i\)th subject

\[ (T_{ik}^*, C_{ik}) (k = 1, 2; i = 1, \ldots, n_j) \]

Total sample size

- Test: \(n_1 = rn\)
- Control: \(n_2 = (1 - r)n\)

Allocation ratio: \(0 < r < 1\)

\[ \psi_k(t) = \frac{\lambda_1^{(k)}}{\lambda_2^{(k)}} \]

Hazard ratio

\[ \varphi(j)(t) = \frac{\lambda_1^{(j)}}{\lambda_2^{(j)}} \]

\[ \lambda_k^{(j)}(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T_{ij}^* < t + dt | t \leq T_{ij}^*, g_i = j)}{dt} \]

\(j = 1\) (test); \(2\) (control)

Study end

Censored

Censored

Test

Control

\(N\)
Logrank test statistics for each endpoint

### Hypothesis for each endpoint

\[
\begin{align*}
H_{0k}: & \psi_k(t) \geq 1, \quad \text{for all } t \\
H_{1k}: & \psi_k(t) < 1, \quad \text{at some } t
\end{align*}
\]

### Logrank test statistics

\[
Z_k = -\hat{U}_k(\tau) / \sqrt{\hat{V}_{kk}(\tau)}
\]

For large sample, each \(Z_k\) is approximately normally distributed
\(Z_k \sim N(0,1)\) under \(H_{0k}\)

\[
\hat{U}_k(t) = \sqrt{n} \int_0^t \hat{H}_k(s) \{d\hat{\Lambda}_k^{(2)}(s) - d\hat{\Lambda}_k^{(1)}(s)\} \quad \text{…Bivariate (weighted) logrank statistics process}
\]

\[
\Lambda_k^{(j)}: \text{Cumulative hazard function } \Lambda_k^{(j)} = \int_0^t \lambda_k^{(j)}(s) ds
\]

\[
\hat{\Lambda}_k^{(j)}(t): \text{Nelson-Aalen estimator of } \Lambda_k^{(j)}
\]

\[
y_k^{(j)}: \text{“At risk” process in group } j \quad y_k^{(j)} = \sum_{i=1}^N 1(g_i = j, T_{ik} \geq t)
\]

\[
\hat{H}_k(s) = n^{-1} \bar{W}_k(t) y_k^{(1)} y_k^{(2)} / \{y_k^{(1)} + y_k^{(2)}\}
\]

\[
\hat{V}_{kk}(\tau) \quad \text{…Well-known conditional variance of } U_k(\tau)
\]
Superiority hypothesis testing for both or at least one endpoint

Co-Primary

\[
\begin{cases}
H_0: \psi_1(t) \geq 1 \text{ or } \psi_2(t) \geq 1, & \text{for all } t \\
H_1: \psi_1(t) < 1 \text{ and } \psi_2(t) < 1, & \text{at some } t
\end{cases}
\]

Intersection-union test

Primary

\[
\begin{cases}
H_0: \psi_1(t) < 1 \text{ or } \psi_2(t) < 1, & \text{for all } t \\
H_1: \psi_1(t) \geq 1 \text{ and } \psi_2(t) \geq 1, & \text{at some } t
\end{cases}
\]

Union intersection test

Weighted Bonferroni adjustment

Rejection region of $H_0$

\[
\{Z_1 > z_\alpha\} \cap \{Z_2 > z_\alpha\}
\]

$Z_k \sim N(0,1)$

$\text{corr}[Z_1, Z_2] = \rho_{Z_1Z_2}$

$\alpha$ \(\cdots\) significant level for hypothesis testing

$z_\alpha$ \(\cdots\) a upper $\alpha$ th percent point of the standard normal distribution

$\gamma_k$ \(\cdots\) weight $\gamma_1 + \gamma_2 = 1$

$z_{\gamma_k \alpha}$ \(\cdots\) a upper $\gamma_k \alpha$ th percent point of the standard normal distribution
Power for detecting the effect on both or at least one endpoint

\[ 1 - \beta = \Pr \left[ \bigcap_{k=1}^{2} \{ Z_k > z_\alpha \} \mid H_1 \right] \]

**Conjunctive power**

\[ 1 - \beta = \Pr \left[ \bigcup_{k=1}^{2} \{ Z_k > z_{\gamma_k}\alpha \} \mid H_1 \right] \]

**Disjunctive power**

Weighted Bonferroni adjustment

The power is calculated by the cumulative distribution function of bivariate standardized normal distribution with correlation \( \rho_Z \)

\[
1 - \beta = \Pr \left[ \bigcap_{j=1}^{2} \{ Z_k > z_\alpha \} \right] \approx \int_{z_\alpha - \sqrt{n\mu_1}/\sqrt{V_{11}}}^{\infty} \int_{z_\alpha - \sqrt{n\mu_2}/\sqrt{V_{22}}}^{\infty} f(z_1^*, z_2^*, \rho_Z) \, dz_1 \, dz_2
\]

\[
z_k^* = \left( U_k - \sqrt{n\mu_k} \right) / \sqrt{V_{kk}}
\]

\[
\rho_Z = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}
\]

\( f(\cdot;\cdot;\rho) \) is the bivariate normal density function with zero mean vector and correlation matrix \( \rho_Z \)
Censoring scheme: dependent or independent censoring

\[ T_{i1}^* \]

\[ T_{i2}^* \]

- Dependent or independent censoring
- One event is subject to censoring by other event?

1. Both non-fatal
   - **E1**: \( T_{i1} < C \)  
     - Time to virologic failure
   - **E2**: \( T_{i2} < C \)  
     - Discontinuation due to toxicity
   - *Sugimoto et al (2013)*

2. One fatal
   - **Cardiovascular Trial/ Oncology**
   - **Semi-competing risk** (Fine et al, 2001)
   - **E1**: \( T_{i1} < \min (T_{i2}, C) \)  
     - Hospitalization  
     - TTS
   - **E2**: \( T_{i2} < C \)  
     - Death  
     - OS

3. Both fatal
   - **Oncology Trial**
   - **E1**: \( T_{i1} < \min (T_{i2}, C) \)  
     - Disease-specific mortality
   - **E2**: \( T_{i2} < \min (T_{i1}, C) \)  
     - Other cause-specific mortality

4. Composite
   - **E1’**: \( T_{i1} = \min (T_{i1}, T_{i2}) \)
   - **E2’**: \( T_{i1} \) or \( T_{i2} \)
   - Composite and its component
     - MACE and Death
     - PFS and OS

Modeling time-dependent association

Joint survival Function \( S^{(j)}(t, s) = \Pr(t < T_{i1}^*, s < T_{i2}^* | g_i = j) \)

Marginal survival function \( S_k^{(j)}(t) = \Pr(t < T_{ik}^* | g_i = j) \)

\[ \rho^{(j)} = \text{corr} \left[ \Lambda^{(j)}_1(T_{i1}^*), \Lambda^{(j)}_2(T_{i2}^*) \right] = \int_0^\infty \int_0^\infty S^{(j)}(t, s) d\Lambda^{(j)}_1(t) d\Lambda^{(j)}_2(s) - 1 > 0 \]

In absence of censoring, \( \rho^{(j)} \) can be estimated replacing functions \( \Lambda^{(j)}_k(t) \) with Nelson-Aalen estimators. If each marginal is exponential distribution, \( \rho^{(j)} = \text{corr}[T_{i1}^*, T_{i2}^*] \)

\[ \rho^{12}_Z = \frac{V_{12}}{\sqrt{V_{11}V_{22}}} \]

\[ \rho^{kk} = \int_0^\tau H_k^2(t) \left\{ \frac{d\Lambda^{(1)}_k(t)}{a^{(1)} y^{(1)}_k(t)} + \frac{d\Lambda^{(2)}_k(t)}{a^{(2)} y^{(2)}_k(t)} \right\} \]

\[ \Lambda^{(j)}(t, s) = \int_0^t \int_0^s \lambda^{(j)}(x, y) dy dx \]

\[ a^{(k)} = \frac{n_k}{N} \]

\[ y^{(j)}(t) = \text{E}[\bar{Y}^{(j)}] / n^{(j)} \]
Modeling time-dependent association by copulas

\[ S^{(j)}(t, s) = C(S_1^{(j)}(t), S_2^{(j)}(s) : \theta^{(j)}) \leftrightarrow \leftrightarrow \leftrightarrow \text{Marginal - Exponential} \]

\( C \) be a function which generates the joint survival functions \( S^{(j)}(t, s) \) from the two marginal \( S_1^{(j)}(t) \) and \( S_2^{(j)}(t) \) with association parameter \( \theta^{(j)} \).

Clayton DG. *Biometrika* 1978; 65:14-151.
Hougaard P. *Biometrika* 1984; 71:75-83.
Total sample size for “co-primary” endpoints: “late-time” dependency

- For **both non-fatal**, the sample size **decreases** as correlation goes toward one, maximum is given when the correlation is zero
- For **one fatal** and **one fatal composite**, the sample size **increases** until some point and then **decreases** as correlation goes toward one, maximum depends on hazard ratios - For is significant effect of censoring by other event on sample size behavior.

ψ_1 = ψ_2 = 0.667
α = 0.025
1 − β = 0.80
r = 0.5

φ_1 = φ_2 = 0.76
S_1 = 0.4 S_2 = 0.5
φ_1 = φ_2 = 0.58
S_1 = 0.3 S_2 = 0.5
φ_1 = φ_2 = 0.43
S_1 = 0.2 S_2 = 0.5
Total sample size for “co-primary” endpoints: “early-time” dependency

- For **both nonfatal, one fatal** and **one fatal composite**, the sample size decreases as correlation goes toward one, maximum is given when the correlation is zero.
- For **one fatal** and **one fatal composite**, there is no significant effect of censoring by other event on sample size behavior.
Total sample size for “primary” endpoints: un-weighted Bonferroni

\[
\psi_1 = \psi_2 = 0.667 \\
\alpha_1 = \alpha_2 = 0.0125 \\
1 - \beta = 0.80 \\
r = 0.5
\]

- \(\varphi_1 = \varphi_2 = 0.76\)
- \(S_1 = 0.4\) \(S_2 = 0.5\)
- \(\varphi_1 = \varphi_2 = 0.58\)
- \(S_1 = 0.3\) \(S_2 = 0.5\)
- \(\varphi_1 = \varphi_2 = 0.43\)
- \(S_1 = 0.2\) \(S_2 = 0.5\)
Summary

**Focus**

- Methods for power and sample size determination for comparing the effect of two interventions in superiority clinical trials with two time-to-event outcomes, when the aim is (i) to evaluate a joint effect on both outcomes, or (ii) to evaluate an effect on at least one outcome

**Findings**

- Sample sizing in clinical trials with two time-to-event outcome is more complex compared with other scale endpoints such as continuous or binary outcomes - many aspects to be considered in sample size determination in clinical trials.
  - **Co-primary**: Assuming zero correlation is not conservative when one event is fatal and the association between the two time-to-event is late-time dependency
  - **Primary**: Assuming one correlation is not conservative when one event is fatal and the association between the two time-to-event is late-time dependency
- The relationship between two time-to-event outcomes including censoring scheme and time dependency associations should be carefully evaluated when sample size is determined
Thank you for your kind attention

If you have any questions, please e-mail to
toshi.hamasaki@ncvc.go.jp