Group-sequential three-arm noninferiority clinical trial designs

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

joint work with

Toshimitsu Ochiai, PhD
Shionogi Co. & LTD.

Scott R Evans, PhD
Harvard T.H. Chan School of Public Health

Koko Asakura, PhD
National Cerebral and Cardiovascular Center

Mayumi Fukuda, MD, MPH
National Cerebral and Cardiovascular Center

Haruko Yamamoto, MD, PhD
National Cerebral and Cardiovascular Center
Presentation outline

1. **Introduction**
   - Issues in two-arm and three-arm NI trials
   - Objectives

2. **Group-sequential designs for three–arm NI trials**
   - Notation and statistical settings
   - Fixed margin and fraction approach
   - Two decision-making frameworks

3. **An illustration**: Rotigotine trial

4. **Summary and findings, and possible extensions**
Issues in noninferiority trials: constancy and assay sensitivity

Two-arm noninferiority (NI)

Test

Active Control

An active intervention which has been shown to be efficacious (e.g., superior to placebo) in a historical trial may considered as the active control in a NI trial.

**Constancy:** assumption that the active control effect over placebo has not changed over time.

*This is not testable* in a trial without a concurrent placebo group.

**Assay Sensitivity (AS):** ability for the trial to be able to detect differences between strategies if they truly exist.

Many factors can affect AS: poor disease diagnosis, endpoint selection and timing, poor adherence, loss to follow-up, prior therapy, inclusion of subgroups, and use of concomitant therapies.

Three-arm NI Trial as a gold standard design

Provide the opportunity of establishing the validity of the assay sensitivity via a comparison of the placebo with the active control intervention within the trial.

Provide challenges:

**Ethical issue** there may be ethical constraints to using a placebo

**Difficulty** there is the added difficulty of evaluating **two distinct co-objectives**: evaluation of (i) the superiority of the active control intervention to placebo (AS) and (ii) the NI of the test intervention to the active control intervention.

**Feasibility** it may result in a trial with too large and impractical of a sample size to conduct due to the two co-objectives
Objectives

To discuss group-sequential designs (GSDs) for three-arm NI clinical trials

GSDs offer the possibility to stop a trial early when evidence is overwhelming and thus offers efficiency- potentially fewer trial participants and minimizing the amount of time that participants receive a placebo, compared to fixed-sample designs.

To extend two existing approaches for evaluating AS and NI into a GS setting

**Fixed margin approach** (Koch and Röhmel, 2004; Hida and Tango, 2011, 2013)

**Fraction approach** (Pigeot et al, 2003)

To discuss a three-arm NI trial that has two co-primary objectives: AS and NI

If the AS assumption does not hold, then there will be uncertainty regarding whether a NI result means that they are similarly effective or similarly ineffective.

When there is a concern about the AS, to make the evaluation of objective (ii) more interpretable, evaluate a direct comparison of the control intervention (C) with the placebo (P)

Notations and statistical settings

Three-arm NI Test (T)
- $Y_{Ti}(i = 1, ..., n_{TL})$
  - $E[Y_{Ti}] = \mu_T$
  - $\text{var}[Y_{Ti}] = \sigma^2$

Active Control (C)
- $Y_{Ci}(j = 1, ..., n_{CL})$
  - $E[Y_{Ci}] = \mu_C$
  - $\text{var}[Y_{Ci}] = \sigma^2$

Placebo Control (P)
- $Y_{Pm}(m = 1, ..., n_{PL})$
  - $E[Y_{Pm}] = \mu_P$
  - $\text{var}[Y_{Pm}] = \sigma^2$

- $N_1 \rightarrow 1$
- $N_l \rightarrow l$
- $N_L \rightarrow L$

$N_L = n_{TL} + n_{CL} + n_{PL} = n_{TL}(1 + r_C + r_P)$

Test statistics for NI
- $Z_1, Z_l, Z_L$

Test statistics for AS
- $Z_1^A, Z_l^A, Z_L^A$

- Each test statistic normally distributed for large sample
- 2L test statistics 2L-variate normally distributed
Fixed margin approach

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AS (Superiority)</strong></td>
<td>$H_0^{AS}: \mu_C - \mu_P \leq \omega$</td>
<td>$H_0^{NI}: \mu_T - \mu_C \leq -\omega$</td>
</tr>
<tr>
<td>$H_1^{AS}: \mu_C - \mu_P &gt; \omega$</td>
<td>$H_1^{NI}: \mu_T - \mu_C &gt; -\omega$</td>
<td></td>
</tr>
<tr>
<td>$Z_{l}^{AS} = \frac{\bar{Y}<em>{Cl} - \bar{Y}</em>{Pl} - \omega}{\sigma \sqrt{\frac{1}{n_{Cl}} + \frac{1}{n_{Pl}}}}$</td>
<td>$Z_{l}^{NI} = \frac{\bar{Y}<em>{Tl} - \bar{Y}</em>{Cl} + \omega}{\sigma \sqrt{\frac{1}{n_{Tl}} + \frac{1}{n_{Cl}}}}$</td>
<td></td>
</tr>
<tr>
<td>$\bar{Y}<em>{Tl} = n</em>{Tl}^{-1} \sum_{i=1}^{n_{Tl}} Y_{Ti}$, $\bar{Y}<em>{Cl} = n</em>{Cl}^{-1} \sum_{j=1}^{n_{Cl}} Y_{Cj}$, $\bar{Y}<em>{Pl} = n</em>{Pl}^{-1} \sum_{m=1}^{n_{Pl}} Y_{Pm}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr[$Z_{l}^{AS}, Z_{l'}^{AS}$] = $-\frac{r_P}{\sqrt{(1 + r_C)(r_P + r_C)}}$ $(1 \leq l \leq l' \leq L)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Both hypotheses are tested at $\alpha$
- Imposes an extra condition on the hypothesis testing for the AS, that is, superiority the C to the P is demonstrated with a NI margin $\omega$
- The inequalities $\mu_P < \mu_C - \omega \leq \mu_T$ hold for any value of $\omega$

Fraction approach

<table>
<thead>
<tr>
<th>AS (Superiority)</th>
<th>NI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0^{AS} ): ( \mu_C - \mu_P \leq 0 )</td>
<td>( H_0^{NI} ): ( \frac{\mu_T - \mu_C}{\mu_C - \mu_C} \leq \theta )</td>
</tr>
<tr>
<td>( H_1^{AS} ): ( \mu_C - \mu_P &gt; 0 )</td>
<td>( H_1^{NI} ): ( \frac{\mu_T - \mu_C}{\mu_C - \mu_C} &gt; \theta )</td>
</tr>
</tbody>
</table>

\[
Z_l^{AS} = \frac{\bar{Y}_{Cl} - \bar{Y}_{Pl}}{\sigma \sqrt{\frac{1}{n_{Cl}} + \frac{1}{n_{Pl}}}}
\]

\[
Z_l^{NI} = \frac{\bar{Y}_{Pl} - \theta \bar{Y}_{Cl} + (1 - \theta) \omega}{\sigma \sqrt{\frac{1}{n_{Pl}} + \frac{\theta^2}{n_{Cl}} + \frac{(1 - \theta)^2}{n_{Cl}}}}
\]

\[
\bar{Y}_{Ti} = n_{Ti}^{-1} \Sigma_{i=1}^{n_{Ti}} Y_{Ti}, \quad \bar{Y}_{Cl} = n_{Cl}^{-1} \Sigma_{j=1}^{n_{Cl}} Y_{Cj}, \quad \bar{Y}_{Pl} = n_{Pl}^{-1} \Sigma_{m=1}^{n_{Pl}} Y_{Pm}
\]

\[
\text{corr}[Z_l^{AS}, Z_{l'}^{NI}] = \frac{\frac{-\theta}{r_C} + \frac{(1 - \theta)}{r_P}}{\sqrt{1 + \frac{\theta^2}{r_C^2} + \frac{(1 - \theta)^2}{r_P^2}}} \left(1 \leq l \leq l' \leq L\right)
\]

- Both hypotheses are tested at \( \alpha \)
- \( \theta (0 < \theta < 1) \) is prespecified and determined by
  \( \theta = 1 - \omega / (\mu_C - \mu_P) \)
  as a fraction of difference between \( \mu_C \) and \( \mu_P \), using the NI margin \( \omega \).
- Hypothesis testing is logically ordered: AS \( \rightarrow \) NI

Features in the two approaches

<table>
<thead>
<tr>
<th>Fixed margin</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>● The correlation is determined by the sampling ratio, but the two test statistics are always negatively correlated</td>
<td>● The two test statistics are positively or negatively correlated depending on sampling ratio and fraction</td>
</tr>
<tr>
<td>● Can indirectly demonstrate the superiority of the experimental intervention relative to the placebo if ( H_0^{AS} ) and ( H_0^{NI} ) are rejected, without direct comparison of the experimental intervention to the placebo.</td>
<td>● Can demonstrate ( \mu_T &gt; \mu_P ) irrespective of ( \theta ) since ( \mu_T - \mu_P &gt; \theta (\mu_C - \mu_P) &gt; 0 ) if both null hypotheses ( H_0^{AS} ) and ( H_0^{NI} ) are rejected</td>
</tr>
<tr>
<td>● Can reject ( H_0^{NI} ) when ( \mu_C - \Delta &lt; \mu_P &lt; \mu_C ) is true</td>
<td>● Cannot reject ( H_0^{NI} ) when ( \mu_C - \Delta &lt; \mu_P &lt; \mu_C ) is true - whether the fraction approach can allow demonstration of noninferiority of the experimental intervention to the control intervention is questionable under ( \mu_C - \Delta &lt; \mu_P ).</td>
</tr>
</tbody>
</table>
Decision-making frameworks for GS designs: DF-A

- NI is evaluated only after the AS is demonstrated.
- A trial stops if the AS and the NI are achieved at any interim analysis, i.e., not necessarily simultaneously.
- If AS is demonstrated but NI is not, then the trial continues and subsequent hypothesis testing is repeatedly conducted only for NI until the NI is demonstrated.

Stopping rule for DF-A

<table>
<thead>
<tr>
<th>At the lth analysis (l = l', ..., L − 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>if $Z_{l'}^{AS} &gt; c_{l'}^{AS}(\alpha)$ for some $l'(1 \leq l' \leq l)$ ($H_0^{AS}$ has been rejected), and $Z_{l}^{NI} &gt; c_{l}^{NI}(\alpha)$, then reject $H_0^{NI}$ and stop the trial</td>
</tr>
<tr>
<td>otherwise, continue the trial</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At the Lth analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>if $Z_{L}^{AS} &gt; c_{L}^{AS}(\alpha)$ for some $l'(1 \leq l' \leq L)$, and $Z_{L}^{NI} &gt; c_{L}^{NI}(\alpha)$, then reject $H_0^{NI}$</td>
</tr>
<tr>
<td>otherwise do not reject $H_0^{NI}$</td>
</tr>
</tbody>
</table>

$c_{l}^{AS}(\alpha)$ and $c_{l}^{NI}(\alpha)$ are critical values separately prespecified, using any GS methods.
### Decision-making frameworks for GS designs: DF-B

A special case of DF-A, NI is evaluated only after the AS is demonstrated,

A trial stops only if AS and NI are demonstrated at the same interim analysis simultaneously.

Otherwise, the trial will continue and the subsequent hypothesis testing is repeatedly conducted for both AS and NI until simultaneous significance is reached.

#### Stopping rule for DF-B

- **At the $l$th analysis ($l = 1, \ldots, L - 1$)**
  - if $Z_{l,AS} > c_{l,AS}^*(\alpha)$ and $Z_{l,NI} > c_{l,NI}^*(\alpha)$, then reject $H_0^{AS}$ and $H_0^{NI}$ and stop the trial
  - otherwise, continue the trial

- **At the $L$th analysis**
  - if $Z_{L,AS} > c_{L,AS}^*(\alpha)$ and $Z_{L,NI} > c_{L,NI}^*(\alpha)$, then reject $H_0^{AS}$ and $H_0^{NI}$ and stop the trial
  - otherwise do not reject $H_0^{AS}$ and $H_0^{NI}$

$c_{l,AS}^*(\alpha)$ and $c_{l,NI}^*(\alpha)$ are critical values separately prespecified, using any GS methods.
## Approaches and decision-making frameworks

<table>
<thead>
<tr>
<th>Fixed Margin</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Flexible, and slightly powerful than DF-B</td>
<td>• Flexible, and slightly powerful than DF-B</td>
</tr>
<tr>
<td>• Allows for dropping placebo group if AS is demonstrated at the interim</td>
<td>• Not allow for dropping placebo group as the test statistics for the NI includes the amount of $\bar{Y}_{pm}$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DF-A</th>
<th>DF-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Simple, but less powerful than DF-A</td>
<td>• Simple, but less powerful than DF-A</td>
</tr>
<tr>
<td>• Not allow for dropping of the placebo group even if AS is demonstrated at the interim</td>
<td></td>
</tr>
</tbody>
</table>
An illustration: Rotigotine trial (Mizuno et al 2014)

**Objective** To evaluate the superiority of transdermal rotigotine to placebo, and to evaluate NI to ropinirole, in Japanese Parkinson’s disease patients on concomitant levodopa therapy.

**Primary endpoints** The change in the unified Parkinson’s disease rating scale (UPDRS) Part III (ON state) sum score from baseline to week 16 of the treatment period

### Rotigotine trial: MSS and ASN based on DF-A and DF-B

<table>
<thead>
<tr>
<th># of analyses and decision-making frameworks</th>
<th>Bound. Func. (AS-NI)</th>
<th>Fixed Margin Approach</th>
<th>Fraction Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSS</td>
<td>ASN1</td>
<td>ASN2</td>
</tr>
<tr>
<td></td>
<td>MSS</td>
<td>ASN1</td>
<td>ASN2</td>
</tr>
<tr>
<td>( L = 1 ) \text{ DF-A}</td>
<td>OF-OF</td>
<td>720</td>
<td>713</td>
</tr>
<tr>
<td></td>
<td>PC-PC</td>
<td>801</td>
<td>681</td>
</tr>
<tr>
<td></td>
<td>OF-PC</td>
<td>759</td>
<td>726</td>
</tr>
<tr>
<td></td>
<td>PC-OF</td>
<td>771</td>
<td>730</td>
</tr>
<tr>
<td>( L = 2 ) \text{ DF-B}</td>
<td>OF-OF</td>
<td>720</td>
<td>713</td>
</tr>
<tr>
<td></td>
<td>PC-PC</td>
<td>810</td>
<td>686</td>
</tr>
<tr>
<td></td>
<td>OF-PC</td>
<td>759</td>
<td>726</td>
</tr>
<tr>
<td></td>
<td>PC-OF</td>
<td>789</td>
<td>743</td>
</tr>
</tbody>
</table>

**MSS:** Maximum sample size  
**ASN1:** Average sample number without dropping P after AS is demonstrated.  
**ASN2:** Average sample number with dropping P after AS is demonstrated.
Summary and findings

1. DF-A and DF-B for the fixed margin and the fraction approaches provide the possibility of stopping a trial early when evidence is overwhelming, thus offering efficiency (e.g., an ASN potentially 4–15 % fewer than the fixed-sample designs with equally sized groups and four analyses).

2. There are no major differences in both MSS and ASN between DF-A and DF-B for the fixed margin and the fraction approaches, although DF-A is slightly more powerful than DF-B. By using the DF-A for the fixed margin approach, the time that participants are exposed to placebo can be minimized as the DF-A allows dropping of the placebo group if AS has been demonstrated at an interim analysis.

3. For the fixed margin approach, selecting the OF-type boundary for both AS and NI could lead to fewer participants for the MSS and the ASN compared with other critical boundary combinations.

4. For the fraction approach, selecting the OF-type boundary for both AS and NI, or the PC-type boundary for AS and the OF-type boundary for NI provides better efficiency with respect to the MSS and the ASN compared with other critical boundary combinations.
Possible extensions for fixed margin approach

Extension 1

- AS
- NI

1 2 3 4

Delaying analyses for NI

Extension 2

- AS
- NI

1 2 3 4

Allocating $\alpha$ adaptively depending on AS

- Can improve the power and decrease MSS, but lose the efficiency (increase ASN)
- Can apply the fraction approach

- Can improve the power and decrease MSS.
- Generally adaptive alpha-allocation inflate the Type I error when the two test statistics are positively correlated, but this is not happen when they are negatively correlated.
- Can not apply the fraction approach
Thank you for your kind attention

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