Non-Inferiority Trial Design Without Placebo Arm

H.M. James Hung, Ph.D.

DB1/OB/OTS/CDER
U.S. Food and Drug Administration

Presented in Biostatistics Day at Rutgers University, April 3, 2009
Collaborators

Sue-Jane Wang, OB/OTS/CDER/FDA
Robert O’Neill, OB/OTS/CDER/FDA

Disclaimer

The views presented in this presentation are not necessarily of the U.S. Food and Drug Administration.
Non-inferiority Design w/o Placebo

T: Test Drug
C: (Active) Control
P: Placebo *(absent from NI trial)*

Endpoint mostly evaluated in NI trial:
- time to clinical event (e.g., mortality)
- clinical event (yes/no)

Risk ratio (RR): hazard ratio, relative risk, odds ratio
Mostly, such an NI trial is to assert that test drug T is efficacious (i.e., would have beaten placebo had the placebo been present), by indirect inference via direct comparison with the selected active control, and retains a substantial proportion of active control effect.

For this objective, the term ‘non-inferiority’ may be very misleading.
Outline

• Challenges
• Essence of fixed margin and synthesis methods
• Back to reality
  - assess adequacy of NI margin
• Remarks
**Parameters**

**Historical trial**
- \( C_0/P_0 \): risk ratio of control vs. placebo

**NI trial**
- \( T/C \): risk ratio of test drug (T) vs. control (C)
- \( C/P \)

\[
100\lambda \% \text{ (what percent?)} \text{ retention}
\]

- \( H_1: \ln(P/T) > \lambda \ln(P/C) \iff \ln(T/C) < (1-\lambda)\ln(P/C) \)
- \( H_0: \ln(T/C) \geq (1-\lambda)\ln(P/C) \)

**NI margin:** \( \delta \equiv (1-\lambda)\ln(P/C) \)

(parameter, value unknown)
Challenge 1

True margin to rule out depends on $C/P$ and $\lambda$ (this is unnecessary)

Need knowledge of $C/P$ to make a subjective selection of $\lambda$

$C/P$ not estimable.
At best, may bridge from historical trial to NI trial to connect $C/P$ with $C_0/P_0$
Challenge 2

How to estimate $C_0/P_0$ from historical PC trials?

- **Fixed effect approach:**
  Estimate “average effect”, what does it mean if there is large between-trial variability?
  Ignore between-trial variability in deriving CI

- **Random effect approach:**
  Account for between-trial variability by making some unverifiable assumption (randomness), but is it harmful?
Challenge 3

Only control’s effect in NI trial is relevant to retain. Thus constancy assumption* (Frequentist model: \( \frac{P}{C} = \frac{P_0}{C_0} \)) is critical.

If the assumption does not hold, the hypothesis of effect retention cannot be tested.

No data to verify this assumption

* A Bayesian model (still needs its version of CA):

\[
\frac{P}{C} = \theta + \eta, \quad \frac{P_0}{C_0} = \theta + \eta_0 \\
\eta, \eta_0, \text{ i.i.d. } \sim (0, \sigma^2) 
\]
Figure 1. Placebo response against year since 1965 of publication. Size of circle is proportional to size of study

Constancy Assumption (CA)?

Estimates available

**Historical trial**

\[
\ln(\tilde{C}_0 / \tilde{P}_0) \sim N(\ln(C_0 / P_0), \sigma_{cp0}^2)
\]

\[
\ln(\tilde{P}_0 / \tilde{C}_0) - 1.96\sigma_{CP0} > 0 \quad [\text{Control is effective}]
\]

**NI trial**

\[
\ln(\hat{T} / \hat{C}) \sim N(\ln(T / C), \sigma_{tc}^2)
\]
Challenge 4: Inference Method

Fixed margin vs. Synthesis methods
Different philosophy/paradigm
Fixed margin method $\Rightarrow$ control NI trial error for direct comparison of $T$ vs. $C$
Synthesis method $\Rightarrow$ control across-trial inference (i.e., integrating NI and historical trials) error for including indirect inference for $T$ vs. $P$
Fixed Margin Method

- Historical Trials
- Assumptions: CA, AS
- Define NI, δ
- Est. \( P_0/C_0 \) & SE
- Clinical Margin
- Stat Margin

J. Hung, Biostat Day, Rutgers U
Fixed Margin Method

Define NI $\delta$

NI hypothesis established

Stat Inference $\Rightarrow$

NI trial

$\geq 95\%$ CI rule out $\delta$?
Fixed Margin Method

Find an estimate $\tilde{\delta}$ (from historical trials only), e.g., worst limit of 95% CI, hoping the target NI margin satisfies $\tilde{\delta} < \delta$ with high probability (the inequality cannot be verified, purely based on subjective judgment).

Note: $\tilde{\delta} < \delta$ factors in some statistical uncertainty, at least from historical data and subjective judgment of assumptions (CA, AS).
Fixed Margin Method

95_{NI} - 95_{H} method for asserting 50% retention

\[ \tilde{\delta} = 0.5[\ln(\tilde{P}_0 / \tilde{C}_0) - 1.96\sigma_{CP0}] \]

\[ \ln(\hat{T} / \hat{C}) + 1.96\sigma_{TC} < \tilde{\delta} \]

\[ \Rightarrow \text{assert 50% retention} \]
The fixed margin method, $95_{NI}-95_H$, is intended to control NI trial type I error rate for testing of 50% retention hypothesis or beating placebo

$$\Pr_{NI}\{ \ln(\hat{T} / \hat{C}) + 1.96\sigma_{TC} < \tilde{\delta} \mid H_0; \tilde{\delta} \} \leq 0.025$$

$H_0$: $\ln(T/C) \geq \delta$, not $\tilde{\delta}$

This error rate is probability of falsely rejecting $H_0$, conditional on the established margin; that is, this error rate is calculated by repeating only NI trial infinitely often, given $\tilde{\delta}$ is fixed and accepted.
Synthesis Method

Historical Trials

Est. $P_0/C_0$ & SE

Assumptions: CA, AS

NI Trial

Est. $T/C$ & SE

Synthesis test

Statistical Inference
Synthesis Method

Synthesis method combines standard errors from both sources (i.e., historical trials and NI trial). The resulting standard error is not the standard error from a randomized comparison.

!!! Clinical margin is not considered !!!
Synthesis Test Method

\[ H_1: \ln(P/T) > 0.5\ln(P/C) \iff \ln(T/C) < 0.5\ln(P/C) \]
\[ H_0: \ln(T/C) \geq 0.5\ln(P/C) \]

\[
Z = \frac{\ln(\frac{\hat{T}}{\hat{C}}) + 0.5\ln(\frac{\bar{C}_0}{\bar{P}_0})}{\sqrt{\sigma_{tc}^2 + (0.5)^2 \sigma_{cp0}^2}}
\]

\[ Z < -1.96 \Rightarrow reject \ H_0 \]

\[ Pr(Z < -1.96 \mid H_0) \leq 0.025, \]

if constancy assumption holds
Synthesis Test Method

If constancy assumption is doubtful, add discounting factors# to numerator and/or denominator of synthesis Z test.

How much to discount is purely a subjective judgment w/o any data to support!

# Snapinn and Jiang (2007)
Note The synthesis method is intended to control across-trial (or meta-analytic) type I error rate
\[
\Pr_{\text{Across-trial}} \left\{ \frac{\ln(\hat{T} / \hat{C}) + 0.5 \ln(\tilde{C}_0 / \tilde{P}_0)}{\sqrt{\sigma^2_{tc} + (0.5)^2 \sigma^2_{cp0}}} \right\} < -1.96 \mid H_0 \leq 0.025 \quad \text{(if constant assumption holds)}
\]
This error rate is calculated by repeating both NI trial and historical trials infinitely often. The calculation incorporates statistical distributions from NI trial and historical trials.
Controversy

Is ‘meta-analytic’ type I error relevant to non-inferiority inference?

Historical trials are already done well before NI trial planning. From the standpoint of frequentist replication, is it sensible to incorporate historical trials in consideration of type I error rate for false NI conclusion?

‘Meta-analytic’ p-value or type I error rate is rarely considered in show-superiority trial.
Controversy

In the across-trial inference paradigm, inferences from two statistically independent NI trials are always statistically dependent*, because they use the same set of historical trials. But in classical paradigm, once the margin is set, inferences from two statistically independent NI trials are statistically independent.

Back to Reality

Need a NI margin (clinical assessment is necessary)

Where to pick for estimating AC effect

Discounting for uncertainty of CA
Is $95_{\text{NI}} \times \text{H}$ method ($X < 95\%$) tenable?

If C/P differs from $C_0/P_0$ only by a location shift, then exploring across trial type I error rate for asserting efficacy (beating putative placebo) may be viable for “guiding selection of confidence level $X$”

Note: primary error rate is the NI trial error rate for comparing T with C
Across error rate is secondary consideration
Estimators

**Historical data**

\[ \ln(\tilde{C}_0 / \tilde{P}_0) \sim N( \ln(C_0 / P_0), \sigma^2_{cp0} ) \]

\[ \ln(\bar{P}_0 / \bar{C}_0) - h \times 1.96 \sigma_{CP0} > 0, \quad \text{for some } h \geq 1 \]

- **h=1:** 95% CI
- **h=2:** 99.99% CI

**NI trial**

\[ \ln(\hat{T} / \hat{C}) \sim N( \ln(T / C), \sigma^2_{tc} ) \]
Across-trial type I error rate of falsely concluding ‘beat imputed placebo’ for ’95_{NI}-X_H’ method aiming at 100\% retention

\[
\Pr \left\{ \log \left( \frac{\hat{T}}{\hat{C}} \right) + 1.96 \sigma_{TC} < (1 - \lambda) (\log \left( \frac{\tilde{P}_0}{\tilde{C}_0} \right) - z_{(1-x/100)/2} \sigma_{PC0}) \right\} | K_0 \}
\]

\[
\leq \Phi \left( \frac{b / \sigma_{tc}}{\sqrt{1 + f}} - \frac{1.96 + (1.96 \lambda h + (1 - \lambda) z_{(1-x/100)/2}) \sqrt{f}}{\sqrt{1 + f}} \right)
\]

\(K_0:\ T/P = 1, \quad f = (\sigma_{CP0}/\sigma_{TC})^2\)

\(b = \log(P_0/C_0) - \log(P/C), \) location shift in act control effect (\(b > 0\) is of concern)
Ex. Suppose that based on the properly selected historical trials, we have

$$\frac{\tilde{C}_0}{\tilde{P}_0} = 0.40, \quad 95\% \text{ CI} : (0.27, 0.60)$$

$$\sigma_{CP0} \approx 0.20687$$

99.999\% CI (i.e., h = 2.25) is also below one
95_{NI}-95_{H} method with 50% retention gives

\[ \tilde{\delta} = 0.5 \log(1/0.60) = 0.25541 \]

Use of this margin to plan NI trial for detecting T = C with 90% power requires

\[ \sigma_{TC} = \frac{\tilde{\delta}}{1.96 + 1.28} = 0.07883 \]

Thus, \( f = (\sigma_{CP0}/\sigma_{TC})^2 \approx 7 \)
Let \( \frac{C}{P} = M \times \left( \frac{C_0}{P_0} \right) \), where \( M > 1 \) of concern

\[ M = \exp(b) \]

\[
\begin{array}{|c|c|}
\hline
M & \text{Unc. type I error} \\
\hline
1.0 & 0.00012 \\
1.1 & 0.00058 \\
1.2 & 0.0022 \\
1.3 & 0.0063 \\
1.4 & 0.015 \\
1.5 & 0.032 \\
1.6 & 0.059 \\
1.7 & 0.098 \\
\hline
\end{array}
\]

95\text{NI}-95\text{H} method

with \( \lambda = 0.5 \), \( f = 7 \),
\[
\sigma_{TC} = 0.07883
\]

\[ M_{\text{max}} = \log(1/0.60) \]

\[ = 1.67 \]
If $M < 1.4$ then explore $95_{NI} - X_H$ method.

<table>
<thead>
<tr>
<th>$M$</th>
<th>Unc. type I error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.00020</td>
</tr>
<tr>
<td>1.1</td>
<td>0.00090</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0031</td>
</tr>
<tr>
<td>1.3</td>
<td>0.0086</td>
</tr>
<tr>
<td>1.4</td>
<td>0.020</td>
</tr>
<tr>
<td>1.5</td>
<td>0.041</td>
</tr>
<tr>
<td>1.6</td>
<td>0.073</td>
</tr>
<tr>
<td>1.7</td>
<td>0.12</td>
</tr>
</tbody>
</table>

$95_{NI} - 90_H$ method with $\lambda = 0.5$, $f = 6$, $\sigma_{TC} = 0.08513$

$M_{max} = 1/0.58 = 1.7$
If $M > 1.4$ then explore $95_{NI}-95_H$ method w/ higher retention

<table>
<thead>
<tr>
<th>$M$</th>
<th>Unc. type I error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.00002</td>
</tr>
<tr>
<td>1.1</td>
<td>0.00013</td>
</tr>
<tr>
<td>1.2</td>
<td>0.00061</td>
</tr>
<tr>
<td>1.3</td>
<td>0.0021</td>
</tr>
<tr>
<td>1.4</td>
<td>0.0060</td>
</tr>
<tr>
<td>1.5</td>
<td>0.014</td>
</tr>
<tr>
<td>1.6</td>
<td>0.030</td>
</tr>
<tr>
<td>1.7</td>
<td>0.055</td>
</tr>
</tbody>
</table>

$95_{NI}-95_H$ method with $\lambda = 0.75$, $f = 28$

$\sigma_{TC} = 0.03942$

$M_{max} = 1/0.60 = 1.67$
Remarks

• Clinical margin is necessary and thus fixed margin method is the most natural method
• Synthesis method cannot generate fixed margin*. In what scenario is this method useful?
  maybe in semi-exploratory manner after data is in.
When it is used, what alpha level should be used? cannot be 0.025 because type I error can be far above 0.025 if constancy assumption is violated

*Hung et al (2003, 2007)
Remarks

- Aiming at controlling across-trial error rate at a fixed level is likely to be a fiction.
- Exploring a range of across-trial error rate as a function of discounting factor might be worthy of pursuit.
Remarks

• Fixed margin and synthesis methods are not comparable
• $95_{NI} - 95_H$ fixed margin method is a starting point for consideration in defining margin
  Can a $95_{NI} - X_H (X < 95)$ method useful?
• Synthesis method
  How to discount properly?
• Focus should be on how to use historical data to guide determination of a NI margin
Selected References

Holmgren (1999, JBS)
Hasselblad & Kong (2001, DIJ)
Wang, Hung, Tsong (2001, CCT)
Hung, Wang, Tsong, Lawrence, O’Neill (2003, SIM)
Temple (2001, SCT talk and DIA talk)
Rothmann, Chen, Li, Chi, Temple, Tsou (2003, SIM)
Lawrence (2005, Biometrical Journal)
Hung, Wang, O’Neill (2007, JBS)
Fleming (2006, SIM)
Selected References

Fleiss (1993, SIMR)
Hung (2007, FDA/Industry Wkshop)
Snapinn, Jiang (2007, SIM)