Non-inferiority Methodology in Drug Clinical Trial*

H.M. James Hung
Division of Biometrics I, OB/OTS/CDER, FDA

Presented in New Jersey Chapter of ASA Symposium, Somerset, NJ, May 31, 2007

*The views presented in this presentation are not necessarily of the U.S. Food and Drug Administration

Collaborators

Sue-Jane Wang (OB/OTS/CDER/FDA)
Robert O'Neill (OB/OTS/CDER/FDA)

The best advice is

Avoid non-inferiority trial

Outline

- 50% effect retention
- Two main methods
  - Fixed margin method
  - Synthesis test method
- Convertibility between the two methods
- Testing superiority and non-inferiority

Non-inferiority Design w/o Placebo

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

Endpoint: time to event (mortality)

Risk ratio (RR): hazard ratio

This is worse than missing data problem

Question

Main objectives of non-inferiority efficacy trial?

- Show efficacy of test drug?
- Demonstrate clinical indifference?
- Demonstrate that test drug retain a specified fraction of active control's effect?
  What is effect retention really intended to prove? Is this another way of defining clinical indifference?

50% Effect Retention

Often the goal of showing effect retention w.r.t. time to event endpoint is unclear:

50% retention on RR scale or
50% retention on log RR scale?

Why 50%?

Historical trial populations
C_0/P_0: risk ratio of control vs. placebo
NI trial population
T/C: risk ratio of test drug (T) vs. control (C)

50% retention on log RR scale

H_1: $\ln(P/T) > 0.5\ln(P/C) \Leftrightarrow \ln(T/C) < 0.5\ln(P/C)$
H_0: $\ln(T/C) \geq 0.5\ln(P/C)$
NI margin: $\delta \equiv 0.5\ln(P/C)$
Historical trials give an estimate for C_0/P_0 only

Constancy assumption is critical

Frequentist model: $P/C = P_0/C_0$

A Bayesian model:

$\ln(P/C) = \theta + \eta$

$\ln(P_0/C_0) = \theta + \eta_0$

η, η_0 , i.i.d. ~ $(0, \sigma_\eta^2)$

No data to verify this assumption

Historical data

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

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

NI trial

$$\ln(\hat{T} / \hat{C}) \sim N(\ \ln(T / C),\ \sigma^2_{tc}\ )$$

Two Main Methods

- Fixed margin method
- Synthesis method

Two methods are very different
- underlying assumption
- statistical error probability to control

Fixed Margin Method

Find an estimate $\tilde{\delta}_0$ (from historical trials only) such that the targeted true NI margin

$$\delta \equiv 0.5\ln(P/C) = \tilde{\delta}_0$$

This is statistically impossible. So, find a conservative estimate $\tilde{\delta}_0$ (e.g, worst limit of 95% CI) and wish that the CI covers the true margin δ most of the times.

95_{NI}-95_{H} method: Use 95% CI for test/control from NI trial to rule out conservative margin $\tilde{\delta}_0$

Fixed Margin Method

95_{NI}-95_{H} method

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

NI trial decision criterion:

$$\ln(\hat{T} / \hat{C}) + 1.96\sigma_{TC} < \tilde{\delta}_0$$

to assert 50% retention

Basis for using worst limit of 95% CI in defining statistical margin: Try to ensure

$$\mathbf{Pr_H}\{\ \delta_0 > \tilde{\delta}_0\ \} = 0.975\ ,$$

where $\delta_0 = 0.5\ln(P_0/C_0)$.
Since it is of concern that $\delta < \delta_0$, wish that choosing $\tilde{\delta}_0$ can offer sufficient protection, that is, wish most often $\delta > \tilde{\delta}_0$; thus, asserting
$\ln(T/C) < \tilde{\delta}_0$ leads to conclude that
$\ln(T/C) < \delta$

The fixed margin method, 95_{NI}-95_{H}, is intended to control classical NI-trial level type I error rate

$$\mathbf{Pr_{NI}}\{\ \ln(\hat{T} / \hat{C}) + 1.96\sigma_{TC} < \tilde{\delta}_0 \mid \boldsymbol{H}_0 ; \tilde{\delta}_0\ \} \le 0.025$$

H_0: $\ln(T/C) \ge \delta \Rightarrow \ln(T/C) \ge \tilde{\delta}_0$ (a wish)

Note This error probability does not incorporate statistical distribution of $\tilde{\delta}_0$ in the calculation. That is, it is percent of falsely rejecting H_0 by repeating only NI trial infinitely often, conditional on the margin $\tilde{\delta}_0$

Synthesis Test Method

H_1: $\ln(P/T) > 0.5\ln(P/C) \Leftrightarrow \ln(T/C) < 0.5\ln(P/C)$
H_0: $\ln(T/C) \ge 0.5\ln(P/C)$

50% retention test

$$Z = \frac{\ln(\hat{T} / \hat{C}) + 0.5\ln(\tilde{C}_0 / \tilde{P}_0)}{\sqrt{\hat{\mathrm{var}}(\textit{numerator})}}$$

$Z < -1.96 \Rightarrow$ *reject* H_0

$\Pr(Z < -1.96 \mid H_0) \le 0.025$,
if constancy assumption holds

The synthesis method starts out with constant assumption which is most often doubtful or at least uncertain in practice.

Discounting for synthesis method is absolutely necessary. How?

Note The synthesis method is intended to control across-trial type I error rate

$$\Pr_{\text{Across-trial}}\{\frac{\ln(\hat{T}/\hat{C})+0.5\ln(\tilde{C}_0/\tilde{P}_0)}{\sqrt{\hat{\text{var}}(\textit{numerator})}}<-1.96 \mid H_0\}$$

≤ 0.025 ⇐ **this is true only under CA**

This error probability incorporates statistical distributions from NI trial and historical trials. That is, it is calculated by repeating both NI trial and historical trials infinitely often.

Synthesis method does not surely control NI-trial level type I error

Lawrence (2005)

$$P(\frac{\ln(\hat{T}/\hat{C})+0.5\ln(\tilde{C}_0/\tilde{P}_0)}{\sqrt{\sigma_{tc}^2+0.25\sigma_{cp0}^2}}<-1.96 \mid \partial H_0; \tilde{C}_0/\tilde{P}_0)$$

≈ 0.50 ,

if $\ln(P/C)\approx\ln(\tilde{P}_0/\tilde{C}_0)-1.96\sigma_{cp0}$ **and**

$$\frac{\sigma_{cp0}}{\sigma_{tc}}\to\infty$$

Question

Which type I error is more relevant to non-inferiority inference?

Note: Historical trials are already done way before NI trial planning. Does it make sense to repeat historical trials in considering type I error rate? **NI-trial level type I error should be the primary error to control.**

Convertibility Between Two Methods

Question

Can synthesis test method generate a fixed margin (i.e., independent of NI trial)? A crude conservative margin might be found but needs proper discounting for CA. How?
How conservative is sufficiently conservative?

$$Z = \frac{\ln(\hat{T}/\hat{C}) + 0.5\ln(\tilde{C}_0/\tilde{P}_0)}{\sqrt{\sigma_{tc}^2 + 0.25\,\sigma_{cp\,0}^2}} < -1.96$$

$$\Leftrightarrow$$

$$\ln(\hat{T}/\hat{C}) + 1.96\sqrt{\sigma_{tc}^2 + 0.25\,\sigma_{cp\,0}^2} < -0.5\ln(\tilde{C}_0/\tilde{P}_0)$$

$$\Leftrightarrow$$

$$\ln(\hat{T}/\hat{C}) + 1.96\,\sigma_{tc} < 0.5\{\ \ln(\tilde{P}_0/\tilde{C}_0) - 1.96\,\sigma_{cp\,0}\underbrace{[\sqrt{4\frac{\sigma_{tc}^2}{\sigma_{cp\,0}^2}+1} - 2\frac{\sigma_{tc}}{\sigma_{cp\,0}}]}_{\text{this term is always} < 1}\ \} \qquad (*)$$

* Rothmann et al (2003)

Let $\delta^* = 0.5\{\ln(\tilde{P}_0/\tilde{C}_0) - 1.96\,\sigma_{cp\,0}[\sqrt{4\frac{\sigma_{tc}^2}{\sigma_{cp\,0}^2}+1} - 2\frac{\sigma_{tc}}{\sigma_{cp\,0}}]\}$

δ^* looks like a fixed margin but it is not.
It depends on design elements (e.g., total # of events) of the NI trial.

$$\delta^* > \tilde{\delta}_0$$

Problem with use of δ* to plan NI trial

Ex: HR(C/P): 0.55 (0.38, 0.80) from historical data

Planned # of events for NI trial	δ* (alpha=0.025)
195	1.28
390	1.25
780	1.23
very large	1.12

A larger NI trial is penalized by having to rule out a tighter non-inferiority margin !

Do not use δ^* as a fixed margin (it isn't fixed margin)

Need to find a fixed margin more conservative than δ^* in each disease area

Remarks

1) In practice, medical colleagues need to see a NI margin specified in order to judge whether the margin is clinically acceptable
2) Need of assigning a fixed NI margin is clear for any methods
3) Fixed margin method has done discounting because the presumption is that constancy assumption is doubtful
4) Synthesis method assumes constancy assumption; thus, discounting is necessary

Sometimes it makes sense to consider 95_{NI}-X_H method (X < 95%).

How to determine X?

Can use both
- across-trial alpha error rate for asserting percent retention

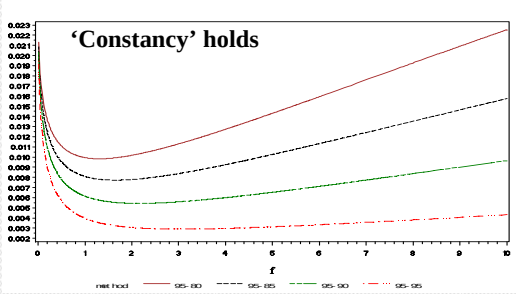
and
- across-trial alpha error rate for asserting efficacy (beating putative placebo)

to "guide selection of confidence level X"

95_{NI}-X_H method ($X \leq 95$)

Across-trial alpha error rate for 50% retention

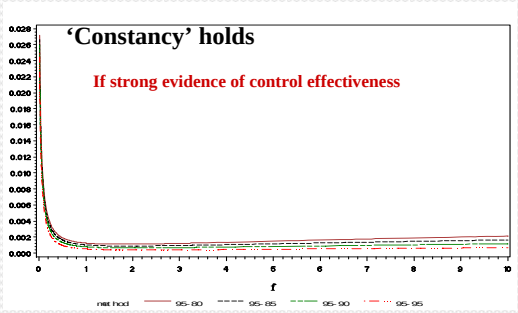
f = (# events in NI)/(# events in historical trials)



95_{NI}-X_H method (X ≤ 95)

Across-trial alpha error rate for Efficacy

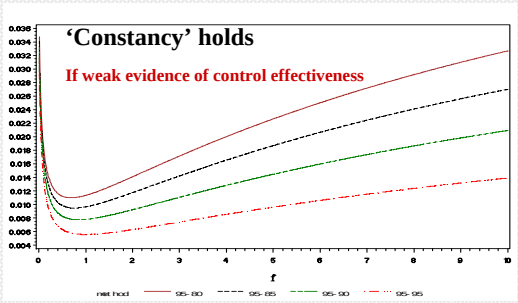
f = (# events in NI)/(# events in historical trials)



95_{NI}-X_H method (X ≤ 95)

Across-trial alpha error rate for Efficacy

f = (# events in NI)/(# events in historical trials)



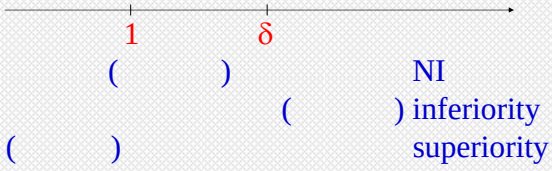
To take account of the impact of violation of 'constancy' assumption, plot across-trial alpha errors versus degree of loss in the control's effect over a plausible range of f in choosing X for the fixed-margin '95_{NI}-X_H' method.

Across-trial type I error is merely used as guidance for selecting X; it need not be controlled at a fixed level

Wang, Hung, Tsong (2001)

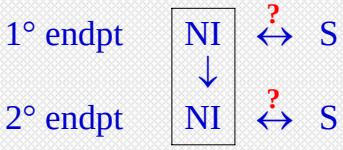
Testing S and NI on Single 1° Endpt

For 1° endpt only, it is asymptotically valid to use same CI ***w/o alpha adjustment*** for testing superiority, H_{0S}: T/C ≥ 1 H_{1S}: T/C < 1 and NI, H_{0NI}: T/C ≥ δ H_{1NI}: T/C < δ



Testing S and NI on Multiple Endpts

1° endpt, 2° endpt
Test 2° endpt for NI using the popular hierarchical testing strategy:
NI testing on 1° endpt at 0.05; if succeeded, NI testing on 2° endpt at 0.05



Can we still test 1° endpt for S at 0.05?

This expanded testing does not control studywise type I error rate strongly, e.g.,
Pr(falsely concluding S on 1° endpt or falsely concluding NI on 2° endpt) ≤ 0.10

For strong control, place superiority testing of 1° endpt in the chain of NI testing of two endpts. Is this sensible?

Superiority testing and NI testing pertain to different questions?
NI testing: test drug would have beaten a placebo had it been in the trial, or test drug would have been not too bad (preserving a desired portion of the active control's effect).
Superiority testing: test drug is superior to active control. Bar for the superiority testing should be higher.

Question

If 1° endpt is positive (i.e., shows either non-inferiority or superiority), how should 2° endpt be tested?

Remarks

- Synthesis method and fixed margin method are not comparable
- Synthesis method cannot generate a fixed margin (in theory).

  In what scenario is this method useful?

  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)?

Remarks

- 95_{NI}-95_{H} fixed margin method is a discounting method but may be too conservative sometimes in practice.

  How to alleviate conservatism?

  Use across-trial type I error as a guidance for selecting a smaller CI for the control effect

Remarks

- Synthesis method using alpha of 0.025 may be too liberal.

  How to discount properly?

  - Include a bias factor in est. of control effect and/or inflate variances in the synthesis test*
  - set alpha level to a much smaller level

*Snapinn and Jiang (2007)

Remarks

- Across-trial inference as primary paradigm of inference renders the following woeful argument

  - Any two 'independent' NI trials relying on the same set of historical trials for inference (e.g., define margin, synthesis analysis) are correlated

  No way to do or discourage more than one NI trials ?

Partial List of References

Holmgren (1999, JBS)
Simon (1999, Biometrics)
Hasselblad & Kong (2001, DIJ)
Fisher (1998, JACC)
Snapinn (2001, ASA talk; JBS, 2004)
Wiens (2002, CCT)
Wang, Hung, Tsong (2001, CCT)
Hung, Wang, Tsong, Lawrence, O'Neill (2003, SIM)
Wang, Hung (2003, SIM)
Wang, Hung (2003, CCT)
Temple (2001, SCT talk and DIA talk)
Rothmann, Chen, Li, Chi, Temple, Tsou (2003, SIM)

Partial List of References

Hung, Wang (2004, JBS)
Hung, Wang, O'Neill (2005, Biometrical Journal)
Chen, Wang, Chi (2004, JBS)
Wang, Chen, Chi (2005, JBS)
Lawrence (2005, Biometrical Journal)
Hung, Wang, O'Neill (2007, JBS)
Hung (2007, DIA)
D'Agostino (2003, SIM)