NON-INFERIORITY TEST BEYOND SIMPLE 2-SAMPLE COMPARISON*

Yi Tsong (tsong@cder.fda.gov)
& Wen Jen Chen
Office of Biostatistics, CDER, U.S. FDA

Presentation at 2005 FDA-Industry Workshop

*Disclaimer: The presentation represents the lecturer's professional opinion. It does not represent the regulatory position of U.S. FDA

OUTLINES

I. Objectives and hypotheses of 2-arm non-inferiority trials

II. Two approaches in design and analysis

III. Issues beyond simple 2-sample comparison

- Switching between superiority and NI
- Group sequential and adaptive designs
- Homogeneity testing
- Change of designs
- Data dependency

I. OBJECTIVES AND HYPOTHESES OF TWO-ARM NON-INFERIORITY TRIALS

(Tsong, Wang, Hung, Cui, JBS, 2003), (Hung, Wang, Tsong, Lawrence, O'Neill, SIM, 2003)

T (new test treatment) vs. C (active control)
P (placebo) - not studied in current AC trial)

A. Efficacy (Required for all new treatment)

Would T have been more effective than P, had P been present?

$$H_0: \mu_T \leq \mu_P \quad \text{vs.} \quad H_a: \mu_T > \mu_P$$

$$H_0: \mu_T - \mu_C \leq (\mu_P - \mu_C)$$

$$\text{vs. } H_a: \mu_T - \mu_C > (\mu_P - \mu_C)$$

μ_P or $(\mu_P - \mu_C)$ = parameter not estimable with AC data

B. Preservation of Certain % of the Control Effect *(Holmgren, J Biophar Stat 1999)*

Does T retain >100λ% of C-effect (i.e. $(\mu_T - \mu_P)/(\mu_C - \mu_P) > \lambda$)?

$H_0: (\mu_T - \mu_P) \leq \lambda (\mu_T - \mu_P)$

vs. $H_a: (\mu_T - \mu_P) > \lambda (\mu_T - \mu_P)$

i.e. $H_0: (\mu_T - \mu_C) \leq (\lambda - 1)(\mu_T - \mu_P)$

vs. $H_a: (\mu_T - \mu_C) > (\lambda - 1)(\mu_T - \mu_P)$

"Reservation of % C - effect" implies Efficacy

Note: $\lambda = 0 \Rightarrow$ Efficacy,

$\lambda = 1 \Rightarrow$ Superior over Control

$(\lambda - 1)(\mu_T - \mu_P)$ – not estimable with the AC data

C. Not much less effective

Is T "not much less effective" than C ?

$H_0: (\mu_T - \mu_P) \leq (\mu_C - \mu_P) - \delta$
vs. $H_a: (\mu_T - \mu_P) > (\mu_C - \mu_P) - \delta$

$H_0: (\mu_T - \mu_C) \leq -\delta$ vs. $H_a: (\mu_T - \mu_C) > -\delta$

$\delta > 0$, a value determined based on data of C vs. P (i.e. $\delta < (\mu_C - \mu_P)$) historical studies and medically judgement

(Blackwelder, Controlled Clinical Trials, 1982)
(Tsong, Higgins, Wang, Hung, Cui, JSM Proceedings, 1999)

(Fisher, Gent, Buller, American Heart Journal, 2001)

In order to have C: “Not much less effective” implying A: “Efficacy”, it requires

$$-\delta \geq (\mu_p - \mu_c),$$

(not estimable from internal data of AC trial)

ICH E9 (1998): Statistical Principles

Smaller than differences observed in superiority trials of the active comparator

ICH E10 (2000): Choice of Control Group

Smaller than that suggested by the smallest expected effect size of the active control

One Set of Hypotheses Fits All

$$H_0: (\mu_T - \mu_c) \leq -\delta \text{ vs. } H_a: (\mu_T - \mu_c) > -\delta$$

or $H_0: (\mu_T - \mu_c) \leq (\lambda - 1)(\mu_c - \mu_p)$

vs. $H_a: (\mu_T - \mu_c) > (\lambda - 1)(\mu_c - \mu_p)$

A: (Efficacy) : $\delta = (\mu_c - \mu_p)$, $\lambda = 0$

B: (Preservation of 100λ% of Control Effect) :

$$\lambda = 1 - \delta/(\mu_c - \mu_p) > 0$$

C: (Not much less effective):

$$-\delta = (\lambda - 1)(\mu_c - \mu_p) > -(\mu_c - \mu_p)$$

- All 3 objectives needs to imply treatment efficacy A.
- In practice, 3 different objectives may actually represent the same objective with adjustment for the uncertainty of ε such that

$$(\mu_c - \mu_p) / (\mu_{c|H} - \mu_{p|H})) \geq \varepsilon$$

(e.g. discounting and proportion preservation).

III. TWO APPROACHES IN DESIGN AND ANALYSIS

A. Generalized historical control Approach

(Non-inferiority margin approach, Fixed margin approach)

- Considers δ, a **fixed value** pre-specified before data collection
- $\delta \leq \delta_M$, medically relevant margin
- Define $d = eL_{C-P}$, L_{C-P}= the lower 1 - α_0 CL of $(\mu_{c|H} - \mu_{p|H})$,
  $(\mu_c - \mu_p)/(\mu_{c|H} - \mu_{p|H}) \geq \varepsilon > 0$ with $\delta \leq \delta_M$

- e.g. $\varepsilon = 0.5$ (*FDA Cardio-Renal Advisory Committee (1992))*
- e.g. $L_{(C - P)}$ = 99.5% lower confidence limit *(Thrombolytics Example)*

– (Ng, Drug Information Journal, 2001), (Wiens, Controlled Clinical Trials, 2002), (Jones, Jarvis, Lewis, Ebbutt, British Medical Journal, 1996)

Test Stat.

$$t(d) = (\hat{m}_T - \hat{m}_C + d)/[s.e.(\hat{m}_T - \hat{m}_C)]$$

$t(\delta)$ is compared with $t(d, 0.975)$ for rejecting H_0.

$d = n_T + n_C - 2$ if $\sigma_T = \sigma_C$.

Otherwise

$$d = (s_T^2/n_T + s_C^2/n_C)^2/[(s_T^2/n_T)^2/(n_T - 1) + (s_C^2/n_C)^2/(n_C - 1)].$$

B. Cross-Trial Comparison Approach

(Synthesis approach, Retention test, Variable margin, etc.)

- δ = a **parameter** to be estimated (with historical data)
- Consider historical C-P trials as part of the data collected independently to the current AC trial
- $\delta \leq \delta_M$, medically relevant margin
- Often define $\delta = (1- \lambda)(\mu_{c|H} - \mu_{p|H})$
- **Active control treatment is used in both P-C and AC trials (3- or 4- parameter study?)**
- Often assume $(\mu_c - \mu_p) = (\mu_{c|H} - \mu_{p|H})$
- In fact $(\mu_c - \mu_p)/(\mu_{c|H} - \mu_{p|H}) = \varepsilon$ (unknown) > 0

Study	Placebo	Active Control	Test
C vs P #1	N_{1P}	N_{1C}	
C vs P #2	N_{2P}	N_{2C}	
C vs P #.	$N_{.P}$	$N_{.C}$	
C vs P #.	$N_{.P}$	$N_{.C}$	
C vs P #K	N_{KP}	N_{KC}	
T vs C #K+1		$N_{(K+1)C}$	$N_{(K+1)T}$

Testing H_0: $(\mu_T - \mu_C) \leq (1 - 1)(\mu_{c|H} - \mu_{p|H})$

vs. H_a: $(\mu_T - \mu_C) \leq (1 - 1)(\mu_{c|H} - \mu_{p|H})$, $0 \leq 1 \leq 1$

Four-arm trial ?

Assume that $e_1 = s.e.(\hat{m}_T - \hat{m}_C)$, $e_2 = s.e.(\hat{m}_{C|H} - \hat{m}_{P|H})$

Test statistic - $z(d) = [(\hat{m}_T - \hat{m}_C) - (l-1)(\hat{m}_{C|H} - \hat{m}_{P|H})] / \sqrt{e_1^2 + (l-1)^2 e_2^2}$

Compare $z(\delta)$ with $Z_{0.975}$ for rejecting H_0 with large n's.

- T test can be derived (Pigeot, Schafer, Rohmel, Hauschke, SIM, 2003)
- Special cases:
  - Assume $\varepsilon = 1$, $\lambda = 0$, but uses $(e_1 + e_2)$ instead of $\sqrt{e_1^2 + e_2^2}$
    (Hauck & Anderson, DIJ, 1999) – "2 – CI Approach"
  - Assume $\varepsilon = 1$, $\lambda > 0$, 2- CI approach using $[e_1 + (1-\lambda)e_2]$ instead of $\sqrt{e_1^2 + (l-1)^2 e_2^2}$. Use pos hoc determined confidence level for the 2nd CI
    (Rothman et al, JBS, 2004)
  - Assume $\varepsilon = 1$ *(Holmgren, JBS, 1999)*
  - Assume $\lambda = 0$, $\varepsilon = 1$ *(Hasselblad & Kong, DIJ, 2001)*

- Control type I error rate for testing

  $\mathbf{H_0}$: $(\mu_T - \mu_C) \leq (1 - 1)(\mu_{c|H} - \mu_{p|H})$

  Not for testing

  $\mathbf{H_0}$: $(\mu_T - \mu_C) \leq (1 - 1)(\mu_c - \mu_p)$
- Interpret the result for testing

  $\mathbf{H_0}$: $(\mu_T - \mu_C) \leq (1/\varepsilon)(1 - 1)(\mu_c - \mu_p)$,

  if $(\mu_c - \mu_p)/(\mu_{c|H} - \mu_{p|H}) = \varepsilon$
- If $\varepsilon < 1 - \lambda$, $(1/\varepsilon)(1 - 1) < -1$, can't imply efficacy

  $\Rightarrow$ invalid NI test
- If $\varepsilon < 1$, avoid $\lambda = 0$ hypothesis

III. Issues beyond simple 2-sample Comparison

i. Switching Between Superiority and Non-Inferiority

Study	Placebo	Active Control	Test
C vs P #1	N_{1P}	N_{1C}	
C vs P #.	$N_{.P}$	$N_{.C}$	
C vs P #K	N_{KP}	N_{KC}	
T vs C #K+1		$N_{(K+1)C}$	$N_{(K+1)T}$

- SUP testing $H_0(\lambda=1)$ - using only AC data
- NI testing:
  - With GHC approach - using AC data and a fixed value δ
  - With X-trial comparison approach - using C-P and AC data

- With GHC approach - switching = simultaneous test ? same type I error rates
- With X-trial comparison approach – switching ? simultaneous test with fixed sample size (equality holds only asymptotically) ? type I error rates change

(Tsong & Zhang, 2005, BiomJ; Tsong & Zhang, 2005, BiomJ, to be submitted)

ii. Group Sequential Design

A. With GHC approach –

Application of group sequential designs has been well described in Wang et al (2001), Li and Tsong (2003), Shih et al (2004), and Tsong et al (2004).

B. With X-trial comparison approach

Consider the data used in the analysis:

Study	Placebo		Active Control		Test	
C vs P #1						
C vs P #2						
C vs P #.						
C vs P #K	$\mathbf{N_H}$	**+**	$\mathbf{N_H}$			$\mathbf{= 2N_H}$
T vs C			$\mathbf{N_{AC(1)}}$	**+**	$\mathbf{N_{AC(1)}}$	$\mathbf{= 2N_{AC(1)}}$
			$\mathbf{N_{AC(2)}}$	**+**	$\mathbf{N_{AC(2)}}$	$\mathbf{= 2N_{AC(2)}}$

Assume

- $X_T \sim N(T, \sigma_1)$, $X_C \sim N(C, \sigma_1)$, $X_{C(H)} \sim N(C_H, \sigma_2)$, $X_{P(H)} \sim N(P_H, \sigma_2)$.
- Sample sizes N_H, $N_{AC} = N_{AC(1)} + N_{AC(2)}$

- How to define information time τ?
  - τ = 0 before interim look ?
  - $\tau = (N_H + N_{AC(1)})/(N_{AC} + N_H)$ at $N_{AC(1)}$?
  - τ = 1 at final analysis
  - In practice, if $N_H >> N_{AC}$,
    and $0.5 << (N_H + N_{AC(1)})/(N_{AC} + N_H) \approx 1$
    $\Rightarrow$ limited usage of interim look.
- Test statistic

$$z(d)|_t = [(\hat{m}_T - \hat{m}_C)|_t - (l-1)(\hat{m}_{C|H} - \hat{m}_{P|H})] / \sqrt{e_{1|t}^2 + (l-1)^2 e_2^2}$$

is not stationary and does not convergent to a Brownian Motion.

- How to adjust for type I error rate?

- Can't not be planned with a sequential or adaptive design

iii. Consistency among centers

A. With GHC approach –

Homogeneity of treatment efficacy in active controlled trials using GHC has been studied by many statisticians

- Quan and Shih (2001); Wiens and Heyes (2003)

B. With X-trial comparison approach –

- With an incomplete unbalanced block design, test treatment and the placebo are not studied within the same center.
- Since the within center (μ_T - μ_C) is estimable, one can examine the homogeneity of treatment effect (μ_T - μ_C) among the centers when testing $H_0(\lambda=1)$.
- ($\mu_{C|H}$ - $\mu_{P|H}$) is estimated using data of historical trials, one can't estimate (μ_T - μ_C) - (λ - 1)($\mu_{C|H}$ - $\mu_{P|H}$) independently within each center
- Homogeneity of (μ_T - μ_C) - (λ - 1)($\mu_{C|H}$ - $\mu_{P|H}$) among the centers can't be tested for the non-inferiority null hypothesis H_0 (λ): (μ_T - μ_C) = (λ - 1)($\mu_{C|H}$ - $\mu_{P|H}$).

iv. Data Transformation & Change of Design

- If $X_{C|H} \sim N(\mu_{C|H}, \sigma_2)$ and $X_{P|H} \sim N(\mu_{P|H}, \sigma_2)$
  - If $X_C \sim N(\mu_C, \sigma_1)$ and $X_T \sim N(\mu_T, \sigma_1)$,
    use t-test or z approximation test
  - Otherwise ? Ward Statistic and approximate Z test
- If $X_{C(H)} \sim F(\mu_{C|H}, \sigma_2)$ and $X_{P(H)} \sim F(\mu_{P|H}, \sigma_2)$
  - With GHC approach
    - Data can be transformed
  - With X-trial comparison approach
    - Data can't be transformed
- Change of study design or analysis method from historical C vs. P trials
  - With GHC approach
    - Parallel arms to paired or crossover, ANOVA to ANCOVA, etc
  - With X-trial comparison approach
    - Not feasible

v. Data independence

- Dependence on the historical C vs. P trials –

A. With GHC approach

  - d is a fixed value determined with both data of historical trials and medical judgement
  - NI testing performed without involving historical data directly

B. With X-trial comparison approach

  - d (i.e. $(\lambda - 1)(\mu_{C|H} - \mu_{P|H})$) is a function of parameters to be estimated with historical trial data
  - NI testing performed with historical trial data involved directly

- Dependency of two NI trials

  (the regulatory requirement of at least 2 positive independent pivotal phase III clinical trials)

  A. With GHC approach –
     - 2 trials share a fixed d
  B. With X-trial comparison approach –
     - 2 trials share data of the same historical A vs. P trials

SUMMARY GHC vs. X-trial Comparison

	Historical Control	Cross-Study
Margin δ	fixed, pre-set	variable, in the study
Null hypothesis	$T - C \leq -\delta$	$T - C \leq (\lambda - 1)(C - P)_H$
NI ? SUP	α won't change	α will change
Homogeneity test	Same as ANOVA	Can't do
Group sequential	Regular	Adjust information time
/adaptive design		Needs new boundary
Transform Data	More complicated	Can't do
Design change	Possible	Can't do
Two phase III	Yes	Dependence

Thank you for your interest!!!

*** The views expressed in this paper are the presenter‘s professional opinions.**

They do not represent the official positions of the U. S. Food and Drug Administration

Interesting Example

Sponsor proposed NI test

H_0: $OR_{TP} \geq 1/\sqrt{OR_{CP|H}}$, i.e. $OR_{TC} \geq (OR_{CP|H})^{3/2}$??

- Estimate 95% CI of $\log OR_{CP|H}$, ($L\log OR_{CP|H}$, $U\log OR_{CP|H}$)
- $\delta = (1/2)\, U\log OR_{CP(H)}$ -- (historical control)
- Estimate 95% CI of $\log OR_{TP}$, ($L\log OR_{TP}$, $U\log OR_{TP}$) with

  $\log OR_{TP} = \log OR_{TC} + \log OR_{CP|H}$ --(cross-study)
- Show that $\min\{|L\log OR_{TP}|, |U\log OR_{TP}|\} > \delta$ -- (hybrid)
- Sponsor declare NI if $\log OR_{TP} < -\delta$
- It is equivalent to show $\exp(\log OR_{TP}) < \exp(-\delta)$