1

Technical Inference Problems in Non-Inferiority Trials

Nick Fieller
&
Nor Afzalina Azmee

Department of Probability & Statistics
University of Sheffield, UK

ISC9, 21st August 2008, Isfahan

2

Introduction

- Topics to be discussed:-
  - Clinical Trials
    - Superiority Trials
    - Non-inferiority Trials
    - 3-arm non-inferiority trials
      - Confidence Interval or Hypothesis Test??
  - Confidence Intervals
    - Construction by pivotal quantities
    - Intervals as roots of quadratic inequalities
    - Fieller's Theorem
  - Application to 3-arm non-inferiority trials
    - Strange problems
      - Exclusive and Imaginary Intervals

3

Clinical Trials

Three principle types of trial:-

- **Superiority trial:**
  - ***Is treatment A better than treatment B?***
- **Non-inferiority trial:**
  - ***Is treatment A no worse than B?***
- **Equivalence Trials:**
  - ***Are treatments A and B equivalent?***

4

Clinical Trials

- **Superiority Trials**
  - Is new treatment A ***better*** than treatment B?
    - (B may be a placebo or a standard treatment)
    - Allocate subjects randomly to A or B and measure response.
      - Do data provide sufficient evidence that response to A is higher than B?
    - Test H_0: $\mu_A = \mu_B$ *vs* H_1: $\mu_A \neq \mu_B$
    - If H_0 is rejected and $\bar{x}_A > \bar{x}_B$ then declare A superior to B
    - Calculate confidence interval for $\mu_A - \mu_B$

5

Clinical Trials

- **Non-inferiority Trials**
  - Is new treatment A ***no worse*** than B?
- Why is this of interest?
  - B may be an ***expensive standard*** treatment
  - Company may want to have its own drug to rival drug B owned by a competitor
  - B may be very effective so a superiority trial may need many subjects & be costly
  - **Showing non-inferiority may be sufficient to enter market**
    - showing superiority may be difficult & expensive

6

Clinical Trials

- **Non-inferiority Trials**
  - Is new treatment A ***no worse*** than B?
    - (B may be an ***expensive standard*** treatment)
    - **Naïve approach** would be allocate subjects randomly to A or B and measure response.
    - Test H_0: $\mu_A = \mu_B$ *vs* H_1: $\mu_A \neq \mu_B$
    - If H_0 is ***not*** rejected then declare A non-inferior to B

not sensible since

Hypothesis tests can not be used to prove a hypothesis true:-

Absence of evidence
IS NOT
Evidence of absence

(and we need a placebo control)

Clinical Trials

7

- Better is to show A is better than a predefined margin below B
  - Test H_0: $\mu_A = \mu_B - \delta$ *vs* H_1: $\mu_A > \mu_B - \delta$
  - δ chosen to be smallest *clinical relevant difference*
- If H_0 rejected then conclude that A is not *clinically relevantly different* from B
  - So **NOT Inferior**

Clinical Trials

8

- More usually work in terms of ***effect size***
  - e.g. *reduction in blood pressure from baseline*

so express relevant difference as a proportion of effect of B, μ_B

- Let $\delta = f\mu_B$ and test H_0: $\mu_A = \mu_B - f\mu_B$ *vs* H_1: $\mu_A > \mu_B - f\mu_B$

note test of ratio ↘ or equivalently test H_0: $\mu_A / \mu_B = \theta$ *vs* H_1: $\mu_A / \mu_B > \theta$ (where $\theta = 1 - f$)

(and we still need a placebo control)

Clinical Trials

9

- Non-inferiority with a Placebo Control
  - 3-arm non-inferiority trials (3 treatments)

Subjects randomised to one of:-

- **P Placebo** treatment
- **R Reference** (or standard treatment)
- **E Experimental** (or new treatment)
- Test proportional difference between E & P to R & P

- Test H_0: $\mu_E - \mu_P = \mu_R - \mu_P - f(\mu_R - \mu_P)$ *vs* H_1: $\mu_E - \mu_P > \mu_R - \mu_P - f(\mu_R - \mu_P)$
  - or H_0: $\mu_E - \mu_P = \theta(\mu_R - \mu_P)$ *vs* H_1: $\mu_E - \mu_P > \theta(\mu_R - \mu_P)$

(where $\theta = 1 - f$, typically θ taken in range 0.5 to 0.8)

Clinical Trials

10

- 3-arm non-inferiority trials
  - (3 treatments E, R, P)

Test H_0: $\mu_E - \mu_P = \theta(\mu_R - \mu_P)$ *vs* H_1: $\mu_E - \mu_P > \theta(\mu_R - \mu_P)$

- Tested by a t-statistic for any specific value θ
  - (we will take θ=0.8)
- Alternatively we can formulate as

Test

$$H_0 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} = \theta \text{ vs } H_1 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} > \theta$$

(with $\theta = 0.8$ say)

(**note ratio again**)

Clinical Trials

11

- 3-arm non-inferiority trials
  - (3 treatments E, R, P)

Test H_0: $\mu_E - \mu_P = \theta(\mu_R - \mu_P)$ *vs* H_1: $\mu_E - \mu_P > \theta(\mu_R - \mu_P)$

- Tested by a t-statistic for any specific value θ
  - (we take $\theta = 0.8$)

$$t_{obs} = \frac{\bar{x}_E - \theta\bar{x}_R - (1-\theta)\bar{x}_P}{s\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}}$$

and non-inferiority is established if $t_{obs} > t_{1-\alpha,r}$

(where s^2 is estimate of variance with r degrees of freedom)

Clinical Trials

12

$$t_{obs} = \frac{\bar{x}_E - \theta\bar{x}_R - (1-\theta)\bar{x}_P}{s\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}}$$

with $\theta = 0.8$ say:

(θ Is the minimum proportion of effect of E compared with R that is clinically irrelevant)

Non-inferiority established if $t_{obs} > t_{1-\alpha,r}$

- However, this does not convey **uncertainty** in exactly 'how non-inferior' E is than R
  - is E very close to R or near the clinically relevant limit?
- Better would be to calculate a confidence interval for θ

Clinical Trials

13

$$\theta = \frac{\mu_E - \mu_P}{\mu_R - \mu_P}$$

Need a confidence interval for θ, (L_α, U_α)

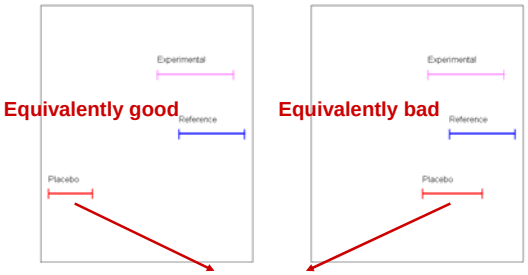
Interpretation based on lower limit of confidence interval L_α :

- if $L_\alpha > 0.8$ then non-inferiority of E to R is established
- if $L_\alpha > 1$ then E is superior to R
- If $L_\alpha < 0.8$ then E not non-inferior to R

♦ However, this does not answer the question of whether the new drug E is effective since we do not know whether the standard reference drug R is [still] effective:→ 1st compare with placebo

Clinical Trials

14



Equivalently good **Equivalently bad**

2 POSSIBILITIES

Clinical Trials

15

So, we first need a test of superiority of R over P:

♦ test H_0: $\mu_R = \mu_P$ *vs* H_1: $\mu_R \neq \mu_P$

- If H_0 is rejected and $\bar{x}_R > \bar{x}_P$ then declare R superior to Placebo and proceed to assess E for non-inferiority to R
  - If H_0 is not rejected then no sense in assessing E in relation to R since insufficient evidence that the Reference R is useful

♦ This is regarded as the "Gold Standard procedure"

- CHMP (2005)

Clinical Trials

16

■ The 3-arm non-inferiority trial:-

test H_0: $\mu_R = \mu_P$ *vs* H_1: $\mu_R \neq \mu_P$

reject H_0 — don't reject H_0

assess non-inferiority — **STOP**

by confidence interval

for $\theta = \frac{\mu_E - \mu_P}{\mu_R - \mu_P}$

Strictly a technical problem with conditional/multiple testing here but universally ignored by practitioners

Clinical Trials

17

■ It can be shown that

♦ if R is superior to P

and

♦ E is not inferior to R

then

♦ E is superior to P

- so ***do not need to test* E *vs* P**

– using size α tests and $(1-\alpha)$ intervals throughout

Confidence Intervals

18

Consider a parameter θ and data X

A $(1-\alpha)$ confidence interval for θ based on data X can be calculated by:

♦ "Set of all θ_0 ***not rejected*** by a size α test of H_0: $\theta = \theta_0$"

– (using data X)

♦ Inversion of a probability statement about a pivotal quantity

- "Pivotal quantity" or "pivot" is a function of X and θ whose distribution is independent of θ
- e.g. $Z=(X-\mu)/\sigma$ when $X \sim N(\mu, \sigma^2)$

Confidence Intervals

19

- "Pivot" is a function of X & θ whose distribution is independent of θ
  - sometimes inversion of probability statement involves extracting roots of a polynomial equation:
    - e.g. a Binomial parameter & use of a Normal approximation: If X~Bin(n, θ) then approx X~N(nθ, nθ(1– θ))

$$\text{so } \frac{X-n\theta}{\sqrt{n(1-\theta)}} \sim N(0,1), \quad \text{so } P\left[\left|\frac{X-n\theta}{\sqrt{n(1-\theta)}}\right| < z_{1-\alpha}\right] = 1-\alpha$$

pivot

Confidence Intervals

20

$$\text{so } \frac{X-n\theta}{\sqrt{n(1-\theta)}} \sim N(0,1), \quad \text{so } P\left[\left|\frac{X-n\theta}{\sqrt{n(1-\theta)}}\right| < z_{1-\alpha}\right] = 1-\alpha$$

- This can be rearranged to leave

$$P[\, Q(\theta) < 0\,] = 1 - \alpha$$

where $Q(\theta)$ is a ***quadratic*** function of θ with coefficients depending only on the data and positive coefficient of θ^2

- It can be shown that $Q(\theta)$ ***always*** has real roots θ_L and θ_U and so we have

$$P[\theta_L < \theta < \theta_U\,] = 1 - \alpha$$

so (θ_L, θ_U) is a $1 - \alpha$ confidence interval for θ

Confidence Intervals

21

- Confidence intervals for ratios can be calculated in a similar way
  - Find a pivot
  - Make a probability statement
  - Inversion of statement involves extracting roots of a *quadratic* in parameter

BUT

- Coefficient of quadratic term can be negative
- Roots of quadratic can be complex

➔ **Exclusive or imaginary confidence intervals**

Confidence Intervals

22

Confidence intervals for ratios:-

- Suppose
- $X_A \sim N(\mu_A, v_{11}\sigma^2)$ and $X_B \sim N(\mu_B, v_{22}\sigma^2)$ with correlation v_{12} and let $m = \mu_A / \mu_B$
  - then 'Fieller's Theorem' gives a confidence interval for m based on samples from A and B and an estimate of variance s^2 on r d.f.
    - (E.C. Fieller, 1932, 1944, **1954**)
  - the construction involves the pivotal quantity $t = (X_A - mX_B)/s$ which has [essentially] a Students t_r distribution

Confidence Intervals

23

Define : -

$$g = \frac{t_{1-\alpha,r}^2 s^2 v_{22}}{x_B}$$

And

$$E = v_{11} - 2\frac{x_A}{x_B}v_{12} + \left(\frac{x_A}{x_B}\right)^2 v_{22} - g\left(v_{11} - \frac{v_{12}^2}{v_{22}}\right)$$

[Note that if g < 1 then E > 0 since when

$$g = 1, \quad E = \left((v_{12} - x_A v_{22}/x_B\right)^2 / v_{22} \geq 0]$$

Confidence Intervals

24

The upper and lower limits are : -

$$m_U, m_L = \frac{1}{1-g}\left[\frac{x_A}{x_B} - g\frac{v_{12}}{v_{22}} \pm \frac{t_{1-\alpha,r}s}{x_B}\sqrt{E}\right]$$

But it is possible that g > 1 and that E < 0

-Exclusive
-Infinite

-Imaginary

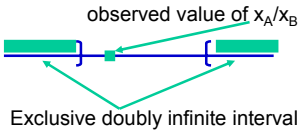
POTENTIAL PROBLEMS

Confidence Intervals

25

The upper and lower limits are : - $m_U, m_L = \frac{1}{1-g}\left[\frac{x_A}{x_B} - g\frac{v_{12}}{v_{22}} \pm \frac{t_{1-\alpha,r}s}{x_B}\sqrt{E}\right]$

If $g > 1$ then



$$g = \frac{t^2_{1-\alpha,r}s^2v_{22}}{x_B} > 1 \quad \text{so} \quad \frac{x_B}{s^2v_{22}} < t^2_{1-\alpha,r}$$

So a test of H_0: $\mu_B = 0$ would ***not*** be rejected at level α, i.e. not "$1-\alpha$ confident that $\mu_B \neq o$"

Confidence Intervals

26

- It is possible to find $\alpha_1 < \alpha_2 < \alpha_3$ so that $g < 1$, $g > 1$ and $E > 0$ and $g > 1$ and $E > 0$, i.e. for the same data we could have
  - $1 - \alpha_1$ confidence interval closed and inclusive
  - $1 - \alpha_2$ confidence interval open and exclusive
  - $1 - \alpha_3$ confidence interval imaginary
- e.g.
  95% interval ok
  97.5% infinite & exclusive
  99% imaginary

Confidence Intervals

27

- It is known that Fieller's confidence interval has the correct coverage probability
  - i.e. it has probability $1-\alpha$ of covering true value of ratio and so must be equivalent to "Set of all values of the ratio ***not rejected*** by a size α test"
- What is interpretation of the confidence interval when it is imaginary or exclusive?
- We can always do the hypothesis test since the coverage probability is correct
  an imaginary or exclusive interval means
  "don't reject"

Application to 3-arm non-inferiority trials

28

- In a 3-arm non-inferiority trial of E, R and P
  - **E**xperimental, **R**eference and **P**lacebo controls

we take A = E − R and B = R − P,

i.e. $\bar{x}_A = \bar{x}_E - \bar{x}_R$ and $\bar{x}_B = \bar{x}_R - \bar{x}_P$

and we take $m = \frac{\mu_E - \mu_P}{\mu_R - \mu_P}$

Recall we have a preliminary test of
H_0: $\mu_R = \mu_P$ *vs* H_1: $\mu_R \neq \mu_P$
and proceed only if H_0 is rejected

Application to 3-arm non-inferiority trials

29

- So in 3-arm non-inferiority trials a test of
  H_0: $\mu_E - \mu_P = \theta(\mu_R - \mu_P)$ *vs* H_1: $\mu_E - \mu_P > \theta(\mu_R - \mu_P)$
  using $t_{obs} = \frac{\bar{x}_E - \theta\bar{x}_R - (1-\theta)\bar{x}_P}{s\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}}$ is equivalent to
  checking if confidence interval for $\frac{\mu_E - \mu_P}{\mu_R - \mu_P}$
  excludes or includes θ
- Because we first test
  H_0: $\mu_R = \mu_P$ *vs* H_1: $\mu_R \neq \mu_P$
  we *should* not have problems
  of imaginary or infinite exclusive intervals

BUT

Application to 3-arm non-inferiority trials

30

- Preliminary test of
  H_0: $\mu_R = \mu_P$ *vs* H_1: $\mu_R \neq \mu_P$
  (1) must use same level α in test as in confidence interval
  (2) must use same estimate of variance s^2 in test as in confidence interval
- Potential problems if for example
  - (1) preliminary test is at level 10% and confidence interval is 95%
  - (2) data from only R and P used to test $\mu_R = \mu_P$ → MUST include data from E in test i.e. must use a pooled variance (E+R+P)

Application to 3-arm non-inferiority trials

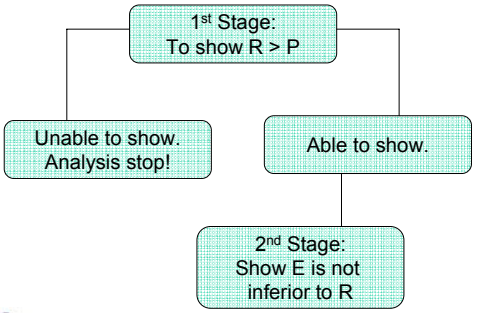
31

- Many possible reasons to use different levels for preliminary test of R *vs* P and for non-inferiority but this can lead to problems
- Also there can be good reasons for not including data from E in test of R *vs* P
  - Note we must also use *pooled variance* t-test
  - Note also that randomisation should be of ***all*** subjects to one of E, R or P

Application to 3-arm non-inferiority trials

32

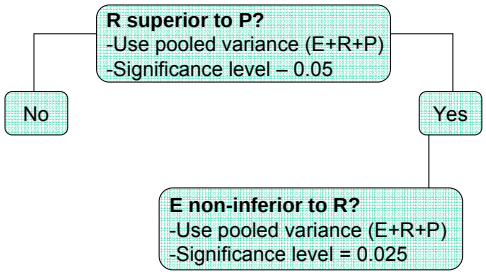
Examples



Application to 3-arm non-inferiority trials

33

Example 1



Application to 3-arm non-inferiority trials

34

Example 1

Data

	Experimental	Reference	Placebo
Mean	1.4	1.53	1.15
Standard Deviation	0.74	0.82	0.72
Sample size	25	25	25

Result

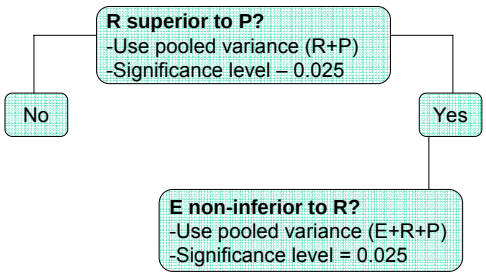
imaginary

Lower CI 1st Test	Upper CI 1st Test	Lower CI 2nd Test	Upper CI 2nd Test
0.02	0.74	IM	IM

Application to 3-arm non-inferiority trials

35

Example 2



Application to 3-arm non-inferiority trials

36

Example 2

Data

	Experimental	Reference	Placebo
Mean	1.81	1.01	0.55
Standard Deviation	1.18	0.77	0.79
Sample size	25	25	25

Result

exclusive & infinite

Lower CI 1st Test	Upper CI 1st Test	Lower CI 2nd Test	Upper CI 2nd Test
0.02	0.91	1.34	-16.17

37

Application to 3-arm non-inferiority trials

- Summary & Conclusions
  - Confidence intervals of ratio give a better understanding of level of non-inferiority
  - Preliminary test of R *vs* P should prevent 'peculiar intervals' but only under restrictive conditions
  - Since calculation of confidence interval is conditional on result of a preliminary test it does not have the declared coverage probability
  - Hypothesis tests are really 'not satisfactory' for non-inferiority problems — **what about a Bayesian approach?**

38

References

Pigeot, I., Schäfer, J., Röhmel, J. and Hauscke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm trial including a placebo. *Statistics in Medicine*, 22:883-899.

Koch, A. and Röhmel, J. (2004). Hypothesis testing in the "Gold Standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics*, 14: 315-325

Fieller, E.C. (1954). Some problems in interval estimation. *Jour. Roy. Statist. Soc.*, B, 16:175-185

http://nickfieller.staff.shef.ac.uk

nick.fieller@sheffield.ac.uk

39

40

41

42