

Issues in Non-Inferiority Trials

Richard Kay

Topics

- Introduction
- Confidence intervals and p-values
- Analysis sets
- Assay sensitivity
- Choice of delta
- Sample size calculations
- Switching between non-inferiority and superiority
- Interim analysis

Introduction

- Increasingly seeing equivalence and non-inferiority trials – why?
- Becoming ethically and practically more difficult to run placebo-controlled studies (Declaration of Helsinki)
- Clinical and commercial reasons; show advantages in for example tolerability/safety or convenience, yet no reduction in efficacy

Introduction

- Still widespread misunderstanding amongst our non-statistical colleagues
- Main issue; a non-significant p-value in a conventional superiority comparison cannot be used to establish 'similarity'

Japanese Statistics Guidelines

'Absence of significant differences does not assure statistical "equivalence"'

Example: 2NN Study

- van Leth, Phanuphak et al (2004) Lancet in a study of first-line antiretroviral therapy in HIV
- 4 treatment groups; main comparison was between nevirapine twice daily and efavirenz (plus stavudine and lamivudine) in terms of 'treatment failure' (a composite of virology, disease progression, therapy change)
- Primary objective was to establish the non-inferiority of nevirapine twice daily ($\Delta=10\%$)

Example: 2NN Study

- Confidence intervals for failure rates (E-N2) were

All data (-12.8%, 0.9%)

Only those starting med. (-14.6%, -0.8%)

Concurrently randomised (-11.9%, 3.4%)

- Non of these intervals are entirely above -10%; one interval also excludes zero

Example: 2NN Study

- Authors also quoted conventional p-values (0.091, 0.03, 0.276)
- Conclusion in the Summary
‘Antiviral therapy with nevirapine or efavirenz showed similar efficacy, so triple-drug regimens with either ... are valid for first-line treatment’

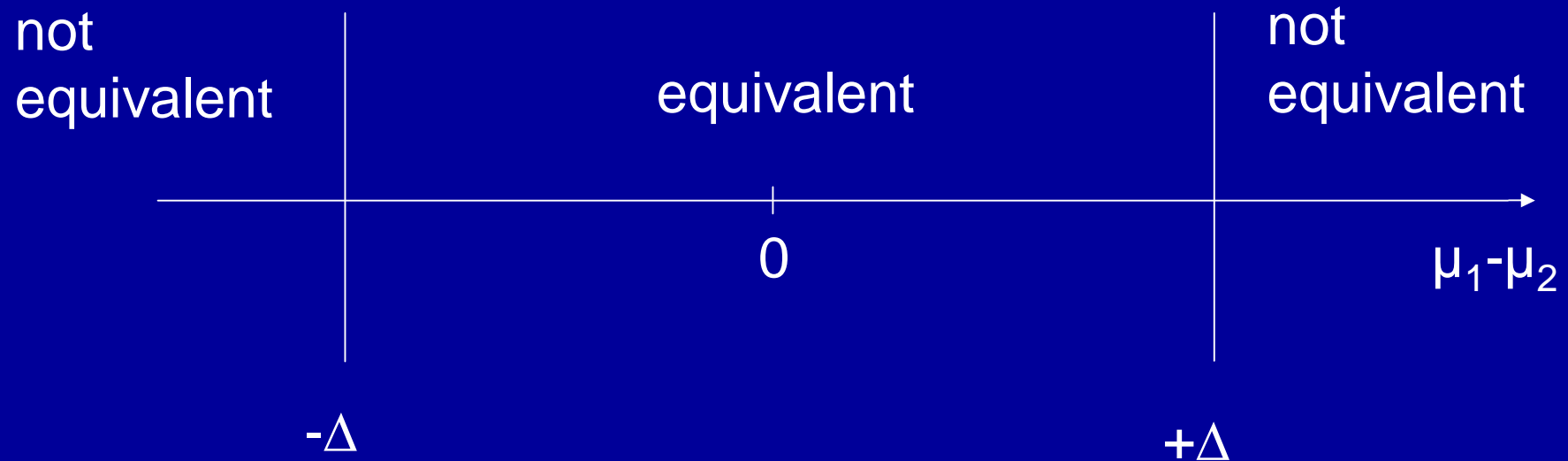
Regulatory Guidance

- ICH E9 'Note for Guidance on Statistical Principles for Clinical Trials', September 1998
- ICH E10 'Note for Guidance on Choice of Control Group', July 2000
- CPMP 'Note for Guidance on the Investigation of Bioavailability and Bioequivalence', July 2001
- CPMP 'Points to Consider on Switching between Superiority and Non-Inferiority', July 2000
- CHMP 'Guideline on the Choice of the Non-Inferiority Margin', July 2005

Confidence Intervals for Equivalence

- First step in establishing equivalence - define 'equivalence margins' ($\pm \Delta$)
- Having run the trial calculate 95% confidence intervals for the difference between the treatments (means or response rates)
- If confidence interval entirely within $\pm \Delta$ then equivalence is established

Confidence Intervals for Equivalence



p-values for Equivalence

- p-value counterpart based upon the following hypotheses

$$H_0: \mu_1 - \mu_2 \leq -\Delta \text{ or } \mu_1 - \mu_2 \geq +\Delta$$

$$H_1: -\Delta < \mu_1 - \mu_2 < +\Delta$$

- Reduces to the evaluation of 2, one-sided tests each at the 2½% level

$$H_{01}: \mu_1 - \mu_2 \leq -\Delta \text{ vs } H_{11}: \mu_1 - \mu_2 > -\Delta$$

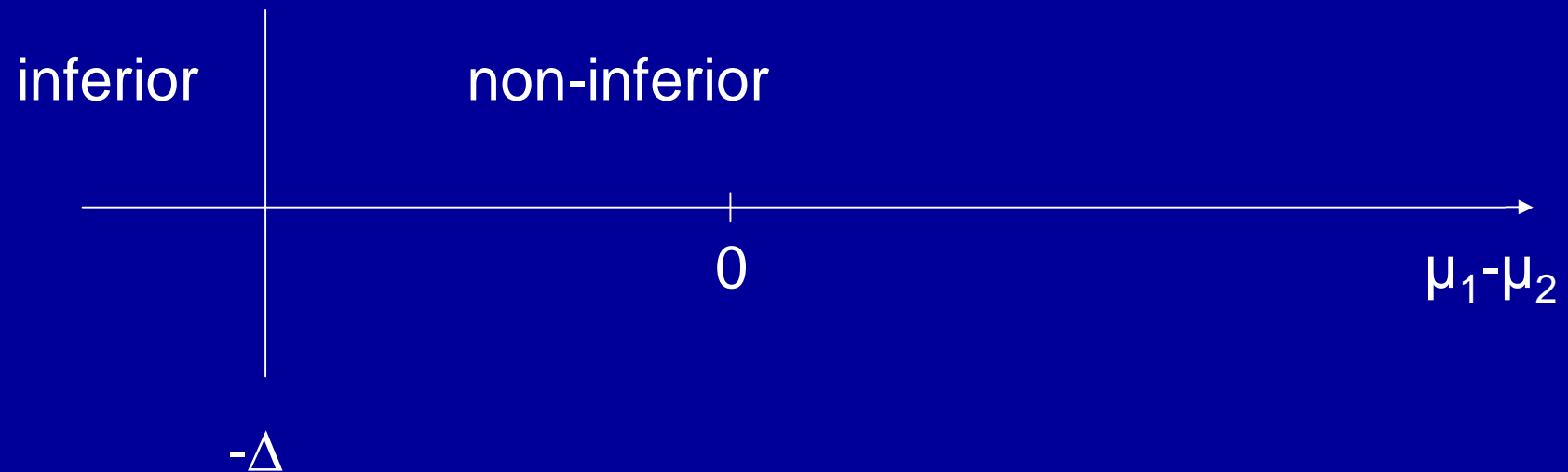
$$H_{02}: \mu_1 - \mu_2 \geq +\Delta \text{ vs } H_{12}: \mu_1 - \mu_2 < +\Delta$$

- Both p-values need to be < 0.025 for 'significant equivalence'

Confidence Intervals for Non-Inferiority

- Again need to define 'non-inferiority margin', $-\Delta$
- Run trial and calculate 95% confidence interval
- Non-inferiority established if confidence interval entirely within the non-inferiority region (for example, entirely above $-\Delta$)
- Only interested in lower end of the confidence interval – could calculate one-sided 97.5% confidence interval and just compare lower end with $-\Delta$

Confidence Intervals for Non-Inferiority



p-values for Non-Inferiority

- p-value counterpart is based upon a single one-sided test of

$$H_0: \mu_1 - \mu_2 \leq -\Delta \text{ vs } H_1: \mu_1 - \mu_2 > -\Delta$$

- Common question; why do we undertake superiority testing at the 5% level yet non-inferiority testing at the 2½% level?
- In superiority; never make a claim for treatment benefit if you are significantly worse than control!
- So in superiority we are in fact conducting one -sided tests at the 2½% level also

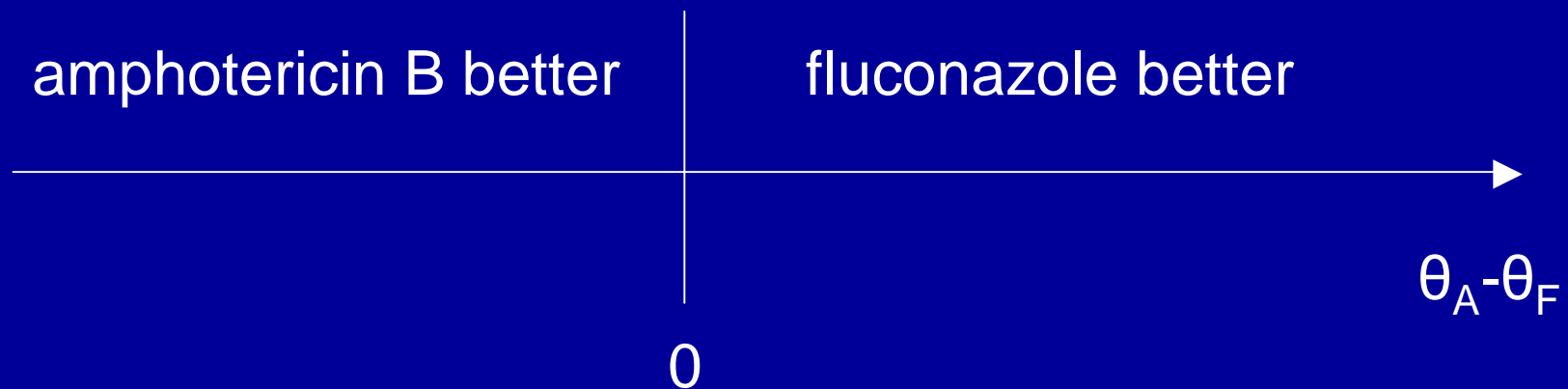
Example

Powderly, Saag et al (1992) NEJM

Primary goal was to determine whether fluconazole would be as effective (or nearly as effective) as amphotericin B in preventing the relapse of cryptococcal meningitis in patients with AIDS. It was thought that reduced toxicity and oral administration of fluconazole might give advantage over amphotericin B, even if fluconazole slightly less effective

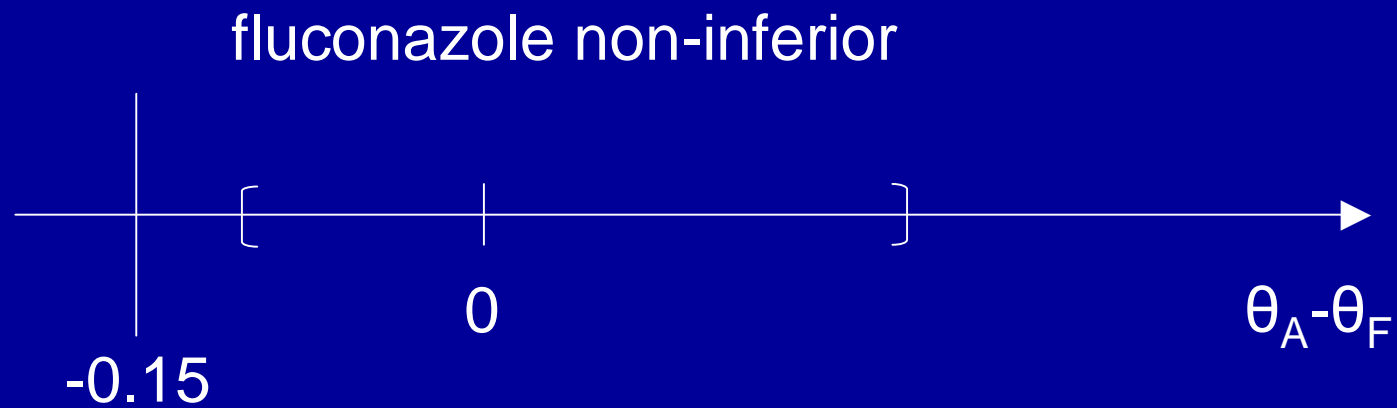
Example

- θ_A = relapse rate on amphotericin B
- θ_F = relapse rate on fluconazole



Example

- Δ chosen to be 15%



Samples for Analysis

- In superiority trials Full Analysis Set – primary analysis
- Approach conservative
- In equivalence/non-inferiority trials Full Analysis Set anti-conservative
- Both Full Analysis and Per-Protocol Sets should be used – lack of sensitivity to choice of analysis set required
- Per-protocol set still subject to bias

Assay Sensitivity

- Concern - insensitive trial unable to detect that treatments really are different
- Design elements for assay sensitivity; inclusion criteria, established endpoints, dose levels
- Internal evidence for assay sensitivity; use of placebo, treatment effects as anticipated
- Conduct of trial effective; few dropouts and missing data, good compliance

Assay Sensitivity

- Most effective way to demonstrate AS is to include a placebo and see active control separating from placebo by 'expected' amount
- Some therapeutic settings where non-inferiority trials (eg depression, allergic rhinitis) difficult unless placebo included
- Effective agents do not consistently demonstrate superiority in these areas

Choice of Delta

- For evaluation of means, Δ chosen to be a proportion of control treatment effect size
- Usually $\frac{1}{2}$ at most (preferably $\frac{1}{3}$) of established effect – considered also for proportions
- Thought now (CHMP Guidance document) to be too prescriptive; if reference treatment has ‘large’ benefit may want to preserve most of this
- Also trade off for efficacy depends on benefits in terms of for example safety

Choice of Delta

- Indirect demonstration that drug works - mixture of statistical and clinical reasoning

Statistical Reasoning

- Look historically at placebo-controlled trials for active comparator (meta-analysis or single large trial) – cure rate; 75% on active control, 55% on placebo - difference 20% with 95% CI (14%, 26%)
- Choose Δ around $\frac{1}{2}$ (14%) = 7% or $\frac{1}{3}$ (14%) = 4.7%
- Suppose based on this set $\Delta = 5\%$

Choice of Delta

- Conclusion of non-inferiority with $\Delta = 5\%$ would say that, at worst, new treatment has a cure rate 5% below that of active control – still comfortably better than placebo
- Method gives statistical confidence coming out of non-inferiority trial that test treatment efficacious

Clinical Reasoning

- Justify the choice from a clinical standpoint
 - difference of 5% is clinically irrelevant
 - or is a price worth paying for other benefits (eg tolerability)

Sample Size Calculations

- For superiority (one-sided test, level α and power $100(1-\beta)\%$) the standard sample size formula for a continuous endpoint is

$$n = \frac{2S^2}{d^2} (Z_{1-\alpha} + Z_{1-\beta})^2$$

where d is the clinically relevant difference to be detected

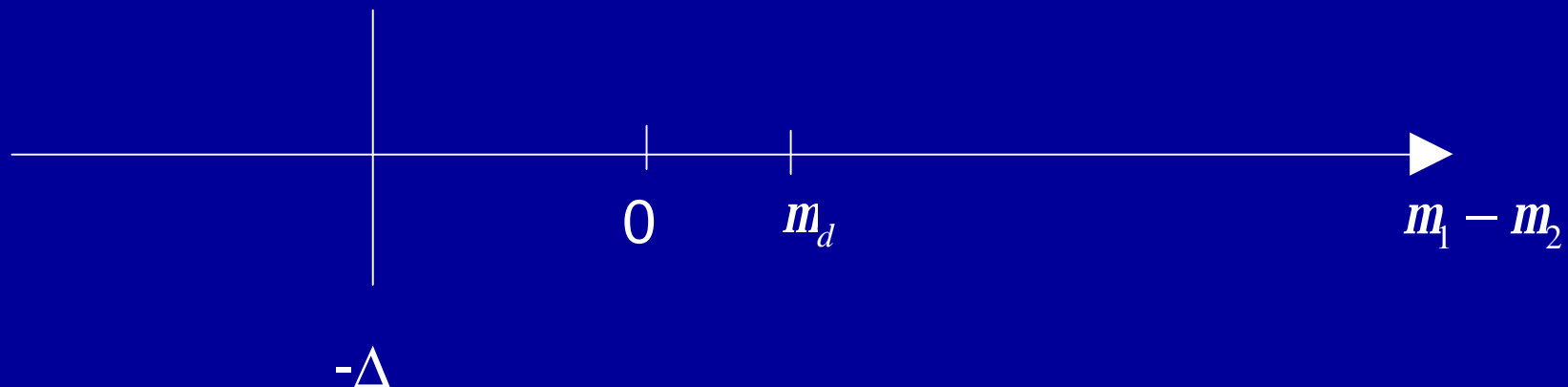
- For non-inferiority d is replaced by $-\Delta + m_d$ where $-\Delta$ is the non-inferiority margin and m_d is the expected difference in the treatment means

Sample Size Calculations

- Usually the non-inferiority margin will be somewhat less than the clinically relevant difference to be detected in a superiority setting
- As a result non-inferiority trials tend to have larger sample sizes under the assumption that the two treatment means are identical
- For example under these conditions if $\Delta = d/2$ then the non-inferiority trial will be 4 times larger

Sample Size Calculations

- If however there is a belief that the new treatment improves the mean response then the required sample size can be somewhat smaller

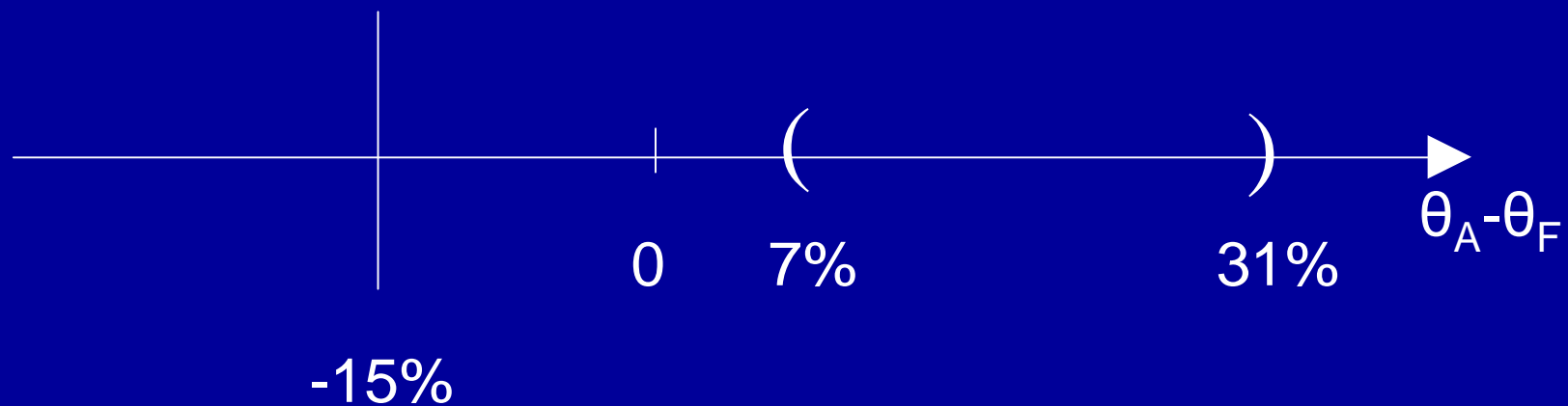


Switching from Non-Inferiority to Superiority

- If 95% confidence interval not only entirely to the right of $-\Delta$ but also to the right of 0, evidence that new treatment superior
- Can claim superiority; calculate conventional p-value and 95% confidence interval gives information on clinical benefit
- No multiplicity concerns

Example

- In amphotericin B vs fluconazole example 95% confidence interval for $\theta_A - \theta_F$ was (7%, 31%)



Example

Powderly, Saag et al (1992) NEJM concluded superiority:

'These data allow us to conclude that fluconazole was at least as effective as weekly amphotericin B ... Indeed the 19% difference in the probability of being relapse-free at one year (95% confidence interval 7% to 31%) suggests that fluconazole was more effective than amphotericin B in preventing a relapse of cryptoccal disease in this population of patients'

Switching from Superiority to Non-Inferiority

- If 95% confidence interval fails to demonstrate superiority may be interest in lesser claim of non-inferiority
- Generally not possible to switch unless non-inferiority margin pre-specified
- May be prudent to pre-specify such a margin if lesser conclusion of value
- Need also to investigate issues with analysis sets and design issues with assay sensitivity

Interim Analysis

- Can build stopping rules for non-inferiority by using adjusted confidence intervals
- O'Brien-Fleming with 2 interims, would split the type I error; 0.0006, 0.0142, 0.0450
- Use respectively 99.94%, 98.58%, 95.50% two-sided confidence intervals to give overall type I error rate of 2.5%

Interim Analysis

- Does it make sense however to consider early stopping for non-inferiority?
- Continuing would lead to narrower confidence intervals; claiming a tighter Δ , establishing superiority
- Nonetheless under some circumstances could be of value
- Example: *Bernard et al (2002) Oral pristinamycin versus standard penicillin regimen to treat erysipelas in adults: randomised, non-inferiority, open trial. BMJ*

QUESTIONS