Methodological aspects of non-inferiority and equivalence trials

Department of Clinical Epidemiology
Leiden University Medical Center
Capita Selecta 09-12-2014
Why RCTs

• Early 1900: Evidence based on medical reports or case series
  • Example penicillin

• Cohorts of patients given the same treatment
  • Natural course of disease
  • Extraneous effects (e.g. lifestyle changes, placebo effect)
  • Observer bias
Why RCTs

• Early 1900: Evidence based on medical reports or case series
  • Example penicillin

• Cohorts of patients given the same treatment
  • Natural course of disease
  • Extraneous effects (e.g. lifestyle changes, placebo effect)
  • Observer bias

→ Need for control group
Why RCTs

• Early 1900: Evidence based on medical reports or case series
  • Example penicillin

• Cohorts of patients given the same treatment
  • Natural course of disease
  • Extraneous effects (e.g. lifestyle changes, placebo effect)
  • Observer bias
→ Need for control group

• 1960s: Randomisation
  • Comparable prognosis of groups

• Reliable estimate of treatment effect
**RCTs**

- Superiority of a new treatment vs old treatment, no treatment or placebo

- Hypothesis testing
  - $H_0$: New treatment and old treatment (or placebo) are equally effective (on average)
  - $H_1$: New treatment is better than old treatment (or placebo) (on average)

- Outcomes
  - Reject $H_0$
  - Do not Reject $H_0$
RCTs

• Superiority of a new treatment vs old treatment, no treatment or placebo

• Hypothesis testing
  • $H_0$: New treatment and old treatment (or placebo) are equally effective (on average)
  • $H_1$: New treatment is better than old treatment (or placebo) (on average)

• Outcomes
  • Reject $H_0$
  • Do not Reject $H_0$

→ Failure to reject $H_0$ does not mean $H_0$ (equivalence) is true
Equivalence and non-inferiority trials

• Nowadays, many established treatments

• New treatment not always more effective, but other advantages
  • Less toxicity / side effects
  • Easier to use
  • Cheaper

• You want to know if $H_0$ is true (equal effective), however this cannot be proven with a superiority trial
Equivalence and non-inferiority trials

- **Objective:**
  - Evaluate the efficacy of new treatments against active controls

- **Equivalence trials**
  - New therapy is not worse and not better than existing therapy

- **Non-inferiority trial**
  - New therapy is not worse than existing therapy

1) Soonawala D – NTvG 2012
Methodological issues

- Why equivalence or non-inferiority trials
- Control group
- Sample size
- Non-inferiority / equivalence margin
- Hypothesis
- Analysis and outcome
Why equivalence or non-inferiority trials

• Treatment has other advantages
  • Less toxicity / less side effects
  • Easier to use
  • Cheaper

• Lower dosing regimens

• Other mechanism of treatment

• Placebo as effective as existing (non-evidence based) treatment

1) ICH E9 guideline; ICH E10 guideline; Soonawala D – NtvG 2012; Christensen E – J Hepatol 2007
Why equivalence or non-inferiority trials

- Treatment has other advantages
  - Less toxicity / less side effects
  - Easier to use
  - Cheaper

- Lower dosing regimens

- Other mechanism of treatment

- Placebo as effective as existing (non-evidence based) treatment

→ Note: Are claimed advantages proven by data?

1) ICH E9 guideline; ICH E10 guideline; Soonawala D – NtvG 2012; Christensen E – J Hepatol 2007
Why equivalence or non-inferiority trials

• Treatment has other advantages
  • Less toxicity / less side effects
  • Easier to use
  • Cheaper

• Lower dosing regimens

• Other mechanism of treatment
  • Placebo as effective as existing (non-evidence based) treatment

→ Note: Are claimed advantages proven by data?

1) ICH E9 guideline; ICH E10 guideline; Soonawala D – NtvG 2012; Christensen E – J Hepatol 2007
Control group

• Active control
  • Widely used and accepted therapy for the indication under study

• Proven effective in superiority trials
  • Meta-analysis
  • Assess the possibility of selection bias / publication bias to prove efficacy
  • Constancy assumption

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Control group

- Active control
  - Widely used and accepted therapy for the indication under study
- Proven effective in superiority trials
  - Meta-analysis
  - Assess the possibility of selection bias / publication bias to prove efficacy
  - Constancy assumption

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Control group

• Active control
  • Widely used and accepted therapy for the indication under study

• Proven effective in superiority trials
  • Meta-analysis
  • Assess the possibility of selection bias / publication bias to prove efficacy
  • Constancy assumption

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Control group

- Active control
  - Widely used and accepted therapy for the indication under study

- Proven effective in superiority trials
  - Meta-analysis
  - Assess the possibility of selection bias / publication bias to prove efficacy
  - Constancy assumption
  - Addition of placebo arm if superiority is not well proven

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Control group

• Active control
  • Widely used and accepted therapy for the indication under study

• Proven effective in superiority trials
  • Meta-analysis
  • Assess the possibility of selection bias / publication bias to prove efficacy
  • Constancy assumption
  • Addition of placebo arm if superiority is not well proven

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Control group

• Copying design features of original trial (e.g. eligibility criteria)
  • Take into account advances in medical and statistical practice

• Adequate dosing regimen and mode of administration

1) ICH E9 guideline; ICH E10 guideline; EMEA guideline – statist med 2006; D’Agostino RB – Statis Med 2003; CONSORT statement – JAMA 2006
Sample size

• Sample size in superiority trial:

\[ N = \frac{(Z_{2\alpha} + Z_\beta)^2 S^2}{\Delta^2} \]

N: Number of patients per treatment arm
\(\alpha\): type 1 error risk (usually 0.05); \(2\alpha=2\)-sided type 1 error risk
\(Z_{2\alpha}\): standardized normal deviates corresponding to the levels of the defined values of \(2\alpha\)
  (for \(2\alpha=0.05 \rightarrow 1.96\))
\(\beta\): Pre-defined value of type II error risk (usually 0.1 or 0.2); Power is 1- \(\beta\)
\(Z_\beta\): standardized normal deviates corresponding to the levels of the defined values of \(\beta\)
  (for \(\beta = 0.2 \rightarrow 0.84\))
\(S\): standard deviation \((S^2 = \text{variance})\)
\(\Delta\): Least relevant clinical difference between new treatment and old treatment

1) Christensen – J Hepatol 2007
Sample size

• Sample size in equivalence and non-inferiority trials:

\[
N = \frac{(Z_{2\alpha} + Z_{\beta})^2 S^2}{\Delta^2}
\]

• True equivalence means dividing by \( \Delta = 0 \) (impossible)
• \( \Delta = 0.00001 \) \( \Rightarrow \) unrealistic large sample size

• Compromise \( \Rightarrow \) Equivalence or non-inferiority margin

1) Christensen – J Hepatol 2007
Sample size

- Equivalence margin
  - Predefined margin for which we accept equivalence

1) Christensen – J Hepatol 2007
Sample size

• Equivalence margin
  • Predefined margin for which we accept equivalence
  • Aim to determine if effect of new treatment lies within this interval

• For example $\Delta = 0.1$

1) Christensen – J Hepatol 2007
Sample size

• Equivalence margin
  • Predefined margin for which we accept equivalence
  • Aim to determine if effect of new treatment lies within this interval

• For example $\Delta = 0.1$

• Formula can be used

$N = \frac{(Z_{2\alpha} + Z_\beta)^2 S^2}{\Delta^2}$

1) Christensen – J Hepatol 2007
Sample size

- Non-inferiority
  - Predefined margin for which we accept non-inferiority
  - Aim to determine if the effect of new treatment lies above this margin

- For example $\Delta = 0.1$

1) Christensen – J Hepatol 2007
Sample size

- Non-inferiority
  - Predefined margin for which we accept non-inferiority
  - Aim to determine if the effect of new treatment lies above this margin

- For example $\Delta = 0.1$

- One sided:

$$N = \frac{(Z_\alpha + Z_\beta)^2 S^2}{\Delta^2}$$

1) Christensen – J Hepatol 2007
Defining the margin

• $\Delta$ too wide? Accepting ‘non-inferiority’ while treatment is inferior

• $\Delta$ too narrow? Unrealistic large sample size

• Statistical reasoning and clinical judgment

1) ICH E9 guideline; EMEA guideline – statist med 2006
Defining the margin

• $\Delta$ too wide? Accepting ‘non-inferiority’ while treatment is inferior

• $\Delta$ too narrow? Unrealistic large sample size

• Statistical reasoning and clinical judgment

• Rules of thumb
  • $\Delta$ smaller than smallest clinically meaningful difference
  • $\Delta$ half the value of the value used in superiority trial
  • Superiority to placebo should remain

• Note: Using a $\Delta$ smaller than would be used in a superiority trial usually leads to larger sample sizes of equivalence and non-inferiority trials

1) ICH E9 guideline; EMEA guideline – statist med 2006
Defining the margin

- Historical data

Defining the margin

- Historical data

- Select $\Delta$ based on clinical relevance

Defining the margin

- Historical data

- Select \( \Delta \) based on clinical relevance

- Note: by using a proportion of the difference between \( C \) and \( P \) \([x(C – P)]\) the \( \Delta \) becomes smaller if the difference between \( C \) and \( P \) is smaller

- Putative placebo comparison

Defining the margin

• Note: the non-inferiority or equivalence margin needs to be predefined and mentioned in the study protocol

• Note: not needed to report on clinical trials.gov?

Hypothesis

• Superiority trials
  • $H_0$: Treatments equally effective (on average)
  • $H_1$: New treatment better (on average)

• Equivalence or non–inferiority trials
  • Objective: to prove new treatment is statistically (and clinically) equal or non-inferior to active control

Hypothesis

• Superiority trials
  • $H_0$: Treatments equally effective (on average)
  • $H_1$: New treatment better (on average)

• Equivalence or non – inferiority trials
  • Objective: to prove new treatment is statistically (and clinically) equal or non-inferior to active control

• Reversal of $H_0$ and $H_1$
  • $H_0$: Control treatment is better than New treatment (on average)
  • $H_1$: New treatment and control treatment are equally effective (on average)

Hypothesis

• True equality can not be proven and equality margin needs to be incorporated
  • $H_0$: Effect of control treatment minus effect of new treatment is equal to or larger than the pre specified margin
  • $H_1$: Effect of control treatment and effect of new treatment is smaller than the pre specified margin

• In formula:
  • $H_0$: $C - N \geq \Delta$
  • $H_1$: $C - N < \Delta$

• (note: can also be defined in terms of means, proportions, ratios successes and so on)

1) D'Agostino RB – Statis Med 2003
Analysis

• Dilution of true differences between treatments
  • Poor adherence
  • Dropouts
  • Crossovers
→ Erroneous accepting non-inferiority

Analysis

• Dilution of true differences between treatments
  • Poor adherence
  • Dropouts
  • Crossovers
  → Erroneous accepting non-inferiority

• Intention-to-treat analysis
  • Recommend for superiority trials
  • In general results in smaller observed differences between treatments
  → Erroneous accepting non-inferiority

• Per-protocol analysis
  • In general less patients
  • Wider confidence intervals
  • Less likely to erroneous accepting non-inferiority
  • Preferred over intention-to-treat analysis

Note: because per-protocol analysis is preferred, dropouts need to be accounted for in sample size calculation

Note: best is per-protocol analysis and intention-to-treat analysis with same results

• In case of true equality
  • 50% positive results and 50% negative results (regardless of sample size)
  • Lower limit of confidence interval would move closer to zero with increasing sample size

• Interpretation of outcome
  • Confidence interval (predominantly lower boundary)
  • Point estimate

1) EMEA guideline – statist med 2006
Outcomes

1) CONSORT statement – JAMA 2006
Outcomes

1) CONSORT statement – JAMA 2006
Outcomes

1) CONSORT statement – JAMA 2006
Outcomes

1) CONSORT statement – JAMA 2006
Outcomes

1) CONSORT statement – JAMA 2006
Claiming superiority

• Can be done when non-inferiority is evident
• Preferably defined a priori
• Using intention-to-treat analysis

1) CONSORT statement – JAMA 2006
Concluding remarks

• Non-inferiority trials or equivalence trials
  • Evaluate the efficacy of new treatments against active controls
  • New treatments must have other advantages

• Treatment effect of active controls needs to be well established

• True equivalence cannot be proven $\rightarrow$ a predefined non-inferiority or equivalence margin is used ($\Delta$)
  • Defined using clinical and statistical reasoning
    • ‘Non-inferior’ to active control
    • Superior to placebo
  • Must be conservative

• Reversal of $H_0$ and $H_1$ compared to superiority trials

• Per-protocol analysis is most conservative
Defining the margin

• Biocreep