Statistical Issues:
Non-inferiority design;
Phase III endpoints

Andrew Nunn
MRC Clinical Trials Unit, London
Background

- Ever since the mid-60s there have been TB regimens which are highly effective under controlled trial conditions.

- It’s next to impossible to improve on 97% efficacy and even if we could we would need a very large trial, but

- We could improve by making regimens simpler, shorter or safer.
Objective of an NI trial

• In contrast to a superiority trial, in which our aim is to show the new treatment to be more effective in a non-inferiority trial the objective is to show that the new treatment is as good as the control.

• It is impossible to show that two treatments are equally effective, only that they differ by no more than a specified amount – this is the margin of non-inferiority.
A significant landmark

• The 1st East African short course study was probably the second most important TB trial after the streptomycin trial.

• Not described as a non-inferiority trial but ...

• In the per protocol analysis the relapse rate of the 6m rifampicin based regimen was 0.4% less than that of the standard 18 month regimen with a 95% confidence interval of -3.5 to +4.2%
The margin of non-inferiority

- If the investigators had required a worst case scenario whereby the difference from the control arm was no more than 5% then a lower limit of 3.4% would have meant they could have declared the regimen to be non-inferior to the control.

- The **margin of non-inferiority** is the lower limit of the confidence interval for the difference from the control regimen (a worse case scenario – *not* the point estimate for the difference)
Difference from control regimen
(1st East African short course study)

-35%  -25%  -15%  -5%  0  5%

S streptomycin, H isoniazid, R rifampicin, Z pyrazinamide
T thiacetazone;  Control 2STH/16TH

margin of non-inferiority (δ)
Choosing $\delta$

- Two important assumptions made when planning a Phase III non-inferiority trial are:
  - the true value of the proportion with a favourable response in the intervention arm compared with that in the control
  - the acceptable margin of non-inferiority

- Most studies assume an equality of response between the two study arms.
Justifying $\delta$

• $\delta$ needs to be justified on both statistical and clinical grounds.
Statistical justification

• When assessing whether the substitution of a new drug for an existing drug can lead to shorter treatment duration we need to know whether the same result could have been achieved by simply shortening the original regimen.

• In REMox we are studying the effect of substituting Moxifloxacin for Ethambutol and reducing the total duration of treatment.

• But do we know what results we would get if we shortened the control regimen to 4 months?
The role of ethambutol

- 2EHRZ/2HR has never been studied.

- However we do have good data on the poor response of 2SHRZ/2HR and it is generally accepted that:
  - a) ethambutol has no apparent sterilizing activity and contributes little to the intensive phase of short course regimens in patients with fully sensitive organisms
  - b) ethambutol substituted for streptomycin in the 2SHRZ/2HR regimen is unlikely to have resulted in a better response.

S streptomycin, H isoniazid, R rifampicin, E ethambutol, Z pyrazinamide
## 4-month regimens in East Africa

<table>
<thead>
<tr>
<th>Regimen</th>
<th>N</th>
<th>Relapses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SHRZ/2HRZ</td>
<td>104</td>
<td>17</td>
<td>16%</td>
</tr>
<tr>
<td>2SHRZ/2HR</td>
<td>104</td>
<td>11</td>
<td>11%</td>
</tr>
<tr>
<td>2SHRZ/2HZ</td>
<td>98</td>
<td>32</td>
<td>32%</td>
</tr>
<tr>
<td>2SHRZ/2H</td>
<td>105</td>
<td>30</td>
<td>30%</td>
</tr>
<tr>
<td>2HRZ/2H</td>
<td>100</td>
<td>40</td>
<td>40%</td>
</tr>
</tbody>
</table>

S Streptomycin, H isoniazid, R rifampicin, Z pyrazinamide

4th EA/MRC SC study
The table below shows the results of a Singapore short course chemotherapy study, comparing different regimens for the treatment of tuberculosis (TB). The regimens are coded as follows: S for Streptomycin, H for isoniazid, R for rifampicin, and Z for pyrazinamide.

<table>
<thead>
<tr>
<th>Regimen</th>
<th>N</th>
<th>Relapses</th>
<th>Relapse Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SHRZ /2HRZ</td>
<td>79</td>
<td>9</td>
<td>11%</td>
</tr>
<tr>
<td>2SHRZ/2HR</td>
<td>77</td>
<td>6</td>
<td>8%</td>
</tr>
<tr>
<td>2SHRZ/4HRZ</td>
<td>78</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>2SHRZ/4HR</td>
<td>80</td>
<td>2</td>
<td>2%</td>
</tr>
</tbody>
</table>

S Streptomycin, H isoniazid, R rifampicin, Z pyrazinamide

Singapore short course chemotherapy study
• Results from studies of the 2SHRZ/4HR and 2SHRZ/2HR(Z) regimen demonstrate the extent of inferiority of the 4 month regimen when compared with the same treatment given for 6 months

<table>
<thead>
<tr>
<th>Location</th>
<th>4 month</th>
<th>6 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Africa</td>
<td>28/208 (13%)</td>
<td>8/215 (3%) non-concurrent</td>
</tr>
<tr>
<td>Singapore</td>
<td>15/156 (10%)</td>
<td>2/158 (1%)</td>
</tr>
</tbody>
</table>

• The difference in relapse rates between the 6-month and 4-month regimens is ~9-10%.
Justifying δ

- In the opinion of the African clinical investigators the benefits of reducing treatment duration to 4 months would have considerable advantages which outweigh a possible increase in relapse rate of up to 6%.
Endpoints
Relapse v reinfection

- In the trials conducted in the 1970s and ‘80s so-called relapse rates were actually recurrence rates since relapse was not distinguished from re-infection.

- If that had been possible relapse rates could well have been lower.

- In the era of HIV/TB coinfection it is particularly important to distinguish re-infections from relapses.
Failures during treatment

• Although failure during treatment was not included in the primary assessments it was invariably less than 1% except in the presence of initial drug resistance.

• This was so even in the case of 4 month regimens.

• In the 4th East African short course study, although relapse rates were high, ranging from 11%-40% only 4 of 555 patients were considered to be treatment failures.
Defining response

• Historically trials in TB were been analysed on per protocol populations resulting in low relapse rates of ~3-5%.

• An NI trial requires a per protocol analysis but also an ITT analysis.

• How should we define the response of patients who do not complete treatment, or interrupt or change treatment for drug toxicity?
Defining failure

• To classify a patient who did not complete treatment as unfavourable may be too conservative in view of what we know of the results from 4 or even 3 month regimens.

• Unfortunately we cannot accurately predict who will relapse – otherwise we could already shorten treatment for large numbers of patients.

• Deaths during treatment also present a problem as we cannot always be sure whether or not TB was a contributory cause.
Defining failure

- In REMox we have chosen a conservative approach to defining failure.
- This will include:
  - All deaths during treatment
  - All patients lost sight of during treatment
  - Patients returning after default requiring a restart of their treatment.
FDA requirements

• In a communication received last week from the FDA even stricter criteria are proposed:
  – Deaths *from whatever cause* during the 18 months of treatment and follow-up are to be classified as unfavourable.
  – Patients not assessable at 18 months are to be excluded from analysis.

• The FDA have also asked for a smaller delta for the ITT analysis.
But …

• In a population with a high proportion of HIV-infected patients a substantial death rate is inevitable.

• Retaining cured patients for 18 months will prove challenging.

• Unclear why a smaller delta is being asked for in the ITT analysis.
Implications for power
Classifying the un-assessable

- If all non-TB deaths and defaulters during treatment are classified as being unfavourable this is likely to result in considerably higher failure/relapse rates than the rate seen in past trials – the rate could easily be doubled.

- The implications for power are serious!
N per arm for 2 arm trial with varying unfavourable response rates and delta

![Graph showing N per arm for 2 arm trial with varying unfavourable response rates and delta. The graph includes lines for different delta values: delta = 4%, delta = 5%, and delta = 6%. The x-axis represents the unfavourable status (%) ranging from 4 to 12, and the y-axis represents N per arm ranging from 0 to 1600. The graph shows how N per arm increases with an increase in the percentage of unfavourable status and for different delta values.](image-url)
Constraints on $\delta$

- If the intervention is *expected to be inferior* to the control regimen this will puts limit on how small the margin of inferiority, $\delta$, can be.

- Retaining a small value of $\delta$ will mean increasing sample size.

- e.g. if the intervention is 2% less effective than the control could increase the number needed per arm by 2.5 to 3 fold.
N per arm for 2 arm trial with differences between unfavourable response rates between the control and intervention (delta = 6%)
The analysis
The analysis

- Superiority trials are analysed by ITT because it is the most conservative and least likely to be biased.

- ITT (intention to treat) principle ignores the fact that patients may not always receive their allocated treatment.
ITT analysis in an NI trial

• ITT analysis of non-inferiority trials is not conservative since the inclusion of patients who violate the protocol will tend to minimise differences between study arms thereby increasing the possibility of declaring non-inferiority.
Per-protocol also biased!

- Per-protocol analyses, in which only those adherent to the protocol are included, are also biased since not all randomised patients are included and although one might expect such an analysis to remove unwanted noise it also has the potential to wrongly conclude there is no difference when a difference exists.
Which analysis?

- The CPMP state that ‘similar conclusions from both an ITT and per protocol analysis are required’ to declare non-inferiority.

- CPMP also states that the primary analysis should be per protocol, ‘since it is most sensitive for the detection of any real difference’.

CPMP = Committee on Proprietary Medical Products
Difference from control regimen
(1st East African short course study)

(a) Per protocol analysis
-35%  -25%  -15%  -5%  0  5%  15%
6SHR  6SHZ  6SHT  6SHT  6SH
S = streptomycin, R = rifampicin, H = isoniazid, T = thiacetazone
Control = 2STH/16TH

(b) Modified ITT analysis
6SHR  6SHZ  6SHT  6SHT  6SH

MRC Clinical Trials Unit
Analysis scenarios

As well as performing a per protocol and ITT analysis

- Sensitivity analyses also enable alternative scenarios to be evaluated.
- This should reduce the possibility of falsely declaring a regimen to be non-inferior.
Sub-group analyses

• Important differences may occur in the response to treatment between HIV-infected and HIV-uninfected patients.

• These sub-groups should be analysed separately even though the comparisons will have limited power.

• Other sub-groups that should be analysed are patients with fully sensitive strains at enrolment and those with isoniazid-resistant strains.
And ....

• We should also compare
  – adverse events - these should also be non-inferior by an agreed criteria.
  – centre differences

  – reasons for exclusions from the per-protocol analysis, these are particularly important as they may reflect important differences between the management of the patients in the different treatment arms
A different approach

• Whereas in a superiority trial we must beware over analysing the data in an attempt to find a beneficial effect,

• In an NI study we should go out of our way to find differences in order not to wrongly ascribe ‘no difference’.

• The conclusion will be most convincing when all the evidence points in the same direction.
Interim analyses - do we need them?

- In a superiority trial we perform an interim analysis to provide the opportunity to close the study early if a substantial benefit has been demonstrated.
- Apart from adverse effects, and the possibility that the new regimen is performing particularly badly there is much less justification for interim analysis in an NI trial.
- If the intervention is doing well it is important to continue so as to better estimate its effectiveness and to possibly demonstrate its superiority to the control.
Possible outcomes

-35%    -25%    -15%    -5%     0      5%

Non-inferiority
Superiority
Inferiority
No conclusion

margin of non-inferiority (δ)
Criticism of the NI design

- In October 2007 an article in the Lancet expressed the view that all NI trials are unethical and should be banned!

- A badly designed NI could be unethical but the same is also true of a badly designed superiority trial.

- Although a shortened TB regimen may have a slightly increased failure/relapse rate it offers the prospect of less exposure to potentially toxic drugs and less interaction with the health services.

- To deny that possibility would be unethical!
Conduct of NI trials

• NI trials should be conducted with a high degree of rigour to avoid the possibility of falsely declaring non-inferiority.

– Only patients likely to comply should be enrolled.

– Losses should be kept to a minimum.

– As far as possible the same protocol and laboratory methods should be used as in the studies that demonstrated the effectiveness of the control regimen.

– If possible the trial should be conducted double blind.
In summary

• To license a new regimen it will be necessary to prove it is of comparable efficacy to the standard chemotherapy.

• Choice of the margin of non-inferiority will be critical

• Careful consideration needs to be given to the definition of endpoints.

• Analyses should be performed both by intention to treat, and on a protocol correct population.

• Non-inferiority could be falsely concluded if a trial was conducted with insufficient rigour