

Design issues in pivotal drug trials for drug sensitive tuberculosis (TB)

Andrew J. Nunn^{a,*}, Patrick P.J. Phillips^a, Stephen H. Gillespie^b

^aMedical Research Council Clinical Trials Unit, 222 Euston Road, London, NW1 2DA, UK

^bCentre for Medical Microbiology, Hampstead Campus, Rowland Hill Street, London, NW3 2PF, UK

KEYWORDS

Tuberculosis;
Non-inferiority;
Clinical trials;
Trial design;
Intention to treat;
Per protocol

Summary

The urgent need for new anti-tuberculosis drugs raises the question of the design, conduct and analysis of the trials that will be required for licensing purposes. Current standard regimens are highly effective under controlled trial conditions with relapse rates of 5% or less. It is very unlikely better results can be achieved with new drugs, a clinically more relevant goal would be a regimen of comparable efficacy to standard treatment but of substantially shorter duration. In order for a new regimen to be licensed, it will be necessary to demonstrate that it is of comparable efficacy to the standard regimen. An important issue will be the choice of the margin of non-inferiority which needs to be justified both on statistical and clinical grounds; non-inferiority could be falsely concluded if a trial was not conducted appropriately, with substantial losses to follow-up or unsatisfactory laboratory procedures. It is particularly important therefore that such trials are conducted with a high degree of rigor. Analyses should be performed both by intention to treat, which is conservative and therefore biased towards no difference, and on a protocol correct population. Similar conclusions would be required from both analyses. Substantial developments have been made in tuberculosis bacteriology in recent years enhancing our ability to diagnose and differentiate strains of *M. tuberculosis*. Many of these techniques affect the design of trials but have yet to be evaluated in that setting. Non-inferiority pivotal trials require that, as far as practicable, the same techniques are used as were employed when the trials assessing the standard regimen were conducted.

© 2008 Elsevier Ltd. All rights reserved

Background

During the 1950s and 1960s standard treatment for newly diagnosed pulmonary tuberculosis in technically advanced countries was based on PAS and isoniazid for a minimum of 18 months, supplemented in the first two to three months by streptomycin. This regimen was too expensive for many developing countries to adopt because of the cost of PAS; however, in the mid-1950s, in a series of

studies, thiacetazone substituted for PAS was found to be equally effective. Regimens based on thiacetazone and isoniazid supplemented by streptomycin became the regimen of choice in many countries. Success rates in excess of 95% achieved under trial conditions were, however, not matched by those under program conditions; in a study conducted in Kenya in 1974 only 72% of patients collected 6 months' supply of drugs and only 24% collected 12 months' supply.¹ A short, effective regimen was required and the advent of rifampicin with promising studies demonstrating its sterilizing activity in murine tuberculosis made this goal a real possibility.

The first East African/British Medical Research Council (BMRC) short course chemotherapy trial proved

*Corresponding author: Tel.: +44 207 670 4703;

fax: +44 207 670 4815

E-mail address: ajn@ctu.mrc.ac.uk (A.J. Nunn).

to be the most significant landmark in the treatment of tuberculosis since the introduction of streptomycin almost 25 years previously.² Four regimens of six months duration were compared with the standard 18 month regimen of isoniazid and thiacetazone, supplemented by streptomycin for the first two months. All four six month regimens contained streptomycin and isoniazid throughout and three of them included either rifampicin, pyrazinamide or thiacetazone. The regimen containing rifampicin proved highly effective with results closely similar to the control regimen in a per-protocol population of patients with fully sensitive strains.

The success of a six month regimen was confirmed by studies conducted by the BMRC, primarily in East Africa but also in Singapore and Hong Kong. Variations on the six month rifampicin based regimen were tested in order to minimize the use of the more expensive drugs with the objective of delivering highly effective treatments affordable in developing countries with limited resources.

The most widely used regimen today is based on six months of isoniazid (H) and rifampicin (R), supplemented in the first two months with pyrazinamide (Z) and ethambutol (E), the 2EHRZ/4HR regimen. In a variety of clinical trial settings, some close to program conditions, this regimen has been found to be highly efficacious with relapse rates of 5% or less. The 2EHRZ/4HR regimen, although too expensive for developing countries at the time it was first investigated, has since been adopted by many National Tuberculosis Programs in developing countries as well as many technically advanced countries. The relapse rates in the early trials are likely to be overestimates since at the time it was not possible to distinguish relapse from exogenous reinfection and they should more accurately be described as recurrence rates (see laboratory methods section below).

The dramatic increase in new cases of tuberculosis which began in the late 1980s, largely associated with the HIV/AIDS pandemic in sub-Saharan Africa, drew attention to the urgent need to develop new drugs. Any new regimen of chemotherapy is very unlikely to achieve higher rates of efficacy than those seen with the standard regimen. Even if it were possible such a study would need to be very large. To detect, with 90% power, a halving of the relapse rate from 5% to 2.5% would require over 2500 assessable patients. However, rather than reducing an already very low rate of recurrence a clinically more relevant goal would be a regimen of comparable efficacy but substantially shorter duration. Such a regimen would have many advantages, adherence to treatment would be better, patients would have less exposure to potentially toxic drugs, they would need less contact with the health services and total costs should be reduced. In order for such a new 'improved' regimen to be licensed, it would be necessary to demonstrate that the estimates of efficacy (prevention of both failure and relapse) obtained are no worse than those of the control gold standard regimen using a *non-inferiority* trial design.

In consideration of what is meant by a regimen of comparable efficacy it is important to remember that it

will never be possible to prove that two regimens have the same effect since there will always be some uncertainty surrounding estimates of treatment effects, and an arbitrarily small difference can never be excluded. It is also wrong to infer that a non-significant difference, which may be due to an under powered study, necessarily implies non-inferiority.

A key issue in the design of non-inferiority trials is the choice of the margin of non-inferiority (δ). This needs to be justified both on statistical and clinical grounds.³ In contrast to superiority trials that are designed to determine whether a new treatment is more efficacious, with non-inferiority we are only interested in determining whether a new treatment is no worse by this agreed amount, δ . If it is better, that would be an added advantage.

Other important considerations include the confidence level to be employed, the population to be studied and the definition of a favorable or unfavorable outcome. An assessment of comparative adverse effects also needs to be carefully considered; it is not enough to prove non-inferiority in terms of efficacy, it would seem a reasonable requirement that a new treatment should be at least as safe if not safer than the old one.

Conduct of earlier trials

The series of studies conducted by the BMRC which established short course chemotherapy for tuberculosis was conducted according to pre-defined protocols but with some important differences from currently accepted criteria. The most important of these was the conduct of the main analysis which was restricted to patients known to have organisms sensitive to both streptomycin and isoniazid who had not missed more than a specified amount of treatment, i.e. a per-protocol population. Early deaths were excluded because in some instances it was clear that to have included them would have distorted the findings,⁴ details of deaths were given in the reports. The published reports also provided separate analyses of patients with drug resistance and gave details of the numbers excluded from analysis on account of default, insufficient bacteriology or other reasons. Failure during treatment was presented separately from relapse after successful treatment. Such failures, however, were rare even on regimens which had unsatisfactorily high relapse rates. In the fourth East African short course study, only 4 of 555 patients were classified as treatment failures, whereas relapse rates were high, ranging from 11–40%.⁵ This study together with the first Singapore short course study⁶ demonstrated that it was not possible to reduce treatment duration from six to four months with the existing drug combinations. No further attempt was made to reduce the duration of treatment.

In 1998 the International Union Against Tuberculosis & Lung Disease conducted a study designed to compare the eight month, WHO recommended, regimen based on ethambutol and isoniazid with the six month regimen based on rifampicin and isoniazid. In contrast to the earlier studies this included all assessable randomized

patients in the primary analysis, apart from those identified as having MDR (multi-drug resistant) disease. Although designed as a non-inferiority trial the intention to treat analysis demonstrated unequivocally that the eight month regimen was significantly inferior to the control.⁷

Non-inferiority trial design

The objective of a non-inferiority trial is to show that the control arm is superior to the intervention arm(s) by no more than a pre-specified amount. This is in contrast to a superiority trial where the objective is to show that there is a difference in efficacy between the regimens being studied.

A case study

The first East African short course chemotherapy trial could be considered to have been a non-inferiority trial, although it was not described as such at the time. Relapse rates in the intervention arms of six months duration were compared with the failure/relapse rate in the eighteen month control regimen, not with the expectation that the shorter duration regimens would be more efficacious but that they would be of similar efficacy. The top half of the Figure shows the 95% confidence intervals for the difference in relapse rates in a per protocol analysis for each of the six month regimens 30 months after randomization when compared with failures of the standard regimen. Three of the four regimens are significantly inferior to the control arm at the 5% level, two of them by a wide margin. For the 6SHR (rifampicin) regimen the point estimate for the difference in efficacy from the control arm (2STH/16TH) arm was 0.4% in favor of the 6SHR regimen (95% CI -3.5%, 4.2%). If the trial had been designed as a non-inferiority trial (without any adjustment for multiple comparisons) a pre-stated margin of non-inferiority (δ) of either 5% or even 4% would have been sufficient for this regimen to be declared non-inferior to the 2STH/16TH regimen. The lower limit of the 95% confidence interval for the difference between the 6SHZ (pyrazinamide) regimen and the control arm was 10.8% in favor of the control arm, a margin too large to be regarded as non-inferior. The two weaker regimens, 6SHT and 6SH were clearly inferior to the control and were stopped early at an interim analysis.

A modified intention to treat (MITT) analysis of these trial data was reported in which patients whose treatment was altered on account of drug toxicity or who defaulted a substantial proportion of their treatment were assigned an unfavorable outcome. This analysis demonstrated a significant benefit to the 6SHR regimen over the control (95% CI for difference in status 3.4%, 16.9%) and, in this instance, a suggestion of superiority of the 6SHZ regimen over the control (95% CI for difference in status -3.5%, 11.7%), lower half of the Figure, whereas the two weaker regimens were confirmed to be inferior.

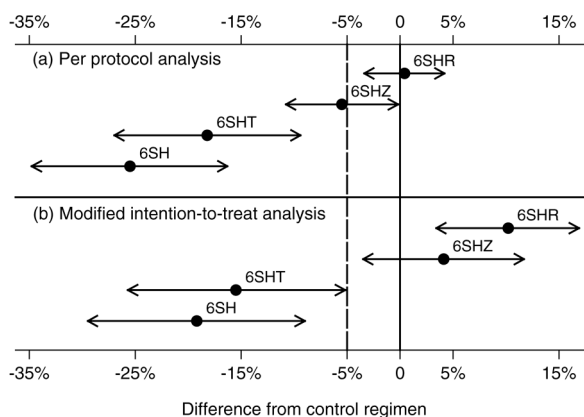


Figure 95% confidence intervals for difference in relapse rates from standard regimen (1st East African short course chemotherapy study, per protocol analysis and modified intention to treat analysis).

Choice of control regimen

It is essential that the design of a non-inferiority trial does not allow for the possibility of establishing as non-inferior a treatment that is no better than a regimen already considered inadequate. Unless there is already robust evidence for the efficacy of the control arm it is recommended that a placebo control should be included.⁸ The efficacy of the 2EHRZ/4HR regimen for newly diagnosed pulmonary tuberculosis is very well established.^{7,9} However, when assessing whether the substitution of a new drug for an existing drug in a multi-drug regimen can lead to shorter treatment duration it is important to be able to demonstrate that the same result could not have been obtained simply by shortening the original regimen.

This can best be illustrated by reference to the design of the three-armed REMoxTB clinical trial (Box). Using the 2EHRZ/4HR regimen as a control, moxifloxacin replaces ethambutol in one intervention arm and replaces isoniazid in the other. Both intervention arms are to be given for only four months. To demonstrate that either of the moxifloxacin containing regimens are not inferior to the control regimen requires that the value of δ is such that the non-inferiority cannot be declared unless there is clear benefit over the results that would have been obtained had the control regimen been given for only four months.

Were it not for ethical considerations, the most convincing way to demonstrate the superiority of one or both of the intervention arms over the control regimen given for four months would be to include that regimen as a second control arm, namely 2EHRZ/2HR. This cannot be done ethically because, although the four month 2EHRZ/2HR regimen has never been studied, there is no evidence to suggest that it would outperform the 2SHRZ/2HR regimen which has already been demonstrated to have an unsatisfactory efficacy under trial conditions.⁵ It has long been acknowledged that ethambutol contributes little if anything to either the bactericidal or the sterilizing activity of short course regi-

Box**REMOxTB trial design****2EHRZ/4HR (control regimen)**

6 months of rifampicin plus isoniazid, supplemented by 2 months of ethambutol plus pyrazinamide.

2MHRZ/2MHR

4 months of rifampicin plus isoniazid plus moxifloxacin, supplemented by 2 months of pyrazinamide.

2EMRZ/2MR

4 months of rifampicin plus moxifloxacin, supplemented by 2 months of ethambutol plus pyrazinamide.

During the 2 month intensive phase all patients receive four active drugs plus one placebo to match the 5th drug. In the first 2 months of the continuation phase they receive three active drugs or two active drugs and a matching placebo and in the last two months they receive two active drugs or two matching placebos.

mens,¹⁰ its role being to protect against the development of further resistance in those patients harboring isoniazid resistant strains.

What do we know about the extent of inferiority of the four month 2SHRZ/2HR regimen when compared to the standard control? Table 1 compares the relapse rates in the 4 and 6 month regimens in East Africa and Singapore. In this analysis the regimens that only differ on account of having pyrazinamide in the continuation phase have been combined since the data indicate that there is no evidence of benefit from pyrazinamide after the initial two months, albeit on relatively small numbers of patients assessed;^{5,6} the results for the 6-month regimen in East Africa comes from the 5th short course study.¹¹ In both East Africa and Singapore the difference in relapse rates between the four and six month regimens is 9–10%. This provides an important guide for the upper limit for the choice of delta.

Power considerations

There are many assumptions underlying the power calculations in a clinical trial. The basis of these assump-

tions vary; in some instances, the supporting data are much less well quantified than others. The efficacy of the control regimen, under per-protocol conditions, is very well documented. However, it is much more difficult to assess what can be expected in an intention to treat analysis when, as required by the regulators, those failing to complete chemotherapy are classified as having an unfavorable response. A low 4–5% failure/relapse rate in the control arm would be doubled if 5% of patients either defaulted before the end of treatment, required a change of treatment for drug toxicity or died from non-tuberculous causes. The latter group may be the most difficult to prevent occurring in populations with a substantial proportion of HIV-infected patients. Classifying patients who do not complete chemotherapy as unfavorable will be very conservative, in view of the fact that over 80% of patients completing no more than 4 months of standard chemotherapy could be expected to have a favorable response.^{9,10} By employing such a classification it is possible that chance differences in the distribution of deaths and defaulters across the regimens could result in falsely declaring or failing to declare non-inferiority.

In view of an expected failure/relapse rate of at least 10% in the control regimen of the REMoxTB trial using the above classification and the difference of 9–10% observed between the relapse rates in the four and six month regimens, the REMoxTB investigators chose a margin of non-inferiority of 6%. It is important to point out that this represents the lower bound of the confidence interval for the difference in failure/relapse rates between the two regimens, not the expected difference. The assumption made in the power calculations for the REMoxTB trial, as in many such trials, is that the point estimates for the failure/relapse rates are the same.

The choice of delta is an important consideration and must, as indicated above, be justified not only on statistical but also on clinical grounds. Would the possibility of an increase in failure/relapse of 6% be acceptable to African physicians? Discussions with the African investigators participating in the REMoxTB trial indicated they would consider that the considerable advantages in reducing the duration of chemotherapy from 6 to 4 months would outweigh a possible increase in relapse rates of up to 6%. Relapse cases can usually be retreated with the same drugs as before and if a relatively small number of patients need to be retreated it would be a small price to pay if the remainder of patients could be

Table 1 Comparison of 4 and 6 month regimens in East African and Singapore short course studies

Country	Duration	Regimen	N assessed	Relapse (N %)
East Africa ^a	4 month	2SHRZ/2HR(Z) ^b	208	28 (13)
	6 month	2SHRZ/4HR	166	4 (3)
Singapore	4 month	2SHRZ/2HR(Z)	156	15 (10)
	6 month	2SHRZ/4HR(Z)	158	2 (1)

^anon-concurrent comparison.

^b(Z) combines patients receiving and not receiving pyrazinamide in the continuation phase.

Table 2 Number of patients needed per regimen to demonstrate non-inferiority with 90% power using one sided 97.5% confidence intervals for varying failure/relapse rates in the control arm, equal or inferior efficacy of the intervention to control and different levels of delta.

Failure/relapse % (control arm)	Equal efficacy δ (margin of non-inf.)			Intervention 1% less effective than control δ (margin of non-inf.)			Intervention 2% less effective than control δ (margin of non-inf.)		
	4%	5%	6%	4%	5%	6%	4%	5%	6%
4	505	323	225	1003	565	362	2491	1107	623
5	624	400	278	1214	683	437	2958	1315	740
6	741	475	330	1419	798	511	3415	1518	854
7	856	548	381	1620	911	583	3862	1717	966
8	967	619	430	1816	1022	654	4298	1911	1075
9	1076	689	479	2007	1129	723	4724	2100	1181
10	1183	757	526	2194	1234	790	5139	2284	1285
11	1286	823	572	2376	1337	856	5543	2464	1386
12	1387	888	617	2554	1437	920	5937	2639	1485

treated for only four months. There would be an even greater benefit to those countries still using the eight-month 2EHRZ/6HE regimen.

Table 2 sets out the number of patients needed per regimen under a variety of conditions keeping power constant at 90% and using a one-sided 97.5% confidence interval. The lower control rates, 4–6%, are applicable to a per-protocol analysis, whereas the higher rates of 8% or 12% correspond to what may be expected if unfavorable response rates are doubled or even tripled as a consequence of classifying those who do not complete treatment as unfavorable.

It is clear that with delta fixed, as the failure/relapse rate in the control arm increases (or the success rate decreases), the number of patients required in each arm increases. The table also shows the sample size required if, for a fixed value of delta, the intervention arm is assumed to be 1% or 2% less effective than the control. Thus, if both the control arm and the intervention arm are assumed to have a 10% unfavorable outcome and delta is set at 5% then 757 patients are needed in each arm. If the intervention has an 11% unfavorable outcome, i.e. a point estimate that is 1% inferior to the control, the number required in each arm increases to 1234. A 12% unfavorable outcome in the intervention would require 2284 per arm – a total of close to 5000 patients in a two arm study.

Trial conduct

Both an ITT and a per-protocol analysis can be biased towards falsely declaring non-inferiority. In a superiority trial there is always a strong incentive to ensure high quality of conduct. In contrast in a non-inferiority trial, the conclusion of non-inferiority could be reached because of poor discriminatory power. This could be a consequence of poor follow-up rates or laboratory failures, such as inability to detect all relapses. It is therefore of particular importance that non-inferiority trials are conducted with a high degree of rigor. Every effort should be made to minimize losses to follow-up

and other missing data. Particular care should be taken to enroll patients likely to comply with the demands of the protocol. This requirement should be balanced against the need to enroll populations which are as similar as possible to those included in earlier studies. The advent of the HIV/AIDS epidemic with large numbers of co-infected patients presents an added challenge which should be addressed when the study is analyzed. The non-inferiority trial with only an active control (i.e. no assessment against a placebo control) can be seen as a special example of a study using historical controls with all the limitations that implies.

As far as possible the same protocol should be used as in the earlier studies which demonstrated the superiority of the current standard regimen: closely similar inclusion and exclusion criteria and, ideally, the same laboratory methods. For this reason although liquid culture media are increasingly used to detect *M. tuberculosis*, solid media will be used in the primary analysis of the REMoxTB trial for comparability with the past BMRC trials. Both methods will be employed for all samples in order to make comparison between the methods.

How should the per-protocol population be defined? The patients included should all be assessable for the primary endpoint, have missed no more than a pre-agreed amount of their chemotherapy regimen and not been subject to important protocol violations. These conditions should all be determined before any data analysis begins.

A particularly important issue with non-inferiority trials is the potential for erosion of the efficacy of the control regimen, referred to by some researchers as biocreep. This occurs when a less effective regimen having been declared to be non-inferior then becomes the standard control in the next non-inferiority trial. As time goes on what is regarded as an acceptable response could slip further away from the original standard. Non-inferiority trials should use as their control the regimen with the best known outcome as recommended by the World Medical Association.¹² The 2EHRZ/6HE regimen would not be appropriate as the

standard control even if it had been shown to be non-inferior to the 2EHRZ/4HR regimen.

In the REMoxTB trial four of the drugs under study are blinded. The main advantage of blinding is an objective assessment of adverse events; assessments of outcome will be objectively based on bacteriological results. However, blinding also removes the possibility that patients on the regimens of shorter duration might be selectively withdrawn because of concerns about a possible poor response which would reduce the chance of demonstrating non-inferiority.

Criticisms of non-inferiority design

There have been recent concerns expressed about the ethics of non-inferiority designs.¹³ While there is no doubt that a badly designed non-inferiority study could indeed be unethical the same could also be said of a badly designed superiority trial. Critics often fail to recognize that given the highly effective regimens currently available for the treatment of a disease such as tuberculosis, statistical superiority in terms of efficacy is unlikely to be demonstrated but regimens of shorter duration, which carry no more than a small increased risk of relapse, offer the prospect of less exposure to potentially toxic drugs and less need for interaction with the health services.¹⁴ Clearly it is essential that before enrollment patients are fully aware of the possible outcomes including the risks of higher relapse rates, risks which will be similar to those experienced when participating in a superiority trial.

Laboratory methodology

In recent years there has been a transformation in the methodology used in microbiology laboratories to diagnose infection with mycobacteria and this will have a profound impact on the conduct of tuberculosis clinical trials. The developments include automated liquid culture systems that improve the speed of diagnosis, molecular diagnostics and speciation methods that enhance our ability to detect and characterize different mycobacteria, and better technology for strain comparison that have shed enormous light on the population genetics and epidemiology of *M. tuberculosis*.

Automated liquid culture systems

Automation of mycobacterial culture has been a revolution in the ease with which mycobacteria can be detected. Instead of the use of solid slopes of Lowenstein Jensen egg based or similar media these systems use a liquid media containing a cocktail of antibiotics to suppress the growth of other organisms. More significantly, the growth of mycobacteria is automated using radiometric, or fluorescence methods to detect the production of carbon dioxide by growing mycobacteria.¹⁵ As well as simplifying the work of the diagnostic laboratory it reduces the labor of checking

for positive cultures with the effect that positive cultures are detected and identified more frequently. Importantly, liquid culture is significantly more sensitive than solid media and the effect of this is that positive cultures are detected usually in a little under two weeks quicker.¹⁶ It also means that there is a lower limit of detection with some specimens that would be negative using solid culture being flagged as positive. As in all cases in microbiology, improvements occur alongside drawbacks and in this case contamination with non-tuberculosis mycobacteria and with other bacteria is more likely. Each of these factors has an impact on trials methodology and will be discussed in turn.

Improved limit of detection

Bacteriological positivity is usually the most important primary endpoint, more sensitive detection means that some samples that would have been previously recorded as negative will be recorded as positive. This is especially important in relation to the culture conversion rate which is widely considered to be related to the risk of relapse.¹⁷ If, in the future, culture conversion rates are used as surrogates for relapse rates leading to licensing in the same way as changes in viral load have been used in antiretroviral treatment it will be important to conduct studies to determine what are the most important predictors, any evidence of positivity or only cultures growing above a certain threshold as measured by time to positivity (TTP). We have already produced preliminary data that demonstrate that this approach can be used to differentiate between patient groups and to monitor response to therapy. Such methods could have a significant impact on the conduct of clinical trials and there is pressing need to study the use of TTP as a measure of treatment response in a clinical trial setting. Trials in which rates of culture conversion are primary endpoints need to use standardized laboratory procedures to avoid difficulties of interpretation when a mixture of methods are employed.

Contamination of cultures

With increased culture sensitivity more cultures will be positive with contaminating organisms. These can be of three types: conventional oral flora that are not fully suppressed by the decontamination process and antibiotics, non-tuberculosis mycobacteria that may have been present in the sample and cross-contamination with *M. tuberculosis* from other specimens processed in the same batch. Contamination with bacteria from the oral flora can be addressed by ensuring that the decontamination methods are adjusted to keep contamination rates near to five percent.¹⁸ Although contaminated cultures can be retreated, there is a risk that important data points will be lost if duplicate cultures are not established in parallel or parallel solid cultures. Contamination with non-tuberculosis mycobacteria (NTM) poses a risk in that these organisms might be falsely assumed to be *M. tuberculosis*, indicating the

importance of effective speciation of all critical positive samples.¹⁹ The growth of a non-tuberculosis mycobacterium may mean the data point will be lost as it will never be certain whether there is a true positive *M. tuberculosis* culture under the NTM and the decontamination process is likely to completely inhibit the tuberculosis isolate causing a false negative. This dilemma emphasizes the need to obtain multiple samples during the follow up period.

New typing techniques

In the past it had been assumed that *M. tuberculosis* was genetically very homogeneous. A range of studies has improved our understanding of the population genetics of the *M. tuberculosis* species¹⁸ The ability to differentiate many different strain types of *M. tuberculosis* has also transformed our understanding of the transmission of tuberculosis. Typing can be achieved by methods that detect the distribution of the insertion sequence IS6110 in the genome (IS6110 typing), the number of mycobacterial intergenic repeat units present at different locations in the chromosome (MIRU) or polymorphism of the chromosomal DR locus, which contains a variable number of short direct repeats interspersed with nonrepetitive spacers (spoligotyping).^{20–22} Previously, an isolate obtained during the post treatment phase of follow-up would have been classified as a relapse. Recent reports indicate that it is less certain whether all so called clinical relapses are due to the organisms present before treatment commenced. In some situations clinical relapse is more commonly due to reinfection than relapse.²³ This is particularly important in the case of HIV-infected patients. The situation may be more complex following the description of infection with multiple strains which may occur in settings where the infection pressure is high.²⁴

Analysis of data

The results from the first East African trial highlight the need to perform both ITT and per-protocol analyses when assessing non-inferiority. Superiority trials are analyzed by ITT because it is the most conservative and least likely to be biased. In contrast, ITT analysis of non-inferiority trials is not conservative since the inclusion in the analysis of patients who violate the protocol will tend to minimize differences between study arms thereby increasing the possibility of declaring non-inferiority.

Per-protocol analyses, in which only those adherent to the protocol are included, are also biased since not all randomized 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. With this in mind, the reason for withdrawal from the study regimens and the pattern of withdrawal over time needs to be analyzed to explore potential differences between the treatments.

It is a requirement of the CPMP (Committee on Proprietary Medical Products) that 'similar conclusions from both the ITT and per-protocol analysis are required' to declare non-inferiority and 'sample size computations should ensure sufficient numbers in the per-protocol population'.²⁵ However, it is also stated that the primary analysis should be per protocol 'since it is most sensitive for the detection of any real difference'.²⁶

Analyzing non-inferiority trials by both intention to treat and per-protocol helps to reduce the possibility of wrongly declaring a regimen to be non-inferior. In addition, further sensitivity analyses should be done to evaluate alternative classification scenarios. Important differences may occur in the response to treatment between HIV-infected and HIV-uninfected patients. These sub-groups should be analyzed separately even though the comparisons will have limited power. Other sub-groups that should be analyzed are patients with fully sensitive strains at enrollment and those with isoniazid-resistant strains.

Comparative assessments of adverse events and reasons for exclusions from the per-protocol analysis are particularly important; these may reflect differences between the management of the patients in the different treatment arms.

Conclusions concerning whether or not non-inferiority has been demonstrated need to take into account a much wider set of scenarios than would be customary with superiority trials. The conclusion will be most convincing when all the evidence points in the same direction.

Acknowledgements

Funding: Medical Research Council, European Developing Country Trials Program, The Global Alliance for TB Drug Development, Health Technology Assessment (UK) and DFID. PPJ Phillips has been funded by a PhD studentship provided by the MRC Clinical Trials Unit.

Competing Interests: A Nunn has served Tibotec DMC and NIH DMC – receiving a fee.

References

- 1 Tuberculosis in Kenya: follow-up of the second (1974) national sampling survey and a comparison with the follow-up data from the first (1964) national sampling survey. An East African and British Medical Research Council co-operative investigation. *Tubercle* 1979;60: 125–149.
- 2 East African/British Medical Research Council. Controlled clinical trial of four short course (6 month) regimens of chemotherapy for treatment of pulmonary tuberculosis. Third report. *Lancet* 1974;2: 237–240.
- 3 Committee for Medicinal Products for human use: Guideline on the choice of the non-inferiority margin. CPMP 2005.
- 4 East African/British Medical Research Council. Controlled clinical trial of four short course (6 month) regimens of chemotherapy for treatment of pulmonary tuberculosis.

- Lancet* 1973;1:1331–1339.
- 5 East African/British Medical Research Council. Controlled clinical trial of five short course (4-month) chemotherapy regimens in pulmonary tuberculosis. Second report of the 4th study. *Am Rev Respir Dis* 1981;123:165–170.
 - 6 Singapore Tuberculosis Service/British Medical Research Council. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis: The results up to 30 months. *Tubercle* 1981; 62:95–102.
 - 7 Jindani A, Nunn AJ, Enarson DA. Two 8-month regimens of chemotherapy for treatment of newly diagnosed pulmonary tuberculosis: international multicentre randomized trial. *Lancet* 2004;364:1244–1251.
 - 8 International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline. Choice of control group and related issues (ICH E10). May 2001.
 - 9 Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council Tuberculosis Units, 1946–1986, with relevant subsequent publications. *Int J Tuberc Lung Dis* 1999;3: S231–S279.
 - 10 Fox W. Whither Short Course Chemotherapy. *Br J Dis Chest* 1981;75:331–357.
 - 11 East African and Central African/British Medical Research Council: Controlled clinical trial of 4 short course regimens of chemotherapy (three 6-month and one 8-month) for pulmonary tuberculosis. Final report. *Tubercle*, 1986;67:5–15.
 - 12 World Medical Association: Declaration of Helsinki. <http://www.wma.net/e/policy/b3.htm> Sept. 10, 2004.
 - 13 Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007; 370:1875–1877.
 - 14 Nunn AJ, Meredith SK, Spigelman MK, Ginsberg AM, Gillespie SH. The ethics of non-inferiority trials. *Lancet* 2008;371:895.
 - 15 Hanna BA, Ebrahimzadeh A, Elliott LB, Morgan MA, Novak SM, Rusch-Gerdes S et al. Multicenter evaluation of the BACTEC MGIT 960 system for recovery of mycobacteria. *J Clin Microbiol* 1999;37:748–752.
 - 16 Rohner P, Ninet B, Benri AM, Auckenthaler R. Evaluation of the Bactec 960 automated nonradiometric system for isolation of mycobacteria from clinical specimens. *Eur J Clin Microbiol Infect Dis* 2000;19:715–717.
 - 17 Aber VR, Nunn A. Short term chemotherapy of tuberculosis. Factors affecting relapse following short term chemotherapy. *Bull Int Union Tuberc* 1978;53:280.
 - 18 American Thoracic Society. Diagnostic standards and classification of tuberculosis in adults and children. *Am J Respir Crit Care Med* 2000;161:1376–1395.
 - 19 Beggs ML, Stevanova R, Eisenach KD. Species identification of *Mycobacterium avium* complex isolates by a variety of molecular techniques. *J Clin Microbiol* 2000;38: 508–512.
 - 20 Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B et al. Strain identification of *Mycobacterium tuberculosis* by DNA Fingerprinting: Recommendations for a Standardized Methodology. *Journal of Clinical Microbiology* 1993;31:406–409.
 - 21 Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2006;44:4498–4510.
 - 22 Kamerbeek J, Schouls L, Kolk A, van AM, van SD, Kuijper S et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;35:907–914.
 - 23 Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW et al. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *Am J Respir Crit Care Med* 2005;171: 1430–1435.
 - 24 van RA, Victor TC, Richardson M, Johnson R, van der Spuy GD, Murray EJ et al. Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns. *Am J Respir Crit Care Med* 2005;172: 636–642.
 - 25 Committee on Proprietary Medical Products Point-to-Consider. Points to consider on switching between superiority and non-inferiority. CPMP, 2000.
 - 26 Committee on Proprietary Medical Products Note for guidance on evaluation of medicinal products indicated for treatment of bacterial infections. CPMP, 2004.