Statistical Analysis Plan

Optimization of the TB treatment regimen cascade

OneRIF

Date: 18-September-2015

Author: J. Buyze
Biostatistician, Institute of Tropical Medicine, Antwerp, Belgium

Approved by: Coordinating Investigator
Dr. A. Van Deun

Date: 18 / 09 / 2015

i.o.v.

Table of Contents

1. Introduction

This Statistical Analysis Plan (SAP) provides a detailed and comprehensive description of the main, pre-planned analyses for the study "Optimization of the TB treatment regimen cascade", including the recently approved interim analysis. The purpose of this study is to compare a high-dose rifampicin standard TB regimen of 6 months duration with the standard TB regimen in respect to adverse treatment outcome (relapse, failure, death and default) and toxicity, in smear-positive pulmonary TB cases. The study conduct is described in the Protocol.

These planned analyses will be performed by the statistician(s) at the Clinical Trials Unit of the Institute of Tropical Medicine (Antwerp) in collaboration with the research consortium. The analysis results will be described in a statistical analysis report, to be used as the basis of the primary research publications. This document describes statistical methods for the primary and secondary objectives of the study as defined by protocol. Additional analyses may be performed but are not covered by the current analysis plan.

Analyses will be performed at three time-points.
1) Assessment of referral for rapid DST plus its outcome based on the auramine and/or FDA decline for the patients randomized in the intervention arm versus persisting fever, for the patients randomized in the control arm, 2 weeks after the first 500 enrolled patients (intervention + controls) reached two weeks follow-up. With this analysis we want to assess whether any of those can really be used as an early indicator of serious rifampicin resistance, and also assess the impact of improvement of staining techniques that were made after the first ITM monitoring visits. The results in terms of resistance missed / correctly suspected will be analysed based on the initial genetic rifampicin resistance data which will by that time also be available for part of these patients.

2) Assessment of drug toxicity and initial treatment response will be performed at the time the last recruited patient finishes treatment. The data will be compiled at 9 months after the end of recruitment to ensure all treatment related data, including 8 months of treatment for retreatment cases in the control arm, are available. This analysis will be primarily performed for the go/no-go decision and design considerations for a possible next study, such as a cluster-randomized trial. The decision on proceeding to the cluster randomized study will be based on the absence of excess toxicity, a trend toward a reduction of unfavourable outcomes (excluding relapse), possible favourable effects overcoming low-resistance mutations.
3) A final analysis of the study endpoints, including relapse, will be performed one year after end of treatment of the last enrolled patients, i.e. about 2.5 years after enrolment start.

This analysis plan will be finalized and approved before database lock of the first interim analysis data. Major changes in statistical methodology used for the main and pre-planned analyses from this SAP, will require detailed description and justification in the statistical analysis report. The final analysis datasets, programs, and outputs are archived following good clinical practice guidelines (ICH E9).

2. Study design and objectives

2.1. Study design

This study is an open-label randomised clinical trial. Consecutively diagnosed smear-positive pulmonary TB patients, which conform to all of the inclusion and none of the exclusion criteria and provide informed consent, will be randomized to the intervention or control arm.

The control arm will receive the standard pulmonary TB treatment regimens recommended by the International Union against Tuberculosis and Lung Disease (The Union). This consists of:

Treatment:

New cases: 2HRZE/4HR, i.e. 2 months of isoniazid, rifampicin, pyrazinamide and ethambutol intensive phase, followed by 4 months continuation phase isoniazid plus rifampicin) (Cat. 1 regimen, 6 months)

Retreatment cases: 2 months of SHRZE (S= streptomycin) followed by 1 month of HRZE and 5 months HRE (Cat. 2 regimen, 8 months)

Any patients with *MDR TB detected* at any time will be switched to 4KHPtoGCEZ/5GCEZ, i.e. 4 months of kanamycin, isoniazid, prothionamide, gatifloxacin high-dose, clofazimine, ethambutol and pyrazinamide, followed by 5 months of gatifloxacin high-dose, clofazimine, ethambutol and pyrazinamide (MDR regimen, 9 months)

Testing:

Monitoring by smear and rapid DST as per National TB Programme (NTP) indications:

- Smears for AFB [by Ziehl-Neelsen brightfield (ZN) or auramine fluorescence (FM) technique] at diagnosis, after 2, 5 and 6 months (Cat. 1) or after 3, 5 and 8 months (Cat. 2); repeat after month 3, respectively 4, if AFB were seen at month 2 or 3
- Vital staining fluorescein diacetate (FDA) smears for AFB to confirm a positive ZN or FM result during Cat. 1
- rapid DST if fever is still present after 2 weeks, or in case an AFB smear (auramine and/or FDA) is positive during Cat. 1, or from the 3rd month of Cat. 2 onwards. Rapid DST is performed also from any retreatment case (relapse, return after default and other previously treated for at least 1 month), and from previously untreated contacts of MDR-TB

Additionally, patients enrolled in the study (control as well as intervention arm) will be followed-up for relapse 12 months after the end of treatment by clinical examination, sputum smear and slow solid culture.

The intervention consists of the standard care as provided to the control arm with the following modifications:

Treatment:

New cases: Compared to standard regimen dosing of rifampicin is doubled, while standard dose isoniazid, pyrazinamide and ethambutol are maintained, i.e. 2 months of HRhZE followed by 4 months of HRh (6 months)

Retreatment cases: retreatment cases randomized to the intervention arm receive the same care as new cases in the same arm.

Any patients with *MDR TB detected* at any time will be switched to 4 months of KHPtoGCEZ followed by 5 months of GCEZ (MDR regimen)

Testing

In addition to standard monitoring (as described above for the control arm), patients in the intervention arm will have sputum tested by auramine and FDA smear after the first and again after the second week of intensive phase treatment. Management decisions will be made after 2 weeks of intensive phase, based on remaining numbers and decline of AFB on FDA staining, but not based on continuing fever. Those with min. 1 log decline in the number of AFB, or who are already negative, would continue the unaltered regimen. Those not meeting these criteria would be referred for rapid DST by slide DST and/or Xpert MTB/RIF, depending on ease of referral.

2.2. Study objectives

An additional secondary objective has been added, compared to the objectives in the protocol.

Primary objective:

To compare a high-dose rifampicin standard TB regimen of 6 months duration with standard TB regimen with respect to adverse treatment outcome (relapse, failure, death and default) and toxicity, in TB cases (new or retreatment, pooled) not shown to have rifampicin resistance before one of the standard treatment outcomes applies.

Secondary objectives:

1. To assess whether the study regimen also cures low-level rifampicin resistant TB
2. To assess the effectiveness of FDA vital staining screening at two weeks of treatment for early switch of non-responding rifampicin resistant TB to MDR-TB treatment, versus clinical evaluation based on fever resolution.
3. To assess the negative predictive value of conversion at 2 weeks for relapse
4. To estimate the proportion of acquired rifampicin resistance among failures and relapses
5. To compare the predictive value of auramine with FDA staining at 2 weeks to predict unfavourable outcomes
6. To compare weight gain and fever resolution between the two regimens

2.3. Study hypothesis

This study is a pilot study to estimate the differences in the incidences in adverse treatment outcomes and excess toxicity.

3. Description of study population

The study population will be described overall and in each hospital separately.

3.1. Patient accounting

Details of patients screened, those who meet the study inclusion criteria, those who are eligible and randomized, those who are eligible but not randomized, those who withdraw from the study after randomization and those who are lost to follow-up will be summarized in a CONSORT flow diagram. The number (%) of patients attending scheduled follow-up visits by study day will be reported.

3.2. Description of study population

Patients in each treatment group, overall and by hospital, will be described with respect to baseline characteristics. The description will be in terms of medians and interquartile ranges for continuous characteristics and using counts and percentages for categorical characteristics. The clinical importance of any imbalance will be noted but statistical tests of significance of baseline imbalance will not be undertaken.

4. Description of patient populations and outcomes

4.1. Patient populations

We will analyse the efficacy data using Intention-to-Treat and Per-Protocol approaches, with Intention-to-Treat as primary approach. Safety analysis will be performed using an all-patients-treated approach, including all patients who received

at least a single dose, and according to the treatment regimen actually received. Patients who receive a second treatment regimen will be included in the Intention-to-Treat analysis with their outcome after the first treatment regimen, but excluded from the Per-Protocol analysis.

4.1.1. Intention to treat (ITT) analysis

In the Intention-to-Treat analysis, all patients will be analysed according to their randomized allocation, even in case they receive another treatment regimen, show protocol violations prior to or during the study. Patients lost-to-follow-up before the 12 month post treatment follow-up or who do not provide a sputum are considered to be adverse treatment outcomes in this analysis, unless they are known to have moved out of the study region.

4.1.2. Per protocol (PP) analysis

In the per-protocol analysis only patients who receive study drug as planned, receive a treatment regimen only once, have complete follow-up (till 2.5 years or development of a study endpoint) or have reached a study endpoint at an earlier time-point, and follow the protocol as planned are included.
In Table 2 the protocol violations are classified as minor and major where minor violations can be included in the PP analysis population and major violations are excluded.

Table 2: The protocol violations classified as minor or major violation

Protocol Violation	Major/Minor Violation	Comments
Inclusion criteria		
1. Smear-positive pulmonary TB	Major	
2. 15 years or older	Major	
3. Able and willing to provide written informed consent	Major	
Exclusion criteria		
1. MDR-TB diagnosed with resistance on rapid DST for rifampicin prior to start of treatment	Major	
2. Extra-pulmonary TB, smear-negative TB	Major	
3. Patients in need of hospitalization because of very bad general condition or complications at screening	Major	
4. Patients with clinically active liver disease at screening	Major	
5. Known HIV-positives	Major	
6. Known hepatitis B or C infected patients	Major	
7. Pregnant women	Major	
Treatment violations		
1. Missing at least a full day of treatment	Minor	
2. Not following the randomized regimen	Major	
Follow-up violations		
1. Missing study endpoint	Major	

4.2. Efficacy study endpoints

Primary endpoint

The primary efficacy endpoint is the incidence of adverse treatment outcomes. Following the WHO guidelines from 2010, as currently in use in the NTP in Bangladesh, an adverse treatment outcome is defined as any occurrence of the following:

- Relapse: Cured previously from TB or completed treatment for TB and now having bacteriologically positive sputum for TB (at 12 months follow-up or at an earlier time point)
- Default: The patient whose treatment was interrupted for 2 consecutive months or more.

- Failure: Sputum positive for TB at 5 months or later during treatment. Failure will also be declared if the regimen has to be changed for at least 2 drugs due to adverse events.
- Death: All-cause mortality between case registration and end of TB treatment (related or not to TB or TB treatment)

Patients who are transferred out will be excluded from the outcome analysis.

In line with current WHO recommendations, patients detected with MDR-TB or rifampicin resistance before any outcome applies and switched to the MDR-TB regimen will be excluded from the outcome analysis.
Analyses will be performed using the NTP smear-based recurrence (failure, relapse) definitions, but as a secondary approach culture-based definitions will be used. In each analysis, if a patient has two events for the analysis of interest, the first endpoint will be taken into account, except for cured/completed followed by relapse which is considered as relapse.

Safety endpoint
The primary safety endpoint will be the occurrence of any of the following:

- Serious Adverse Events
- Grade 3-4 Liver Toxicity following NIH common toxicity criteria (CTC), defined as transaminase increases to >225 for men or >175 for women/transgender (grade 3), or > 900 for men or >700 for women/transgender (grade 4)

Secondary endpoints
Secondary objectives are assessed with the adverse treatment outcomes defined above.

5. Interim analysis

An interim analysis will be performed after the first 500 enrolled patients have finished the week 1+2 follow up visit. This interim analysis will have no effect on the continuation or conduct of the trial. Therefore no adaptation of sample size, multiplicity, … is required.

Following analyses will be performed.
In the intervention group
The decrease from enrollment until week 2 in Auramine and FDA will be described using median and interquartile range. The number (%) of referrals to rapid DST will be presented. Counts (%) of the rapid DST result will be presented.

In the control group
The number (%) of referrals to rapid DST (based on body temperature) will be presented. Counts (%) of the rapid DST result will be presented.

The percentage of resistance missed will be compared between groups using the Chi-square or Fisher's exact test.

All analyses will be stratified on new vs retreatment cases. Analysis in the intervention group will be additionally stratified on period (2014 vs 2015).

6. Analysis of main efficacy outcomes: adverse treatment outcome

Counts and proportions (and 95% CI) of patients with any adverse treatment outcome, and by adverse treatment outcome category, will be presented. The confidence intervals for proportions will be estimated using Wilson's score method. Odds ratio's (ORs) of treatment effect on any adverse treatment outcome and 95% CIs for the OR will be estimated from a logistic regression model with effects for arm (intervention vs control), hospital, and stratum (new/re-treatment case). The p-value for the test of no difference among arms will be presented. If the number of events is low a logistic regression model with an effect for hospital may be used instead of the full model proposed.

6.1. Subgroup analyses

If the number of events is sufficient, subgroup analyses will be performed on the primary endpoint to assess consistency of response across patient populations. The following subgroup analyses will be performed:

- New and retreatment TB cases

The interaction between treatment differences and subgroup will be tested in the logistic regression model described above.

6.2. Other aspects

a. Multiplicity

As this is a pilot estimation study with a single primary efficacy and safety endpoint, no multiplicity adjustments are needed. An interim analysis will be performed, but this analysis will not influence the continuation of the study.

b. Missing data

In the intention-to-treat analysis, patients who are lost-to-follow-up before the 12 month post treatment follow-up will be counted as adverse treatment outcomes, unless they are known to have moved outside of the area.
Missing values will be imputed based on a logistic regression model with factors for hospital and treatment group. We will randomly impute 10 datasets, perform the analysis on each imputed data set, and obtain final effect estimates by combining results using Rubin's rules.

7. Analysis of secondary objectives

7.1. To assess whether the study regimen also cures low-level rifampicin resistant TB

The interaction between treatment differences and initial rifampicin resistance mutations detected will be tested in the logistic regression model from the main analysis. Initial rifampicin resistance mutations detected is categorized as:

- None
- Low-level
- High

7.2. To assess the effectiveness of FDA vital staining screening at two weeks of treatment for early switch of non-responding rifampicin resistant TB to MDR-TB treatment, versus clinical evaluation based on fever

The proportions of initial rifampicin resistant cases started on MDR treatment within the first 2 months of primo-treatment as well as the proportion of initially present resistance never switched to the MDR regimen and not cured will be compared between the two arms using the chi-squared or Fisher's exact test.

7.3. To assess the negative predictive value of conversion at 2 weeks for relapse

The complement of the Negative Predictive value (with 95% Wilson CI) of conversion in the intervention arm will be estimated as the % of relapses among those who are already completely negative on AFB smear (auramine or FDA) at week 2.

7.4. To estimate the proportion of acquired rifampicin resistance among failures and relapses

Proportions per treatment arm will be estimated together with 95% Wilson confidence intervals, using the number of failure / relapse cases without mutation detected at diagnosis as the denominator.

7.5. To compare the predictive value of auramine with FDA staining at 2 weeks to predict unfavourable outcomes

The ROC curves of auramine resp. FDA staining at 2 weeks to predict adverse treatment outcome will be plotted and compared using the nonparametric approach of DeLong, DeLong, and Clarke-Pearson (1988). In addition, the proportion of referrals to rapid DST will be compared between auramine and FDA using the chi-squared or Fisher exact test.

7.6. To compare weight gain and fever resolution between the two regimens

Weight gain between enrollment and end of treatment will be compared between the two regimens using the t-test or Wilcoxon rank sum test, as appropriate. Fever resolution after two weeks of treatment will be compared between the two regimens using the chi-squared of Fisher's exact test.

8. Safety analyses

8.1. Safety analysis population

For the analysis of safety outcomes, all patients who effectively received any drug (i.e. at least one treatment dose) are included in the safety analysis in the intervention group they actually received (all-patients-treated approach).

8.2. Safety outcomes and analyses

8.2.1. Serious adverse events

General aspects: Serious adverse events (SAEs) will be coded using the Medical Dictionary for Regulatory Activities (MEDDRA) and will be reported based on MEDDRA preferred terms and body systems. All SAEs will be analyzed based on counts of patients with a specific category and not on counts of individual adverse events. The relationship between SAEs and treatment is determined by the local investigator and categorized as "drug-related" if possibly, probably or definitely related to treatment.

The following summaries will be present by treatment group:

a. description of all deaths of patients up to 12 months after treatment stop;

b. the total number of patients with any SAE by preferred term and body system up to 12 months after treatment stop;

c. the total number of patients with any drug-related SAE by preferred term and body system up to 12 months after treatment stop;

Statistical tests of association, using logistic regression, adjusted for hospital and stratum (new/ retreatment case), between treatment and occurrence of SAE will be performed for the following categories:

a. number (%) of patients with any SAE, any drug-related SAE;

b. number (%) of patients with the most common SAEs and drug-related SAEs (most common: with an incidence of 10% or more in at least one treatment group).

8.2.2. Grade 3-4 Liver Toxicity

The total number of patients with grade 3, grade 4 and grade 3-4 liver toxicity up to 12 months after treatment stop will be presented. Statistical tests of association, using logistic regression, adjusted for hospital and stratum (new/ retreatment case), between treatment and liver toxicity will be performed.

Due to the multiplicity of tests performed, the statistical tests of association between treatment and safety outcomes will not be interpreted as definite proof of more or less harm of any of the treatment but will be used as a screening tool. The use of a statistical tests will allow the exclusion of any observed difference which may be due to chance.

9. References

Rubin DB (1976) "Inference and Missing Data," Biometrika, 63, 581–592.

DeLong ER, DeLong DM, Clarke-Pearson DL (1988) "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach " Biometrics 44, 837-845.