



# Efficacy Statistical Analysis Plan

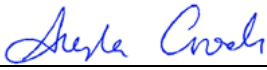
**Protocol Title:** A Phase 3, Open-Label Partially Randomized Trial to Evaluate the Efficacy, Safety and Tolerability of the Combination of Moxifloxacin plus PA-824 plus Pyrazinamide after 4 and 6 months in Adult Subjects with Drug-Sensitive Smear-Positive Pulmonary Tuberculosis and after 6 months of Treatment in Adult Subjects with Multi-Drug Resistant, Smear Positive Pulmonary Tuberculosis.

**Protocol Number:** NC-006-(M-Pa-Z)


**Protocol Name:** STAND (Shortening Treatments by Advancing Novel Drugs)

**Version:** 1.1

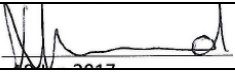
Author name: Angela Crook  
Author position: Trial Statistician

Author signature:   
Date: 18 Jan 2017

Approval name: Carl Mendel  
Approval position: Senior Vice President, Research and Development

Approval signature:   
Date: 18 Jan 2017

Approval name: Leandra Lombard  
Approval position: Global Clinical Project Manager

Approval signature:   
Date: 18 Jan 2017

## Table of Contents

<b>1. Introduction</b>	<b>4</b>
<b>2. Primary Efficacy Endpoint</b>	<b>4</b>
<b>3. Definitions and data handling issues</b>	<b>5</b>
3.1. Definitions	5
3.2. Inability to produce sputum	6
3.3. Isolated positive cultures	6
3.4. Timing of events	6
<b>4. Analysis populations</b>	<b>6</b>
4.1. Exclusions from ITT analysis (late screening failures)	7
4.2. Additional exclusions from MITT analysis	7
4.3. Additional exclusions from PP analysis	7
4.4. Lost to Follow-up or Early Withdrawal	8
4.5. Definition of adequate treatment	9
4.6. Determining cause of death	9
<b>5. Baseline comparisons of key characteristics</b>	<b>10</b>
<b>6. Classification of primary endpoint status</b>	<b>10</b>
6.1.1. Favourable status (all analyses)	10
6.1.2. Unfavourable status in ITT population	10
6.1.3. Unfavourable status in MITT population	10
6.1.4. Unfavourable status in PP population	10
<b>7. Determination of statistical non-inferiority</b>	<b>11</b>
<b>8. Sensitivity analyses of primary endpoint analysis</b>	<b>11</b>
<b>9. Secondary efficacy analyses of primary endpoint</b>	<b>12</b>
9.1. Unadjusted results	12
9.2. Additional covariate adjustment	12
9.3. Time to event unfavorable outcome analysis	12
9.4. Status at end of treatment	12
9.5. Status at 12 months in those who are favourable at end of treatment	12
<b>10. Secondary efficacy endpoints</b>	<b>13</b>
10.1. Incidence of bacteriologic failure or relapse at 24 months	13
10.2. Rate of change in time of sputum culture positivity (TTP)	13
10.3. Time to sputum culture conversion to negative status	13

Efficacy Statistical Analysis Plan

10.4. Culture conversion status at 4, 8, 12 and 17 weeks ..... 13

10.5. TB symptoms ..... 13

10.6. Patient reported health status ..... 13

**11. Sub-group analyses ..... 13**

**12. Reasons for treatment failure as determined by the local PI..... 14**

**13. Appendix 1 ..... 15**

**Version History:**

<b>Version Number/Date</b>	<b>Change</b>
0.1 18Jul2014	-
0.2 29Aug204	Incorporated comments from DE and circulated for wider review
0.3 06Oct2014	Incorporated changes following ST meeting 17Sep2014
0.4 21Oct2014	Incorporated changes following email to ST 06Oct2014
1.0 25Nov2014	Finalised version 1.0
1.1 18Jan2017	Updated to reflect revised bacteriology algorithm; changed INH-mono resistant patients to be classified as late exclusions; other minor clarifications; updated names for signature on front page

## 1. Introduction

This document outlines the efficacy statistical analysis plan (SAP) for the protocol STAND, a phase 3 trial of three experimental regimens of the combination regimen of moxifloxacin, PA-824 and pyrazinamide (MPaZ).

The primary efficacy analysis will be conducted using culture results from liquid culture (MGIT) and will evaluate the hypothesis that in the patients with drug-sensitive tuberculosis (DS-TB) treated with any of the experimental MPaZ regimens, the incidence of bacteriologic failure or relapse or clinical failure at 12 months from the start of therapy is non-inferior to the proportion observed in patients who are treated with the 6 month standard isoniazid, rifampicin, pyrazinamide, ethambutol (HRZE/HR) regimen.

The three experimental regimens detailed below will be compared to the control regimen in the order specified. For the second and third comparison, statistical inference will only be made if non-inferiority is declared on the preceding comparison:

*First comparison:* MPaZ with 200mg PA-824 (MPa<sub>200</sub>Z) for 6 months vs HRZE/HR

*Second comparison:* MPaZ with 200mg PA-824 (MPa<sub>200</sub>Z) for 4 months vs HRZE/HR

*Third comparison:* MPaZ with 100mg PA-824 (MPa<sub>100</sub>Z) for 4 months vs HRZE/HR

300 patients per randomised arm will be recruited. See Appendix 1 for details of the sample size. An additional non-randomised group of up to 300 patients with multi-drug resistant (MDR) TB will be treated with regimen MPa<sub>200</sub>Z for 6 months. No formal statistical analysis is planned for this group and unless otherwise stated all analyses outlined in this document refer to the analysis of the 4 randomised groups of the DS-TB patients.

While STAND is an open label study, only members of the Data Safety Monitoring Committee (DSMC) and the unblinded statistician(s) will have access to data **grouped by study arm**. Access to data by investigators, CRO and TB Alliance staff will not allow any access to data grouped by study arm. In addition, site staff will be strongly discouraged from attempting to aggregate by treatment arm at a site level.

## 2. Primary Efficacy Endpoint

The primary efficacy endpoint will be the incidence of bacteriologic failure or relapse or clinical failure within 12 months from start of therapy. See section 6 for the detailed definition of an “unfavourable response” at 12 months.

There will be three main analyses of the primary efficacy endpoint: A per-protocol (PP) analysis, a modified intent to treat (MITT) analysis and an intent to treat (ITT) analysis. As the true bacteriologic treatment failure and relapse rate in the control group is expected to be between 3% and 5%, “unfavourable” rates in any defined ‘ITT’ population will likely be increased considerably by factors other than bacteriologic or clinical treatment failure and relapse. The PP and MITT analyses will be considered primary; results of ITT and additional sensitivity analyses (listed in section 8) will need to be consistent to draw a strong conclusion.

### 3. Definitions and data handling issues

#### 3.1. Definitions

**Positive culture** refers to the culture being positive for M.tb. N.B.: In the absence of the speciation data, the unconfirmed result will be treated as positive for M.tb. False positive or contaminated will be treated as missing. Specimens classified as non-tuberculous mycobacteria (NTM) and negative for M.tb will be treated as contaminated (adequate conditions for growth of M.tb exist if NTM grow). The bacteriology algorithm for reporting MGIT results can be found in STAND Laboratory Manual. Two sputum samples per visit are collected at each visit throughout treatment and follow-up. The culture result for a given visit is established using all samples obtained for that visit. A positive culture takes precedence over a negative culture at the same visit.

**Culture negative status** is achieved when a patient produces at least 2 negative culture results at different visits (at least 7 days apart) without an intervening positive culture result for M.tb. The date of the first negative culture of these two is the date at which culture negative status was obtained. Once obtained, culture negative status continues until there are two positive cultures at different visits (at least 7 days apart), without an intervening negative culture, or until there is a single positive culture not followed by two negative cultures. Culture negative status can be achieved at any time during treatment or follow-up but before any re-treatment and can be re-established.

Patients with two contaminated or missing samples at a given visit (from Month 2 onwards) will be asked to return to produce two more sputum samples. A single culture result or inability to produce sputum (see section 3.2) at 12 months will be sufficient to count that patient as having data for primary endpoint classification.

**Treatment failure or Bacteriologic failure** is defined as being declared an unfavourable status (as defined in section 6) at or before the end of treatment (either 4 or 6 months) or failing to attain culture negative status and being declared an unfavourable outcome.

**Relapse or Bacteriologic relapse** is defined as failing to maintain culture negative status or being declared an unfavourable outcome after the end of treatment (either 4 or 6 months) in those patients who attained culture negative status by the end of treatment, and were re-infected with the **same** (as per the STAND Laboratory Manual) *Mycobacterium tuberculosis (MTB)* strain.

**Reinfection or Bacteriologic Reinfection** is defined as failing to maintain culture negative status or being declared an unfavourable outcome after the end of treatment (either 4 or 6 months) in those patients who attained culture negative status by the end of treatment and were infected with a *Mycobacterium tuberculosis (MTB)* strain that is **different** (as per the STAND Laboratory Manual) from the infecting strain at baseline.

**Clinical failure** is defined as a change from protocol-specified TB treatment or retreatment for TB during follow up as determined by the treating physician, or TB-related death.

The **treatment period** is defined as either 4 or 6 months from start of therapy depending on allocated treatment regimen and the **follow-up period** is defined as the period after the end of treatment to the end of follow-up.

### 3.2. Inability to produce sputum

In general, inability to produce sputum is treated as being equivalent to having a negative (favourable) culture result. This includes the rare situation where a patient who never achieves culture negative status due to inability to produce sputum, after TB confirmation via culture has been confirmed on the applicable baseline sample, but completes 12/24 months follow-up without clinical or microbiological evidence of relapse. Such a patient will be considered to have a favourable outcome.

### 3.3. Isolated positive cultures

It is known that occasionally patients produce sputum samples that are isolated positives, that is a positive result preceded by a series of negative cultures and followed thereafter by at least 2 negative cultures without an intervening positive result. This phenomenon may be the result of a sealed cavity breaking down or laboratory contamination and does not in itself signify that the patient is relapsing. In the event of a single positive culture result occurring in a patient who has previously been classified as having culture negative status (in the absence of any retreatment), the patient will not be classified as a recurrence unless a second positive culture result is obtained at a separate visit (at least 7 days apart) without an intervening negative culture or unless the patient is lost to follow up or completes the study (and is unable to be brought back) before two negative cultures are obtained. The clinical condition of the patient will also be taken into account in deciding whether re-treatment is indicated.

Most of the experience with isolated positives has been with solid culture. Because liquid culture is more sensitive, it is possible that more than one isolated positive may occasionally occur. Therefore, the clinical condition of the patient will also be taken into account when deciding whether re-treatment is indicated and in determining the outcome. For example, if a patient after being culture negative has two positive cultures in a row, but is deemed to be doing well clinically, the investigator may choose to leave the patient untreated on clinical grounds. In such a case, so long as two consecutive negative cultures are eventually obtained in the absence of treatment, the patient will not be classified as a recurrence.

### 3.4. Timing of events

In all analyses, visit date rather than day or week number will be used to define the timing of events. The 4 month regimens will be taken as 17 weeks (119 days) and the 6 month regimens as 26 weeks (182 days) from the start of therapy. For the end of treatment visit (months 4/6), a  $\pm 1$  week window will be applied (as per the protocol). For the 12 and 24 month visits, a window of  $\pm 2$  weeks will be applied (as per the protocol).

## 4. Analysis populations

The analysis populations for efficacy analyses are:

- The **All Randomised** population defined as all patients randomised with no exclusions
- The **Intent to treat (ITT)** population is defined as All Randomised excluding late screening failures (see 4.1)
- The **Modified intent to treat (MITT)** population is defined as the ITT population with extra exclusions (See 4.2)
- The **Per-protocol (PP)** population is defined as the MITT population with extra exclusions (see 4.3)

#### **4.1. Exclusions from ITT analysis (late screening failures)**

1. Patients randomised into one of the DS arm who on subsequent receipt of the results of the MGIT drug sensitivity tests are found to be rifampicin and/or pyrazinamide and/or fluoroquinolone resistant and/or INH mono-resistant) documented from Day 1 (baseline) sputum samples (or screening or out to Week 4 if the baseline is contaminated or negative) (late exclusions from the study). These patients will be withdrawn from the study.
2. Patients without culture confirmation of M. tb at Day 1 (baseline) sputum samples (or screening or out to Week 4 if the baseline is contaminated or negative) (late exclusions from the study)
3. Patients withdrawn from treatment because of a protocol violation at enrolment (late exclusions from the study, based on data collected prior to randomisation).

#### **4.2. Additional exclusions from MITT analysis**

1. Patients who, having completed treatment, are lost to follow-up or withdrawn from the study, their last status being culture negative and their last positive culture result (“isolated positive culture”) followed by at least two negative culture results at different visits (at least 7 days apart, without an intervening positive culture)
2. Women who become pregnant during treatment and stop their allocated treatment
3. Patients who died during treatment from violent or accidental cause (e.g. road traffic accident). N.B.: This does not include death from suicide, which will be considered an unfavourable outcome.
4. Patients who died during follow-up (after the end of treatment) with no evidence of failure or relapse of their TB, their last status being culture negative and their last positive culture result (“isolated positive culture”) followed by at least two negative culture results at different visits (at least 7 days apart), and who have not already been classified as unfavourable.
5. Patients who, after being classified as having culture negative status, are re-infected with a new strain from that with which they were originally infected. Reinfection will be defined specifically as a patient infected with a strain that is genetically different from the initial strain (see Mycobacteriology Laboratory Manual).
6. Patients who are able to produce sputum at 12 months, but whose 12-month visit sputum samples are all contaminated or missing, who cannot be brought back for repeat cultures, provided they have not already been classified as unfavourable and provided their last positive culture was followed by at least two negative cultures. N.B.: This does not apply to patients who are unable to produce sputum at 12 months, or to patients who are able to be brought back subsequently and produce negative cultures.

Patients in categories 1-5 above who had already been classified as having an unfavourable outcome will not be excluded.

#### **4.3. Additional exclusions from PP analysis**

1. Patients lost to follow-up or withdrawn before the end of treatment, unless they have already been classified as having an unfavourable outcome.

2. Patients whose treatment was modified or extended for reasons (e.g. an adverse drug reaction or pregnancy) other than an unfavourable therapeutic response to treatment, unless they have already been classified as having an unfavourable outcome.

3. Patients not meeting the definition of having received an adequate amount of their allocated study regimen (see section 4.5 for definition), provided they have not already been classified as having an unfavourable outcome.

4. Patients who are classified as “major protocol deviations” (see below), unless they have already been classified as having an unfavourable outcome on the basis of data obtained prior to the protocol deviation.

A list of all protocol deviations will be compiled throughout the course of the study.

A **Major Protocol Deviation** is defined as a serious protocol deviation which is likely to affect to a significant degree the safety or physical or mental integrity of the subjects of the trial or the scientific value of the trial. These patients will be included in the ITT and MITT analyses, but not in the Per Protocol analysis. A list of all major protocol deviations will be approved by the study Coordinating Investigator before database lock.

A **Minor Protocol Deviation** is defined as a technical deviation which does not result in harm to the trial subjects or significantly affect the scientific value of the reported results of the trial.

#### **4.4. Lost to Follow-up or Early Withdrawal**

Lost to Follow-up or Early Withdrawals *before* the end of the treatment (month 4 or 6) are considered as unfavourable outcomes for ITT and MITT. However these patients will be excluded from the Per Protocol analysis. The MITT and Per Protocol analyses will consider Lost to Follow-up *after* end of treatment (month 4 and 6) as unassessable unless at the time of default from follow-up the patient a) was already classified as having an unfavourable outcome, b) did not have culture negative status, or c) had a positive culture result (“isolated positive culture”) not followed by at least two negative culture results at different visits (at least 7 days apart), in which cases the patient will be classified as having an unfavourable outcome. We believe this is the most appropriate approach for the primary analysis because together with the non-tuberculosis deaths, this group is likely to considerably out-number the bacteriological failures and relapses. These patients will be considered as having an unfavourable outcome in the ITT analysis.

There is a clear precedent for this analytic approach in other TB trials, and these trials also provide examples of why the inclusion of the losses to follow-up as unfavourable greatly affects the results.

Data from the Priftin trial which led to accelerated approval of rifapentine and a trial conducted by the International Union Against TB & Lung Disease (IUATLD) in African and Asian sites illustrate the problems associated with classifying all losses to follow-up and deaths as having an unfavourable outcome.



In the Priftin trial bacteriological relapses occurred in 5% of patients on the rifampicin based regimen compared to 11% on the rifapentine based regimen. Approximately one third of patients were lost to follow-up and when this group combined with patients unassessable for other reasons were added to the bacteriological failures, the rates increased to 53% and 57% respectively. The true bacteriological relapses were greatly outnumbered by these other groups. At the time of the licensing submission to the FDA it was recognised that because there were a substantial number of patients likely to be unassessable the main focus should be on the relapse rates. In the final statistical report the results were first reported excluding those unassessable and then assuming all losses had an unfavourable outcome and finally assuming all losses had a favourable outcome.

In the study conducted by the IUATLD the published failure/relapse rates 12 months after stopping treatment based on 1044 assessable patients were 4% for the control regimen and 10% and 14% in each of the experimental arms. If the 311 unassessable patients were considered to have an unfavourable outcome these rates would increase to 24%, 32% and 35% respectively. The 311 unassessable patients were not evenly distributed across the three trial arms. There were 42 deaths, of which 20 occurred in one of the experimental arms (the more efficacious of the two) and 11 in each of the other, a difference which was not considered to be due to the treatment, but due to chance. There were also imbalances among those without a bacteriological assessment (7 in one arm versus 19 and 22 in the other two arms) and in the distribution of losses to follow-up.

#### **4.5. Definition of adequate treatment**

The definition of adequate treatment sets a limit for the amount of treatment missed. Patients not taking the adequate amount of treatment by this definition will be excluded from the PP analysis.

For patients allocated to a 6 month regimen, to meet the definition of adequate treatment they must have taken at least 146 doses (80%) of their allocated 182 day (26 weeks) treatment regimen within 203 days of starting therapy (i.e. 26 weeks plus an allowable 3 weeks extension ).

For patients allocated to a 4 month regimen, to meet the definition of adequate treatment they must have taken at least 96 doses (80%) of their allocated 119 day (17 weeks) treatment regimen within 140 days of starting therapy (i.e. 17 weeks plus an allowable 3 weeks extension ).

Patients in the control arm are additionally required to have taken at least 80% of their allocated intensive treatment.

#### **4.6. Determining cause of death**

A list of all ***TB-related*** and ***non-TB-related deaths*** will be generated and approved by a review committee **blind to randomised arm** before database lock. Similarly a list of violent or accidental deaths will be generated.

## **5. Baseline comparisons of key characteristics**

The following baseline characteristics of patients on each of the five study arms (including the non-randomised MDR group) will be tabulated and compared for each analysis population: age, gender, race, region, site, smoking status, HIV status/CD4 count, cavitation (based on blinded, central assessment of the chest radiograph (refer to Central Chest X-ray Manual), initial bacterial load in sputum as indicated by baseline Time to Positivity (TTP) result from MGIT, resistance to first and second line antibiotics and whether or not the patient is taking efavirenz.

## **6. Classification of primary endpoint status**

Patients will be classified as having a favourable, unfavourable or unassessable status at 12 months from the start of therapy.

### **6.1.1. Favourable status (all analyses)**

Patients with a negative culture status at 12 months from start of therapy (from 50 weeks to 54 weeks), who had not already been classified as having an unfavourable outcome, and whose last positive culture result ("isolated positive culture") was followed by at least two negative culture results.

### **6.1.2. Unfavourable status in ITT population**

Patients in the ITT analysis population who do not have a favourable outcome by 12 months will be considered to have an unfavourable response in the ITT analysis.

### **6.1.3. Unfavourable status in MITT population**

1. Patients not classified as having achieved or maintained culture negative status when last seen, or
2. Patients previously classified as having culture negative status who, following the end of treatment, have two positive cultures without an intervening negative culture, (however, see Section 3.3 for an exception), or
3. Patients who had a positive culture not followed by at least two negative cultures when last seen, or
4. Patients dying from any cause during treatment, except from violent or accidental cause (e.g. road traffic accident), not including suicide (eg, suicide will be considered an unfavourable outcome) or
5. Patients definitely or possibly dying from TB related cause during the follow-up phase or
6. Patients requiring an extension of their treatment beyond that permitted by the protocol, a restart or a change of treatment for any reason except reinfection or pregnancy, or
7. Patients failing to complete an adequate course of treatment, as defined in section 4.5, who were unassessable at 12 months, or
8. Patients lost to follow up or withdrawn from the study before the end of treatment

### **6.1.4. Unfavourable status in PP population**

1. Patients not classified as having achieved or maintained culture negative status when last seen, or
2. Patients previously classified as having culture negative status who, following the end of treatment, have two positive cultures without an intervening negative culture, (however, see Section 3.3 for an exception), or
3. Patients who had a positive culture not followed by at least two negative cultures when last seen, or
4. Patients dying from any cause during the 6 month treatment phase, except from violent or accidental cause (e.g. road traffic accident), not including suicide (e.g., suicide will be considered an unfavourable outcome), or
5. Patients definitely or possibly dying from TB related cause during the follow-up phase, or
6. Patients requiring a restart or a change of treatment because of an unfavourable outcome with or without bacteriological confirmation, i.e. on bacteriological, radiographic or clinical grounds

## **7. Determination of statistical non-inferiority**

The first comparison to be tested will be the MPa<sub>200</sub>Z 6 month experimental regimen against the standard regimen. The second comparison to be tested will be the MPa<sub>200</sub>Z 4 month experimental regimen against the standard regimen. The third comparison to be tested will be the MPa<sub>100</sub>Z 4 month experimental regimen against the standard regimen. To preserve the type I error rate, the second comparison will only be tested if non-inferiority is demonstrated with the 6 month experimental regimen. Similarly the third hypothesis will only be tested if non-inferiority of the PA-824 200mg 4 month regimen has been demonstrated. This strategy means no adjustments are required for multiple comparisons.

Non-inferiority will be assessed using the upper bound of the two-sided 95% confidence interval for the difference between the proportion of participants who are classified as having an unfavourable status on the intervention and the control regimen.

If the upper bound of the two-sided 95% confidence limit for the difference, proportion with unfavourable status on intervention arm less the proportion with unfavourable outcome on the control arm, is less than 12% (the margin of non-inferiority) the intervention will be considered to be non-inferior to the control arm on that comparison.

The PP and MITT analyses will be considered primary. However, in order to declare an intervention to be non-inferior, consistent results should be obtained from the ITT analysis and from sensitivity and subgroup analyses, although not all of these will necessarily satisfy the strict definition of non-inferiority.

All primary efficacy analyses of the three main analysis populations will be adjusted for HIV status and cavitation, using a stratified analysis of the risk difference from each stratum using Cochran-Mantel-Haenszel weights.

When interpreting the results, particularly if they are close to the margin, the STAND study team will weigh all the data including the unfavourable rates with their component parts, safety and adherence in considering the benefit of any experimental arm(s). However, if the control arm unfavourable rate is observed to be much lower than what is anticipated we acknowledge there may be difficulty in interpreting the results and there may be situations where it may not be possible to draw positive conclusions from the trial. A table and commentary of hypothetical observed results with variable N assessable and % unfavourable in the control arm is given Appendix 1.

## **8. Sensitivity analyses of primary endpoint analysis**

In addition to analysing the primary endpoint data by per protocol and MITT and ITT, it is planned to conduct the following sensitivity analyses:

1. An analysis of all randomised patients
2. An analysis of the ITT population, with missing outcomes data treated as unfavourable
3. An analysis of the ITT population, with missing outcomes data treated as favourable

4. An analysis of the ITT population, with missing outcomes data treated according to status at the last observation (last two samples)
5. An analysis of patients in the MITT and PP populations where reinfections are classified as unfavourable outcomes
6. An analysis of the MITT and PP populations treating all deaths as unfavourable
7. An analysis of the MITT population in which patients who changed treatment for reasons other than failure or relapse were classified according to their status at the end of follow-up.

In these sensitivity analyses, reinfections are considered as missing outcome data (except for point 5).

## **9. Secondary efficacy analyses of primary endpoint**

### **9.1. Unadjusted results**

Unadjusted difference in the proportion of unfavourable for each pairwise comparison (intervention arm-control arm) with 95% confidence intervals will be presented.

### **9.2. Additional covariate adjustment**

Additional baseline covariates listed in section 6 will be considered for a secondary adjusted analysis for the primary endpoint. This will be a secondary analysis, assuming baseline covariates are well balanced across the treatment arms.

### **9.3. Time to event unfavorable outcome analysis**

Time to an unfavourable outcome will be analysed with Kaplan Meier plots and Cox's proportional-hazards regressions analysis. These analyses will be performed according to PP, MITT and ITT endpoint classifications. Patients withdrawn from the study during treatment will be censored at the time last seen in the study and classified as having an unfavourable endpoint at that time. Those changing treatment or restarting treatment for whatever reason will be classified as having an unfavourable endpoint at the time of change. Patients lost during the follow-up phase who have not already been classified as having an unfavourable endpoint will be classified according to their status at the time they are last seen, thus a patient who had not yet achieved culture negative status, or a patient with a positive culture even of a low colony growth not followed by two negative cultures would be classified as having attained an unfavourable endpoint.

### **9.4 Status at end of treatment**

Status at end of treatment based on assessable patients based on culture result at end of chemotherapy, 4 or 6 months, using the penultimate month culture if the culture result at end of chemotherapy is unavailable will be compared across randomised groups.

### **9.5 Status at 12 months in those who are favourable at end of treatment**

Status at 12 months after the start of therapy in those patients who have a favourable response at the end of treatment and are assessable at 12 months or have already met the definition of unfavourable response.

## **10. Secondary efficacy endpoints**

### **10.1. Incidence of bacteriologic failure or relapse at 24 months**

Efficacy analyses as described for the 12 month endpoint will be repeated for the 24 month endpoint.

### **10.2 Rate of change in time of sputum culture positivity (TTP)**

Speed of decline of sputum viable counts as assessed by non-linear mixed effects (NLME) modelling of the time to positivity (TTP) in MGIT will be performed by the Statistics Team performing the safety analyses for STAND. Details of this will be documented elsewhere.

### **10.3 Time to sputum culture conversion to negative status**

Time to culture negative status (first of two negative cultures without an intervening positive culture) will be analysed using survival analysis techniques, Kaplan Meier plots and Cox proportional hazard regression.

### **10.4. Culture conversion status at 4, 8, 12 and 17 weeks**

Patients will be classified as being culture positive, culture negative, dead or unassessable at 4, 8, 12 and 17 weeks. Every effort will be made to obtain a sputum sample from all patients but it is recognised that some patients may not have produced any sputum in the preceding week and may be unable to do so when requested. Patients who cannot produce sputum will be classified as being culture negative at that time point. The proportion culture negative will be those classified as being culture negative divided by the total considered culture negative, culture positive or have died.

In a separate sensitivity analysis patients who are unassessable at this time point for whatever reason (e.g. culture contaminated) will be considered to be culture positive unless they are culture negative on at least two previous assessments in which case they will be considered to be culture negative. This does not include patients who could not produce sputum at 4, 8, 12 and 17 weeks, who are considered to be culture negative.

Differences from the control regimen in culture negativity at weeks 4, 8, 12 and 17 will be expressed as a difference in proportions with corresponding 2-sided 95% confidence intervals.

### **10.5. TB symptoms**

Analyses of symptoms will include time to resolution of symptoms. This will be defined as the first of two consecutive attendances when the patient is free from TB related symptoms.

Please refer to Safety SAP for other analyses of TB symptoms.

### **10.6. Patient reported health status**

Please refer to Safety SAP.

## **11. Sub-group analyses**

To assess consistency of outcome, the following sub-group analyses (with tests for interaction) of the primary endpoint on the per-protocol and MITT analysis populations will be performed according to: age; gender; race; region; smoking status; HIV status/CD4 count; cavitation (based on blinded, central assessment of the chest radiograph), initial bacterial load in sputum as indicated by baseline TTP result from MGIT; whether the patient is taking efavirenz or not during the treatment period.

**12. Reasons for treatment failure as determined by the local PI**

Reason(s) that led the site investigator to conclude that an individual patient failed treatment or relapsed will be classified as a) bacteriology alone, b) clinical deterioration alone, c) radiological deterioration alone, d) bacteriology plus clinical deterioration, e) bacteriology plus radiological deterioration, f) clinical deterioration plus radiological deterioration, or g) bacteriology plus clinical deterioration plus radiological deterioration. These classifications will be tabulated and compared by treatment arm.

### 13. Appendix 1

#### Details of sample size and issues relating to choice of delta, the non-inferiority margin

The following assumptions have been made in calculating the numbers of subjects needed for the primary endpoint analysis.

- The unfavorable rate in the control regimen will be 16%, consistent with the control arm unfavourable rate in a recent Phase 3 TB trial.
- An unassessable rate of 15%, consistent with the overall unassessable rate in a recent Phase 3 TB trial.
- An additional 8% excluded for isoniazid mono-resistance, consistent with the overall percentage observed in a recent Phase 3 TB trial.
- An acceptable non-inferiority margin (delta) is 12% (see below)
- The power to demonstrate non-inferiority is 90% and the significance is 2.5% (1-sided).

Under these assumptions we would require 255 subjects for each treatment arm, giving a total sample size of 1,020 DS-TB subjects. Although these are considered to be reasonable assumptions based on recent data, in order to allow for the possibility of higher unfavorable and/or unassessable rates, or somewhat greater than 8% of subjects with INH mono-resistance, 300 DS-TB subjects will be recruited to each randomized arm giving a total of 1,200 subjects with DS-TB. Note that the primary analysis of the DS-TB subjects will exclude the subjects with isoniazid mono-resistance.

Examples of maximum observed effects which still meet the 12% margin for non-inferiority are given in Table A1 below for varying levels of the number N assessable and the proportion unfavourable in the control arm. Scenarios with observed differences of 6% or more are considered highly unlikely given the sample size for STAND is based on an expected MITT control arm unfavorable rate of 16% based on a recently completed phase 3 efficacy trial.

**Table A1. Scenarios demonstrating statistically non-inferior results**

Values are the maximum % difference between the intervention unfavourable rate minus the control, which if observed would be compatible with a 12% margin

N assessable per group	5% unfavourable in control arm	10% unfavourable in control arm	15% unfavourable in control arm	20% unfavourable in control arm
150	5.3%	4.0%	3.3%	2.6%
200	6.0%	5.0%	4.0%	3.5%
250	6.8%	6.0%	4.8%	4.0%