Statistical Analysis Plan

endTB-Q (Evaluating Newly Approved Drugs in Combination Regimens for Multidrug-Resistant TB with Fluoroquinolone Resistance)

Protocol Number: NCT03896685 (ClinicalTrials.gov)

Statistical Analysis Plan (SAP)	**Effective Date** : 04-Oct-2024
Version Number and Date: 1.0, 02-Oct-2024	

Revision	
1.0 Date 02-Oct-2024	Initial version

Name and function	**Date**	**Signature**
Authors		
Maelenn Gouillou, Biostatistician	03/10/2024	Maelenn Gouillou Maelenn Gouillou (Oct 3, 2024 19:38 GMT+2)
Reviewed by		
Maryline Bonnet, Co-Investigator	03/10/2024	Maryline Bonnet Maryline Bonnet (Oct 3, 2024 13:44 GMT+2)
Elisabeth Baudin, Co-Investigator, Biostatistician	03/10/2024	Elisabeth BAUDIN
Approved by		
Lorenzo Guglielmetti, Co-Principal Investigator	03/10/2024	Lorenzo Guglielmetti Lorenzo Guglielmetti (Oct 3, 2024 12:11 GMT+2)
Carole Mitnick, Co-Principal Investigator	03/10/2024	Carole Mitnick

LIST OF ABBREVIATIONS

AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
Be	Bedaquiline
BMI	Body Mass Index
C	Clofazimine
CA++	Calcium
CBC	Complete Blood Count
CD4	CD4 Cell Count
CI	Confidence Interval
De	Delamanid
DR	Drug-resistant
DS	Drug-Susceptible
DSMB	Data and Safety Monitoring Board
DST	Drug Susceptibility Test
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EEO	Exploratory Efficacy Objective
ESO	Exploratory Safety Objective
endTB-Q	Study name
EVG	Event Validation Group
FQ	Fluoroquinolones
HbA1c	Hemoglobin A1c
HBc	Hepatitis B Core
HBsAg	Hepatitis B Surface Antigen
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HCVAb	Hepatitis C virus antibody
HIV	Human Immunodeficiency Virus
ITT	Intent-to-treat
IWRS	Interactive Randomization System
IP	Investigational Product
K+	Potassium
Li	Linezolid
LJ	Löwenstein Jensen
LTFU	Loss to Follow Up
MDR-TB	Multidrug-resistant Tuberculosis
MedDRA	Medical Dictionary for Regulatory Activities
MG++	Magnesium
MGIT	Mycobacterial Growth Indicator Tube
mITT	Modified Intent-to-Treat
MTB	Mycobacterium Tuberculosis
PI	Principal Investigator
PO	Primary Objective
PP	Per-protocol
PT	Preferred Term
QTcF	QT Interval Heart-Rate-Corrected by Fridericia Formula
RIF	Rifampicin
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCC	Sputum culture conversion
SD	Standard Deviation
SEO	Secondary Efficacy Objective
SSO	Secondary Safety Objective
SOC	System Organ Class
SOP	Standard operating procedure
TB	Tuberculosis
TSH	Thyroid-stimulating Hormone
TTP	Time to (culture) positivity
WHO	World Health Organization

1. INTRODUCTION

This standard methodology document, the statistical analysis plan, explains the rules and conventions to be used in the presentation and analysis of efficacy and safety results for the clinical study "endTB-Q, evaluating newly approved drugs in combination regimens for multidrug resistant TB with fluoroquinolone resistance". It describes the data and endpoints to be summarized and analyzed, including specifics of the statistical analyses to be performed. SAP version 1.0 was based on protocol Version 4.3 dated July 05 2022. Any changes made after the approval of this document will be recorded in the cover page and in the appropriate section of the clinical study report. Any post hoc analyses that are performed will be detailed in the same section of the clinical study report.

2. STUDY OBJECTIVES

The primary objective of the study is to assess whether the efficacy of the experimental arm at Week 73 is non-inferior to that of the control (PO).

The secondary objectives are:

1. To compare the efficacy of the experimental arm at Week 104 to that of the control (SEO1).
2. To compare the frequency of and time to early treatment response (culture conversion) in the experimental arm to that of the control (SEO2).
3. To compare the efficacy of the experimental regimen at Week 39 to that of the control (SEO3).
4. To compare, at Week 73 and Week 104, the proportion of participants who experienced failure or recurrence in the experimental arm to that in the control arm (SEO4).
5. To compare, at Week 73 and Week 104, the proportion of participants who died of any cause in the experimental arm to that in the control arm (SSO5).
6. To compare, at Week 73 and Week 104, the proportion of participants who experienced grade 3 or higher AEs or SAEs of any grade in the experimental arm to that in the control arm (SSO6).
7. To describe, at Week 73 and Week 104, the proportion of participants who experienced AESIs in the experimental arm to that in the control arm (SSO7).

The exploratory objectives are:

1. To estimate the frequency of drug resistance amplification among treatment failures and relapses occurring by Week 104 (EEO1).
2. To compare treatment adherence in the experimental arm to that in the control (EEO2).
3. To compare the frequency of and time to severe linezolid-related toxicity between linezolid dose-reduction strategies (ESO3).
4. To compare, at Week 73 and Week 104, efficacy endpoints across linezolid dose-reduction strategies: 300 mg daily or 600 mg thrice weekly (EEO4).
5. To assess for modification of effect of experimental regimen (vs. control) by baseline extent-of-TB-disease phenotype (EEO5).

3. STUDY DESIGN

This is a randomized, controlled, open-label, multi-country Phase III trial evaluating the efficacy of a new combination regimen and strategy for treatment of fluoroquinolone-resistant MDR-TB. The study enrolls in parallel across 1 experimental and 1 standard-of-care control arms, in a 2:1 ratio. The experimental arm is a strategy in which people with extensive disease (described in Table 3) receive 39 weeks of treatment with bedaquiline, delamanid, clofazimine, and linezolid (BDCL) and people with non-extensive disease receive the same treatment for 24 weeks. Randomization is stratified by country and baseline extent-of-TB-disease phenotype, using an interactive randomization system (IWRS). Trial participation in both arms is at least until Week 73 and up to Week 104. Study follow-up ends after the scheduled Week 73 visit for the last participant randomized. In the experimental arm, treatment is for 24 or 39 weeks; duration is assigned according to baseline extent-of-TB-disease phenotype and treatment response. In the control arm, treatment is delivered according to WHO guidance (and local practice); duration is variable, typically lasting approximately 86 weeks for the conventional regimen.

4. RANDOMISATION PROCEDURE

Under protocol Version 2.2, patients were randomized to one of three arms on a 1:1:1 basis: standard of care, experimental regimen for 24 weeks, or experimental regimen for 39 weeks. Starting with protocol Version 3.0, participants were randomized to either the experimental or the control arm, in a 2:1 ratio. In the experimental arm, treatment duration was assigned according to baseline extent-of-TB-disease phenotype and treatment response as described in sections 10.8 and 10.9.

For participants randomized to an experimental arm under protocol Version 2.2, who had not completed

(or discontinued) study treatment before Version 3.0 or higher was implemented, treatment duration was revised according to baseline extent-of-TB-disease phenotype and treatment response as described in sections 10.8 and 10.9 and according to the following:

- a participant who had been randomized to 24 weeks (Version 2.2) and who met criteria for 39 weeks (Version 3.0 or higher), had their ongoing treatment extended to 39 weeks;
- a participant who had been randomized to 39 weeks (Version 2.2) and who met criteria for 24 weeks (Version 3.0 or higher) had their treatment stopped at 24 weeks or as soon as possible if already treated for more than 24 weeks.

'Applicable protocol version' is used to refer to the last protocol version under which study treatment duration has been assigned, e.g. if treatment duration was assigned by randomization under Version 2.2 and revised under Version 3.0 or higher, then Version 3.0 or higher is the applicable protocol version.

5. SAMPLE SIZE CALCULATION

The sample size calculation required assumptions about the primary outcome frequency at Week 73 for the experimental arm and the control arm, the type I error, and the non-inferiority margin.

We assumed a 73-week treatment response of 78% (the lower bound of the 95% CI around the point estimate of treatment success in NiX-TB) in the experimental arm and 75% (corresponding to the upper bound of the 95% CI around the point estimate of treatment success for the longer conventional regimens containing newly approved drugs) in the control arm. This is conservative in that it assumes a relatively small difference in treatment response.

With a 12% non-inferiority margin and a type I error set to 2.5% (one-sided), assumed loss of 6% of subjects between the randomized population and modified intent-to-treat (mITT) population and further loss of 10% between the mITT and per-protocol (PP) populations, and a 2:1 allocation ratio between experimental and control arm, a sample size of 324 randomized participants provides power greater than 80% to demonstrate the non-inferiority in both the mITT and PP populations. The analysis populations are described in section 8.

Exploratory analysis of toxicity related to linezolid dose reduction

We performed a secondary randomization to a linezolid dose-reduction strategy (1:1, 300 mg daily vs 600 mg thrice weekly) in the experimental strategy. We assumed a sample size of 216 patients randomized to the endTB-Q experimental strategy. We assumed that the proportion of AEs observed in the 300 mg daily arm would be 69% as previously reported in the only other clinical trial to investigate the 300 mg dose[i], and 52% for the 600 mg thrice weekly group. The latter estimate is extrapolated, conservatively, from a meta-analysis among patients who received linezolid doses less than or equal to 600 mg daily. It exceeds the upper limit of the 95% CI reported for AEs in that study: 46%. The frequency point estimate was 34% (95% CI, 23-46%)[ii]. Based on these effect sizes, and assuming 10% of patients are not assessable, a time-to-event analysis with a sample size of n = 216 patients randomized 1:1 to the two linezolid dosing strategies would have 80% power to detect a hazard ratio of 1.50 for severe linezolid-related toxicity with 300 mg daily compared to 600 mg thrice weekly in a two-sided, two-sample log-rank test with 5% significance.

6. TREATMENT GROUPS

The control regimen uses standard of care regimen for fluoroquinolone-resistant MDR-TB according to local practice and consistent with WHO guidance.

The endTB-Q experimental regimen is BDCL for 24 weeks or 39 weeks and is described in Table 1.

Table 1: endTB-Q regimens

Trial Regimens	Bedaquiline	Delamanid	Clofazimine	Linezolid	Duration
endTB-Q Experimental BDCL	B	D	C	L	24 or 39 weeks (duration assigned according to participant's extent-of-TB-disease phenotype* and treatment response**)
endTB-Q Control	Standard of care control, composed according to WHO Guidelines, including the possible use of D, B, or both.				

* as defined in section 10.8;
** as defined in section 10.9

Treatment duration is approximately 78 weeks and up to 104 weeks in the control group.

7. STUDY PLAN

Table 2 is the Schedule of Events, which details the timing and content of all assessments performed during the study.

Table 2: endTB-Q Clinical Trial Schedule of Events

		Treatment																				Follow-Up												
	Screening	Baseline	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W16	W20	W24	W28	W32	W36	W39	W43	W47	W53	W59	W65	W73	W81	W89	W97	W104	Sub total	Early Termination	Post-termination follow-up
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31			
Window Period			+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/- 2[#]	+/- 2	+/- 2	+/- 2	+/- 2	+/- 7	+/- 7	+/- 7	+/- 7	+/- 7	+/- 7	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 30	+/- 14	+/- 14	+/- 14	+/- 30			
Eligibility																																		
Subject Consent	X	X																														2		
Demographics	X	X																														2		
Medical History	X																															1		
Inclusion/ Exclusion	X	X																														2		
Clinical Evaluation																																		
Vital Signs	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	31	X	X
Interval Medical History		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	30	X	X
Physical Exam	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	31	X	X
TB Symptom Assessment		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	30	X	X
ECOG Assessment		X															X				X						X				X	5	X	X
Neurologic Exam		X				X				X				X	X	X	X	X	X	X	X	X	X				X				X	15	X	
Ophthalmologic Exam		X				X				X				X	X	X	X	X	X	X	X	X	X				X				X	15	X	
Mental Health Assessment		X																									X				X	3	X	
Treatment																																		
Randomization		X*													X**																	(1)		
Treatment Duration Decision																	X																	
Treatment Compliance			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	19	X[C]	
Adherence Counseling		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	X[C]	20		
Concomitant Medication Review	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	31	X	X
AE Review			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	29	X	X
Laboratory Testing																																		
CBC	X[1]					X				X				X	X	X	X	X	X	X	X	X	X									14	X	
Total Ca++, K+, Mg++	X[1]					X				X				X	X	X	X	X	X	X	X	X	X									13	X	

	Screening	Baseline	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W16	W20	W24	W28	W32	W36	W39	W43	W47	W53	W59	W65	W73	W81	W89	W97	W104	Sub total	Early Termination	Post-termination follow-up
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31			
Window Period			+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/- 2	+/-2#	+/-2	+/-2	+/-2	+/-2	+/-7	+/-7	+/-7	+/-7	+/-7	+/-7	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 30	+/- 14	+/- 14	+/- 14	+/- 30	+/-2	+/- 2	
Creatinine	X[1]					X				X				X	X	X	X	X	X	X	X	X	X									13	X	
AST and ALT2	X[1]					X				X				X	X	X	X	X	X	X	X	X	X									13	X	
Total & Direct Bilirubin	X[1]					X				X				X	X	X	X	X	X	X	X	X	X									13	X	
Albumin	X[1]																															1		
TSH	X[3]	X[8]															X						X									3	X	
HbA1c	X[4]	X[8]															X[5]						X[5]				X[5]					(1)	X[5,17]	X[5,17]
HIV test	X[6]																															1		
CD4 count and HIV viral load (in HIV+)	X[7]	X[8]															X[9]						X[9]				X[9]					(0)	X[9,17]	X[9,17]
Hepatitis B serology (anti-HBc total and HbsAg)	X	X[8]																														1		
Hepatitis C Serology: HCVAb	X	X[8]																														1		
Pregnancy test11	X	X[10]																														(2)	X	
Sputum Specimen Testing																																		
Smear Microscopy	X			X		X				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	22	X	X
Culture (LJ & MGIT)	X			X		X				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	22	X	X
Rapid Molecular Test for RIF Resistance	X[12]	X[8]																														1		
Rapid molecular test for FQ resistance	X[12]	X[8]													X[13]																	(1)		
DST (conventional 1st and 2nd line)	X														X[13,14]																	(1)	X[17]	X[17]
DST (new drugs)	X[14]														X[14]																	(0)	X[17]	X[17]
Genotyping14	X														X																	(0)	X[17]	X[17]
Specimen Storage	X[15]														X[15]																	(1)	X	
Special Assays or Procedures																																		
Audiometry		X				X				X				X	X	X	X	X	X	X	X	X	X				X				X	15	X	
ECG	X[1]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X					25	X	
Chest X-ray		X[16]								X							X				X						X				X	6	X[17]	X[17]

#. W8 visit window should occur within 8 weeks (+/-2 days) of study treatment initiation.
*: Randomization for treatment regimen assignment.

**: Only for patients receiving linezolid, randomization to linezolid dose reduction strategy may occur at W16 or earlier (concomitant to dose reduction decision).

1. Test with abnormal result at screening visit may be repeated within screening window.
2. ALT/AST tests to be symptom driven after Week 47.
3. Optionally, documented results of TSH (performed less than 4 weeks prior to screening visit date) may substitute for screening/baseline test. Investigators are encouraged to repeat TSH if they deem it useful for patient management or to have a reliable pre-treatment result.
4. Optionally, documented results of HbA1c (performed less than 3 months prior to screening visit date) may substitute for screening/baseline test. Investigators are encouraged to repeat HbA1c if they deem it useful for patient management or to have a reliable pre-treatment result.
5. HbA1c test (or 2 fasting blood glucose test if HbA1c cannot be done) to be repeated every 6 months only if abnormal at screening/baseline.
6. HIV serology will be offered, unless patient is known to be positive or a documented negative result is available from less than one month before screening visit date.
7. CD4 count and HIV viral load will be performed if the patient is known or found to be HIV positive. Optionally, documented results of CD4 count (performed less than six months prior to screening visit date) and viral load (performed less than 4 weeks prior to screening visit date) may substitute for screening test. Investigators are encouraged to repeat CD4 count and HIV viral load if they deem it useful for patient management or to have a reliable pre-treatment result.
8. Complete only if not done at screening.
9. CD4 count and HIV viral load to be monitored every 6 months in HIV-positive patients.
10. Serum pregnancy test at baseline visit may be repeated if the blood specimen was not drawn within 72 hours prior to treatment start.
11. When required, pregnancy tests may be performed on urine or serum samples after baseline visit until the end of study treatment.
12. Rapid molecular test results for RIF and FQ resistance, performed at designated study lab from sputum collected within 3 weeks prior to screening visit date can be used to assess eligibility for randomization.
13. Done locally on first positive culture or any culture reversion at or after Week 16.
14. Done at ITM on first positive culture or any culture reversion at or after Week 16 and on corresponding screening/baseline strain.
15. Store the screening/baseline sample and isolate from the corresponding culture. When possible, store samples and isolates from subsequent positive cultures.
16. Not necessary if previous adequate imaging results within 3 weeks prior to baseline visit date are available.
17. May be indicated, based on previous results available.

c. Only for participants still on prescription.

8. STUDY ANALYSIS POPULATIONS

8.1 Efficacy populations

8.1.1. Modified Intent To Treat population

The modified intent-to-treat population (mITT) will contain all randomized participants except those in whom any of the following occur:

1) There is no rapid molecular test result documenting rifampicin resistance at baseline (or screening);
2) There is no rapid molecular test result documenting fluoroquinolone-resistant TB at baseline (or screening);
   a. and, among participants from Kazakhstan, Lesotho, Peru, Vietnam with at least one month of prior exposure to any fluoroquinolone, there is no fluoroquinolone-undefined TB at baseline (or screening) on rapid molecular test;
   b. and, among participants from India, Pakistan, there is no fluoroquinolone-undefined TB at baseline (or screening) on rapid molecular test;
   c. and, among all participants with a fluoroquinolone-undefined TB at baseline (or screening), there is any subsequent test (phenotypic or genotypic) finding fluoroquinolone susceptibility before the end of study treatment, in the absence of any test finding fluoroquinolone resistance before the end of study treatment;
3) Any DST results performed on any screening or baseline sample at the designated study lab indicate resistance to bedaquiline, delamanid, clofazimine and/or linezolid;
4) There is no positive sputum culture before randomization;
5) There is no post baseline data (i.e. no study visits performed, no reported events occurred after randomization);
6) Participant was randomized in error (did not fulfill eligibility criteria) and, upon investigator awareness, was discontinued from study participation.

Only results from DST and rapid molecular tests performed in a designated study laboratory will be considered.
Participants in the mITT population will be analyzed in the arm to which they were randomly allocated by the IWRS (as randomized) regardless of treatment received.

8.1.2. Per Protocol Population

The per protocol (PP) population will include participants in the mITT population, except participants who, for reasons other than treatment failure or death, do not complete a protocol-adherent course of treatment. A protocol-adherent course of treatment contains at least 80% of expected doses within 120% of the nominal regimen duration. In participants who were expected to receive 24 weeks of treatment duration according to the applicable protocol version, as defined in section 4, a protocol-adherent course of treatment contains no more than 120% of the expected 24 weeks of doses. Receipt of either a prohibited concomitant medication or an IP not prescribed according to protocol for more than 7 days constitutes a non-protocol-adherent course of treatment and results in exclusion from the PP population. No other deviations resulted in exclusion from PP.

8.1.3. Other efficacy Populations

Additional efficacy populations will be defined for sensitivity analyses for the Week 73 and Week 104 endpoints. Use of these populations is specified in section 11.3:

1) V3.0-mITT: this population will comprise the mITT population excluding participants who were not assigned the correct treatment duration based on protocol Version 3.0 or higher.
2) All-culture mITT: this population will comprise the mITT population plus participants with baseline culture-negative TB.
3) All-FQ mITT: this population will comprise the mITT population plus participants who had a fluoroquinolone-susceptible subsequent test before the end of study treatment, in the absence of any test finding fluoroquinolone resistance before the end of study treatment, among those with an undefined fluoroquinolone resistance test result at baseline (or screening).

8.2 Safety population

The safety population will include all enrolled participants who received at least one dose of study treatment (exposed) of any drug. Safety analyses will be based on the observed treatment regimen received (as treated) regardless of randomization.

9. ASSESSEMENT OF OBJECTIVES

9.1 Efficacy endpoints

9.1.1. Primary efficacy endpoint

The primary efficacy endpoint used for the primary objective (PO) is a favorable outcome at Week 73.

A favorable outcome can only be established if all possible unfavorable outcomes are ruled out and one of the following is true:

- The last two culture results are negative. These two cultures must be from sputum samples collected on separate visits, the latest between Week 65 and Week 73;
- The last culture result (from a sputum sample collected between Week 65 and Week 73) is negative; and either there is no other post-baseline culture result or the penultimate culture result is positive due to laboratory cross contamination; and bacteriological, radiological and clinical evolution is favorable;
- There is no culture result from a sputum sample collected between Week 65 and Week 73 or the result of that culture is positive due to laboratory cross contamination; and the most recent culture result is negative; and bacteriological, radiological and clinical evolution is favorable.

Unfavourable outcome is established by any of the following situations:

- Replacement or addition of one or more investigational drugs in the experimental arm (failure);
- Replacement or addition of two or more investigational drugs in the control arm (failure); [1]
- Initiation of a new MDR-TB treatment regimen after the end of the allocated study regimen and before Week 73 (recurrence);
- Death from any cause (death);
- At least one of the last two TB cultures, the latest being from a sputum sample collected between Week 65 and Week 73, is positive in the absence of evidence of laboratory cross contamination, as defined in section 10.4 (failure/recurrence);
- The most recent (i.e., last) TB culture result (from a sputum sample collected between Week 65 and Week 73) is negative; AND

[1] Addition or replacement of two or more drugs to the control arm in order to conform to a new WHO treatment guidance —rather than in response to emerging participant data — does not lead to establishment of an unfavourable outcome.

- there is no other post-baseline culture result or the penultimate culture is positive due to laboratory cross contamination; and bacteriological, radiological or clinical evolution is unfavorable (failure/ recurrence);

- There is no TB culture result from a sputum sample collected between Week 65 and Week 73 or it is positive due to laboratory cross contamination; AND
  - the most recent (i.e., last) culture is negative; and bacteriological, radiological or clinical evolution is unfavorable
  - or, the most recent (i.e., last) culture result is positive in the absence of laboratory contamination (failure/recurrence);
- There is no culture result from a sputum sample collected between Week 65 and Week 73 or it is positive due to laboratory cross contamination; AND
  - there is no other post-baseline culture result or the most recent culture is positive due to laboratory cross contamination; or,
  - the most recent culture is negative and bacteriological, radiological and clinical evolution is not assessable (not assessable);
- Previously classified as unfavorable in the present study[2].

The Week 73 efficacy endpoint will not be determined from raw data at the time of analysis. It will have been previously assigned through the event-validation process (according to the above definitions and described in SOP SP-031-CT), validated, and entered into the study database. The validated outcome will be used for the analysis.

The primary efficacy endpoint will also be used for the exploratory objectives EEO4 and EEO5.

9.1.2. Secondary efficacy endpoints

The main secondary efficacy endpoints are:

1. Favorable outcome at Week 104 (used for SEO1).

A favorable outcome can only be established if all possible unfavorable outcomes are ruled out and one of the following is true:

- The last two cultures are negative. These two cultures must be from: sputum samples collected on separate visits, the latest between Week 97 and Week 104;
- The last culture result (from a sputum sample collected between Week 97 and Week 104) is negative; and either there is no other post-baseline culture result or the penultimate culture result is positive due to laboratory cross contamination; and bacteriological, radiological and clinical evolution is favorable;
- There is no culture result from a sputum sample collected between Week 97 and Week 104 or the result of that culture is positive due to laboratory cross contamination; and the most recent culture result is negative; and bacteriological, radiological and clinical evolution is favorable.

Unfavorable outcome is established by any of the following situations:

- Replacement or addition of one or more investigational drugs in the experimental arm (failure);
- Replacement or addition of two or more investigational drugs in the control arm (failure);
- Initiation of a new MDR-TB treatment regimen after the end of the allocated study regimen and before Week 104 (recurrence);
- Death from any cause (death);

[2] Exception: a participant whose outcome is unfavorable because it is unassessable at Week 39 is eligible for re-evaluation at Week 73.

- At least one of the last two cultures, the latest being from a sputum sample collected between Week 97 and Week 104, is positive in the absence of evidence of laboratory cross contamination (failure/recurrence);
- The last culture result (from a sputum sample collected between Week 97 and Week 104) is negative and
  - there is no other post-baseline culture result or the penultimate culture is positive due to laboratory cross contamination; and bacteriological, radiological or clinical evolution is unfavorable (failure/recurrence);
- There is no culture result from a sputum sample collected between Week 97 and Week 104 or it is positive due to laboratory cross contamination and
  - the most recent culture is negative; and bacteriological, radiological or clinical evolution is unfavorable (failure/recurrence);
- There is no culture result from a sputum sample collected between Week 97 and Week 104 or it is positive due to laboratory cross contamination and
  - there is no other post-baseline culture or it is positive or,
  - the most recent culture is negative and bacteriological, radiological and clinical evolution is not assessable (not assessable);
- Previously classified as unfavorable in the present study[3];
- Lost to follow-up, as defined in section 10.3.

The Week 104 efficacy endpoint will not be determined from raw data at the time of analysis. It will have been previously assigned through the event-validation process (according to the above definitions and described in SOP SP-031-CT), validated, and entered into the study database. The validated outcome will be used for the analysis.

Favorable outcome at Week 104 will also be used for the exploratory objectives EEO4 and EEO8.

2. Culture conversion endpoints (used for SEO2):

   a. Initial sputum culture conversion (SCC) by Week 8. This is defined as 2 consecutive negative sputum culture results from specimens collected at 2 different visits, the first being at or before Week 8. If there is a single missing or contaminated (not AFB/contaminated) culture between 2 negatives, the definition of conversion is still met.
   b. Time to initial culture conversion. This is defined as the interval between the date of study treatment initiation and the date of collection of sputum resulting in the first of at least 2 consecutive negative cultures (as defined above). If there is a missing or contaminated (not AFB/contaminated) culture between 2 negatives, the definition of conversion is still met.
   c. Change in time to MGIT culture positivity (TTP) from baseline through 8 weeks (defined in section 10.7).
   Participants who discontinue study participation or die before conversion remain in the denominator of those eligible for conversion and count as not having converted. Participants with only contaminated or missing results will be counted as not having converted.

3. Favorable outcome at Week 39 (used for SEO3).

A favorable outcome can only be established if all possible unfavorable outcomes are ruled out. Unfavorable outcome is established by any of the following situations:

[3] Exception: a participant whose outcome is unfavorable because it is unassessable at Week 73 is eligible for re-evaluation at Week 104.

- In the experimental arm, addition or replacement of one or more drugs;
- In the control arm, addition or replacement of two or more drugs;
- Death from any cause;
- At least one culture result (from a sample collected between Week 36 and Week 39) is positive;
- The outcome is not assessable because the last available culture result is from a sample collected before Week 36.

If an unfavorable outcome has not been assigned and all culture results from samples collected between Week 36 and Week 39 are negative, the outcome is favorable..

The Week 39 efficacy endpoint will not be calculated from raw data at the time of analysis. It will have been previously assigned through the event-validation process (according to the above definitions and described in SOP SP-031-CT), validated, and entered into the study database. The validated outcome will be used for the analysis.

4. Treatment failure or recurrence (used for SEO4):
   a. at Week 73 (as defined in list of unfavorable outcomes in section 9.1.1).
   b. at Week 104 (as defined in list of unfavorable outcomes in section 9.1.2).

9.1.3. Exploratory efficacy endpoints

The additional exploratory efficacy endpoints are:

1. Drug resistance amplification by week 104. This is defined as change in any DST result from susceptible on baseline DST to resistant on any post-baseline DST, on same (≤12 SNPs in the distance matrix) or likely same strain (without whole genome sequencing of both strains and strains of the same lineage, similar genotype [absence of change in mutation, or reversion to wildtype, in specific genes: rpoB, gyrA, and pncA, only if fixed >90%])[4]: used for EEO1.
2. Treatment adherence of ≥80% as defined in section 10.10: used for EEO2.

9.2 Safety endpoints

1. Death from any cause (SSO5).

2. Occurrence of (SSO6):
   a. AE of grade 3 or higher as defined in section 10.11 by Week 73.
   b. AE of grade 3 or higher as defined in section 10.11 by Week 104.
   c. SAE as defined in section 10.12 by Week 73.
   d. SAE as defined in section 10.12 by Week 104.
   e. AE of grade 3 or higher or SAE by Week 73.
   f. AE of grade 3 or higher or SAE by Week 104.

3. Occurrence of (SSO7):
   a. AESI as defined in section 10.13 by Week 73.
   b. AESI as defined in section 10.13 by Week 104.

4. Severe linezolid-related toxicity (as described in section 10.16) (ESO3):
   a. Occurrence of severe linezolid-related toxicity
   b. Time to first severe linezolid-related toxicity.

[4] Data related to drug resistance amplification and relapse outcomes will be provided by ITM.

9.3 Summary

N°	Objective	Endpoint
	Primary Efficacy	
PO	To assess whether the efficacy of the experimental regimen at Week 73 is non-inferior to that of the control	Favorable outcome at Week 73
	Secondary efficacy	
SEO1	To compare the efficacy of the experimental regimen at Week 104 to that of the control.	Favorable outcome at Week 104
SEO2	To compare the frequency of and time to early treatment response (culture conversion) in the experimental regimen to that of the control.	Culture conversion endpoints: a. Initial sputum culture conversion (SCC) by Week 8. b. Time to initial culture conversion c. Change in time to MGIT culture positivity (TTP) from baseline through 8 weeks.
SEO3	To compare the efficacy of the experimental regimen at Week 39 to that of the control.	Favorable outcome at Week 39
SEO4	To compare, at Week 73 and Week 104, the proportion of participants who experienced failure or recurrence in the experimental arm to that in the control arm	Treatment failure or recurrence: a. at Week 73 b. at Week 104
	Exploratory efficacy	
EEO1	To estimate the frequency of drug resistance amplification among treatment failures and relapses occurring by Week 104	Drug resistance amplification by week 104
EEO2	To compare treatment adherence in the experimental arm to that in the control	Treatment adherence of ≥80%
EEO4	To compare, at Week 73 and Week 104, efficacy endpoints across linezolid dose-reduction strategies: 300 mg daily or 600 mg thrice weekly	Favorable outcome at Week 73 Favorable outcome at Week 104
EEO5	To assess for modification of effect of experimental regimen (vs. control) by extent-of-TB-disease phenotype	Favorable outcome at Week 73
	Safety	
SSO5	To compare, at Week 73 and Week 104, the proportion of participants who died of any cause in the experimental arm to that in the control arm	Death from any cause

SSO6	To compare, at Week 73 and Week 104, the proportion of participants who experienced grade 3 or higher AEs or SAEs of any grade in the experimental arm to that in the control arm	Occurrence of : a. AE of grade 3 or higher by Week 73. b. AE of grade 3 or higher by Week 104. c. SAE by Week 73. d. SAE by Week 104. e. AE of grade 3 or higher or SAE by Week 73. f. AE of grade 3 or higher or SAE by Week 104
SSO7	To compare, at Weeks 73 and 104, the proportion of participants who experienced AESIs in the experimental arm to that in the control arm	Occurrence of: a. AESI by Week 73. b. AESI by Week 104.
ESO3	To compare the frequency of and time to severe linezolid-related toxicity between linezolid dose-reduction strategies	Severe linezolid-related toxicity: a. Occurrence of severe linezolid-related toxicity b. Time to first severe linezolid-related toxicity.

10. DEFINITIONS

10.1. Scheduled visits, visit window definition, and unscheduled visits

Scheduled study visits occur at screening, baseline (Week 0), at weekly intervals until Week 12, then at 4-weekly intervals until Week 36 and then at Weeks 39, 43, 47, 53, 59, 65, 73, 81, 89, 97 and 104, all calculated from date of randomization. Each scheduled visit had a predefined allowable window of days before and after the target scheduled date during which a protocol-conforming study visit could occur.

For the purpose of analysis, the scheduled visit results will be used preferentially. If the scheduled visit did not occur, or if the scheduled visit result is missing, the value measured at the date closest to the target scheduled visit date and within the window period (as defined in Table 2) will be used for analyses.

If the scheduled visit result is missing and there is no result from an unscheduled visit in the window period, the result for that visit will be classified as missing. Exceptions are described for outcome assignment at Weeks 73 and 104 in section 10.2.

The following results from evaluations at pre-screening visits are also allowable to characterize participants at randomization and establish eligibility for analysis populations:

- TSH performed less than 4 weeks prior to screening visit date,
- HbA1c performed less than 3 months prior to screening visit date,
- HBV, HCV and HIV serology performed less than one month before screening visit date,
- CD4 count performed less than six months prior to screening visit date,
- HIV viral load performed less than 4 weeks prior to screening visit date
- Molecular tests for rifampicin and fluoroquinolone resistance performed by a designated study lab on pre-screening samples collected no more than 3 weeks prior to screening visit,
- Chest X-ray of adequate quality performed less than 3 weeks prior to baseline visit date.

For longitudinal analyses (e.g., change in time to positivity), all available values (from scheduled and unscheduled visits) will be used.

10.2. Data Handling Convention for Week 73 and 104 outcomes

For assignment of outcomes at Weeks 73 and 104, data collected up to 30 days after the close of the window period around the study visit (i.e., up to 60 days after the target visit date) are included among those used to inform outcome classification. This applies to data on clinical, radiological, and bacteriological evolution, culture results, and death.

For the Week 73 endpoint, data outside the window period are only used if relevant results were not available between 65 to 73 Weeks. For the 104 Week endpoint, these data are only used if relevant results were not available between 97 to 104 Weeks.

10.3. Loss to follow-up

Participants a) who did not complete their scheduled final study visit and b) for whom study staff have no information and whom they have been unable to contact, in-person or by phone for more than 14 weeks before the last study visit per study schedule will be considered lost to follow-up.
The “loss to follow-up” designation is made only after the scheduled final study visit.

10.4. Recurrence, relapse, and reinfection

Recurrence is a composite outcome of relapse and reinfection:

- Relapse is recurrence that occurs after the end of the assigned study treatment and in the absence of genotypic evidence of a new infection.
- Reinfection is recurrence that occurs during or after study treatment and in the presence of genotypic evidence of a new infection.

In primary analysis, recurrences (relapse and reinfection) will be considered unfavourable outcomes.

10.5. Culture result

A culture result is called positive for TB if *Mycobacterium tuberculosis* (MTB) grows in MGIT or LJ medium. If more than one culture result is available from sputum collected at a single visit, results will be collapsed into a single (positive-dominant) result for the purposes of all analyses as follows:

1) **Positive**, if culture result from sample A and/or sample B is positive for MTB.
2) **Negative**, if no culture result is positive for MTB and sample A and/or sample B is negative.
3) **Contaminated,** if culture results from sample A and/or sample B are contaminated or AFB/Contaminated and no culture result is negative or positive for MTB.
4) **Missing**, if no sputum sample was collected or no result is available from all samples collected at a single visit.

Outcome assignment relied on available genotypic evidence and consulted the study reference laboratory to establish or rule out cross contamination.

10.6. AFB smear result

A smear result is called positive if it is graded as 'scanty' or 1+ or more.
If more than one smear result is available from sputum collected at the same visit, results will be collapsed into a single (positive-dominant) smear result for the purposes of all analyses, with the following overall result:

1) **Positive**, if at least one of the smear results is positive; positive smears will be further classified as scanty, 1+, 2+, 3+ in accordance with the highest reported value among smear results a given visit;
2) **Negative,** if at least one of the smear results is negative and none of the additional smear results are positive;
3) **Missing,** if no sputum sample was collected or no result is available from all samples collected at a single visit.

10.7. Time to positivity (TTP) in MGIT

Time to positivity (TTP) in MGIT is "time to detection" of *M. tuberculosis* in the BACTEC 960 system. It is reported in days and hours.

- When a MGIT culture result is positive, the TTP is expected to be lower than 42 days.
- When MTB is not detected within 41 days and 23 hours, the MGIT culture result is assumed to be negative and the resulting TTP set to the maximum of 42 days.
- When a MGIT culture result is positive with a TTP greater than 41 days and 23 hours, the MGIT culture is considered negative and TTP is set to 42 days.
- When a MGIT culture result is contaminated or missing, the value for TTP is set to missing.

If more than one TTP result is available from sputum collected at the same visit, results will be collapsed into a single TTP result for the purposes of analyses, with the following overall result:

- If both TTP results are positive, the mean value of 2 TTP is used.
- If only one is positive, the TTP from the positive sample is used.
- If both are negative, TTP is set to 42.
- If one is negative and the other is neither negative nor positive, then TTP is set to 42.

10.8. Extent-of-TB-disease phenotype & provisional assignment of treatment duration at baseline

In the experimental group, 39 weeks of treatment is assigned at baseline if one of the following is true (see Table 3):

- no cavity is present on allowable baseline X-ray and the highest smear grade from a screening sputum sample is 2+ or higher (extensive);
- a cavity is present on allowable baseline X-ray and the highest smear grade from a screening sputum sample is 1+ or higher (extensive);

If none of the above is true at baseline (not extensive), the participant is assigned 24 weeks of treatment at baseline and re-evaluated at 24 weeks.

Table 3: Duration of treatment according to extent-of-TB-disease phenotype at screening/baseline

	Smear Neg or Scanty	**Smear 1+**	**Smear 2+**	**Smear 3+**
Cavity absent	24 weeks of treatment*	24 weeks of treatment*	39 weeks of treatment**	39 weeks of treatment**
Cavity present	24 weeks of treatment*	39 weeks of treatment**	39 weeks of treatment**	39 weeks of treatment**

** Not extensive ** Extensive*

10.9. Re-evaluation of treatment duration in the experimental arm at 24 weeks

Treatment duration is re-evaluated only among those who were assigned 24 weeks of treatment at baseline (as defined in section 10.8).

Treatment duration is extended to 39 weeks if one or more of the following is true:

- there is ≥1 positive culture result from sputum specimens collected at Week 8 or all culture results from sputum specimens collected at Week 8 are missing or contaminated; OR,
- if any culture is positive from a sputum specimen collected after Week 8, with result available at the time of the Week 24 visit.

10.10. Protocol-adherent treatment

Treatment adherence is defined as the percentage of total doses taken compared to the expected number of doses to be taken, based on assigned regimen. It does not discriminate among reasons for missed intakes. For each drug expected number of doses is calculated using: expected duration, number of daily doses, and the number of times per week at treatment initiation.

The following formula will be applied for each participant:

$$\frac{\sum_{i=1}^{n}(Total\ doses\ taken\ drug\ i)}{\sum_{i=1}^{n}(Number\ of\ daily\ doses\ drug\ i * Number\ of\ days\ per\ week\ drug\ i * (Expected\ duration\ drug\ i/7))} \times 100$$

where:

- n= number of drugs prescribed at treatment initiation
- Total doses taken drug *i* = number of times the drug *i* was taken under DOT + number of times drug *i* was self-administered
- Number of daily doses drug *i* = number of prescribed doses in assigned regimen per day
- Number of days per week drug *i* = number of prescribed days of administration per week
- Expected duration drug *i* = expected number of days that the participant will receive drug *i* based on study protocol for experimental arm and WHO recommendations for control arm (see below).

The expected duration for each drug will be:

- In the experimental arm: 182 days (6 months) or 273 days (9 months) for all drugs;
- In control arm: 456 days (15 months) for oral drugs;

According to the study protocol, at least 80% of expected doses (per applicable protocol version, defined in section 4) must be taken for a protocol-adherent treatment. Participants expected to receive 24 weeks of experimental treatment may take as long as 32 weeks (224 days) to complete this minimum. Participants expected to receive 39 weeks of experimental treatment may take as long 47 weeks (329 days) to complete this minimum.

According to the study protocol, a maximum of 120% of expected doses must be taken by participants who were expected to receive 24 weeks of treatment duration (per applicable protocol version, defined in section 4) for a protocol-adherent treatment. There is no maximum time when calculating the number of doses for the evaluation of treatment duration for this purpose.

10.11. Adverse event

AE is any untoward medical occurrence in a study participant after administration of an investigational drug. It does not necessarily have a causal relationship with the study treatment. A pre-existing condition that deteriorates at any time during the study (e.g., increase of severity) is an AE.

All events are graded according to v.5.0 of the MSF Severity Scale. Grade 3 events are generally those which result in marked limitation in activity. Some assistance and medical intervention/therapy are usually required, hospitalization is possible.

10.12. Serious adverse event

An SAE is defined as an AE meeting at least one of the following conditions:

- An event leading to death.
- A life-threatening event (defined as a subject at immediate risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe).
- An event requiring hospitalization or prolongation of existing hospitalization. The following will not meet the definition of SAE: hospitalization for elective surgery planned prior to subject enrollment; admission to remove barriers to care in the ambulatory environment; admission to perform ECG per protocol; hospitalization for infection control or smear positivity; or visit to the hospital (e.g. emergency room) that lasted less than 24 hours and did not result in admission.
- An event resulting in a persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions.
- An event resulting in congenital anomaly or birth defect.
- Any other important medical event that may not result in death, be immediately life-threatening, or require hospitalization may be considered a serious adverse experience when, based upon appropriate medical judgment, the event may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed above. Examples of such medical events include allergic bronchospasm requiring intensive treatment in an emergency room or at home, blood dyscrasias or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse.

10.13. Adverse event of special interest

An AE of special interest (AESI) is a medical concern specific to the investigational medicinal product(s) for which close monitoring is required. In this study, the following AEs, regardless of their seriousness or causal relationship to treatment, are considered of interest:

- Grade 3 or above “electrocardiogram QT corrected interval prolonged”;
- Grade 3 or above leukopenia, anemia or thrombocytopenia;
- Grade 3 or above peripheral neuropathy;
- Grade 3 or above optic neuritis;
- Grade 3 or above increase in alanine aminotransferase (ALT) or aspartate aminotransferase (AST).

10.14. Non-TB microbiology laboratory variables

Laboratory variables routinely collected are: complete blood cell count (including WBC count, neutrophils, lymphocytes, eosinophils, hemoglobin, hematocrit, platelet count, RBC), biochemistry and electrolytes (including ALT/AST, total and direct bilirubin, albumin, creatinine, total calcium, potassium and magnesium).

10.15. Clinical signs, symptoms, evaluations

These include: weight/height, TB signs and symptoms, functional status (ECOG), mental health (anxiety by GAD-7, depression by PHQ-9), neurologic (subjective sensory neuropathy grade), ophthalmologic (visual acuity, color blindness), audiometry assessments, chest radiograph, and at least duplicate electrocardiogram (Qt interval, heart rate, respiratory rate; mean of QTcF is used for analysis).

Reporting includes presence/absence and grading according to the MSF Severity Scale.

10.16. Severe linezolid-related toxicity

Any of the following events established as linezolid related:

1) Grade 3 or higher leukopenia, anemia, thrombocytopenia, peripheral neuropathy, or optic neuropathy;
2) SAEs; or
3) AEs requiring linezolid discontinuation.

11. STATISTICAL METHODS AND DATA ANALYSIS

11.1. General rules

Epicentre will perform final statistical analysis after the database lock and the final data review.

Data will be presented in summary tables by treatment arm or as a total group and by visit if applicable. Continuous variables will be summarized for each treatment arm using the number of observations available (N), number of missing observations, mean, standard deviation (SD), minimum, maximum, median and 25th (Q1) and 75th (Q3) percentiles. Categorical data will be summarized for each regimen group using counts of non-missing and missing data and percentages.
All statistical tests will be two-sided unless otherwise specified.

Statistical analyses will be performed using STATA version 18, R version 4.4 or higher versions of either.

11.2. Data summaries/description

11.2.1. Patient/Participant disposition

The total number of patients/participants for each of the following categories will be presented in a table and in a study flow diagram (see in appendix 15.1):

- Screened patients and reason for non-inclusion
- Randomized participants
- Safety population (exposed participants)
- mITT population
- PP population

Although exclusion can be attributed to multiple reasons, only the primary reason will be displayed, in the following order of priority:

1. FQ resistance/RIF resistance or no MTB
    a. MTB not detected or inconclusive in 1st-line molecular test, RIF susceptible, or RIF indeterminate: RIF resistance not demonstrated
    b. MTB not detected or inconclusive in 2nd-line molecular test, FQ susceptible, or FQ indeterminate: FQ susceptibility not ruled out.
2. Age <inclusion age approved in the country
3. Cardiac risk factor
4. Lab value outside acceptable range
5. Contraindicated drug
6. Prior exposure/resistance

7. Drug allergy
8. Pregnant or breastfeeding
9. Unwilling to use contraception
10. Did not consent to study participation or withdrew before inclusion
11. Investigator discretion
12. Participation in other trial
13. Screening not completed within 14 days: 1st or 2nd-line molecular tests results missing or screening otherwise not complete in 14 days

Frequency and reason for discontinuation of treatment and/or study will be presented in a table.

11.2.2. Demographic and baseline characteristics

Demographic characteristics:

- Sex
- Age (years) at baseline

Clinical laboratory baseline (or screening) characteristics:

- Complete blood cell count at screening
- ALT/AST, total and direct bilirubin, albumin, creatinine, electrolyte testing at screening
- TSH
- HbA1C or 2 fasting blood glucose tests if HbA1C not done
- CD4 cell count and Viral load for HIV+ participants
- Hepatitis serology
- Pregnancy test

Microbiological laboratory baseline (or screening) characteristics:

- Sputum smear microscopy (Negative, Scanty, 1+, 2+, 3+, missing) at screening-collapsed
- Phenotypic drug susceptibility testing results
- Sputum culture result (Positive for MTB, Negative, Contaminated (contaminated + AFB/contaminated + positive for NTM) at screening-collapsed across MGIT & LJ
- Sputum culture result on MGIT and LJ separately in sites where both were used at screening
- Time to positivity (TTP) in MGIT at screening

Clinical and other baseline (or screening) characteristics:

- Medical history at screening
- TB history at screening
- TB symptom assessment
- Vital signs (body weight, height, BMI, oxygen saturation, temperature, pulse, respiratory rates and blood pressure)
- Physical exam (Symptoms/signs) (Grade 0-4)
- ECOG assessment
- Neurologic exam results (Grade 0-3)
- Ophthalmologic exam results (Grade 0-4)
- Mental health assessment results (Anxiety grade 0-4, Depression grade 0-4)
- Audiometry results (Grade 0-4)
- Concomitant medication (None, any; anti-retroviral therapy among HIV co-infected; other major immunomodulators [e.g. corticosteroids])
- Chest X ray (normal in both lungs/abnormal, presence of cavity and other lung lesions, extent of disease limited/moderate/extensive)

- ECG (Mean of the 2 highest QTcF intervals, Normal/Abnormal, Grade)
- Risk factors (history of homeless, unemployment, drugs use, cigarette smoking and alcohol use)

Descriptive statistics will be displayed to summarize the demographic and baseline characteristics data on the mITT population. No statistical test will be performed at baseline between treatment arms.

11.2.3. Regimen characteristics (control arm)

Composition and duration of control arm regimens will be described.

Changes in control regimen guidance over time will be described, with the list of changes, date each started, affected countries, and implications for analysis.

11.2.4. Experimental arm and extent of TB disease phenotype

We will describe the frequency and percentage of durations of treatment, assigned, received, and expected per protocol versions 2.2 and 3.0 or higher.

We will also describe baseline characteristics of participants based on provisional and re-evaluated treatment duration.

11.3. Efficacy analyses

Analyses for the efficacy endpoints will be performed in the mITT and PP populations.

11.3.1. Analysis of the primary efficacy endpoint (PO)

The primary analysis is adjusted by stratification factors (country and baseline extent-of-TB-disease phenotype). Analysis of the primary endpoint will compare the proportions of participants with a favorable outcome at Week 73 between the experimental and control arms. The absolute difference in proportions (risk difference) adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) will be estimated with corresponding 2-sided 95% confidence interval (CI) using a binomial regression model (generalized linear model for a binomial outcome with an identity link function). In case of convergence failure, risk difference will be estimated using a modified Poisson regression model with robust standard errors. If convergence is still an issue, then stratification by baseline extent-of-TB-disease phenotype status will be prioritised. A 2-sided 95% CI of the proportion of participants with a favorable outcome will also be estimated for each regimen.

The non-inferiority of the experimental arm compared to the control will be established if the difference in proportion with favorable outcome (proportion of participants with favorable outcome in the experimental arm minus proportion of participants with favorable outcome in the control arm) at Week 73 is greater than the lower non-inferiority margin i.e. if the lower bound of the one-sided 97.5% CI (which correspond to the lower bound of the 2-sided 95% CI) is greater than or equal to -12% in mITT and confirmed in PP populations.

If non-inferiority of the experimental arm is demonstrated, superiority compared to the control will then be tested at the 5% level of significance.

No adjustments will be made to control the type I error.

11.3.1.1 Subgroup analyses of the primary efficacy endpoint (PO)

Sub-group analyses of the primary efficacy endpoint will be performed (in both mITT and PP populations). Subgroups will be established on the following baseline characteristics:

- Country (India, Kazakhstan, Lesotho, Pakistan, Peru, Vietnam)
- HIV (presence/absence)
- Hepatitis C (presence/absence)
- Diabetes (presence/absence)
- BMI (<18.5 kg/m2, 18.5-24.9 kg/m2, ≥25 kg/m2)
- Age (<18, 18-45, ≥45)
- Sex (male/female)
- Smear result (positive/negative)
- Cavitation (presence/absence)
- Baseline extent-of-TB-disease phenotype (not extensive/extensive)
- Prior exposure to TB treatment (none/first line only/at least 2nd line)
- WHO recommendations implemented (by implementation date).

11.3.1.2 Sensitivity analyses of the primary efficacy endpoint (PO)

The primary efficacy analysis will be repeated:

- in the following populations (described in section 8.1.3): V3.0-mITT, all-culture mITT and all-FQ mITT populations;
- Unadjusted for any covariates;
- Adjusted for randomisation stratification factors and BMI, HIV, diabetes;

11.3.2. Analysis of the secondary efficacy endpoints

For the analysis of the secondary efficacy endpoints, proportions will be calculated by treatment group and corresponding 95% confidence intervals will be presented.

For all comparison of proportions between the experimental and the control arms, the risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) will be estimated with corresponding 95% confidence interval using a binomial regression model. In case of convergence failure, risk difference will be estimated using a modified Poisson regression model with robust standard errors. If convergence is still an issue, then stratification by baseline extent-of-TB-disease phenotype status will be prioritised.

No adjustments in significance levels will be made for multiple comparisons.

Analyses will be performed for:

1) SEO1: Proportion of participants with a favorable outcome at Week 104.

2) SEO2: Sputum culture conversion (as defined in 9.1.2):
    a. Proportion of participants with initial sputum culture conversion by Week 8,
    b. Time to initial sputum culture conversion assessed in MGIT system will be compared between the experimental and the control arms using Kaplan-Meier analysis. Data for participants who discontinued treatment will be censored at the last assessment. The logrank test will be used for the comparison of the median time to sputum culture conversion. Kaplan-Meier curves will be presented. A cox proportional hazards model will be used to estimate hazard ratios with corresponding two-sided 95% confidence intervals

and p-value. Adjusted hazard ratios for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) with corresponding 95% confidence interval will be calculated using a Cox proportional hazards model.

c. Mean change from baseline in time to positivity (TTP) in MGIT over 8 weeks will be compared between the experimental and the control arms using a linear mixed-effects regression model, with change from baseline as the outcome. Visit, treatment group, interaction visit*treatment group and baseline TTP value will be fixed effects. Study participant will be treated as a random effect to account for both heterogeneity among participants and correlation among measurements taken on the same participant over time. To evaluate the approach of managing negative culture, this will also be analyzed in a mixed-effects tobit regression model where negative culture will be right-censored.

3) SEO3: Proportion of participants with a favorable outcome at Week 39.

4) SEO4: Proportion of participants with a treatment failure or recurrence at Week 73 and at Week 104.

11.3.3. Sensitivity analyses of the secondary efficacy endpoints

Analysis of the risk differences in proportions of participants with a favorable outcome at Week 104 (SEO1) and participants with treatment failure or recurrence at Week 104 (SEO4) will be repeated:

- in the following populations (described in section 8.1.3): V3.0-mITT, all-culture mITT and all-FQ mITT populations;
- Unadjusted for any covariates;
- Adjusted for randomisation stratification factors and BMI, HIV, diabetes

11.3.4. Analysis of the exploratory efficacy endpoints

Analyses of exploratory endpoints will be performed in the mITT population.

1) EEO1: Proportion of participants with drug resistance amplification among participants with failure or relapse outcome at Week 104 will be calculated by treatment arm as number of participants with failure or relapse at Week 104 (denominator) and number with failure or relapse at Week 104 with drug resistance amplification (numerator). The risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) between the experimental and the control arms will be estimated with corresponding 95% confidence interval using a binomial regression model.

2) EEO2: Proportion of adherent participants: The risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) between the experimental and the control arms will be estimated with corresponding 95% confidence interval using a binomial regression model.

3) EEO4: Proportion of participants with a favorable outcome at Week 73 and 104 compared between participants in the experimental arm randomized to linezolid dose-reduction strategies: the risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) will be estimated with corresponding 95% confidence interval using a binomial regression model.

Linezolid dose-reduction data from endTB and endTB-Q will be pooled and the same analyses will be done on the pooled data: Proportion of participants with a favorable outcome at Week 73 and 104 compared between participants in the experimental arm randomized to linezolid dose-reduction strategies: the risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) and study will be estimated with corresponding 95% confidence interval using a binomial regression model.

4) EEO5: Proportion of participants with a favorable outcome at Week 73 compared between participants in the experimental and the control arms by baseline extent-of-TB-disease phenotype: the risk difference will be estimated with corresponding 95% confidence interval using a binomial regression model. An interaction term between treatment arm and baseline extent-of-TB-disease phenotype will be added to the model to test for a differential effect of treatment. See subgroup analyses in section 11.1.

11.4. Safety analyses

The safety analysis will be based on the reported adverse events (AEs) in the case report forms, and other safety information, such as laboratory evaluations, clinical evaluations including neurological assessments, physical exams, mental health, X-ray, and ECG performed during the follow-up of the study.

The analyses of safety data will be performed on the safety population during the intervals of 73 weeks and 104 weeks post-randomization.

The Pharmacovigilance Database (PVDB) will be consulted as necessary to supplement safety data captured in the Clinical Database, OpenClinica (OC).

Missing data will not be imputed for safety analyses.

11.4.1. Analysis of adverse events

Each AE is coded to a “Preferred Term” (PT) and primary associated “System-Organ Class” (SOC) according to the MedDRA dictionary (Medical Dictionary for Regulatory Activities, version 19.1). AE grade 3 or greater, SAEs and AESIs are also coded to a category, as described in appendix 15.2.

The summaries of AE by treatment group and grade will include:

- The number and percentage of participants with at least one AE
- The number and percentage of participants with at least one AE in a specific SOC/PT.

Number and percentage of participants with at least one of the following events will also be provided, by SOC, PT and category:

- AEs of Grade 3 or higher
- Serious AEs regardless of grade
- AEs leading to death of any cause (overall) regardless of grade
- AE of special interest.

Counts will be provided by regimen for each PT within each SOC, and by category, and regardless of relationship to the study drug.

For all comparisons between the experimental and the control arms, the risk difference adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) will be estimated with corresponding 95% confidence interval using a binomial regression model for:

- SSO5: the proportion of participants who died of any cause up to Week 73 / Week 104

- SSO6: a) the proportion of participants with AEs of Grade 3 or higher by Week 73 / Week 104;
  b) the proportion of participants with SAE up to Week 73 / Week 104; and
  c) the proportion of participants with AEs grade 3 or higher or SAEs of any grade up to Week 73 / Week 104
- SSO7: the proportion of participants with AESI up to Week 73 / Week 104.

11.4.2. Analysis of laboratory variables

Descriptive statistics of laboratory parameters grades and changes in grades from baseline will be presented by regimen at each time point they were measured, following the schedule of events. If additional measurements were performed in between, leading to abnormal results, they will be analysed through the safety reporting of AEs, not in the descriptive statistics; nevertheless such values will be reported in the listings. Summaries will be displayed for patients having at least one baseline and one post baseline values.

Figures will present the mean change of grade overtime for each treatment arm.

11.4.3. Analyses of clinical signs and assessments

Descriptive statistics of parameters and changes from baseline will be presented by regimen at each time point they were measured, following the schedule of events. When available, grades will be used rather than parameter value (physical exam: cough, hemoptysis, chest pain, fever (from severity scale), dyspnea, ECOG, mental health assessment (anxiety and depression grades), neurological, ophthalmological, neurological, ophthalmological and audiometry assessments, BMI, chest X-ray (including normal/abnormal, cavitation, extent of disease). For ECG, mean of the two highest QTcF intervals will be presented as well as grade. For other clinical signs, we do not plan to report descriptively.

Results from unscheduled visits will not be reported in the descriptive summaries or in tables of changes from baseline. Results from unscheduled visits will be included in the reporting frequency, time, and grade of AEs, SAEs, and AESIs and in listings of abnormal values.

Figures will present the mean change overtime by treatment arm.

11.4.4. Analysis of the exploratory safety endpoints

ESO3: The risk difference in proportions of participants with severe linezolid-related toxicity occurring after linezolid dose-reduction between participants randomized to each linezolid dose-reduction strategy (300 mg daily or 600 mg thrice weekly) adjusted for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) will be estimated with corresponding 95% confidence interval using a binomial regression model.

Median time from linezolid dose reduction randomization to first severe linezolid-related toxicity occurring after linezolid dose-reduction will be compared between participants randomized to linezolid dose-reduction strategies using a two-sided log-rank test the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) at the 5% significance level. Kaplan-Meier curves will be presented. A Cox proportional hazards model will be used to estimate hazard ratios with corresponding two-sided 95% confidence intervals and p-value. Adjusted hazard ratios for the randomization stratification factors (country and baseline extent-of-TB-disease phenotype) with corresponding 95% confidence interval will be calculated using a cox proportional hazards model.

11.5. Major protocol deviations

Major protocol deviations (as defined in SOP SM-005-CT) will be tabulated by site with information such as type of deviations, time of occurrence and reason. The percent of subjects with major deviations by type will also be summarized by site and study arm.

12. INTERIM ANALYSIS

A data safety and monitoring board (DSMB) independently assessed the safety and risk/benefit semi-annually during recruitment and delivery of experimental treatment. Analyses were performed by an external statistician.

13. REFERENCES

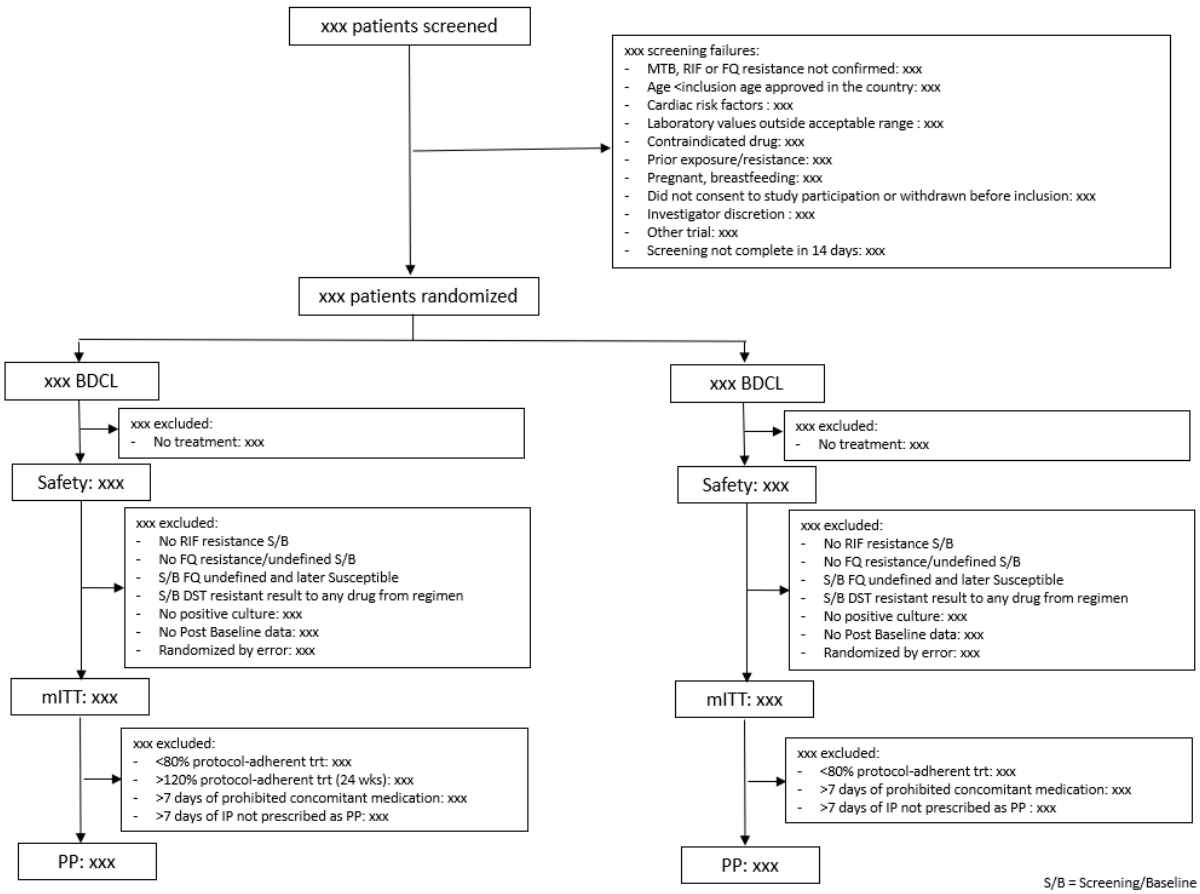
None

14. SUPORTING DOCUMENTS

None

15. APPENDIX

15.1 Study flow diagram



15.2 AE categories

In addition to "Preferred Term" (PT) and primary associated "System-Organ Class" (SOC) according to the MedDRA dictionary (Medical Dictionary for Regulatory Activities, version 19.1), AE grade 3 or higher, SAE and AESI will be coded to one of the following categories, partially based on the Standardized MedDRA Queries:

- Blood glucose abnormalities
- Breast conditions
- Cardiac abnormalities
- Death of unknown cause
- Diabetes-related events
- Electrolytes abnormalities
- Eye/vision abnormalities
- Female reproductive system abnormalities
- Fractures
- Gastrointestinal abnormalities
- General signs and symptoms
- General system disorders
- Hearing abnormalities
- Hematologic abnormalities
- Immune reactions
- Injuries
- Joint disorders
- Liver-related abnormalities
- Male reproductive conditions
- Musculoskeletal symptoms
- Neoplasms
- Optic nerve abnormalities
- Other ear signs and symptoms
- Other infections
- Other investigations
- Other lab abnormalities
- Other metabolic abnormalities
- Other nervous system abnormalities
- Pancreatic abnormalities
- Peripheral neuropathy and related symptoms
- Pneumonias
- Pregnancy and Pregnancy-related events
- Psychiatric events
- Renal abnormalities
- Reproductive system abnormalities
- Respiratory events
- Seizure disorders
- Skin conditions
- Substance abuses

- TB/TB progression
- Tendon and ligaments abnormalities
- Thyroid conditions
- Urinary conditions
- Vascular events

[i] Lee, M. et al. Linezolid for Treatment of Chronic Extensively Drug-Resistant Tuberculosis. N. Engl. J. Med. 367, 1508–1518 (2012).

[ii] Cox, H. & Ford, N. Linezolid for the treatment of complicated drug-resistant tuberculosis: a systematic review and meta-analysis. Int. J. Tuberc. Lung Dis. 16, 447–454 (2012).