STREAM

The evaluation of a standard treatment regimen of anti-tuberculosis drugs for patients with MDR-TB

An international, multi-centre, open-label, parallel-group, randomised, controlled trial

STAGE 1 STATISTICAL ANALYSIS PLAN

v.1.1 | June 2017

Authorised by:

Name:	Prof. Andrew Nunn
Role:	Co-Chief Investigator
Signature:	Date: 6/5/2017
Name:	Dr. Sarah Meredith
Role:	Co-Chief Investigator
Signature:	Date: 9 June 2017
Name:	Dr. Patrick Phillips
Role:	Lead Statistician
Signature:	Date: 31/5/2017

STREAM

The evaluation of a standard treatment regimen of anti-tuberculosis drugs for patients with MDR-TB

An international, multi-centre, open-label, parallel-group, randomised, controlled trial

STAGE 1 STATISTICAL ANALYSIS PLAN

v.1.1 | June 2017

Authorised by:

Name:	Prof. Andrew Nunn
Role:	Co-Chief Investigator
Signature:	Date:
Name:	Dr. Sarah Meredith
Role:	Co-Chief Investigator
Signature:	Date:
Name:	Dr. Patrick Phillips
Role:	Lead Statistician
Signature:	Date:

GENERAL INFORMATION

This document describes and substantiates the statistical principles and methods used for the analysis of data from Stage 1 of the STREAM trial. This document is designed to support the STREAM protocol. This Statistical Analysis Plan (SAP) supersedes version 0.1 of the SAP. Every care was taken in the drafting of this SAP, but corrections or amendments may be necessary. The final version of the SAP will be signed off before database lock for final Stage 1 analysis.

The STREAM trial consists of two stages. Stage 1 involves the comparison of two treatment regimens: Regimen A and Regimen B. Stage 2 involves two additional regimens, Regimen C and Regimen D, and makes two comparisons between Regimen B and Regimen C, and Regimen B and Regimen D for the analysis of the primary endpoint. All treatment regimens are described in detail in the STREAM protocol, Section 2.1.3. Stage 1 and Stage 2 of the STREAM trial each have SAPs listed below. Each SAP has differences, but the fundamental statistical principles will be consistent across all SAPs.

Document	**Description**
Stage 1 SAP	All analyses relating to stage 1
Core Stage 2 SAP	Core analyses for stage 2 relating to analyses after the Week 76 database lock
Extensive Stage 2 SAP	Expanded analyses for stage 2 relating to analyses after the Week 76 database lock
Core Stage 2 Week 132 SAP	Core analyses for stage 2 relating to further analyses conducted after the final (Week 132) database lock
Extensive Stage 2 Week 132 SAP	Expanded analyses for stage 2 relating to further analyses conducted after the final (Week 132) database lock

Compliance:

The trial will be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, in accordance with the principles of Good Clinical Practice (GCP) as laid down by the ICH topic E6 (Note for Guidance on GCP), and the applicable regulatory requirements in the participating countries.

Sponsor:

The International Union Against Tuberculosis and Lung Disease (The Union North America)
61 Broadway, Suite 1720
New York, NY
10006 USA
Tel (main): +1 212 500 5720
Fax: +1 212 480 6040
Email: STREAM@theunion.org

Funders:

The United States Agency for International Development (USAID)

UK Medical Research Council (MRC) / Department for International Development (DFID)

Janssen Research & Development, LLC

Main Contacts:

Co-Chief Investigator

Prof. Andrew Nunn
MRC Clinical Trials Unit at UCL
Institute of Clinical Trials and Methodology
Aviation House
125 Kingsway
London, WC2B 6NH
Tel: +44 (0)20 7670 4703
Fax: +44 (0)20 7670 4829
Email: a.nunn@ucl.ac.uk

Co-Chief Investigator

Dr. Sarah Meredith
MRC Clinical Trials Unit at UCL
Institute of Clinical Trials and Methodology
Aviation House
125 Kingsway
London, WC2B 6NH
Tel: +44 (0)20 7670 4787
Fax: +44 (0)20 7670 4829
Email: s.meredith@ucl.ac.uk

LIST OF CHANGES

Changes from v1.0 to v1.1

Section	Change	Reason for change
1.2 and 1.4	Clarification that Regimen A is in accordance with 2011 WHO guidelines.	In line with text in the protocol. WHO guidelines were amended in 2016.
1.3	Patient eligibility criteria are removed and reference made to relevant section of the trial protocol.	Eligibility criteria were slightly amended in v7.0 of the protocol.
3.1	Definition of the Week 76 analysis window for the primary analysis is added.	This was previously unclear. Now in line with SAP for Stage 2.
3.2	Inclusion of culture media other than Ogawa for primary analysis.	At a small number of patient visits, Ogawa results were unavailable and other culture media had to be used.
3.2	Definition of unfavourable expanded.	This change is to bring it in line with protocol v7.0. Previous wording was ambiguous.
3.2	Patients unable to produce sputum at Week 132 can be favourable rather than not assessable.	This change is to ensure the text is consistent with the definition of favourable.
4.1	Only patients from Stage 1 are included in the Stage 1 analysis population	Previous text allowed for the possibility of an overlap between Stage 1 and Stage 2. Now that Stage 2 has started, no overlap occurred.
4.4	Inclusion of other culture media and Week 4 culture result for defining the MITT population.	As above, other culture media have been included to limit the inclusions from the MITT population where the Ogawa result is not available. In addition, cultures up to Week 4 are allowed to increase the number of patients in the analysis populations.
5.3	Removal of text referring to the visit schedule in Stage 2.	No overlap occurred between Stage 1 and Stage 2, so no reference to Stage 2 analyses or visit schedules is required.
5.8	Addition of new section specifying that the definition of treatment extensions and restarts is based on data from the treatment log.	This detail of how treatment extensions and restarts are defined was previously missing.
6.2	Addition of text 'or not' in '…further sub-classified by	Clarification

	whether or not the patients subsequently died before or during the Week 76 window'.	
6.3	Addition of weight band and smear grade at baseline for subgroup analyses	Additional subgroup analyses of interest.
6.4	Addition of 'Using the methods described in Section 6.1.1'	Clarification.
7.1.2	Addition of 'No sputum produced at Week 132' category.	Allows for distinguishing between favourable outcomes based on negative cultures only and those based on no sputum produced at Week 132.
7.3.3	Addition of '(including changes for QT prolongation)'	Clarification
8.4.1	Replacement of linear mixed effects model with simple analysis presenting mean and SD. Addition of ECG subgroup analyses by weight band and choice of fluoroquinolone in control arm.	Presentation of raw means and SD was considered more appropriate than a linear mixed effect models for this secondary outcome. Further subgroup analyses are of clinical interest to understand differences in QT prolongation.
8.4.5	Change from 'SAE and NE' to 'AE'	Clarification

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND GLOSSARY

AE	Adverse Event
AFB	Acid Fast Bacilli
AR	Adverse Reaction
AST	Aspartate aminotransferase
ALT	Alanine aminotransferase
BDQ	Bedaquiline
ICF	Informed Consent Form
CI	Chief Investigator
CFZ	Clofazimine
CRF	Case Report Form
CTA	Clinical Trials Authorisation
DAIDS	Division of Acquired Immunodeficiency Syndrome
DCF	Data Clarification Form
DOT	Directly Observed Treatment
DST	Drug Susceptibility Test
ECG	Electrocardiogram
EMA	European Medicines Agency
EMB	Ethambutol
EQA	External Quality Assurance
FDA	Fluorescein diacetate staining
US FDA	United States Food and Drugs Administration
GCP	Good Clinical Practice
GLC	Green Light Committee
HE	Health Economics
HIV	Human Immunodeficiency Virus
IDMC	Independent Data Monitoring Committee
IRB	Institutional Review Board
ISRCTN	International Standard Randomised Controlled Trial Number
ITM	Institute of Tropical Medicine
ITT	Intention To Treat
KM	Kanamycin
INH	Isoniazid
LFX	Levofloxacin
LPA	Line Probe Assay
LQAS	Lot Quality Assurance Sampling
M2	Metabolite 2
MDR	Multi-Drug Resistant
MFX	Moxifloxacin
Genotype MTBDRPlus	Rapid test for *M. tuberculosis* Complex and its resistance to Rifampicin and/or Isoniazid
Genotype MTBDR*sl*	Rapid test for *M. tuberculosis* Complex and its resistance to fluoroquinolones and/or second-line injectables/cyclic peptides and/or ethambutol
MIC	Minimal Inhibitory Concentration
MIRU-VNTR	Mycobacterial Interspersed Repetitive Units–Variable Number of Tandem Repeats
MRC CTU	Medical Research Council Clinical Trials Unit
NE	Notable Event
NTP	National Tuberculosis Programme
PK	Pharmacokinetics
PI	Principal Investigator
PIS	Patient Information Sheet
PTO	Prothionamide

PZA	Pyrazinamide
QA	Quality Assurance
QT Interval	A measure of time between the start of the Q wave and the end of the T wave in the ECG complex
QTc	QT interval corrected for heart rate
QTcF	QT interval corrected for heart rate using the Fridericia correction
REC	Research Ethics Committee
RMP	Rifampicin
SAE	Serious Adverse Event
SAR	Serious Adverse Reaction
SOP	Standard Operating Procedures
SPC	Summary of Product Characteristics
SSA	Site Specific Assessment
STREAM	The Evaluation of a Standardised Treatment Regimen of Anti-Tuberculosis Drugs for Patients with MDR-TB
SUSAR	Suspected Unexpected Serious Adverse Reaction
TB	Tuberculosis
TM	Trial Manager
TMG	Trial Management Group
TMT	Trial Management Team
TREAT TB	Technology, Research, Education, and Technical Assistance for Tuberculosis
TSC	Trial Steering Committee
UAR	Unexpected Adverse Reaction
ULN	Upper limit of normal
The Union	International Union Against Tuberculosis & Lung Disease
USAID	United States Agency For International Development
WHO	World Health Organisation
XDR	Extensively Drug Resistant
ZN	Ziehl-Neelsen

Note. In this statistical analysis plan, time (in weeks) refers to the time from randomisation, e.g. Week 132 refers to 132 weeks from randomisation.

1 TRIAL OVERVIEW

1.1 Study design

The STREAM study is an international, multi-centre, parallel-group, open-label, randomised, controlled trial.

Patients with multidrug-resistant tuberculosis (MDR-TB) are studied in the STREAM trial.

In Stage 1 of the STREAM trial, the comparison being made is between Regimen A and Regimen B.

Regimen A: The locally-used WHO-approved MDR-TB regimen forms the control treatment regimen.

Regimen B: Regimen B is the study regimen, and is based on the regimen described by Van Deun 2010[1] (updated results[2]) consisting of clofazimine, ethambutol, moxifloxacin, and pyrazinamide given for 40 weeks, supplemented by isoniazid, kanamycin, and prothionamide for the first 16 weeks.

All patients in Stage 1 of the study will be followed up to Week 132.

Under versions of the protocol prior to version 6.0 which describe only Stage 1 of the trial, patients are allocated to either Regimens A or Regimen B. Stage 2 of the trial includes two additional arms, Regimens C and D and is implemented in protocol version 6.0 and subsequent versions which also include minor changes to the eligibility criteria, visit schedule and components of the composite primary outcome.

A sensitivity analysis will be conducted to repeat the primary analysis under the definition of the primary outcome as described in version 5.2, the last version of the protocol prior to Stage 2 (see section 9.3.1).

1.2 Trial objectives

The primary objectives of Stage 1 of the STREAM trial are:

1. To assess whether the proportion of patients with a favourable efficacy outcome on Regimen B is not inferior to that on Regimen A (WHO 2011 long MDR-TB regimen), the control regimen for Stage 1, at Week 132, using a 10% margin of non-inferiority
2. To compare the proportion of patients who experience grade 3 or greater adverse events during treatment or follow-up in Regimen B as compared to Regimen A.

The secondary objectives of Stage 1 of the STREAM trial are:

1. To determine the proportion of patients with a favourable efficacy outcome on the Regimen B in each country setting
2. To compare the economic costs incurred by patients and by the health system during treatment on Regimen B as compared to Regimen A.

1.3 Patient eligibility criteria

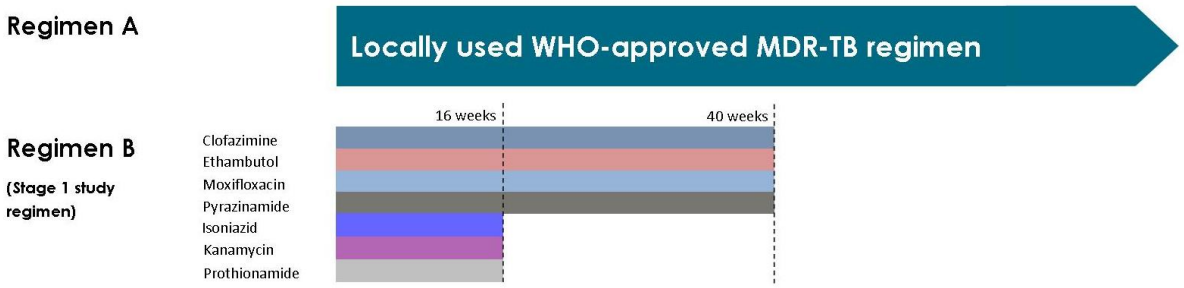
Patient eligibility criteria are listed in section 5 of the trial protocol.

1.4 Study interventions

The control regimen, Regimen A, is the locally-used WHO-approved MDR-TB regimen in accordance with 2011 WHO MDR-TB treatment guidelines. Country- or site-specific regimens are described in the STREAM Patient Management Guide.

The investigative regimen is Regimen B, and consists of moxifloxacin, clofazimine, ethambutol and pyrazinamide given for 40 weeks, supplemented by kanamycin, isoniazid and prothionamide in the first 16 weeks (intensive phase).

Figure 1: Regimen A & Regimen B



In Regimen B, all drugs are given daily (seven days a week), except for kanamycin which is initially given daily and then thrice-weekly from Week 12 onwards.

The intensive phase may be extended from 16 to 20 or 24 weeks for patients whose smear has not converted by 16 or 20 weeks, respectively, as described below.

Table 1: Regimen B doses

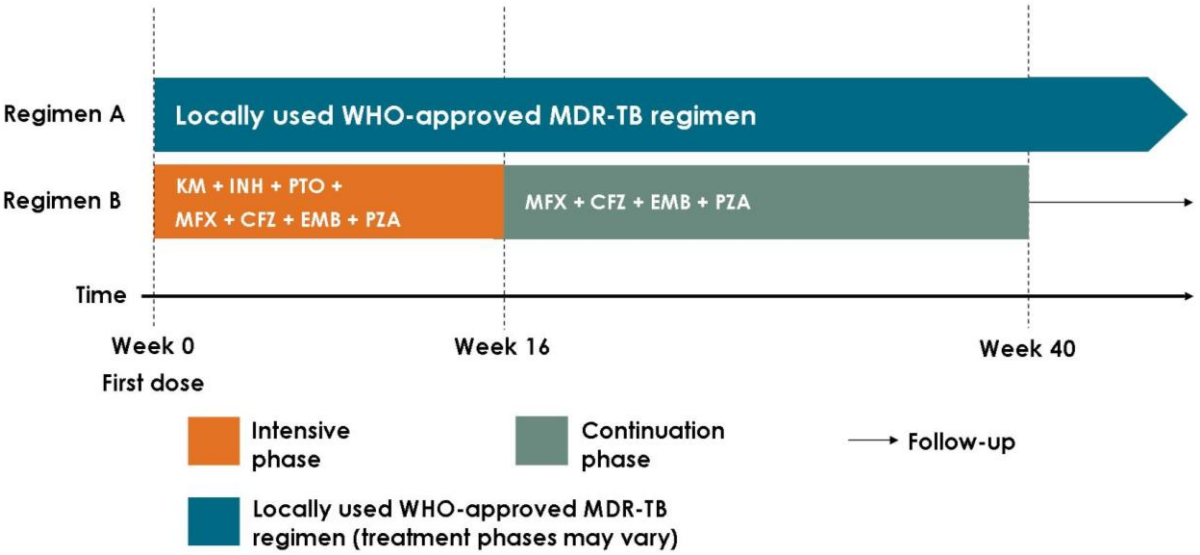
Product	Weight group		
	Less than 33 kg	33 kg to 50 kg	More than 50 kg
Moxifloxacin	400 mg	600 mg	800 mg
Clofazimine	50 mg	100 mg	100 mg
Ethambutol	800 mg	800 mg	1200 mg
Pyrazinamide	1000 mg	1500 mg	2000 mg
Isoniazid	300 mg	400 mg	600 mg
Prothionamide	250 mg	500 mg	750 mg
Kanamycin	15 mg per kilogram body weight (maximum 1g)		

Patients randomised to Regimen B will receive 40 weeks of treatment (16 weeks intensive phase plus 24 weeks continuation phase), as shown in Figure 1.

1.5 Treatment phases

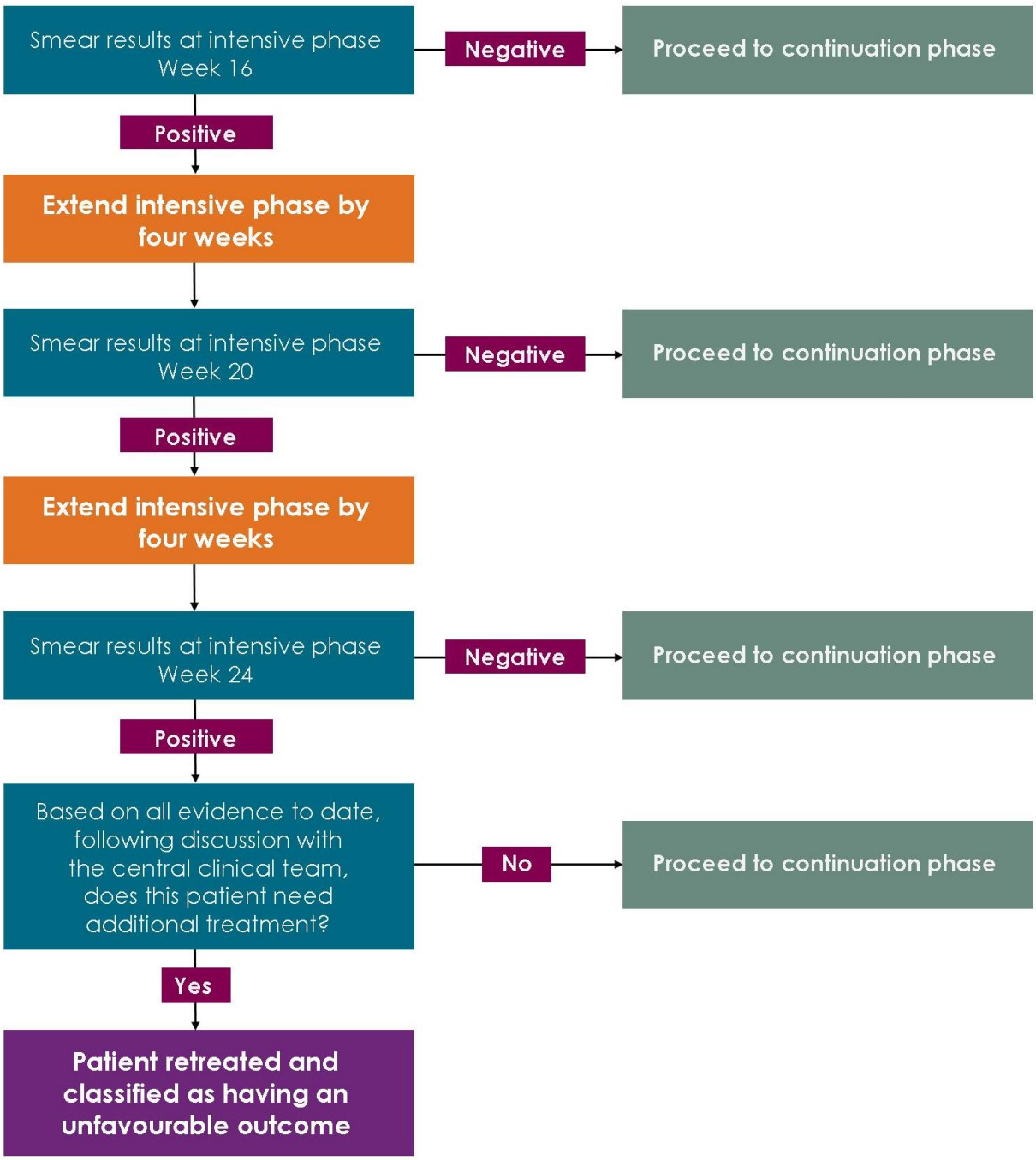
The study regimen, Regimen B, consists of 2 phases; an intense phase followed by a continuation phase, as shown in Figure. 1.

Figure 2: Treatment phases



For patients randomised to Regimen B the following algorithm will be used to determine when a patient can proceed from the intensive to the continuation phase.

Figure 3: Transition from intensive to continuation phase for patients on Regimen B



Patients randomised to Regimen B will be prescribed 40 weeks of treatment (16 weeks intensive phase and 24 weeks continuation phase). In the event of a positive (at least "scanty" on the IUATLD/WHO scale) AFB smear at Week 16, the drugs in the intensive phase of this regimens may be extended by 4 weeks, if the smear is still positive at 20 weeks the intensive phase may be extended by a further 4 weeks allowing a maximum intensive phase of 24 weeks, and hence a maximum total duration of 48 weeks treatment.

1.6 Randomisation procedure

Patients will be randomised to Regimen A or Regimen B. Randomisation will be in a 1:2 ratio in favour of Regimen B to allow more data on efficacy and safety to be collected on this regimen. Randomisation will be stratified by (1) site, (2) HIV status for sites with high TB-HIV co-infection rates.

Separate randomisation lists for each combination of strata will be prepared in advance by a statistician independent of the study, using varying block sizes. Should web access not be available at the time of randomisation, a manual alternative using sealed envelopes will be provided.

Patients will be randomised using a web-based randomisation system. Access to the web-based system will be controlled through an authorised username and password. Before treatment allocation the patient's eligibility will need to be confirmed, and their site, HIV status, and CD4 count entered into the database.

2 SAMPLE SIZE

2.1 Power to demonstrate non-inferiority in the primary efficacy outcome

A 10% margin of non-inferiority is considered an acceptable reduction in efficacy given the considerably reduced pill burden and duration and the expected increase in adherence in reducing a treatment regimen from 104 weeks (as with Regimen A), to 40 weeks (as with regimen B).

A meta-analysis of treatment outcome in patients with MDR-TB found an overall favourable outcome of 64% (95% CI 59-68) in patients given individualised treatment and 54% (95% CI 43-68) in patients given standardised treatment[3]. A reasonable estimate of the efficacy of regimen A in the STREAM trial would therefore be 70%.

Based on the experience with regimen B[1], a reasonable estimate of its efficacy in the STREAM trial would be between 75% and 85%. The lower estimate is used for the sample size calculations below.

Based on a 2:1 allocation ratio in favour of Regimen B to Regimen A, Table 2 gives the total number of patients required to demonstrate non-inferiority under the specified scenarios using a margin of non-inferiority of 10%. These totals allow for 20% of patients being classified as not assessable in a per-protocol analysis and are based on a one-sided level of significance of 2.5%.

Table 2: Power to demonstrate non-inferiority in the primary efficacy outcome

Power	Percentage favourable outcomes in Regimen A	Difference in percentage favourable outcomes in Regimen B compared to Regimen A		
		0%	5%	10%
80%	60%	1060	464	255
	65%	1005	435	238
	70%	928	398	214
90%	60%	1419	620	340
	65%	1345	583	318
	70%	1242	533	287

Therefore, 398 patients would be required (rounding to 400 gives: 267 on Regimen B and 133 on Regimen A) to demonstrate non-inferiority with 80% power assuming 70% favourable outcomes in Regimen A and 75% in Regimen B and 20% not assessable. A larger difference in response rates of 10% would require fewer patients and could also be demonstrated with greater than 90% power with a total enrolment of approximately 400 patients.

A 10% margin of non-inferiority is considered an acceptable reduction in efficacy given the considerably reduced pill burden, duration, and resource utilisation, and the expected increase in adherence in reducing a treatment regimen from 104 weeks (as with Regimen A), to 40 weeks (as with Regimen B).

If the difference in response rates in favour of Regimen B is more than 10% it may be possible to demonstrate superiority of that regimen over the control for stage 1, Regimen A.

At least 400 patients will need to be enrolled across all countries to give sufficient power to demonstrate non-inferiority. Patients will be randomised to Regimen B and Regimen A in the ratio 2:1.

2.2 Power to demonstrate non-inferiority in the primary safety outcome

Assuming a sample size of 400 on a 2:1 allocation ratio in favour of Regimen B to Regimen A, Table 3 gives the power available to demonstrate non-inferiority in the primary safety outcome under different proportions of grade 3 or 4 events on Regimen A and Regimen B. These calculations assume a margin of non-inferiority of 10% and a one-sided level of significance of 2.5%. All randomised patients who have received at least one dose of study medication will be included in the safety analysis.

Table 3: Power to demonstrate non-inferiority in the primary safety outcome

Proportion grade 3 or 4 on Regimen A	Assuming same proportion in Regimen A and Regimen B	Assuming an absolute 5% lower proportion on Regimen B than Regimen A
10%	88%	99%
15%	75%	99%
20%	65%	96%
25%	58%	93%
30%	53%	89%
35%	50%	86%
40%	48%	83%

3 PRIMARY OUTCOMES

3.1 Primary analysis Week 132 window

The Week 132 window is defined as the time period from six weeks before 132 weeks since randomisation with no upper bound, i.e. from Week 126 with no upper bound.

For the purposes of defining the primary efficacy outcome, the Week 76 analysis window is defined as the time period from six weeks before 76 weeks since randomisation to six weeks after 76 weeks since randomisation, i.e. from Week 70 to Week 82. This definition is used for consistency with STREAM Stage 2, but any tabulations of secondary outcomes by visit will use the visit windows as defined in section 5.3 below.

3.2 Primary efficacy outcome

The primary efficacy outcome measure is the proportion of patients with a favourable outcome (as defined below) at Week 132.

Culture results obtained using acidified Ogawa (Kudoh medium) will be used in the primary efficacy analysis, although results from other culture media will be used if the Ogawa result is missing.

A positive culture on Ogawa is defined as at least one colony and a negative culture is defined as absence of growth (no colonies).

Favourable

A patient's outcome will be classified as **favourable** if their last two culture results are negative unless they have previously been classified as unfavourable. These two cultures must be taken on separate visits (on different days); the latest of which being within the Week 132 window.

Patients that don't have a culture result within the Week 132 window because they were unable to produce sputum, will be classified as favourable if their last two cultures before the Week 132 window are negative and they have not previously been classified as unfavourable; such patients will be identified separately in tables (see section 6.2).

Unfavourable

A patient's outcome will be classified as **unfavourable** if:

1. They are discontinued from their allocated study treatment and subsequently restarted on a different MDR-TB regimen
2. Treatment is extended beyond the scheduled end of treatment for any reason other than making up of days when no treatment was given (missed treatment) for a maximum of eight weeks. A maximum of 14 days of extra treatment (irrespective of reason) is acceptable before it is classified as treatment extension. In addition, if the intensive phase of treatment has been extended for delayed sputum conversion (maximum 8-week extension permitted) the scheduled end of treatment will also be extended by the same amount, in accordance with Section 7.3.2 of the protocol.
3. They are restarted on any MDR-TB treatment after the scheduled end of treatment, but before 132 weeks after randomisation.
4. They change their allocated study treatment for any reason other than (1) the replacement of a single drug or (2) for patients allocated to Regimen A when the change is as a result of changes in local guidelines and not related to any change in the patient's circumstances or condition.
5. Bedaquiline is started where the allocated regimen did not originally contain that drug (Regimen A or Regimen B).

6. A drug from the class of nitroimidazoles is started
7. They die at any point during treatment or follow-up
8. At least one of their last two culture results, from specimens taken on separate occasions, is positive
9. They do not have a culture result within the Week 76 window or thereafter

Providing none of the other criteria above are met, starting a single drug is not considered to be a substantial change to the regimen and therefore does not result in an unfavourable outcome, with the exception of adding bedaquiline or a drug from the class of nitroimidazoles.

An extension of the intensive phase of treatment in any study arm does not constitute an unfavourable outcome, as long as the extension is in accordance with either the algorithms described in section 7.3.2 for patients on Regimen B, or the locally-used WHO 2011 long MDR-TB regimen for patients on Regimen A. Similarly, the discontinuation of drugs that are not replaced does not constitute an unfavourable outcome.

Changes of treatment in patients allocated to Regimen A that result from a change in local guidelines not related in any way to any change in the patient's circumstances or condition will not be classified as unfavourable. A sensitivity analysis will be conducted where these changes *are* classified as unfavourable. However, this sensitivity analysis can only result in an increase in unfavourable outcomes on Regimen A, thereby increasing the chance of demonstrating the non-inferiority of Regimen B and therefore the primary analysis described here is more conservative.

All re-infections with a different strain are classified as **not assessable**.

A patient who has a culture result within the Week 76 window or thereafter, but not within the Week 132 window, having not otherwise been classified as unfavourable (based on the definitions above) will be regarded as **not assessable** and will be excluded from the primary analysis provided their last two cultures, from specimens taken on separate occasions, are negative. Such patients that don't have a culture result within the Week 132 window because they were unable to produce sputum will be instead classified as **favourable**. Any patient who does not have a culture result within the Week 132 window and does not fulfil these criteria will be classified as **unfavourable**. These definitions apply to both Regimen A and Regimen B.

3.3 Primary safety outcome

The primary safety outcome measure is the proportion of patients experiencing a grade 3 or greater adverse event, as defined by the DAIDS criteria[4], at any time during treatment and follow-up.

4 ANALYSIS POPULATIONS

4.1 Stage 1 analysis population

Only patients randomised in Stage 1 of the STREAM trial will be included in the Stage 1 analysis population.

4.2 Intention-to-treat (ITT)

All randomised patients will be included in the ITT analysis population.

4.3 Safety population

All randomised patients that have taken at least one dose of treatment will be included in the safety analysis population.

4.4 Modified intention-to-treat (mITT)

The mITT population is defined as all randomised patients that have a positive culture for *M. tuberculosis* on acidified Ogawa (Kudoh medium) or other culture media if the Ogawa result is not available, at screening or randomisation or up to Week 4, with the exception of patients with isolates taken before randomisation that are subsequently found to be susceptible to rifampicin, and patients with isolates taken before randomisation that are subsequently found to be resistant to both fluoroquinolones and second-line injectables (i.e. XDR-TB) on phenotypic DST. Results from the central reference laboratory will take priority over any results from local laboratories where available.

4.5 Per protocol (PP)

The PP population will be the same as the mITT population with the exclusion of patients not completing a protocol-adherent course of treatment, other than for treatment failure or death. Treatment failure is defined as failure to attain and maintain culture negativity until the end of allocated treatment.

4.5.1 Definition of a protocol-adherent course of treatment

Patients will be excluded from the per-protocol analysis if they do not complete a protocol-adherent course of treatment, other than for treatment failure or death.

A patient will have completed a protocol-adherent course of treatment when they have taken 80% of doses within 120% of the minimum duration in both the intensive phase and in the whole treatment period. For this purpose, a dose is defined as all the study medications at the correct dose for that particular day.

For Regimen B, **with or without** an extension of the intensive phase, a patient will have completed a protocol-adherent course of treatment if they have taken:

- 90 doses (80% of 16 weeks) within 134 days (120% of 16 weeks) in the intensive phase, and
- 224 doses (80% of 40 weeks) within 336 days (120% of 40 weeks) over the whole treatment period (i.e. the combined intensive and continuation phases) regardless of treatment extensions.

The same algorithm will apply for Regimen A, the control regimen; the exact number of doses and days depends on the duration of the intensive and continuation phases of Regimen A.

5 GENERAL ANALYSIS PRINCIPLES

5.1 Analysis populations

The analyses of the primary outcomes will be based on both the mITT and the PP populations. All patients included in the analysis will be analysed in the treatment group to which they were originally assigned. Further sensitivity analyses are planned (see Section 9 Sensitivity Analyses).

5.2 Treatment and follow-up phase definitions

For the purpose of analysis, the screening, treatment, and follow-up phases for an individual patient will be defined as follows:

- **Screening phase**
  - Start: date of screening consent
  - End: day before randomisation
- **Treatment phase**
  - Start: date of randomisation.
  - End: date of last dose of any TB treatment defined as last dose of any TB treatment (including retreatment for relapse), plus 7 days.
- **Follow-up phase**
  - Start: the day after the end of the treatment phase.
  - End: date of the last patient contact (scheduled or unscheduled, or other contact e.g. phone call).

The treatment phase includes any extension of treatment or retreatment, and so the Allocated Treatment phase is defined as follows:

- **Allocated Treatment phase**
  - Start: date of randomisation.
  - End: date of last dose of trial treatment defined as last dose of allocated regimen or last dose before the addition of a new drug, whichever happens sooner, plus 7 days.

5.3 Visit window definitions

During Stage 1, patients will be assessed at screening, randomisation (Week 0), Week 1, Week 2, Week 3, Week 4, and at 4-weekly intervals throughout the study, until the end of follow-up, irrespective of whether on treatment or in the post-treatment follow-up phase.

For the purpose of analysis, each scheduled visit will have a window before and after the target date, calculated from date of randomisation. For the purpose of analysis, each scheduled visit will have a window before and after the target date, calculated from date of randomisation. When referring to a visit hereon, this implies within the defined visit window as specified below.

Visit	target date, days after randomisation +1	Analysis window
Screening / Baseline	1	Date of screening consent - 1
Week 4	29	2-42
Week 8	57	43-70
Week 12	84	71-98
Week 16	113	99-126
...		
Week a	b = 1+ (a x 7)	(b-14) – (b+13)
...		
Week 120	841	827-854
Week 124	869	855-882
Week 128	N/A (included within 132 week analysis window)	
Week 132	925	833-no upper bound

Any visit, scheduled or unscheduled, that falls into the analysis window will be assigned to that visit for the purpose of analysis. If two visits fall within the same interval, the one closest to the target date will be used for analyses by visit, so that there is only one unique visit for each patient and analysis time-point.

There are additional study visits at Weeks 1, 2 and 3 only for ECG monitoring. For the analysis of ECG data only, there will be additional visit windows: Week 1 (2-11), Week 2 (12-18), Week 3 (19-25) and the Week 4 visit window will be modified to (26-42).

5.4 Definition of a culture result

A culture result will be called positive for *M. tuberculosis* if the culture tests positive for the presence of microorganisms, at least one colony, and the microorganisms present are then identified as being *M. tuberculosis*. However, if an identification test is not carried out for a particular culture, then for analysis purposes a culture will still be considered positive for *M. tuberculosis* if the culture tests positive for the presence of microorganisms and if that culture result is obtained seven days or more since the start date of sputum processing and incubation of the inoculated Ogawa. If the culture result is obtained less than seven days since the start date of sputum processing and incubation of the inoculated Ogawa, the culture result will not be considered as positive for *M. tuberculosis*, and the culture result will be considered missing in the analysis.

If more than one culture result is available from sputum collected on the same day, this will be regarded as a single culture result for the purposes of all analyses with the following overall result:

i. **Positive**, if at least one of the culture results is positive
ii. **Negative**, if at least one of the culture results is negative and none of the culture results are positive
iii. **Contaminated** if at least one of the culture results is contaminated and none of the culture results are positive or negative.
iv. **Missing**, if no culture result is available.

5.5 Definition of a smear result

A smear result will be called positive if it is graded as 'scanty' or 'rare AFB' or at least 1+.

If more than one smear result is available from sputum collected on the same day, this will be regarded as a single smear result for the purposes of all analyses with the following overall result:

i. **Positive**, if at least one of the smear results is positive
ii. **Negative**, if at least one of the smear results is negative and none of the smear results are positive
iii. **Missing**, if no smear result is available.

5.6 Reference laboratory bacteriology

A number of clinical isolates will be sent from the STREAM sites to a reference laboratory at the Institute of Tropical Medicine (ITM) in Antwerp, Belgium. Drug sensitivity results from the reference laboratory will be used in all analyses in preference to those obtained from local site laboratories where available.

5.7 Adverse events

For all analyses of adverse events, only those occurring after randomisation will be included.

5.8 Defining treatment extensions and restarts

For the purposes of classifying the primary outcome, only data from the treatment logs (CRF 18) will be used to determine treatment extensions, changes or restarts.

6 ANALYSIS OF PRIMARY OUTCOMES

6.1 Primary efficacy analyses

6.1.1 Modelling technique used in analysis

For the primary efficacy analysis the difference in proportions of favourable outcome between two specified trial regimens with corresponding 95% confidence intervals and p-values will be estimated using a stratified analysis of the risk difference from each stratum using Cochran-Mantel-Haenszel weights.[5] The analysis will be stratified only by HIV status: HIV negative and HIV positive. Where there is a difference between data used for stratification and correct data (if randomisation was inadvertently done on incorrect data), the correct data will be used for adjustment in the analysis.

6.1.2 Primary efficacy analysis: non-inferiority of Regimen B

Non-inferiority will be demonstrated if the upper bound of the 95% confidence interval of the difference in proportion of favourable outcomes between Regimens A and B is less than the 10% margin of non-inferiority in both the mITT and PP populations.

6.1.3 Superiority of Regimen B

If Regimen B is declared non-inferior to Regimen A, then superiority of Regimen B compared to Regimen A will be assessed.

If the upper bound of the 95% confidence interval of the difference in proportion of favourable outcomes between Regimens A and Regimen B is less than zero, then superiority of Regimen B compared to Regimen A will be declared. For this analysis, the mITT population will be primary and the PP population will be one of several secondary analyses.

6.2 Tabulation of primary endpoint classification

Since the primary endpoint is a composite of various components, the actual reason (component) for outcome will also be tabulated by treatment arm.

Patients will be classified by the first event that made the patient unfavourable (see section 3.3) and further sub-classified by their microbiological outcome at the time that this outcome occurred (see section 7.1 below) and further sub-classified by whether or not the patients subsequently died before or during the Week 76 window. For example, a patient that has their treatment regimen changed during the treatment phase but subsequently has a positive culture during the Week 76 window will be classified as having had their regimen changed and further sub-classified by whether they had achieved culture conversion when their regimen was changed.

6.3 Subgroup analyses

This primary efficacy analysis will be repeated in subgroups according to HIV infection status, baseline drug resistance patterns (i.e. resistance to pyrazinamide, a fluoroquinolone, a second-line injectable, and isoniazid), BMI (<18, 18-<20, 20-<25, ≥25), cavitation (presence, absence), study centre, age (<45, 45-<65, ≥65), sex, smoking history (current smoker, ex-smoker and never smoked), weight band, smear grade at baseline, and ethnicity.

In addition, to evaluate any effect of the minor differences in the protocol after the initiation of Stage 2, the primary efficacy analysis will be repeated in the subgroup of patients enrolled under protocol 5.2 and prior versions, and in the subgroup of patients enrolled under protocol 6.0 (Stage 2) and later versions.

6.4 Primary safety analysis

The primary safety outcome is the occurrence of a Grade 3 or greater adverse events.

The difference in proportion of patients experiencing a grade 3 or greater adverse event, as defined by the DAIDS criteria, during the treatment and follow-up phases, between Regimen B and Regimen A with corresponding two-sided 95% confidence intervals and p-values will be estimated (using the methods described in Section 6.1.1).

This analysis will be conducted on the whole study period, and separately for each phase (Treatment, Follow-up and Allocated treatment).

7 ANALYSIS OF SECONDARY OUTCOMES

7.1 Microbiological outcome

Sputum culture negative status is defined as two consecutive negative cultures from sputa collected on different days without an intervening positive. Culture negative status is lost when a culture result is positive, but can subsequently be re-achieved if two consecutive cultures from sputa collected on different days are negative without an intervening positive

7.1.1 Microbiological outcome at unfavourable outcome

The microbiological outcome at unfavourable outcome is defined using culture results up to and including the date of the first event that made their primary efficacy outcome unfavourable (the 'unfavourable outcome event'). It is defined as follows:

- **Culture negative.** Culture negative status was satisfied at the date of the unfavourable outcome event.
- **Never culture converted**. The patient never achieved culture negative status at any time during the study prior to the unfavourable outcome event.
- **Culture positive**. Culture negative status was achieved at some point during the study, but was not satisfied at the date of the unfavourable outcome event. Culture positive will be further classified as **Culture positive: Reinfection** if it has been shown that the M. tuberculosis strain of the positive culture is different to baseline; and **Culture positive: Relapse** otherwise.

Patients will be classified by the first event that made the patient unfavourable and further sub-classified by their microbiological outcome at unfavourable outcome and further sub-classified by whether the patients subsequently died (see Section 6.2).

7.1.2 Microbiological outcome at Week 132

The microbiological outcome at Week 132 will be defined as follows:

- **Culture negative at Week 132**. Culture negative status was satisfied when last seen with a negative culture within the Week 132 window.
- **No sputum produced at Week 132**. Culture negative status was satisfied when last seen but there were no culture results during the Week 132 window because they were unable to produce sputum.
- **Culture negative: did not complete follow-up.** There were no culture results during the Week 132 window (and this was not because no sputum was produced) and culture negative status was satisfied when the patient was last seen.
- **Never culture converted**. The patient never achieved culture negative status at any time during the study up to Week 132.

- **Culture positive**. Culture negative status was achieved at some point during the study, but was not satisfied when the patient was last seen (at least one of the last two non-missing culture results was positive). Culture positive will be further classified as **Culture positive: Reinfection** if it has been shown that the M. tuberculosis strain of the positive culture is different to baseline; and **Culture positive: Relapse** otherwise.

Microbiological outcome at Week 132 will be tabulated by regimen. Patients that die will be classified as above based on their available culture results when last seen, but classified separately from patients that did not die.

7.2 Efficacy outcomes

Secondary efficacy outcomes will be analysed on both the mITT and PP analysis populations.

7.2.1 Time to sputum smear and culture conversion

Time to sputum smear conversion is defined as the time from randomisation to the first of two consecutive negative sputum results, collected on separate days. All patients in the respective analysis population will be included in this analysis, except those with no positive smear result at screening or randomisation. Patients that never achieve smear conversion will be censored at the date of collection of sputum that yielded their last smear result.

Time to sputum culture conversion is defined as the time from randomisation to the first of two consecutive negative culture results, collected on separate days. Patients that never achieve culture conversion will be censored at the date of collection of sputum that yielded their last culture result.

Median time to sputum smear and culture conversion will be calculated for Regimen A and Regimen B.

A hazard ratio with corresponding two-sided 95% confidence intervals and p-value will be estimated using a Cox Proportional Hazards model will be used, adjusted for the stratification factors.

The equality of survivor functions for time to sputum conversion for Regimen A and Regimen B will be compared using a (Wilcoxon) Log rank test, stratified by the randomisation stratification factors.

The assumption of proportional hazards will be tested using the proportional hazards test based on the Schoenfeld residuals after fitting the Cox Proportional Hazards model.

Even when Kaplan-Meier curves of time to culture conversion have been shown to diverge in the presence of an effective drug (such as bedaquiline), they tend to converge later in follow-up potentially violating the assumption of proportional hazards. In the case where there is adequate evidence that the proportional hazard assumptions are violated at the 5% level (i.e. $p<0.05$), methods where proportional hazards is not a necessary assumption will be used, such as restricted mean survival time.

The analyses above of time to sputum smear conversion and time to sputum culture conversion will be repeated with the alternative definition as time from randomisation to the first negative culture or smear result respectively (without the need for a second negative culture or smear to confirm).

7.2.2 Time to unfavourable efficacy outcome

Time to unfavourable efficacy outcome is defined as the time from randomisation to the first event that results in the definition of an unfavourable efficacy outcome for that patient (as defined in Section 3.2). Patients that do not culture convert during the treatment and follow-up phases (i.e. fail to have 2 consecutive culture negative results), and have not otherwise been called unfavourable, will be called unfavourable at the date of the last visit when a culture positive result was obtained.

Patients classified as favourable or not assessable will be censored in this analysis at the date of collection of sputum that yielded their last negative culture result.

Median time to unfavourable efficacy outcome will be calculated for Regimen A and Regimen B.

A hazard ratio with corresponding two-sided 95% confidence intervals and p-value will be estimated using a Cox Proportional Hazards model will be used, adjusted for the stratification factors.

The equality of survivor functions for time to unfavourable efficacy outcome for Regimen A and Regimen B will be compared using a (Wilcoxon) Log rank test, stratified by the randomisation stratification factors.

The assumption of proportional hazards will be tested using the proportional hazards test based on the Schoenfeld residuals after fitting the Cox Proportional Hazards model.

In the case where there is adequate evidence that the proportional hazard assumptions are violated at the 5% level (i.e. $p<0.05$), methods where proportional hazards is not a necessary assumption will be used, such as restricted mean survival time.

7.2.3 Time to cessation of clinical symptoms based on PI assessment

Time to cessation of clinical symptoms is defined as the time from randomisation to the first of two consecutive visits where cessation of **all three** of the TB symptoms: a productive cough, fever, and night sweats, as reported by the patient. Patients with none of the TB symptoms at screening and none of the TB symptoms at baseline with be excluded from this analysis. This definition matches the definition of time to culture conversion as the first of two consecutive symptom-free months.

Median time to cessation of clinical symptoms will be calculated for Regimen A and Regimen B.

A hazard ratio with corresponding two-sided 95% confidence intervals and p-value will be estimated using a Cox Proportional Hazards model will be used, adjusted for the stratification factors.

For patients who do not cease clinical symptoms, cessation of clinical symptoms will be censored at the patients' last visit.

The equality of survivor functions for time to cessation of clinical symptoms for Regimen A and Regimen B will be compared using a (Wilcoxon) Log rank test, stratified by the randomisation stratification factors.

The assumption of proportional hazards will be tested using the proportional hazards test based on the Schoenfeld residuals after fitting the Cox Proportional Hazards model.

In the case where there is adequate evidence that the proportional hazard assumptions are violated at the 5% level (i.e. $p<0.05$), methods where proportional hazards is not a necessary assumption will be used, such as restricted mean survival time.

7.3 Safety outcomes

Safety outcomes will be analysed using the safety analysis population.

7.3.1 Placement of events by study phases

Adverse events are placed in study phases (see section 6.1 for definitions) based on the start date. If the start date of an event falls between (or on) the start and stop date of a phase, the AE is attributed to that phase.

In case of partial start dates, the following approach is used:

- **Missing day only**: The event is placed in all phases that overlap the given month and year for the event, excluding any phases that start after the end date of the AE (if specified).
- **Missing day and month only**: The event is placed in all phases that overlap the given year for the event, excluding any phases that start after the end date of the AE (if specified).
- **Missing start date:** The event is placed in the treatment phase, unless the end date of the AE is specified and is before randomisation, in which case the event is placed in the screening phase.

7.3.2 All-cause mortality during treatment or follow-up

All-cause mortality is defined as a patient who has died from any-cause (both TB- or non-TB-related) while in the trial either during treatment or during follow-up.

The number of patients who die during treatment and follow-up will be tabulated by treatment arm.

Survival analysis will be conducted for time to death.
A hazard ratio with corresponding two-sided 95% confidence intervals and p-value will be estimated using a Cox Proportional Hazards model will be used, with no stratification.

For patients that do not die, time will be censored at their final visit.

The equality of survivor functions for time to death for Regimen A and Regimen B will be compared using a (Wilcoxon) Log rank test, stratified by the randomisation stratification factors.

7.3.3 Change of regimen for adverse events

A change of regimen for an adverse event is defined as when a patient's regimen is modified in any way (including stopping a drug, changing the dose of a drug or starting a new drug) with the main reason being an adverse event (including changes for QT prolongation).

The difference in proportion of patients who have a change of regimen for adverse events between Regimen B and Regimen A will be calculated with 95% confidence intervals.

7.3.4 Proportion of patients experience treatment-related grade 3 or greater adverse events occurring on treatment and during the follow-up period

The proportion of patients with treatment-related grade 3 or greater adverse events that occur on treatment and during the follow-up period is defined as the number of grade 3 or greater adverse events considered to be possibly, probably or definitely related to treatment.

The difference in proportion of treatment-related adverse events between Regimen B and Regimen A will be calculated.

7.3.5 Adherence to treatment

Adherence to treatment is defined as either **adherent**; if a patient has taken at least 80% of doses within 120% of the time (as defined above), or **non-adherent**; if a patient has not met these conditions.

The difference in proportion of those who have been adherent to treatment between Regimen B and Regimen A will be calculated.

7.4 Acceptability outcomes

In selected sites, acceptability of Regimen A and B to stakeholders will be analysed in terms of:

- Costs to the health system
- Household costs
- Patient treatment and support experiences
- Health worker experiences.

The analyses of health and household costs and patient and health worker experiences will be described in a separate document.

8 DATA SUMMARIES

8.1 Recruitment and baseline characteristics

8.1.1 Recruitment, screening, & eligibility

The number of patients screened, randomised and treated will be tabulated by centre and treatment arm. The number of patients who failed screening, and the reasons for ineligibility will be presented by randomised group.

8.1.2 Exclusions from analysis

The number of patients excluded from the mITT and PP analysis populations will be tabulated by treatment arm and by reason for exclusion.

8.1.3 Baseline characteristics

All eligible patients randomised will be included in tables of baseline comparisons by treatment group. Characteristics will include sex, age, ethnicity, height, weight, BMI, and laboratory parameters such as, HIV status, CD4 count (if applicable), smoking status (current smoker, ex-smoker, never smoked) smear and culture status, and drug susceptibility status for a number of TB drug types. The baseline characteristics table will be repeated for each of the ITT, safety, PP and mITT populations.

8.2 Efficacy and adherence

Each analysis will be repeated using the mITT and PP analysis populations.

8.2.1 Sputum smear and culture

Sputum smear and culture results (positive or negative) will be tabulated by visit and treatment arm.

8.2.2 Adherence

Adherence will be summarised by treatment arm as the percentage of each of the intensive and continuation phase doses completed and overall across both phases.

8.2.3 Drug resistance

Drug resistance at screening or baseline will be tabulated by treatment arm, with separate tables for genotypic and phenotypic DSTs. Acquired resistance to any drugs will also be described and tabulated by treatment arm using the last available DST result for each drug for each patient.

In addition, acquired resistance to any drugs will also be described and tabulated by treatment arm using any available post-randomisation DST result only from the reference laboratory at the Institute of Tropical Medicine (ITM) in Antwerp (i.e. ignoring any results from local site laboratories) for each drug for each patient.

In a further analysis, acquired resistance to any drugs will also be described and tabulated by treatment arm using any available post-randomisation DST result (i.e. classifying as resistant if any result is resistant from ITM or local site laboratories) for each drug for each patient.

Acquired resistance for each definition will also be tabulated by category of primary endpoint and microbiological outcome to determine any cases of acquired resistance that didn't result in an unfavourable outcome.

8.3 Retention and description of follow-up

8.3.1 Description of follow-up and populations

Completion of treatment and completion of scheduled follow-up will be summarised by treatment group including reasons for failure to complete treatment or follow-up. This analysis will be using the ITT, PP, safety, and mITT analysis populations.

8.4 Safety outcomes

Safety outcomes will be analysed using the safety analysis population.

8.4.1 Electro-cardiology

Both mean (and SE) QT, QTcF and heart rate (HR) by visit and treatment arm, and mean (and SE) QT, QTcF and HR change from baseline by visit (within visit window) and treatment arm will be tabulated.

QT and QTcF will be categorised (<450, 450-479, 480-499, ≥500) and tabulated by visit and treatment arm, and highest post-randomisation value overall by treatment arm. Change from baseline of QT and QTcF will also be categorised (<30, 30-59, ≥60) and tabulated by treatment arm, and highest post-randomisation value overall by treatment arm.

These tables will be done for the whole study period and repeated for the treatment phase only.

Time to first QTcF over 450ms and first QTcF over 500ms and QTcF increase from baseline by 30ms and by 60ms analyses will be conducted. Number of each of these events (i.e. whether a threshold was exceed or not) will be tabulated by treatment arm. Hazard ratios with corresponding two-sided 95% confidence intervals will be estimated using a Cox Proportional Hazards model will be used, with no stratification.

The outcomes will be censored at the patients' last visit.

The equality of survivor functions for time to QTcF over 450ms and over 500ms and QTcF increase from baseline by 30ms and by 60ms for Regimen A and Regimen B will be compared using a (Wilcoxon) Log rank test, with no stratification.

The assumption of proportional hazards will be tested using the proportional hazards test based on the Schoenfeld residuals after fitting the Cox Proportional Hazards model.

In the case where there is adequate evidence that the proportional hazard assumptions are violated at the 5% level (i.e. $p<0.05$), methods where proportional hazards is not a necessary assumption will be used, such as restricted mean survival time.

QTcF will be summarised by visit and by treatment arm using means and standard deviations. Mean and +/- 1 SD will be plotted by visit and treatment arm. This will be repeated for change in QTcF from baseline.

It is likely that treatment and dose changes will impact on QTcF and so this analysis will be repeated ignoring any results after discontinuation or change of dose of any drug.

All of the electro-cardiology analysis will be repeated separately by HIV status, by sex, by weight band, and by choice of fluoroquinolone in the control arm (levofloxacin or moxifloxacin). An interaction between covariates and QTcF will be tested by including an interaction term in the linear mixed models for QTcF and change in QTcF from baseline.

8.4.2 Liver function

ALT, and AST will be categorised (<1xULN; 1-<3xULN, 3-<5xUlLN; 5ULN-<10ULN; ≥10ULN) and tabulated by visit and treatment arm.

Mean ALT, and AST will be presented by visit and treatment arm. The number of patients experiencing more than or equal to five times above the upper normal limit will be tabulated by arm.

8.4.3 Hearing impairment

The number (and proportion) of patients reporting experiencing clinically significant hearing loss (unilateral or bilateral) during the combined treatment and follow-up period will be tabulated by treatment arm.

8.4.4 Weight gain

Patient weight will be tabulated by treatment arm and visit in addition to change from baseline weight by visit and treatment arm.

8.4.5 Adverse Events

AE data will be tabulated as follows:

i. Event grade by treatment arm, with details of type of AE listed with frequencies for each event grade
ii. Event relatedness to study drugs by treatment arm
iii. Number of patients experiencing Grade 3 or higher adverse events by treatment arm
iv. Number of Grade 3 or higher adverse events by treatment arm.

9 SENSITIVITY ANALYSES

9.1 Additional adjusted and unadjusted primary efficacy analyses

All primary efficacy analyses will be repeated:

1. Unadjusted for any covariates.
2. Adjusted for randomisation stratification factors HIV status and centre. Small strata with fewer than 10 patients will be combined within geographical regions.
3. Adjusted for randomisation stratification factors and any additional important covariates such as cavitation at baseline or baseline bacillary load.

9.2 Additional analysis populations for primary efficacy analysis

In addition to the mITT and PP analysis populations, the primary efficacy analyses will be repeated for the (1) ITT analysis population, (2) the safety analysis population, and (3) the mITT analysis population excluding patients that didn't start treatment.

9.3 Reclassification of primary efficacy endpoint

9.3.1 Classification using pre-Stage 2 primary outcome definitions

A sensitivity analysis will be conducted to repeat the primary analysis under the definition of the primary outcome as described in version 5.2, the last version of the protocol prior to Stage 2.

9.3.2 Classification including treatment changes due to changes in local guidelines as unfavourable

A sensitivity analysis will be conducted where any treatment changes due changes in local guidelines are classified as unfavourable (rather than not assessable). However, this sensitivity will only result in more unfavourable outcomes on Regimen A (if any), thereby increasing the chance of demonstrating the non-inferiority of Regimen B.

10 DATA SHARING

Results concerning time to sputum culture conversion will be shared with the TREAT-TB transmission modelling team in order that the longer term impacts of reducing treatment times may be assessed. Any data sharing will follow the MRC CTU SOP 61 on Data Sharing.

11 REFERENCES

1. Van Deun A, Maug AKJ, Salim MAH, et al. Short, Highly Effective, and Inexpensive Standardized Treatment of Multidrug-resistant Tuberculosis. *Am J Respir Crit Care Med* 2010; **182**(5): 684-92.

2. Aung KJ, Van Deun A, Declercq E, et al. Successful '9-month Bangladesh regimen' for multidrug-resistant tuberculosis among over 500 consecutive patients. *Int J Tuberc Lung Dis* 2014; **18**(10): 1180-7.

3. Orenstein EW, Basu S, Shah NS, et al. Treatment outcomes among patients with multidrug-resistant tuberculosis: systematic review and meta-analysis. *The Lancet Infectious Diseases* 2009; **9**(3): 153-61.

4. DAIDS RSC. Division of AIDS table for grading the severity of adult and pediatric adverse events. Version 1.0, December 2004 (Clarification August 2009) ed. DAIDS Regulatory Support Centre; 2009.

5. Mohamed K, Embleton A, Cuffe RL. Adjusting for covariates in non-inferiority studies with margins defined as risk differences. *Pharmaceutical statistics* 2011; **10**(5): 461-6.