National Institute of Allergy and Infectious Diseases
United States National Institutes of Health

Sponsor	National Institute of Allergy and Infectious Diseases, National Institutes of Health, United States of America	
Principal Investigator (NIAID):	Clifton E. Barry III, Ph.D. National Institute for Allergy and Infectious Diseases (NIAID), Laboratory of Clinical Immunology and Microbiology (LCIM), Tuberculosis Research Section (TRS), Building 33, Room 2w10, Bethesda, MD 20892 Phone: (301) 693-4665 Email: cbarry@niaid.nih.gov	
Local Principal and Associate Investigators:	See Key Study Personnel document.	
Protocol Number:	16-I-N133	
Version Description:	13.0	
Version Date:	September 21, 2020	
Protocol Title:	Using Biomarkers to Predict TB Treatment Duration	
Abbreviated Title:	Predict TB Trial	
IRB/Institute Name:	NIH/NIAID	
Proposed Dates:	Start: June 2017	End: December 2022
Accrual Ceiling/Enrollment Goal	Ceiling: 1200 Enrollment Goal: 310 into Arms B and C; randomization into Arm C halted on September 14, 2020, per DSMB recommendation	
Ionizing Radiation Use:	Research Purposes	
Is Tissue Being Collected for Research Purposes?	Yes	
Location of the Study	Zhengzhou, China (multiple sites) Cape Town, South Africa (multiple sites)	
Investigational New Drug/Device:	None	
Research Contact/ Lead Associate Investigator:	Laura Via, PhD NIAID, LCIM, TRS Building 33, Room 2W20G (301) 451-9554 (p) (301) 480-5713 (f) lvia@niaid.nih.gov	
Medical Advisory Investigator:	Ray Chen, MD NIAID, LCIM, TRS Building 33, Room 2w10 (301) 443-5816 (office) (301) 451-5492 (f) RChen@niaid.nih.gov	

Enrollment sites:

Site Name	Country	City
Henan Provincial Chest Hospital (HPCH)	China	Zhengzhou, Henan Province
Kaifeng City Institute of Tuberculosis Prevention and Control (KCITPC)	China	Kaifeng, Henan Province
Xinmi City Center for Disease Control and Prevention	China	Xinmi, Henan Province
Zhongmu County Station for Disease Control and Prevention	China	Zhongmu, Henan Province
Xinxiang Institute for the prevention and control of tuberculosis (XXCITPC)	China	Xinxiang, Henan Province
University of Cape Town (UCT) Lung Institute	South Africa	Cape Town
Stellenbosch University	South Africa	Tygerberg
TASK Applied Science, Inc	South Africa	Bellville/Delft
Khayelitsha Site B	South Africa	Khayelitsha
South African Tuberculosis Vaccine Initiative (SATVI)	South Africa	Cape Town/Worcester

TABLE OF CONTENTS

Contents

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Text
AE	adverse event
AFB	acid-fast bacilli
AI	Associate Investigator

Abbreviation	Text
ALT (SGPT)	alanine transaminase
AST (SGOT)	aspartate transaminase
ATS	American Thoracic Society
β-hCG	human chorionic gonadotropin
BCG	Bacillus Calmette-Guérin
BID	Twice daily
BMRC	British Medical Research Council
CDC	Centers for Disease Control and Prevention
CFU	colony forming unit
CI	confidence interval
CNS	central nervous system
CRF	case report form
CRP	C-reactive protein
CT	Computed Tomography
CV	curricula vitae
CXR	Chest Radiograph
DHHS	Department of Health and Human Services
DIR	Division of Intramural Research
DS-TB	Fully Drug Susceptible Tuberculosis
DSMB	Data and Safety Monitoring Board
DST	drug susceptibility testing
EMB	Ethambutol
EBA	Early Bactericidal Activity
ELISA	enzyme-linked immunosorbent assay
ESAT-6	Early Secretory Antigenic Target 6
FDA	Food and Drug Administration
FDG	[^{18}F]-fluoro-2-deoxy-D-glucose
FQ	Fluoroquinolone
GBCA	Gadolinium-Based Contrast Agent
GCP	Good Clinical Practice
HIV	Human Immunodeficiency Virus
HR	isoniazid, rifampicin
HRCT	high-resolution computed tomography
HRZE	isoniazid, rifampicin, pyrazinamide, and ethambutol
HU	Hounsfield Units
IEC	Institutional Ethics Committee
IFN-ɣ	interferon gamma
IGRA	interferon gamma release assay
INH or H	isoniazid
IRB	institutional review board
IU	International Units
IV	Intravenous drug administration
LCIM	Laboratory of Clinical Immunology and Microbiology
LTBI	Latent Tuberculosis Infection
M or Mth (s)	Month(s)
MDR-TB	multi-drug resistant tuberculosis
MERM	medication event reminder monitor
MGIT	Mycobacteria Growth Indicator Tube

Abbreviation	Text
MIC	Minimum inhibitory concentration
MOHW	Ministry of Health and Welfare
MRC	Medical Research Council, UK
MRI	Magnetic resonance imaging
MTA	Material Transfer Agreement
M.tb	*Mycobacterium tuberculosis*
MXF	Moxifloxacin
NHP	Non-human primate
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institutes of Health
NIMR	National Institute for Medical Research, UK
NLME	Non-linear mixed effects
NMC	National Medical Center
NNT	Number needed to treat
OFLOTUB	A multicenter, randomized, control trial of ofloxacin-containing, short-course regimen for the treatment of pulmonary TB
OHRP	Office for Human Research Protections
OHSR	Office of Human Subjects Research
PA	Posterio-anterior
PBMC	Peripheral blood mononuclear cells
PET	Positron Emission Tomography
PI	Principal Investigator
po	Oral drug administration
PPD	Purified Protein Derivative
PZA or Z	Pyrazinamide
QD	Once daily
QFT	QuantiFERON
RCA	Research Collaboration Agreement
RCHSPB	Regulatory Compliance and Human Subjects Protection Branch
RIF or R	Rifampin
RSA	Republic of South Africa
SAE	Serious adverse event
SAHPRA	South African Health Products Regulatory Authority
SMS	Short Message Service
SOC	Standard of Care
SPSS	Statistical Package for Social Sciences
TB	Tuberculosis
TGA	Total Glycolytic Activity
TRS	Tuberculosis Research Section
TTD	Time to detection
TTP	Time to positivity
TST	Tuberculin Skin Test
ULN	Upper limit of normal
US NRC	United States Nuclear Regulatory Commission
W or Wk (s)	Week(s)
WBC	White blood cell
WHO	World Health Organization

STUDY OVERVIEW

Title	Using Biomarkers to Predict TB Treatment Duration
Précis	Shortening the duration of treatment for patients with drug sensitive tuberculosis from 6 to 4 months has been attempted many times in clinical trials but thus far all have failed. These failures reveal our incomplete understanding of factors driving the need for such extensive treatments. Consistently, trials have demonstrated that 80-85% of patients are successfully cured after 4 months of therapy, including the extensive set of studies from the British Medical Research Council (BMRC) in the 1970s and 1980 [1-3], the Tuberculosis Research Unit (TBRU) treatment shortening study in non-cavitary patients who achieve early culture conversion [4], and the more recent treatment shortening trials using fluoroquinolones like REMoxTB [5-7]. The current standard of care is to over-treat all patients for a total of 6-months to avoid relapse in a small subset of patients at higher risk for incompletely understood reasons. For decades, clinical investigators have attempted to establish culture conversion as a predictor of treatment success. Despite the appealing logic, the real correlation of culture conversion as a surrogate endpoint has been consistently disappointing. In the REMoxTB trial, in particular, the intensive microbiological data collected revealed unambiguously that clearance of bacteria from the sputum did not sufficiently correlate with relapse risk to be a useful surrogate for durable cure [8]. An important subset of patients, despite clearing their sputum of TB quickly and complying with all of their medications, still remained at high risk of relapsing with active disease after stopping treatment. Likewise there are patients who clear their sputum of bacteria slowly that nontheless go on to achieve durable cure. Intuitively this makes sense: only those bacteria at the surface of a cavity are directly open to the airways to seed the sputum. Yet this is not the full story as there are also heterogeneous lesions within each individual patient which respond differently to treatment with chemotherapy. This protocol builds upon the historical trials and several successful small studies that suggest that directly monitoring lung pathology using [^{18}F]-FDG PET/CT correlates better with treatment outcome than culture status [9, 10]. We will prospectively identify patients at low risk based on their baseline radiographic extent of disease, and further refine this risk score by evaluating the rate of resolution of the lung pathology (CT) and inflammation (PET) at one month as well as checking an end-of-treatment GeneXpert test for the sustained presence of bacteria. Patients classified as

	low risk will be randomized to receive a shortened 4-month or a full 6-month course of therapy. The NIAID Data Safety Monitoring Board conducted its 7th interim review of unblinded data on September 11, 2020. Following this review, the DSMB recommended halting randomization to Arms B and C. This recommendation is based on the result of the interim analysis in section 6.7.1 of the study protocol, when 1/3 of randomized participants have been followed for 72 weeks from study entry. The protocol stopping guideline for inferiority of the treatment shortening arm was met. The DSMB was also presented with conditional power calculations relating to the futility interim analysis; although the protocol-specified time for this analysis had not been reached, the results of the analysis were consistent with a decision to terminate randomization into Arms B and C, but to continue enrollment into Arms A and B at the discretion of the investigators. Pre-emptive re-treatment of participants who have completed the course of treatment in Arm C is not recommended and should be left to the discretion of the treating clinician. If successful, this trial will both offer a badly needed alternative to culture status as a trial-level surrogate marker for outcome as well as provide critical information for preclinical and early clinical efforts to identify new agents and combinations with the potential to shorten therapy. Hypothesis: A combination of radiographic characteristics at baseline, the rate of change of these features at one month, and markers of residual bacterial load at the end of treatment will identify patients with tuberculosis who are cured with 4 months (16 weeks) of standard treatment.
Objectives	Primary Objective To demonstrate that the 18-month treatment success rate of standard treatment stopped early at week 16 (Arm C) is not inferior to standard treatment stopped at week 24 (Arm B), in participants classified as low risk for disease failure and relapse by radiographic and bacterial load markers. Note that per the 7th NIAID DSMB review on September 11, 2020, the protocol stopping guideline for inferiority in treatment-shortening arm (Arm C) was met, and randomization to Arm C was discontinued. Secondary Objectives 1) To compare the treatment success rates between a representative 6-month standard of care population vs. the strategy of shortening treatment in low-risk participants. 2) To evaluate the association of demographic, radiographic, bacterial load, microbiologic, and immunologic markers (at

	baseline and during treatment) for predicting treatment failure in the following participant cohorts: • Pooling arms A and B, participants receiving the same duration of therapy, to evaluate the risk criteria. • Between arms B and C, to evaluate whether there are any covariates which predict greater rates of failure under treatment shortening. 3) To evaluate the association of demographic, radiographic, bacterial load, microbiologic, and immunologic markers (at baseline and during treatment) for predicting subsequent relapse in the following participant cohorts: • Pooling arms A and B, participants receiving the same duration of therapy, to evaluate the risk criteria. • Between arms B and C, to evaluate whether there are any covariates which predict greater rates of failure under treatment shortening. 4) To evaluate the ability of PET scans to predict treatment outcomes as described in secondary objectives 2) and 3). 5) To evaluate whether bacterial load markers (TTP and Xpert cycle threshold) collected at later time points are better markers of ultimate treatment outcomes than ones collected earlier 6) To evaluate sub-breakpoint MICs as a significant predictor of treatment outcome in 16 and 24 week treatment regimens 7) To evaluate the ratio AUC/sub-breakpoint MIC as a significant predictor of treatment outcome in (a subset of participants selected from the) 16 and 24 week treatment regimens; to compare this with #6 8) To evaluate the Cmax/sub-breakpoint MIC as a significant predictor of treatment outcome in (a subset of participants selected from the) 16 and 24 week treatment regimens; to compare this with #6 9) To compare MRI with PET/CT to quantitate extent of disease and monitor changes over time in a subset of participants. 10) To optimize 3T lung MRI acquisition and develop quantitation tools for pulmonary TB patients in a subset of participants. 11) To explore provider and participant satisfaction with the medication event reminder-monitor (MERM) and compare different adherence monitoring methods. 12) To explore whether LAM (lipoarabinomannan) measured by an investigational immunoassay that quantitates LAM concentration in sputum can be used as a bacterial load biomarker. Correlation between LAM concentrations and culture results will be examined. 13) To store biological samples for future analysis of potential biomarkers of treatment efficacy.

	14) To investigate the effects of SARS CoV-2 and TB co-infection on the difference in the treatment success rates between the 16- and 24-week treatment regimens.
Design	This is a prospective, randomized, phase 2b noninferiority trial in pulmonary DS-TB participants. Eligible participants will initially receive HRZE for 8 weeks, then switch to HR. Early treatment completion criteria will be evaluated for each participant up through week 16. Those who do not meet the early treatment completion criteria will be put into Arm A (2HRZE/4HR). Those who meet early treatment completion criteria will be randomized at week 16 either to continue therapy to week 24 (Arm B) or to complete therapy early at week 16 (Arm C). Per DSMB recommendations, randomization to Arm C was halted on September 14, 2020. All participants will be followed until approximately 18 months from the start of the study, with the primary endpoint evaluated at 18 months.
Population	Participants will be recruited from clinics in and around Cape Town, South Africa and Henan Province, China. Approximately 310 participants will be enrolled into Arms B&C. Per DSMB recommendations, randomization to Arm C was halted on September 14, 2020.
Participant Duration	Approximately 18 months
Estimated Time to Complete Enrollment	Approximately 3 years

1 Introduction: Background and Rationale

1.1 Biomarkers for tuberculosis

Multiple potential tuberculosis biomarkers, surrogate biological markers of clinically meaningful outcomes, are under active study [11]. From a microbiological standpoint, sputum culture conversion at 2 months of treatment is the biomarker most commonly used to predict non-relapsing cure [12] but its true predictive ability is not great, with one meta-analysis showing a pooled sensitivity and specificity for predicting relapse of 40% (95% CI 25%-56%) and 85% (95% CI 77%-91%), respectively [13]. A review of data from the British Medical Research Council (BMRC) trials from the 1970s and 1980s found only a weak correlation (R^2=0.36) for this marker as a surrogate for treatment failure and relapse, depending on factors such as geographic location, baseline disease and cavity status, and concomitantly used medications [14]. Further evidence against using culture conversion as a surrogate for predicting treatment outcome was recently demonstrated from the phase 3 TB treatment shortening trial REMoxTB [5], where subsequent analyses of the culture data collected demonstrated poor correlation with treatment outcome whether analyzed at a single time point (2 months) or over time (time to culture conversion or time to culture positivity) [8]. Furthermore, a small proportion of participants had poor outcomes despite clearing bacilli from their sputa quickly. Clearly, other factors independent of sputum culture conversion affect treatment outcomes and 2-

month culture conversion is no longer an appropriate surrogate marker of treatment outcome for clinical trials.

A number of additional factors have been associated with treatment outcome. In a study of 1004 pulmonary DS-TB participants from the Tuberculosis Trials Consortium (TBTC), a positive sputum culture at 2 months (adjusted HR 2.8, 95% CI 1.7-4.7) and a cavitation on initial chest x-ray (adjusted HR 3.0, 95% CI 1.5-5.8) were both independent predictors of treatment relapse [15]. In addition, the sterilizing efficacy of the overall treatment regimen is also dependent on the regimen used after 2 months, as demonstrated by significantly differing relapse rates in trials with arms using the same 2-month intensive phase regimen but differing continuation phase regimens [16] and trials using differing durations of continuation phase treatment [4].

The ability of early radiographic changes to predict subsequent treatment outcomes in TB has been recognized for over 50 years [17]. More recent analyses of radiographic biomarkers have moved beyond chest x-ray to using 2-deoxy-2-[18F]-fluoro-d-glucose (FDG) positron emission tomography/computed tomography (PET/CT) as an early marker of treatment response and possibly as a marker for relapse at the end of treatment. In macaques, changes on PET/CT correlate with TB disease activity and treatment response [18, 19]. Our group has analyzed human PET/CT data from a randomized clinical trial using metronidazole in the treatment of pulmonary multi-drug resistant tuberculosis (MDR-TB) participants [20]. As a substudy within the overall MDR-TB study, we performed PET/CT scans at 0 and 2 months and high resolution CT scans at 0, 2, and 6 months of treatment and correlated these changes with final treatment outcomes 30 months after treatment start (6 months after the end of therapy, Table 1).

Table 1: Sensitivity and specificity of 2-month sputum culture conversion compared to CT and PET scan changes for predicting treatment outcomes in MDR-TB in humans.

Modality	Sensitivity	Specificity
PET (2 months)	0.96 (23/24)*	0.75 (3/4)*
Automated CT (6 months): HU −100 to 200	0.96 (23/24)*	0.75 (3/4)*
Automated CT (2 months): HU −100 to 200	0.79 (19/24)*	0.75 (3/4)*
Culture—solid (2 months)	0.79 (19/24)	0.5 (2/4)
Smear (2 months)	0.75 (18/24)	0.5 (2/4)
Culture—liquid (2 months)	0.58 (14/24)	0.5 (2/4)

*Estimates have been corrected for bias in selection of optimal threshold using cross-validation.

PET changes at 2 months and CT changes at 6 months appeared to be more sensitive to predict final treatment outcomes than sputum culture conversion at 2 months, although these differences were not statistically significant. These results support the potential of PET/CT imaging biomarkers as possible surrogate endpoints in clinical trials and larger cohorts are needed to confirm these results [10].

In addition to the above study, which was of MDR-TB participants, 100 DS-TB participants from Cape Town received PET/CT scans at baseline, 1 month, and 6 months while receiving standard therapy through the national program. Adherence was directly observed for the initial 2 weeks, then monitored with monthly pill counts. The participants were then followed through 18 months for final treatment outcomes. In total, 92 participants had complete PET/CT scan data, Xpert MTB/RIF cycle thresholds (see Section 1.2), and treatment outcomes. Table 2 describes the sensitivity and specificity of the early treatment completion criteria to be used in this trial (see Section 3.1, Table 6) for predicting treatment outcome:

Table 2: Sensitivity and Specificity of Early Treatment Completion Criteria for Predicting Treatment Failure or Programmatic Treatment Restart

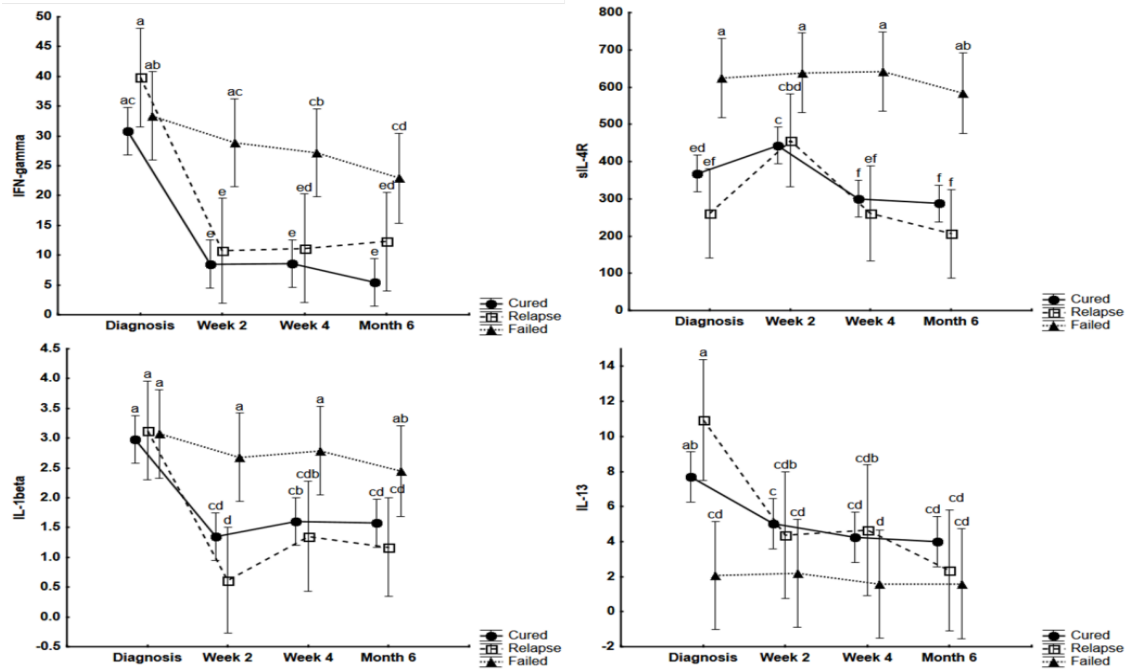
	Treatment Outcome (Numbers of Participants)			
Risk Categorization	Cure	Failure	Programmatic Treatment Restart	Total
Low Risk	55	2	8	65
High Risk	18	6	3	27
Total	73	8	11	92

Of these 92 participants, 73 were considered cured, 8 failed treatment, and 11 were restarted programmatically on TB treatment, defined by the participant restarting treatment during follow-up for any reason. Confirmatory culture data were only available for 1/11 participants and true relapse was not differentiated from re-infection. A previous analysis from Cape Town retrospectively analyzed TB cases from 1996-2008 and identified 130 recurrent cases with paired DNA fingerprinting results available. Among these, 64 (49%) were relapses and 66 (51%) were reinfections, with reinfection accounting for 9/44 (20%) cases within the first year [21]. Consequently, with the vast majority of the programmatic treatment restart participants in our analysis not microbiologically confirmed and reinfection likely playing a significant but undefined role, this category of participants was not weighed heavily in the development of the early treatment completion criteria.

Given these limitations, the Predict Trial early treatment completion criteria, which incorporates baseline and one-month PET/CT results and a week 16 Xpert cycle threshold, correctly classified 75.3% (55/73) of the cured participants as low risk and 6/8 failures (75.0%) as high risk (week 16 Xpert cycle threshold interpolated for this analysis). With adherence not closely monitored, the reasons for failure were not clear. In addition, the one failure participant classified as low risk had mixed microbiological results so the final determination was not clear-cut. Overall, 65/92 (70.7%) of all patients were classified as low risk.

Host immunological biomarkers that indicate an early response to treatment and reflect restoration of the balance of pro-and anti-inflammatory markers would be useful for

Phase 2 trials and may also predict response to shortened therapy. In a separate study, we used the Luminex technology to evaluate the levels of 72 host inflammatory markers in serum samples from 83 subsequently cured patients, 15 patients where treatment failed, and 12 patients with bacterial strain confirmed relapse. The levels of 36 of the 72 markers changed significantly during treatment in successfully cured patients with different response patterns, with most changes observed within 2 weeks of onset of therapy. The marker signature obtained in the cured patients was distinctly different from that obtained in patients with treatment failure and relapse outcomes.



IFN-γ, IL-1β, sIL-4R and IL-13 levels at diagnosis, weeks 2, 4, and month 6 of treatment are shown. Data (pg/ml) were analysed by mixed model repeated measures ANOVA. Cured patients are indicated by solid lines and solid circles, failed patients by dotted lines and solid triangles and relapse patients by dashed lines and open squares. Values with different letters indicate that they are significantly different from each other (p≤0.05). Mean (log transformed) and 95% Confidence intervals are shown.

Figure 1: Representative plots for the different patterns observed in serum cytokine levels in cured, failed and relapse patients during treatment.

The differences observed in cytokine kinetics during treatment of non-responsive patients may reflect ongoing bacterial replication and inflammation, whereas differences in pre-treatment levels of some markers, e.g. sIL-4R, IL-13, IL-5, MMP-2 and MMP-9, might reflect an underlying increased risk of treatment failure, possibly due to increased extent of disease or poor host response. Prediction models comprised of clinical (body mass index, BMI), microbiological (time to detection, TTD) and inflammatory markers measured at the time of diagnosis and during early treatment classified the three different study groups with accuracy up to 95% (95% CI 77%-97%). As these data are based on relatively small participant numbers, we therefore propose to collect additional host samples, including serum, for refinement of predictive host marker signatures and to evaluate them as potential surrogate markers for the PET/CT measurements.

1.2 Surrogate markers of sputum bacterial load

There is currently no direct measure of TB sputum bacterial load available. Some markers, however, may serve as a surrogate. The time to positivity (TTP) or TTD of a positive culture on liquid mycobacterial culture systems has been correlated with various outcome measures, with the shorter TTP indicating a higher bacterial load. In one analysis among 263 smear positive pulmonary DS-TB patients, a very short TTP at baseline (≤3 days) was significantly correlated with delayed 2-month sputum culture conversion and relapse at 24 months [22]. In another analysis of 146 patients including only 5 failures, the mean baseline TTD among the failures was shorter than the cures (16.3 vs 25.4 days, p=0.003). However, 2-month culture conversion, although 100% sensitive for predicting cure, was only 49% specific, meaning that many cured patients were still positive on culture at 2 months. Changing from culture negative to TTD >21 days at 2 months (allowing a low bacterial load) did not change the sensitivity (100%) but improved the specificity to 82% [23]. So although there is some correlation between Mycobacteria Growth Indicator Tube (MGIT) TTP and sputum bacterial load, the poor specificity of TTP with subsequent treatment outcomes is evident in unpublished data from the REMoxTB trial [5].

Table 3: TTP on week 8 MGIT stratified by per protocol treatment outcome status

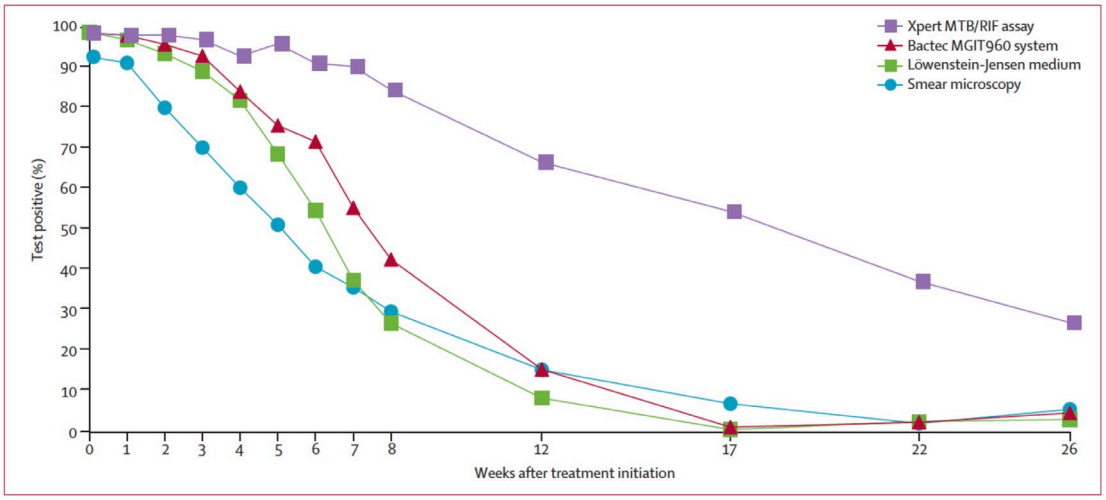
	Week 8 Time to Positivity (days)				
Treatment outcome (per protocol)	**0-4**	**5-14**	**15-24**	**≥25**	**Total**
Favorable	7	45	72	30	154
Unfavorable	1	5	10	2	18
Excluded	2	5	12	1	20
Total	10	55	94	33	192

As table 3 shows, there is no correlation between MGIT TTP at week 8 and treatment outcome (Wilcoxon rank sum test p=0.23), with most participants having favorable outcomes no matter their TTP. Thus, although time to culture conversion and MGIT TTP do correlate independently with treatment outcome, these markers do not discriminate well between high and low risk patients and therefore have only a limited role in predicting treatment outcomes of individual patients [8].

Another marker of sputum bacterial load is the Xpert assay, which is an automated rapid molecular diagnostic test for *M. tb* and resistance to rifampin with results provided directly from sputum within 2 hours [24]. The test is run using a polymerase chain reaction and the number of cycles (cycle threshold) required to obtain a positive result is recorded, with a lower cycle threshold suggestive of a higher bacterial load. Three studies have correlated Xpert MTB/RIF cycle threshold with sputum smear status, with varying sensitivity and specificity levels depending on the cycle threshold cut point used [25-28]. Another analysis by Friedrich, et al correlated Xpert MTB/RIF results with sputum smear and culture results, not only at baseline but also longitudinally over time as a sputum biomarker of treatment response [29]. The reduction in detection of quantitative Mtb DNA with Xpert MTB/RIF correlated with smear grades, solid culture grades, and liquid

culture TTP (all p<0.0001) but Xpert MTB/RIF sputum positivity rates decline more slowly during treatment than sputum smear and culture results.

Figure 2: Qualitative data for noted tests at baseline and follow-up



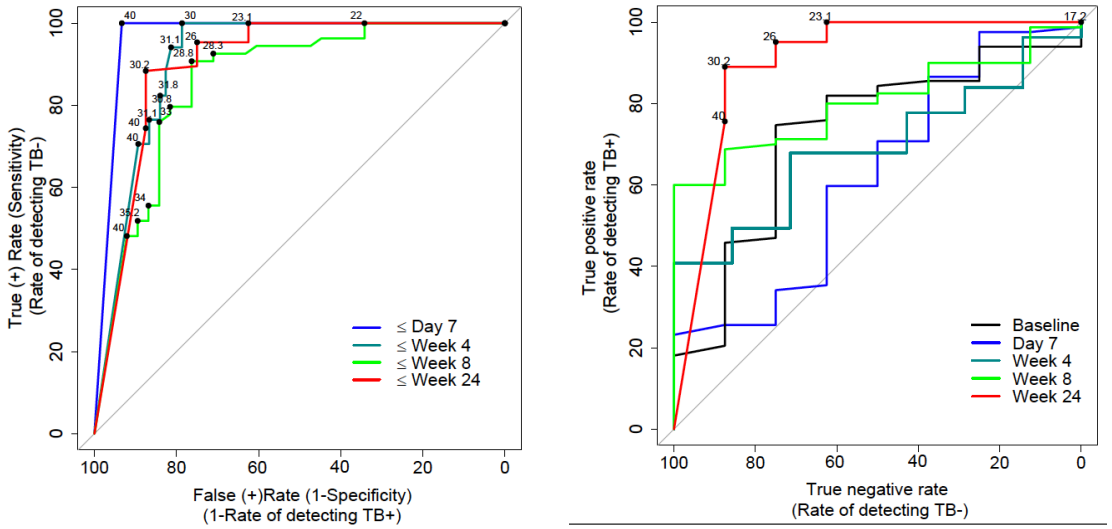
Using the combined binary smear and culture results as a reference standard, sensitivity of Xpert MTB/RIF was excellent at 97.0% (95% CI 95.8-97.9) but specificity was poor at 48.6% (95% CI 45.0-52.2) as the assay is unable to differentiate viable, dormant, and non-viable MTB bacteria.

In the study mentioned previously of 100 pulmonary DS-TB participants in Cape Town, sputa collected during 6 months of treatment were cultured on liquid medium (MGIT) and tested by Xpert MTB/RIF with an Xpert MTB/RIF cycle threshold recorded. Similar to the Friedrich study [30], serial testing with Xpert MTB/RIF negativity (defined by Xpert MTB/RIF cycle threshold ≥30) correlated with MGIT culture negative status at the same time points (Figure 3a). All AUCs were significantly greater than 0.5, indicating the ability to differentiate TB from no TB, with no significant differences in AUCs between the different time points when adjusted for multiple comparisons. When Xpert MTB/RIF cycle threshold was used to predict treatment failure status at the end of treatment, week 24 cycle threshold produced the largest AUC (Figure 3b, p=0.086, when compared to baseline cycle threshold). [31] A week 24 Xpert MTB/RIF cycle threshold of ≥30 predicted failure with higher sensitivity and specificity than earlier time points. This is in contrast to the week 8 culture (Table 4), which has lower sensitivity (for cure) than Xpert MTB/RIF cycle threshold at week 24: 61% (week 8 culture) vs 89% (Xpert MTB/RIF cycle threshold week 24, p<0.01). Estimates of specificity (for failure) for Xpert MTB/RIF cycle threshold week 24 was higher than week 8 culture (88% vs 50%), but the

improvement was not statistically significant. Note that the 13 "unevaluable" cultures make direct comparison problematic. The poor performance of week 8 culture data as a predictor of treatment outcome is similar to what was observed in the REMoxTB trial [8].

The observation that a later test predicts outcomes better than an earlier test may be similar to findings in HIV infection, where baseline CD4 cell count is a strong predictor of mortality over time but current CD4 cell count is even stronger [32]. This has been seen in TB too, where culture conversion at month 6 predicts final treatment outcome significantly better than culture conversion at month 2 [33]. Taken together, these results suggest that Xpert MTB/RIF cycle thresholds collected later may be able to replace an earlier microbiological culture in predicting treatment outcomes, with the major advantage of Xpert MTB/RIF over culture being the time to test result, with Xpert requiring 2 hours and culture up to 6 weeks. Thus, Xpert cycle threshold may be able to be used as a point-of-care test whereas culture cannot.

Figure 3: Receiver-operating characteristic (ROC) curves of Xpert MTB/RIF cycle threshold with MGIT liquid culture negativity (3a) and patient treatment failure (3b).



In this data set, the sensitivity and specificity of Xpert MTB/RIF cycle threshold ≥30 at week 24 (sensitivity 89%, specificity 88%) is better than MGIT culture negative at week 8 (sensitivity 61%, specificity 50%) to predict cure (Table 4).

Table 4: Xpert MTB/RIF vs. MGIT: Sensitivity and specificity for predicting outcome

	Outcome*	
	Cure	**Failure**
Week 24 Xpert ≥30	89% (68/76)	13% (1/8)
Week 24 Xpert <30	11% (8/76)	88% (7/8)
Week 8 MGIT TB culture negative	61% (46/75)	25% (2/8)
Week 8 MGIT TB culture positive	24% (18/75)	50% (4/8)

Week 8 MGIT unevaluable**	15% (11/75)	25% (2/8)

* 12 participants who programmatically restarted treatment not included due to low confidence in their true treatment outcome

** One participant was missing week 8 MGIT

The above data compared Xpert results with MGIT liquid culture data. The Predict TB study uses LJ solid culture as our primary endpoint. The TBTC-29 study collected Ct values and LJ culture results. In evaluating the change from Ct ≥30 used above, we consider the chance of missing an LJ+ result, as well as the sensitivity and specificity of various Ct cutpoints. In contrast to positive and negative predictive values, sensitivity and specificity do not depend on the underlying proportion of culture positive results, which varies over time and from study to study. That said, patient safety is a driving factor so we consider how many positive cultures might be missed for various cutpoints. This was defined as the probability of being LJ+ given a Xpert Ct value less than the threshold, i.e., P(LJ+ | Ct-). We assume what we consider are high proportions of LJ+ cultures (i.e., 10% and 5% at week 16 of treatment in a lower risk cohort of arm B/C) when making this decision. In contrast to the TBTC-29 study, which randomized all-comers and did not stratify participants by risk, the Predict TB study further excludes poorly adherent participants with too severe disease at baseline and not responding appropriately to treatment at one month. As a result, the expected LJ+ rates of 10% and 5% are thought to be on the high-side. Appendix table 2 (section 13) describes these proportions for the sensitivity and specificity estimates from TBTC-29. Based on these estimates, a Ct of 30 is expected to miss 2.1% of LJ+ results, while a threshold of 28 would miss 2.5%, assuming a 10% LJ+ rate. This translates to an increase in less than one participant being missed amongst those randomized to arm C. That is, if the background LJ+ rate is 10%, 3.3 (of 155 randomized to stop treatment at week 16) true LJ+ participants may be missed with Ct=30, and 3.9 may be missed with Ct=28. If the underlying LJ+ rate is 5%, this becomes 1.6 missed LJ+ participants with Ct=30 and 1.9 missed LJ+ participants with Ct=28. If the true underlying LJ+ rate is even lower, the difference between the two Ct values becomes even smaller. Based on this analysis, we decided to use Ct threshold ≥ 28.

1.3 Treatment Shortening

Shortening the treatment duration of drug sensitive (DS)-TB and multidrug-resistant (MDR)-TB is a major goal of TB research and several clinical trials have evaluated shortened treatment regimens. For DS-TB, based on promising murine data [34, 35] and preliminary results in human clinical trials [36-40], 3 recent trials tried to shorten treatment from the standard 6 months (SOC) to 4 months by substituting a fluoroquinolone (FQ; moxifloxacin or gatifloxacin) for either isoniazid (INH) or ethambutol in varying combinations [5-7]. These results are summarized in Table 5.

Table 5: Summary of Recent Trials Aimed at Shortening TB Therapy

	Months 1-2	Months 3-4	Months 5-6	Unfavorable Outcome per protocol (adjusted difference from control)
RIFAQUIN [6]				
Control	HRZE	HR	HR	4.9% (18 mo post randomization)
Arm 1	MRZE	MP (1x/wk)	MP (1x/wk)	3.2% (-1.8%, 95% CI -6.9 to 3.3)

Arm 2	MRZE	MP (2x/wk)	--	18.2% (13.6%, 95% CI 7.0-20.2)
OFLOTUB [7]				
Control	HRZE	HR	HR	11.3% (24 mo post randomization)
Intervention	HRZG	HRG	--	17.7% (5.5%, 95% CI 1.6-9.4)
REMoxTB [5]				
Control	HRZE	HR	HR	8% (18 mo post randomization)
INH arm	HRZM	HRM	--	15% (6.1%, 95% CI 1.7-10.5)
ETH arm	MRZE	MR	--	20% (11.4%, 95% CI 6.7-16.1)

In secondary analyses, OFLOTUB patients with cavities on baseline chest x-ray had significantly worse outcomes in the 4-month arm compared to the 6-month arm (P=0.04). Patients in REMoxTB with cavities on baseline chest x-ray also did worse in the two 4-month arms but this difference did not reach statistical significance (P=0.058 for INH arm, P=0.118 for ethambutol arm).

These three FQ treatment-shortening trials suggest that 4 months of treatment will cure around 80%-85% of sputum smear positive patients. Historically, 4-month treatment regimens similar to the SOC used today had relapse rates ranging from 10%-16% [1, 2]. One review found a relapse rate for 4-month regimens of 12% (95% CI 9-16) at 2 years of follow-up [3]. All of these 4-month regimen outcomes, which span multiple countries over 30 years, have been remarkably consistent. Therefore, it is likely that when considering all TB patients as a whole, the current SOC regimen will cure >95% of patients with 6 months of treatment but only 80%-85% of patients with 4 months of treatment. Identifying those cured at 4 months, then, may be the key to successful treatment shortening.

A 4th treatment shortening study in DS-TB attempted exactly this strategy. Recognizing that the presence of a pulmonary cavity at baseline and the lack of sputum culture conversion at month 2 were risk factors for relapse disease, this study shortened treatment to 4 months only for those who were without pulmonary cavities at baseline and were sputum culture negative by month 2. Although the DSMB stopped this trial early due to a higher relapse rate in the treatment-shortened arm compared to the 6-month SOC arm (7.0% vs 1.6%, P<0.01) [4], this study suggests that shortening treatment only among those with less severe disease resulted in a lower relapse rate compared to previous studies which did not employ this strategy (7% vs. about 15-20%). Refining this treatment-shortening algorithm beyond just lack of baseline cavity on chest x-ray and sputum culture conversion at month 2, such as additional baseline and follow-up radiological markers, may successfully identify those fully treated by month 4.

1.4 Goals of the Predict Trial

The immediate goal of this study is to determine if our criteria for patient-specific radiographic and bacterial load biomarkers successfully identify patients who are cured with 16 weeks of treatment as a proof of concept study. We understand that PET/CT scanners are not available in many resource-limited settings globally and thus we do not expect PET/CT scans to be used as a global public health tool. We expect that, if successful, this treatment-shortening concept can be applied to other clinical trials as a trial-level surrogate biomarker of treatment outcome. Our primary analysis and treatment shortening criteria are based on CT scans, which provide structural lung pathology data,

PET scans, which provide functional data about lung inflammation, and Xpert cycle threshold, a point-of-care measurement of remaining bacterial load. By excluding participants at baseline with a high disease burden (baseline PET/CT scan), participants at 4 weeks with an insufficient treatment response (week 4 PET/CT scan), and participants with residual bacterial burden at the end of treatment (week 16 Xpert cycle threshold), we hypothesize that we will identify a low risk cohort of participants cured with 16 weeks of chemotherapy. Sputum LAM concentration will be analyzed against sputum culture results. Additional secondary analyses will also be performed to determine if immunological parameters that are more readily available also correlate and serve as equivalent surrogate biomarkers, allowing these criteria to be applied more broadly in resource-limited settings.

1.5 Sub-breakpoint MIC

Minimum inhibitory concentrations (MIC) are used in standard drug susceptibility testing assays to determine at what drug concentrations a given "resistant" bacteria will grow in culture. This is typically defined as the antibiotic concentration at which 1% of bacterial isolates grow, compared to cultures without any antibiotics [41-43]. Conventional MIC testing will determine growth at one or two drug concentration "breakpoints." The results of this testing are important because drug resistance is a well-known determinant of treatment outcome [44].

Our collaborator, David Alland, has shown that just as high-level MIC resistance correlates with treatment outcome, low-level or sub-breakpoint MIC resistance also correlates with treatment outcome [45]. In the TBTC Study 22 trial [15], 1004 DS-TB patients were randomized after 2 months of standard isoniazid (H), rifampin (R), pyrazinamide (Z), and ethambutol (E) (HRZE) to 4 months of continuation phase with either once weekly rifapentine/isoniazid or twice weekly rifampin/isoniazid. Using data from this trial, sub-breakpoint MICs were tested among 57 HIV negative relapsed cases and 64 cured participants and found highly reproducible results with triplicate testing. There was in addition a significant association between baseline elevated sub-breakpoint MIC to INH ($p=0.014$) or RIF ($p<0.001$) and relapse. In multivariable analysis, INH MIC, RIF MIC, baseline cavity, bilateral disease on CXR, and low body weight were independent predictors of relapse. In a ROC curve of the TBTC study 22 features independently associated with relapse, the AUC for the ROC curve of only INH and RIF MICs (0.785) was comparable to the ROC of the significantly associated clinical features of cavitary disease, bilateral disease, low body weight, and 2-month sputum culture conversion (AUC 0.776). When baseline INH and RIF MICs were added to the clinical features, the resulting AUC was 0.888 to predict relapse.

In the previously mentioned TBRU treatment-shortening study that found significantly higher relapse rate in the 4-month arm compared to the 6-month arm [4], sub-breakpoint RIF MIC testing was done on nine available *M.tb* isolates from 4-month arm patients who relapsed and nine 4-month arm patients who were cured. Similar to the TBTC study 22 results, sub-breakpoint RIF MICs were significantly higher in the relapse cases compared to the cured cases.

These data, although preliminary due to small numbers, show that baseline sub-breakpoint MIC data may be independently predictive of treatment outcome and/or treatment shortening and need to be further validated. As part of this study, we will batch sub-breakpoint MIC testing at the end of the study using saved samples. We will also collect blood for PK analyses, as described further in section 7.0.

1.6 Lung magnetic resonance imaging scans

The PredictTB trial stratifies drug-sensitive pulmonary TB participants into higher risk and lower risk cohorts based on baseline and week 4 PET/CT scans and a GeneXpert MTB/RIF cycle threshold at week 16, to test whether low-risk patients can successfully be treated within 4 months. More recent data suggest that magnetic resonance imaging (MRI) scans may be more effective than PET/CT scans to measure baseline disease burden and treatment response. Two studies including 113 patients [46, 47] and a single case report [48] compared MRI with CT in pulmonary TB patients. All three studies consistently found that MRI characterizes TB lesion features better than CT, with MRI able to detect heterogeneous patterns that appear homogeneous on CT (Figure 4). Another study evaluated the histopathology of pulmonary nodules that were subsequently resected and found varying MRI intensity patterns on T1 vs T2-weighted images that correlated with caseating vs noncaseating granulomas, solid vs. liquid caseation, and fibrocalcific vs not fibrocalcific granulomas [49]. Thus, in these small studies, MRI image features correlate better with histopathologic findings of TB lesions than CT scan features, allowing more precise determination of different TB lesion types. Necrotic lesions are likely associated with worse outcomes, so accurately measuring the extent of caseous necrosis may guide treatment duration. If MRI can identify these types of higher risk lung lesions more accurately than CT, MRI may perform better than PET/CT in a TB treatment shortening stratification algorithm such as in PredictTB. MRI biomarkers could also potentially be applied as a precision medicine tool to guide individualized treatment, monitor treatment response and evaluate novel interventions.

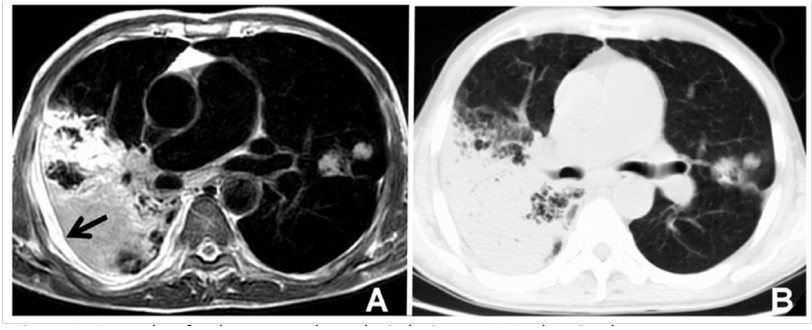


Figure 4: Example of pulmonary tuberculosis on MRI showing heterogenous texture (A) that appears homogeneous on CT (B). From Zeng J, et al. Int J Infect Dis 2019;82:138-146

This substudy administers baseline and week 4 lung MRI scans to a subset of 60 PredictTB participants who provide informed consent. Additional lesion information provided by the MRI scan will likely be most useful in those with more severe disease, with large, dense TB lesions that may be less well characterized by the PET/CT scan. Subsequently, participants with a range of disease severity will be enrolled but weighted towards the more heavily diseased. Only participants with a GeneXpert semiquantitative reading of medium and high will be included in this substudy. Quantitative analysis of MRI scans will be used to compare with the PET/CT scan data collected to determine if MRI biomarkers may provide additional baseline or treatment response information useful for risk stratification of participants. This would be an important advance because:

1) MRI scans do not involve radiation so are safer than PET/CT scans; 2) MRI scans are cheaper than PET/CT scans and therefore may become more widely available in resource limited settings than PET/CT scans. If the Predict TB trial is successful, blood, sputum, or urine-based biomarkers that correlate with the PET/CT signature will be sought but adding MRI scans provides an additional potential route to global applicability and scalability of the treatment shortening algorithm. For these reasons, a successful MRI-based algorithm may have a larger global impact than a PET/CT-based algorithm. In addition to performing MRI scans in a subset of PredictTB trial participants, this study will also develop MRI scanning protocols for better visualization and conduct image analysis to improve quantitation of different types of TB lung lesions. This is important as there are many different MRI scanning and analysis techniques available so identifying the best one for our purposes will improve our chances of success.

1.7 COVID-19 Infection

The coronavirus disease 2019 (COVID-19) is a respiratory viral infection caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and was declared a global pandemic by the WHO on March 11, 2020. COVID-19 can present with influenza-like symptoms, including fever, chills, cough, shortness of breath, myalgias, and headache. These symptoms may overlap with those of pulmonary tuberculosis. How co-infection with SARS-CoV-2 affects the clinical presentation or treatment outcome of pulmonary TB patients is not currently known. There are no published, peer-reviewed reports of SARS-CoV-2 and TB co-infected patients yet, although a pre-print manuscript case-control study of 36 SARS-CoV-2 positive cases from China suggests that latent TB infection may increase COVID-19 disease severity (https://www.medrxiv.org/content/10.1101/2020.03.10.20033795v1).

During the original SARS outbreak in 2002-2004, three case reports were published including 6 patients with SARS-TB coinfection, all of whom recovered [58-60]. Two additional cases of Middle East Respiratory Syndrome coronavirus (MERS-CoV)-TB coinfection have also been reported [61]. No conclusions can be drawn from such small numbers.

To understand better the effect of SARS-CoV-2 and TB co-infection, a new exploratory objective will be added to the study: to explore the effect of SARS-CoV-2 infection on pulmonary TB treatment outcomes and recurrence. Study procedures related to this objective will be implemented only at the clinical sites in South Africa because there have been no recent cases at the clinical sites in Henan, China. Additionally, this testing is contingent on enough SARS-CoV-2 testing capacity and personal protective equipment for study staff. Not conducting the procedure pertaining to SARS-CoV-2 testing will not be considered a protocol deviation. This will be an exploratory objective because we do not know how many enrolled participants may subsequently develop SARS-CoV-2 infection. Co-infected patients identified during screening will be excluded from the study due to the unpredictable course of the disease and the requirement for SARS-CoV-2 strict isolation and the risk of nosocomial transmission to study staff, in a time when there are frequent and essential study contact visits. Participants who become infected with COVID-19 after enrollment will be maintained in the study. As far as possible, face-to-face visits will be rescheduled or conducted telephonically. If quarantine is required

when a visit essential for randomization (W4 scan, W16 sputum collection) is scheduled, and strict transmission control requirements at the study site or PET Center cannot be maintained, the participant may be withdrawn.

2 Study Hypothesis and Objectives

2.1 Hypothesis

A combination of radiographic characteristics at baseline, the rate of change of these features at one month, and markers of residual bacterial load at the end of treatment will identify patients with tuberculosis who are cured with 4 months (16 weeks) of standard treatment.

2.2 Primary Objective

To demonstrate that the 18-month treatment success rate of standard treatment stopped early at week 16 (Arm C) is not inferior to standard treatment stopped at week 24 (Arm B) in patients classified as low risk for disease failure and relapse by radiographic and bacterial load markers. Note that per the 7th NIAID DSMB review on September 11, 2020, the protocol stopping guideline for inferiority in treatment-shortening arm (Arm C) was met, and randomization to Arm C was discontinued on September 14, 2020.

2.3 Secondary Objectives

1) To compare the treatment success rates between a representative 6-month standard of care population vs. the strategy of shortening treatment in low-risk participants.
2) To evaluate the association of demographic, radiographic, bacterial load, microbiologic, and immunologic markers (at baseline and during treatment) for predicting treatment failure in the following patient cohorts:
    a. Pooling arms A and B, participants receiving the same duration of therapy, to evaluate the risk criteria.
    b. Between arms B and C, to evaluate whether there are any covariates which predict greater rates of failure under treatment shortening.
3) To evaluate the association of demographic, radiographic, bacterial load, microbiologic, and immunologic markers (at baseline and during treatment) for predicting subsequent relapse in the following participant cohorts:
    a. Pooling arms A and B, participants receiving the same duration of therapy, to evaluate the risk criteria.
    b. Between arms B and C, to evaluate whether there are any covariates which predict greater rates of failure under treatment shortening.
4) To evaluate the ability of PET scans to predict treatment outcomes as described in secondary objectives 2) and 3).
5) To evaluate whether bacterial load markers (TTP and Xpert cycle threshold) collected at later time points are better markers of ultimate treatment outcomes than ones collected earlier
6) To evaluate sub-breakpoint MICs as a significant predictor of treatment outcome in 16 and 24 week treatment regimens.

7) To evaluate the ratio AUC/sub-breakpoint MIC as a significant predictor of treatment outcome in (a subset of participants selected from the) 16 and 24 week treatment regimens; to compare this with #5
8) To evaluate the Cmax/sub-breakpoint MIC as a significant predictor of treatment outcome in (a subset of participants selected from the) 16 and 24 week treatment regimens; to compare this with #6
9) To compare MRI with PET/CT to quantitate extent of disease and monitor changes over time in a subset of participants.
10) To optimize 3T lung MRI acquisition and develop quantitation tools for pulmonary TB patients in a subset of participants.
11) To explore provider and patient satisfaction with the medication event reminder-monitor (MERM) and compare different adherence monitoring methods.
12) To explore whether LAM (lipoarabinomannan) measured by an investigational immunoassay that quantitates LAM concentration in sputum can be used as a bacterial load biomarker. Correlation between LAM concentrations and culture results will be examined.
13) To store biological samples (sputum, saliva, blood, serum, urine) for future analysis of potential biomarkers of treatment efficacy.
14) To investigate the effects of SARS CoV-2 and TB co-infection on the difference in the treatment success rates between the 16- and 24-week treatment regimens.

3 Study Design and Population

3.1 Design and Arms

This is a prospective, randomized, phase 2b noninferiority trial in pulmonary DS-TB participants. Eligible participants who sign the informed consent will start on HRZE. At week 8, Z and E will be discontinued. Participants who meet randomization criteria will be randomized at week 16. Randomization will be stratified by site. Some participants will complete 16 weeks of TB therapy and others will complete 24 weeks of TB therapy, as determined using early treatment completion criteria (Table 6) and randomization. Note that per the 7th NIAID DSMB review on September 11, 2020, the protocol stopping guideline for inferiority in treatment-shortening arm (Arm C) was met, and randomization to Arm C was discontinued, and all participants will receive 24 weeks of TB therapy.

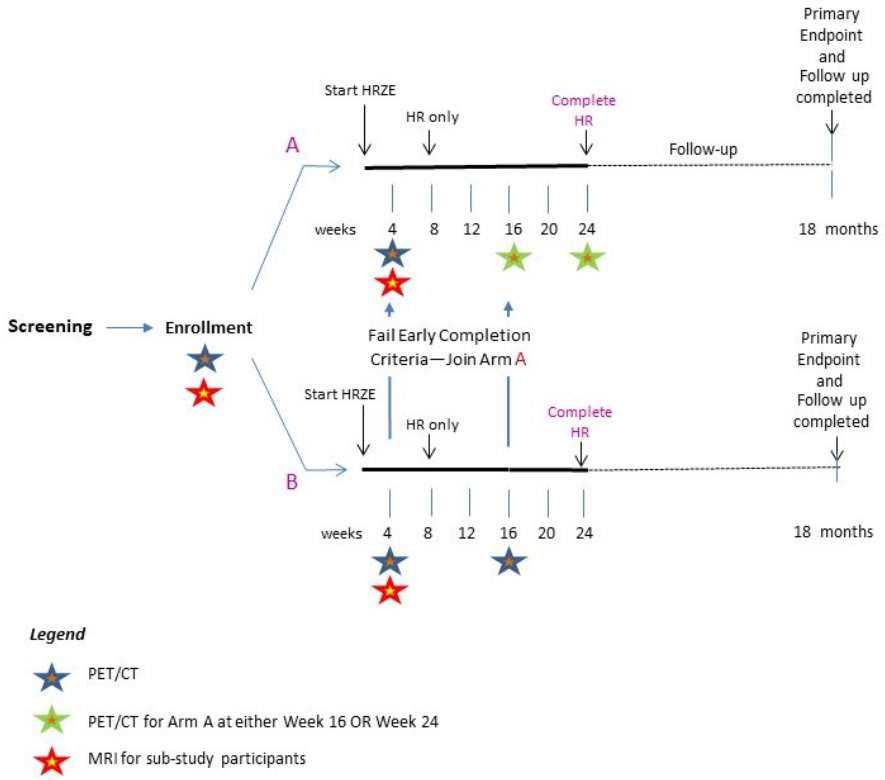


Figure 5: Study Flow

Early treatment completion criteria (Table 6) will be evaluated at baseline (radiology), week 4 (radiology), and Week 16 (bacterial load and treatment adherence) of the study. Per DSMB recommendations, randomization to Arm C was halted on September 14, 2020 and all participants will receive 24 weeks of TB therapy.

Participants will receive a baseline PET/CT scan and those who do not meet baseline PET/CT early completion criteria (Table 6) will be assigned to Arm A. At week 4, participants still eligible for early completion will receive a 2nd PET/CT scan, and all others will receive a scan at this time as well if resources allow. Those who do not meet the radiographic early stopping criteria (Table 6) at this point will be assigned to Arm A. Participants still eligible for early treatment completion at week 16 will be tested by Xpert and, if their cycle threshold is $\geq$28, will be randomized either to stay on treatment to week 24 (Arm B) or to complete treatment early at week 16 (Arm C). Per DSMB recommendations, participants will no longer be randomized to complete treatment early at week 16 (Arm C) as of September 14, 2020. Participants who do not meet the early treatment completion criteria will be assigned to Arm A. Xpert testing will also be conducted on Arm A participants at week 16. Participants on Arms B and C will receive a week 16 PET/CT scan. Participants on Arm A will be randomized to receive a week 16 or week 24 PET/CT scan, stratified by when they were placed in Arm A (baseline or week 4). All participants will be followed to month 18 for cure vs. failure (relapse, death). Participants who develop recurrent disease during follow-up will also receive a PET/CT scan and will be referred for retreatment per local SOC. Participant follow-up is to 18 months, as it has been shown that the vast majority of patients who will relapse do

so by one year after the end of therapy [50, 51]. Participants who develop recurrent TB will be referred to their local TB clinic for retreatment. These participants will be followed observationally until the end of retreatment to determine their retreatment outcome, even if this is longer than 18 months after the start of their participation in the trial.

Table 6: Early Treatment Completion Criteria: All early treatment criteria must be met for the participant to be eligible for randomization

Early Completion criteria:	**Determined at Week 16 – unless known to have failed a radiographic criterion at baseline or week 4.**
Radiographic criteria	Baseline PET/CT: • No total lung collapse of a single side, AND • No pleural effusion, AND • No single cavity air volume on CT scan >30 mL, AND • CT scan hard volume (≥-100 HU density) <200 mL OR PET total lesion glycolysis <1500 units Week 4 PET/CT: • All individual cavities decrease by >20% (unless cavity <2 mL), AND • CT scan hard volume does not increase by >10% unless the increase is <5 mL OR PET total lesion glycoysis does not increase by >30% unless the increase is <50 units
Bacterial load criterion	Week 16 Xpert cycle threshold ≥28*
Adherence criterion	Minimum of 100 doses received by week 16 (Sec 5.2)

*If the week 16 solid medium sputum culture is subsequently found to be positive for *Mtb* in a participant randomized to Arm B or C, this participant will be called in for evaluation and to provide sputum for a repeat culture. If the initial positive culture is confirmed by a second culture positive for *Mtb*, this participant will be considered to have met the study endpoint as a treatment failure and will be referred for continued treatment.

For the research purposes of this study, all PET/CT scans will be read using the best available tools at the time the trial starts. As this is an evolving technology, the reading methods may include but are not limited to:

- A group of 2 or more readers with consensus determination of the lesions to be read and standardized reading methodology.

- An automated reading algorithm using a custom developed software program that can objectively measure PET and CT characteristics of interest.

3.2 Participant Characteristics/Eligibility Criteria

Participants will be recruited from clinics in and around Cape Town, South Africa (RSA) and Henan Province, China. Approximately 620 participants will be enrolled. Previous data estimated that 80-85% would be cured by 16 weeks. For this study, we estimate that about 50% will meet the early treatment completion criteria and undergo randomization. This will result in a sample size of approximately 155 per arm for Arms B and C. Enrollment will continue until 310 patients total are randomized into Arms B and C. Note that enrolled participants subsequently found to have resistance at baseline to the drugs used will be withdrawn from the study and replaced. If Arm A enrolls more quickly than expected, additional baseline factors (e.g. chest x-ray) may be used to screen high risk participants out from enrolling onto the study. If Arm A reaches 310 participants and the study is still enrolling, we will stop enrollment into Arm A.

3.2.1 Inclusion Criteria

1) Age 18 to 75 years with body weight from 35 kg to 90 kg
2) Has not been treated for active TB within the past 3 years
3) Not yet on TB treatment
4) Xpert positive for *M.tb*
5) Rifampin-sensitive pulmonary tuberculosis as indicated by Xpert
6) Laboratory parameters within previous 14 days before enrollment:
   a. Serum AST and ALT <3x upper limit of normal (ULN)
   b. Creatinine <2x ULN
   c. Hemoglobin >7.0 g/dL
   d. Platelet count >50 $x10^9$ cells/L
7) Able and willing to return for follow-up visits
8) Able and willing to provide informed consent to participate in the study
9) Willing to undergo an HIV test
10) At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff, willing to undergo COVID-19 testing: viral RNA PCR testing for SARS-CoV-2 to determine active infection and antibody testing for SARS-CoV-2 to determine prior infection
11) Willing to have samples, including DNA, stored
12) Willing to consistently practice a highly reliable, non-hormonal method of pregnancy prevention (e.g., condoms) during treatment if participant is a premenopausal female unless she has had a hysterectomy or bilateral tubal ligation or her male partner has had a vasectomy. If hormonal contraception is used an additional method of pregnancy prevention (as above) should also be used.

If Arm A reaches 310 before Arms B/C are full, we will stop enrolling into Arm A by:

1) Excluding on screening CXR (or CT scan if done) those with:
   a. Thickened pleura suggesting pleural (extrapulmonary) TB
   b. Complete right or left lung collapse

c. Large pleural effusion indicated by blunting of the costophrenic angle on posteroanterior (PA) or anteroposterior (AP) CXR (or either exceeds the first AP quartile or measures >= 3 cm of the hemithorax on CT) [52]

2) Withdrawing those stratified to Arm A based on the baseline PET/CT scan

3.2.2 Exclusion Criteria

1) Clinical suspicion of or confirmed extrapulmonary TB, including pleural TB
2) Pregnant or desiring/trying to become pregnant in the next 6 months or breastfeeding
3) HIV infected
4) Currently COVID-19 infected
5) Unable to take oral medications
6) Diabetes as defined by point of care HbA1c ≥6.5%, random glucose ≥200 mg/dL (or 11.1 mmol/L), fasting plasma glucose ≥126 mg/dL (or 7.0 mmol/L), or the presence of any anti-diabetic agent (including traditional medicines) as a concomitant medicine
7) Disease complications or concomitant illnesses that may compromise safety or interpretation of trial endpoints, such as known diagnosis of chronic inflammatory condition (e.g. sarcoidosis, rheumatoid arthritis, connective tissue disorder)
8) Use of immunosuppressive medications, such as TNF-alpha inhibitors or systemic or inhaled corticosteroids, within the past 2 weeks
9) Use of any investigational drug in the previous 3 months
10) Substance or alcohol abuse that in the opinion of the investigator may interfere with the participant's adherence to study procedures
11) Any person for whom the physician feels this study is not appropriate

4 Study Schedule

4.1 Participant Screening

Interested patients who sign the main study informed consent form (ICF) (and in addition the Genetic Testing ICF and the Stored Samples ICF where applicable) will undergo screening. The following will then be performed/obtained:

1) Complete medical history and physical exam, including vital signs.
2) One or more sputum samples (total sputum volume at least 3 mL) for Xpert, smear and culture (liquid and solid)
3) Blood draw for
    a. Complete blood count, chemistries, liver function tests, and HbA1c
    b. Serum pregnancy testing, if applicable, before the PET/CT
    c. HIV test
4) For sites not doing serum pregnancy testing, urine will be taken for pregnancy test.

The following will be conducted at sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff. Not conducting the below will not be considered a protocol deviation:

5) Viral RNA PCR testing for SARS-CoV-2 to determine active infection: Samples collected may include nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, sputum, and/or saliva. Patients who test positive will be excluded.
6) Antibody testing for SARS-CoV-2 to determine prior infection: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established. Patients who test positive will not be excluded based on this test.
7) Prior test results done by the local healthcare system within the previous 7 days may be accepted but may still be repeated.

Screening tests may be repeated if necessary. Available labs (excluding sputa) within the previous 2 weeks do not need to be repeated. Phenotypic DST will be conducted from isolates grown using sputa samples collected during the early study visits to confirm the molecular DST results. If there are mixed, conflicting, or inconsistent DST results, any positive culture (or culture from available stored sputa) prior to conversion, may be used for additional DST testing.

4.2 Enrollment

If eligibility is confirmed, the participant will be considered enrolled. This can occur with Day 0.

4.3 Day 0 (Baseline Visit)

The following will be performed/obtained at the Day 0 visit prior to medication:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers
    a. Storage raw
    b. Storage as 1:1 mix with Trizol for MTB mRNA
4) One or more sputum samples (total sputum volume at least 3.5 mL) for repeat smear and culture, and possibly Xpert, if needed,
5) Blood will be collected for biomarkers
6) Urine will be collected for
    a. Biomarkers
    b. A pregnancy test, if indicated (performed prior to the scan); if within 2 days of the screening pregnancy test, this test may be skipped.
7) A finger prick will be done for glucose assessment
8) FDG-PET/CT scan, within 7 days of treatment initiation (providing pregnancy test is negative)
9) Saliva sample

Treatment of TB will be initiated on this day (Day 0) with standard treatment using HRZE at routine doses, per local dosing standards; TB drugs are not considered study drugs. If an electronic adherence monitoring system is used, participants will be issued a

system and provided with training on how to use the system. A questionnaire will be provided to some MERM users to evaluate user experience with this adherence method. The questionnaire will be provided at any time up to Week 2 and again at the end of the participant's treatment course.

If a participant's baseline PET/CT does not meet the radiographic early completion criteria, that participant will be placed in Arm A after their Day 0 visit.

Note that at this and all subsequent visits, if a sputum culture is found to be contaminated or is otherwise unevaluable, the participant may be called back to provide another sputum sample. If for some reason a participant needs to come back the next day to provide sputum, this will not be a deviation. In addition, it will not be a protocol deviation if a participant is unable to provide sufficient sputa to perform all sputum-related testing at any study visit.

4.4 **Week 1**

The following will be performed/obtained at the Week 1 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 3.5 mL) for smear and culture (liquid and solid)
4) Blood will be collected for biomarkers
5) Urine will be collected for biomarkers
6) Adherence monitoring

Participants unable to provide sputa spontaneously at any point in the study may be induced.

4.5 **Week 2**

The following will be performed/obtained at the Week 2 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 2 mL) for smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers
    a. Storage raw
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood will be collected for biomarkers
6) Urine will be collected for biomarkers
7) Adherence monitoring

4.6 **Week 4**

The following will be performed/obtained at the Week 4 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated

3) One or more sputum samples (total sputum volume at least 3.5 mL) for Xpert testing, smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers
    a. Storage raw and
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood will be collected for biomarkers
6) A finger stick will be collected for blood glucose levels (prior to PET/CT scan)
7) Urine will be collected for
    a. Biomarkers
    b. Pregnancy test, if applicable (prior to PET/CT scan)
8) FDG-PET/CT scan (taken 4 weeks after the date of the previous scan, with a -3/+7 day window) for those still qualifying for randomization and also Arm A participants if resources allow.
9) Saliva sample for those scanned
10) Adherence monitoring

At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff (not conducting the below will not be considered a protocol deviation):

11) Viral RNA PCR testing for SARS-CoV-2 to evaluate for possible confounding effects on interpretation of the week 4 PET/CT scan. Patients who test positive will be evaluated on a case-by-case basis to determine if continuation on the study will be possible. Patients unable to adhere to essential study procedures (e.g. due to isolation requirements) may be withdrawn from the study.

Participants who do not meet the early completion criteria at this point will be assigned to Arm A.

4.7 **Week 8**

The following will be performed/obtained at the Week 8 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 3.5 mL) for Xpert testing, smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers
    a. Storage raw and
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood will be collected for biomarkers
6) Urine will be collected for biomarkers
7) Adherence monitoring

At week 8, PZA and ethambutol will be discontinued for all participants, per SOC.

4.8 **Week 12**

The following will be performed/obtained at the Week 12 visit:

1) Vital Signs

2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 2 mL) for smear and culture (liquid and solid)
4) Adherence monitoring

4.9 **Randomization into Arms B and C**

Randomization assignment to determine if a participant is put into Arm B or C will occur at the participant's Week 16 visit. All participants not already assigned to Arm A will be reviewed for early treatment completion criteria eligibility. Those who are not eligible will be assigned to Arm A and will complete treatment with HR through 24 weeks, although treatment may be extended for some as clinically indicated. Those who meet eligibility criteria for early treatment completion will be randomized to Arm B or C. Randomization will be stratified by site. Participants randomized to Arm B will continue on HR therapy through week 24. Those randomized to Arm C will complete treatment at 16 weeks. Note that per DSMB recommendations, randomization of participants to Arm C at Week 16 visit has been halted as of September 14, 2020.

Participants in Arm A will also be randomized to receive a PET/CT scan either at week 16 or week 24, stratified by when they were placed in Arm A (baseline, week 4, or week 16).

4.10 **Week 16**

The following will be performed/obtained at the Week 16 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (at least 3.5 mL) for Xpert testing, smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 6 mL) for biomarkers
    a. Storage raw and
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood will be collected for biomarkers
6) A finger stick will be collected for blood glucose levels prior to PET/CT scan; *Arms B and C only*
7) Urine will be collected for
    a. Biomarkers
    b. Pregnancy test, if applicable, prior to PET/CT scan; *those eligible for Arms B and C and those in Arm A randomized to receive a Week 16 scan*
8) FDG-PET/CT scan *for those eligible for Arms B and C and Arm A participants randomized to receive the scan at this time point* (scan should be within +14 days, but ideally within +7 days of randomization)
9) Saliva sample for those scanned
10) Adherence monitoring
11) MERM questionnaire, may be provided to participating users randomized to stop drug at Week 16

At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff (not conducting the below will not be considered a protocol deviation):

12) Antibody testing for SARS-CoV-2 to determine prior infection: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established.

4.11 **Week 20**

The following will be performed/obtained at the Week 20 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 2 mL) for smear and culture (liquid and solid)
4) Adherence monitoring (Arms A and B only)
5) Blood will be collected with plasma batch tested for isoniazid and/or rifampin levels to confirm that Arm A and B participants are still taking drug and Arm C participants are no longer taking drug

4.12 **Week 24**

The following will be performed/obtained at the Week 24 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (at least 3.5 mL) for Xpert testing, smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 6 mL) for biomarkers
    a. Storage raw and
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood will be collected for biomarkers
6) A finger stick will be collected for blood glucose levels prior to scan; *Arm A only*
7) Urine will be collected for
    a. Biomarkers
    b. Pregnancy test, if applicable, prior to scan; *Arm A only*
8) FDG-PET/CT scan *for those on Arm A randomized to receive the scan at this time point* (scan should be within +14 days, but ideally within +7 days from stopping drug)
9) Saliva sample for those scanned
10) Adherence monitoring (Arms A and B only)
11) MERM questionnaire may be provided to participating users randomized to stop drug at Week 24

At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff (Not conducting the below will not be considered a protocol deviation):

12) Antibody testing for SARS-CoV-2 to determine prior infection: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established.

Participants in Arms A and B will complete treatment at Week 24. An Arm A participant may be treated longer than 24 weeks at the discretion of the treating physician.

4.13 **Week 36**

The following will be performed/obtained at the Week 36 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 2 mL) for smear and culture (liquid and solid)

4.14 **Week 48**

The following will be performed/obtained at the Week 48 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (total sputum volume at least 2 mL) for smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers for storage
5) Blood will be collected for biomarkers
6) Urine will be collected for biomarkers

4.15 **Week 60**

Week 60 will be a phone call visit. A focus history will be discussed. If a participant reports TB-related symptoms, s/he will be asked to come for an in-person visit and sputum will be collected for smear and culture.

4.16 **Week 72**

The following will be performed/obtained at the Week 72 visit:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (at least 3.5 mL) for smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers for storage
5) Blood will be collected for biomarkers
6) Urine will be collected for biomarkers

At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff (not conducting the below will not be considered a protocol deviation):

7) Antibody testing for SARS-CoV-2 to determine prior infection: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established.

Participants with a positive culture at this visit may be asked to return for sputum culture confirmation. In the event that the participant is unable to return for follow-up, every effort will be made to contact the participant by telephone to determine his/her clinical status. If possible, health authorities local to where the patient is located may be asked to collect a sputum sample for culture. Note that the window for this visit is ±30 days. Every effort will be made to have the participant return during this window. However, if the

participant is interested but unable to return (e.g. has moved far away for work), this window may be extended for *up to* another 11 months (total up to 12 months) to allow the participant to return for the final study follow-up visit. The trial itself will not be extended by 11 months, and therefore the last participant will still only be allotted a 30 day window for his/her Week 72 visit.

4.17 Recurrence

The following will be performed/obtained in the event that someone on study has a recurrence of TB:

1) Vital Signs
2) Focused medical history/focused physical exam, if indicated
3) One or more sputum samples (at least 3.5 mL) for Xpert testing, smear and culture (liquid and solid)
4) Two or more sputum samples or induced sputum (total sputum volume at least 4 mL) for biomarkers
    a. Storage raw and
    b. Storage as 1:1 mix with Trizol for MTB mRNA
5) Blood draw for
    a. Complete blood count, chemistries (including HbA1c), and liver function tests
    b. HIV test
    c. Biomarkers
6) Finger stick will be collected for glucose level prior to PET/CT scanning
7) Urine will be collected for
    a. Biomarkers
    b. Pregnancy test, if applicable, prior to scanning
8) FDG-PET/CT scan
9) Saliva sample

At sites with sufficient SARS-CoV-2 testing capacity and personal protective equipment for study staff (not conducting the below will not be considered a protocol deviation):

10) Viral RNA PCR testing for SARS-CoV-2 to determine active infection: Samples collected may include nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, sputum, and/or saliva. Patients who test positive will be evaluated on a case-by-case basis to determine if continuation on the study will be possible. Patients unable to adhere to essential study procedures (e.g. due to isolation requirements) may be withdrawn from the study.
11) If appropriate, antibody testing for SARS-CoV-2 to determine prior infection: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established.

    Prior test results done by the local healthcare system within the previous 7 days may be accepted but may still be repeated.

National/local TB records may be reviewed for unreported TB recurrence and for TB retreatment outcomes.

An unscheduled visit may be conducted at any time per discretion of the treating physician. Tests conducted at these visits will be defined by the treating physician based on the clinical assessment of the participant.

4.18 Study Timeline

Table 7: Study Timeline

	Screening	D0	W1[V] (D7)	W2 (D14)	W4 (D28)	W8[D] (D56)	W12 (D84)	W16 (D112)	W20 (D140)	W24 (D168)	W36 (D252)	W48 (D336)	W60 (D420)	W72 (D504)	TB Recurrence
Main Study Informed Consent (Plus Genetic and HIV in RSA)	X														
Medical History/Focused History	X	X	X	X	X	X	X	X	X	X	X	X	phone call	X	X
Physical Exam/Focused Physical Exam	X	X	X	X	X	X	X	X	X	X	X	X		X	X
Sputum Collection[E]															
Smear/Culture	X[H]	X	X	X	X	X	X	X	X	X	X	X		X	X
GeneXpert	X	(X)[S]			X	X		X		X					X
Biomarkers		X		X	X[B]	X[B]		X[B]		X[B]		X[B]		X[B]	X[B]
Saliva		X			X			X[%]		X[%]					X
Blood collection															
CBC/Chems/LFT[G]	X														X
Biomarkers		X	X	X	X	X		X		X		X		X	X
Pregnancy Test[PT] (serum at screening, urine for others)	X[h] (serum)														
HIV Testing	X														X
Plasma drug levels									X						
Finger Stick		X			X			X[%]		X[%]					X
Urine collection															
Biomarkers		X	X	X	X	X		X		X		X		X	X
Pregnancy Test[P] (serum at screening, urine for others)	X (can do if desired before CXR)	X			X[%]			X[%]		X[%]					X
FDG-PET/CT[W]		X			X			X*		X*					X
Adherence monitoring			X	X	X	X	X	X	X	X					
PK sub-study															
Plasma drug levels								X[P]	X[P]						X[P]
MRI sub-study															
MRI		X[W]			X[W]										
COVID-19 testing[CV]															
Viral RNA PCR testing	X				X										X
Antibody testing	X							X		X				X	X

[ξ] *Additional sputum may be collected if contaminated or otherwise compromised*

[A] *subject could be called back for this additional sputa collection, if necessary.*

[H] *Sputum at screening will also be used for screening Xpert*

[S] *May not be done if within 7 days of screening*

[h] *Pregnancy testing from screening may be used for the D0 PET/CT scan if D0 is within 2 days of the screen*

[G] *These will be performed at any visit if clinically significant.*

[PT] *Before any PET/CT scan, MRI or CXR is done, a pregnancy test will be done for applicable females. If the pregnancy test is positive, the PET/CT scan will not be performed*

[CV] *At sites with sufficient SARS-CoV-2 testing capacity*

[%] *For participants receiving PET/CT scan at this timepoint*

[D] *At week 8, ethambutol and pyrazinamide will be discontinued*

[W] ***PET/CT/MRI scan windows** : baseline w/i 7 days after treatment initiation; W4 must be at least 4 wks after baseline scan with a -3/+7 d window; W16 and 24 scan w/i 14 d of visit; relapse ASAP, but w/i 2 wks of recurrence*

[V] ***Visit windows:** Week 1-2: +/- 3 days; Week 4-24: +/- 7 d, noting that Weeks 16 and 24 should be as close as possible to actual date; Week 36-72: +/- 30 days.*

[g] *If sputum is not available for biomarkers, it will not be a protocol deviation.*

[P] *PK substudy visits can occur anytime after W16, the timeline above may be used as a guide*

**Participants will receive 3rd scan at either wk16 or wk24, as determined by study*

5 Details of Research Procedures & Data Collection Methods

5.1 Consent Methodology

Informed consent is a process where information is presented to enable persons to voluntarily decide whether to participate as a research participant. It is an ongoing conversation between the participant and the researchers, which begins before consent is given and continues until the end of the participant's involvement in the research. Discussions about the research will provide essential information about the study and include: purpose, duration, procedures, alternatives, risks, and benefits. Participants will be given the opportunity to ask questions and have them answered.

There will be a main study informed consent (ICF) given. Some South African sites might also have a Genetic Testing and Stored Sample Informed Consent (the information needed from these is included in the Main ICF for China). If the person is potentially eligible and interested in joining the trial, the main study informed consent will be given. The informed consent process will be conducted by research staff, or qualified investigator. The person obtaining informed consent will ask questions to assess the participant's understanding of the ICF and state that participation is voluntary and that participants may refuse participation or withdraw at any time without prejudice to their clinical care.

The consent forms will be prepared in English and translated to Mandarin for China, and Afrikaans and Xhosa for South Africa. The participant will sign the informed consent documents prior to undergoing any procedures. The researcher will also sign the ICF and document the consent process in the participant's research and/or medical record. Participants will also be informed that study results will be made available to them in the future. Consent forms will be stored securely at study sites.

5.2 Treatment adherence

Treatment adherence may be monitored by any of the following methods (but is not limited to only these), depending on local capacity:

- Directly observed therapy either by study staff, local outreach/healthcare worker, participant family member, or friend
- Pill counts at study visits
- Electronic adherence monitoring
- Cell phone card incentives
- Urine drug level or color monitoring (including for those in Arm C after week 16)

Total treatment duration will be determined by dose counts. Participants will receive either 16 weeks of treatment (Arm C) or 24 weeks of treatment (Arms A and B). (Arm A participants may be treated longer.) Participants receiving 16 weeks of treatment will receive 112 doses with a minimum total of 100 doses. Participants who do not meet this minimum dosing requirement within the 16 week visit window will not be eligible for randomization and will be moved to Arm A. Approximately 90% adherence is used, as missing more than this has been associated with an increased risk of poor outcomes [53]. Note that per DSMB recommendations, randomization to 16 weeks of treatment (Arm C) has been halted as of September 14, 2020. Participants receiving 24 weeks of treatment will receive 168 doses with a minimum total of 150 doses. Missed doses during the initial 8 week intensive phase

will be added on to the end of the intensive phase, replacing continuation phase dosing. Arm B participants who do not achieve the minimum 150 doses within the Week 24 visit window will be allowed to complete a minimum of 150 doses even if this exceeds the visit window.

For sites using the Medication Event Reminder Monitor (MERM) pill box, a questionnaire will be given to participants at the beginning of the study and the end of their time using the MERM, as detailed in section 4. In addition, study staff issuing and working with the pill box during the study will be provided a provider questionnaire about the pill box.

5.3 Medical History

5.3.1 Complete Medical History

A complete medical history will be taken and recorded in a structured format. The complete medical history will cover current symptoms, past medical history, occupational history, family and social history, drug history, allergies, and review of systems.

5.3.2 Focused Medical History

A focused medical history will be taken for cause and recorded in a structured format. The focused medical history will cover new medications, new symptoms, medication compliance (not at study entry), and review of systems, as needed.

5.4 Physical Examination

5.4.1 Complete Physical Examination

A detailed complete physical exam will include examination of lymph nodes (cervical, axillary, inguinal), respiratory system, abdomen, skin, vital signs (temp, BP, pulse, respiratory rate), weight and height. Assessment will be recorded in a structured manner.

5.4.2 Focused Physical Examination

A focused physical exam will be directed only at new symptoms and complaints of the participant to find evidence of active disease. This targeted physical exam will include vital signs and weight and may include examination of lymph nodes, respiratory system, and abdomen. The assessment will be recorded in a structured manner.

5.5 Sputum Collection and Mycobacterial Testing Procedures

5.5.1 Sputum Induction

Sputum induction, if necessary during the study, will take place in a designated area. Hypertonic saline will be administered via the mouthpiece of a nebulizer to induce expectoration. Staff members assisting participants will wear N95 respiratory protection masks.

5.5.2 Sputum Specimen Use

5.5.2.1 *Xpert*

For screening, initial drug susceptibility testing (DST) and *M.tb* infection confirmation will be performed by Xpert (Cepheid, Sunnyvale, CA) and/or newer generation tests if available. Xpert and/or newer generation tests will be used to measure cycle threshold (the amplification cycle when Mtb DNA becomes detectable in the Xpert assay) as an indicator of changing bacterial load at visits indicated in the Study Timeline but are only required at screening and at Week 16 visit. Both Xpert MTB/RIF and Xpert Ultra may be done if sputum or sputum sediment is available.

5.5.2.2 *Sputum Smear and Culture*

Sputum will be cultured on solid and in liquid media at each time point to determine time to sputum culture conversion to negative. The time to negative will be determined on Lowenstein-Jensen solid agar for the primary assessment. The date of culture conversion will be defined as the date of the first of two consecutive negative cultures on solid medium over at least 4 weeks. MGIT liquid cultures will be used for the primary assessment of time to sputum culture conversion for patients with positive baseline or early MGIT cultures but negative baseline or early LJ cultures. Contaminated cultures prior to the week 60 culture will be considered "no test", i.e. neither culture positive nor negative, and will be repeated to the extent possible. Cultures week 60 and after that are contaminated but do not grow *M.tb* will be considered "no test" but will not be repeated. To note, if a culture is contaminated or otherwise not evaluable, another sputum sample may be requested prior to the next study visit. Participants found during follow-up to have a sputum culture positive for *M.tb* that is confirmed on a subsequent culture will be referred to their physician for retreatment. Solid culture results that are missing will be noted as "unavailable".

Decontaminated and concentrated sputum sediment will be inoculated onto LJ medium and into the BACTEC MGIT liquid medium for measurement of time to positivity (TTP) as previously described [54]. The remaining sediment may be frozen for further examination.

Staining for acid-fast bacilli (AFB) on collected sputum will be performed in the diagnostic laboratory according to standard site and/or National Health Laboratory Service (NHLS) SOPs.

5.5.2.3 *Biomarkers in Sputum*

Additional sputa will also be collected and saved for assessing LAM concentration, bacterial mRNA copies, and metabolites. Sputum may also be saved for detecting mixed MTB infections and/or heteroresistance (e.g. drug resistant mutants in the drug susceptible population) as well as attempts to resuscitate possible persister populations. Sediment may be tested for MTB biomarkers.

5.5.3 Isolates and Genotypic Tests

Isolates (during the study and at relapse) will be saved for further work such as strain typing, whole genome sequencing, minimal inhibitory concentration determination, and single nucleotide polymorphism (SNP) typing.

Participants who develop recurrent disease will have DNA extracted from their recurrent TB isolate and their original TB isolate for comparison. We will use one of the several recognized genetic testing methods to determine the level of identity between two isolates. Examples include mycobacterial interspersed repetitive unit-variable number tandem repeats patterns (MIRU-VNTR), or whole genome sequencing to identity single nucleotide polymorphism (SNP) patterns [55]. Other bacterial strain studies to characterize strains may also be performed.

5.5.4 Drug Susceptibility Testing of Isolates

Phenotypic drug susceptibility testing for at least RIF and INH will be performed on one of the first-available culture isolates in real time. Saved early and later isolates including any relapse isolates may be tested for drug resistance by various methods including sub-breakpoint MICs by the research laboratories involved in the study.

5.6 Saliva Collection for Biomarkers

Saliva will be collected by the use of salivettes. Participants will chew on salivettes and the saliva will be collected into a tube. Approximately 2-6 mL of saliva will be collected for immunological marker and metabolomics research tests.

These samples will be used for discovery or validation of a biomarker set that may include approximately 10-15 cytokines, soluble cytokine receptors, ligands, and extracellular matrix proteins. The factors that we expect to measure include, but are not limited to, IL-13, IL-5, IFN gamma, TNF alpha, TGF beta, sIL-4R, sIL-2Ra, sIL-6R, sCD40L, MMP-2, MMP-9, and C-reactive protein.

5.7 Blood Collection

Approximately 10-60 mL of blood will be drawn at collection time points, depending on the necessary testing and cultural acceptability.

5.7.1 CBC, Chemistries, Liver Function Tests, HbA1c and HIV

Up to 10 mL of blood will be collected in a clot activator tube (with or without gel) for serum collection. This serum can be used for performance of HIV test, liver function tests (AST [SGOT], and ALT [SGPT]), creatinine and pregnancy testing.

Up to 5 mL of blood will be collected in an EDTA tube for hematology (complete blood count [CBC]) and HbA1c. Samples will be processed in the hematology and biochemistry laboratories of participating sites or accredited lab service (e.g. National Health Laboratory Service in South Africa) according to local SOP.

Testing may be repeated to check for or follow-up on abnormal results when needed, for example, at screening or after learning about an adverse event.

5.7.2 Biomarkers

5.7.2.1 *Blood Immunological Markers*

Blood (approximately 10 to 42 mL) will be drawn for immunological marker and metabolomic research tests. The blood will be used for discovery or validation of a biomarker set that may include approximately 10-15 cytokines, soluble cytokine receptors, ligands, and extracellular matrix proteins. The factors that we expect to measure include, but are not limited to, IL-13, IL-5, IFN gamma, TNF alpha, TGF beta, sIL-4R, sIL-2Ra, sIL-6R, sCD40L, MMP-2, MMP-9, and C-reactive protein. In addition, serum will be used for metabolomics studies to identify small metabolites that are representative of biologic pathways that are related to differential outcomes. Blood volumes that may be collected include 8 mL whole blood for serum, 1 mL whole blood for FACS, and 30 mL whole blood for whole blood assay or PBMC isolation.

5.7.2.2 *Blood for host mRNA*

Approximately 2.5 mL blood will be collected in PAXgene tubes for determining host mRNA signatures.

5.7.2.3 *Blood for host DNA*

Susceptibility to MTB has not been isolated to a particular gene or immune pathway. Many studies of single nucleotide polymorphisms (SNPs) have identified genetic variations that may contribute to disease progression. SNPs in genes such as NRAMP1, MCP1, TIRAP, P2X7 and TLR2 have been associated with either susceptibility to or protection against developing tuberculosis [56]. However, these and other SNPs often differ among the ethnic groups analyzed making it difficult to decipher the influence of environmental factors verses gene interactions in tuberculosis severity. Since this will be a large cohort study followed for treatment response rate and treatment outcome, DNA samples will be saved for future use in studies of genetic vulnerability.

Approximately 1 mL of heparinized whole blood collected at the first blood biomarkers draw will have the red cells lysed and white cells pelleted for future DNA preparation and saved at -80C. Additional blood cell pellets from other assays above may also be saved for use in DNA preparation.

5.7.3 Plasma for drug levels

Approximately 1 mL of blood will be collected in an EDTA tube and spun down for plasma, which will be stored frozen. Isoniazid and/or rifampin plasma levels will be batch tested.

5.8 Urine Collection and Testing

5.8.1 Pregnancy Testing

Urine will be collected for pregnancy testing before each PET/CT scan (except for the initial pregnancy test which will be from blood). A commercial human chorionic gonadotropin (β-hCG) determination assay will be performed in accordance with manufacturers' guidance.

5.8.2 Biomarkers in Urine

Urine will also be collected and saved for future biomarker research and for research metabolomics studies. Approximately 1 cup will be collected mid-stream.

5.9 **Imaging**

5.9.1 FDG-PET/CT scan

The FDG-PET/CT scan will be performed at a facility local to the study site. The CT portion of the scan will be done without contrast and will be limited to the chest. Participants will consent to receive a maximum of 4 FDG-PET/CT scans during the study (baseline, Week 4, Week 16 or 24, and at recurrence); however, the vast majority of participants will only receive 3 scans. If resources do not allow for all Arm A participants to receive a week 4 scan, they may receive only 2 scans. Participants will be fully briefed with regard to what to expect and any precautions highlighted. Participants will be asked not to eat for about 6 hours prior to the scan but to drink plenty of water. Participants will have an assessment of blood sugar. Following the blood test, a venous cannula will be inserted and approximately 7 mCi of radiolabeled ^{18}F-FDG administered. After about 50 minutes participants will void urine, and at about 60 minutes after injection, the participants will undergo a FDG-PET/CT scan of the chest. For radiation dosimetry details of the scanners used, please refer to the appendices.

5.9.2 Chest Radiograph (CXR)

CXRs may be taken at participating sites if clinically indicated. They will be performed according to local SOP.

5.9.3 Magnetic Resonance Imaging (MRI)

Pulmonary MRIs with a macrocyclic gadolinium-based contrast will be used in a subset of participants who meet the substudy inclusion and exclusion criteria. Gadolinium highlights fibrosis, the development of which may affect treatment outcomes. Participants in this substudy will consent to receive a pulmonary MRI scan at baseline and again at Week 4 (two total scans). They will be performed according to local SOP. There is no ionizing radiation associated with MRI scanning.

Nephrogenic systemic fibrosis (NSF) is a rare, progressive, usually fatal disease involving fibrosis of the skin and internal organs and occurs exclusively in patients with renal failure who receive gadolinium. To reduce the risk of NSF:

- Participants with any evidence of renal insufficiency, defined as an estimated glomerular filtration rate (eGFR) <60 mL/min, will be excluded from the MRI substudy.
- Macrocyclic chelate preparations of gadolinium are associated with lower levels of free gadolinium than linear chelates. Free gadolinium (Gd^{3+}) can precipitate in tissues and is believed to be the toxin that causes NSF in those with renal failure. A macrocyclic gadolinium-based contrast agent will be used for all patients.

Gadolinium retention: There is increasing evidence that trace amounts of gadolinium are retained in the body for months to years, including in the bone, brain, and other organs. Retention seems to occur with all forms of gadolinium but is less with macrocyclic gadolinium contrast agents. The clinical significance of this retention is not clear. No adverse health effects in patients with normal kidney function have been directly linked to gadolinium retention and thus the FDA has concluded that the benefit of approved gadolinium based contrast agents outweighs any potential risks [57].

5.10 COVID-19 Testing

Viral RNA PCR testing for SARS-CoV-2 may be conducted to determine active infection: Samples collected may include nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, sputum, and/or saliva. Antibody testing for SARS-CoV-2 to determine prior infection may also be conducted: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established. Prior test results done by the local healthcare system within the previous 7 days may be accepted but may still be repeated.

5.11 Participants moved to Arm A

During the trial, participants may be moved to Arm A for the following reasons:

1) Not meeting early treatment completion criteria (including incomplete adherence)
2) Found to be pregnant. No further PET/CT scanning will be done on these participants.
3) Experiencing adverse drug effects severe enough to require permanent discontinuation of any individual TB drug. Drug pauses are also allowed without a change in arm as long as treatment adherence requirements are met (see Sec 5.2).

Divided daily dosing or changing fixed-dose combination tablets to single-drug tablets is permitted without a change in arm as long as all standard drug components for that time point are maintained.

5.12 Withdrawal or Termination From the Study

1) Participants with evidence of pleural TB or large pleural effusion on baseline PET/CT will be withdrawn and replaced.
2) Participants with significant incidental findings on PET/CT that require immediate diagnostic procedures or treatment may be withdrawn from the study if in the opinion of the investigator, continuing on the study may not be in the participant's best interests.
3) Participants identified to have resistance on molecular or phenotypic DST will be withdrawn and replaced, if possible.
4) Participants enrolled in the study based on a positive Xpert but subsequently found to be culture negative will be withdrawn and replaced.
5) Participants on Arm A who are still culture positive at Week 24 will be considered to have treatment failure and will be discontinued from the study as having reached a study endpoint.

During any stage of the study, participants may be withdrawn if:
1) Participant withdraws consent
2) There is any reason deemed appropriate by the investigator or attending physician.

5.13 Study Completion

Participants will be considered completed if they complete the final, 18-month follow-up visit and have not developed recurrent TB. For those who develop recurrent TB, Arm A participants will be considered to have completed the study at the time of recurrence confirmation (study endpoint) and will be referred for treatment. Arms B/C participants who develop recurrent TB will complete observational follow-up to the end of re-treatment for retreatment outcome even if this exceeds 18 months. Note that participants who consent to participate in the PK substudy will continue in the overall study until completion of the PK study.

6 Statistical Methods and Justification

6.1 Study Hypotheses

6.1.1 Primary Hypothesis

A combination of radiographic characteristics at baseline, the rate of change of these features at one month, and markers of residual bacterial load at the end of treatment will identify patients with tuberculosis who are cured with 4 months (16 weeks) of standard treatment.

6.1.2 Secondary Hypotheses

1) The treatment success rate among low-risk participants with shortened treatment will be similar to that of a representative 6-month standard of care population.
2) In univariate and multivariate analyses, demographic, radiographic, bacterial load, microbiologic, and immunologic markers are associated with treatment failure.
3) In univariate and multivariate analyses, demographic, radiographic, bacterial load, microbiologic, and immunologic markers are associated with subsequent relapse.
4) PET scans predict treatment outcomes in univariate and multivariate analyses that include relevant variables based on secondary hypotheses 2) and 3).
5) Bacterial load markers (TTP and Xpert cycle threshold) collected at later time points are better markers of ultimate treatment outcomes than markers collected at earlier times.
6) Pharmacokinetic and sub-breakpoint MIC measurements will significantly predict treatment outcome in 16 and 24 week treatment regimens.
7) Baseline and week 4 MRI imaging biomarkers will stratify TB patients into higher risk (require longer treatment) and lower risk (cured with shorter treatment) cohorts in accordance with the PredictTB early treatment completion criteria.

6.2 Definitions

6.2.1 Treatment Success

Treatment success will be defined as a participant with at least 2 consecutive negative cultures on solid medium over a span of at least 4 weeks, achieved before the end of therapy, with no subsequent confirmed positive cultures during follow-up. Two consecutive negative cultures on liquid medium over a span of at least 4 weeks will be used for patients who were liquid medium positive but solid medium negative at the baseline or early cultures.

6.2.2 Treatment Failure

Participants who remain culture positive on solid medium at Week 24 in Arm A will be considered treatment failures. Arm A participants who become culture negative on 2 consecutive cultures over at least 4 weeks who subsequently become culture positive again while still on treatment, confirmed on a subsequent culture, will also be considered treatment failures. This is a study endpoint. The participant will be discontinued from the study and will be referred to continue treatment per the local SOC. Participants who convert to solid culture negative who subsequently have a single solid culture positive for *Mtb* before or at week 24 need to have a subsequent culture positive for *Mtb* to be confirmed as treatment failures.

Participants in Arm A only before Week 24 who develop new clinical or laboratory signs or symptoms of TB worsening may have their intensive phase treatment continued or restarted at the discretion of the treating physician. If the treating physician believes that additional drugs are warranted for patient safety before culture results are known, multiple additional sputum samples over >1 day should be collected before additional treatment is started. The participant will be considered a clinical treatment failure and discontinued from the study.

Solid culture results will be used for all primary endpoint analyses with the exception that the liquid culture result may be used at the final week 72 study visit if the solid culture result is contaminated or missing and the participant cannot be brought back to repeat the sputum sample. In this instance, if the liquid culture result is negative, the sputum sample will be considered negative. For other study visits, only solid culture results will be used for the primary analysis. Solid culture results that are missing or contaminated will be classified as unavailable. Liquid culture results may be used for secondary analyses.

Participants randomized to Arms B or C who are subsequently found to have a positive culture for *Mtb* on solid medium between and including weeks 16-24 that is confirmed on a subsequent culture will be considered treatment failures. These participants will be referred to continue treatment per local SOC and followed observationally until the end of their treatment to determine outcome. TB DNA strain typing may be done (sec 5.5.3) to confirm whether or not this is the same strain of DNA as the participant had at baseline.

Single positive cultures with no other corroborating clinical or laboratory evidence are not considered failures. Single positive cultures may arise from clerical error or laboratory contamination [58].

Participants determined to be treatment failures may be evaluated for COVID-19 infection. Viral RNA PCR testing for SARS-CoV-2 may be done to determine active infection: samples collected may include nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, sputum, and/or saliva. Additionally, antibody testing for SARS-CoV-2 to determine prior infection may be conducted: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established. Prior test results done by the local healthcare system within the previous 7 days may be accepted but may still be repeated.

6.2.3 Treatment Recurrence and Relapse

Participants who convert their sputum to culture negative (2 consecutive negatives over ≥4 weeks) and who subsequently become culture positive for *M.tb* again on solid medium, during follow-up after week 24, confirmed by a second (on another day) sputum culture positive for *M.tb*, will be considered recurrences. Single positive cultures that are negative on follow-up culture will not be considered recurrences. Participants with a positive, contaminated, or unevaluable culture on the final month 18 (week 78) follow-up visit may be asked to return for sputum culture confirmation.

Relapses will be distinguished from re-infections by DNA strain typing (sec 5.5.3) and only relapses will be considered a study endpoint.

Relapses on Arms B and C will have observational follow-up until the end of retreatment.

Participants determined to be treatment recurrences or relapses may be evaluated for COVID-19 infection. Viral RNA PCR testing for SARS-CoV-2 may be done to determine active infection: samples collected may include nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, sputum, and/or saliva. Additionally, antibody testing for SARS-CoV-2 to determine prior infection may be conducted: Samples collected may include blood and/or saliva and may be stored for batched testing at a time when serological testing has been established. Prior test results done by the local healthcare system within the previous 7 days may be accepted but may still be repeated.

6.3 **Primary Endpoint**

The primary endpoint will be a comparison of the rate of treatment success at 18 months (after treatment initiation) between Arms B and C. Final study treatment outcome data from participants who are unable to return at 18 months but do return during the 1 year following will be imputed back to the 18 month time point for the primary endpoint. Imputation methods based on the event rates in Arms B and C will be utilized for

participants who successfully complete treatment, but are later lost to follow-up without evidence of Mtb. The imputation method will assume an exponential model that will account for the duration of missingness. Extensive sensitivity analyses will explore the robustness of the imputation method. The details of these analyses are provided in the statistical analysis plan.

6.4 Secondary Endpoints

The endpoint relating to the first secondary endpoint is treatment success, as defined for the primary endpoint. Other endpoints relating to secondary hypotheses are based on imaging, microbiology, and immunology. The endpoints from images include: PET total glycolytic activity in regions of interest, total volume of hard CT lesions (-100 to 100 HU), total volume of soft CT lesions (-500 to -100 HU), and cavity air (volume of air in cavities). Endpoints relating to immunologic markers will be based on serum cytokine levels as described in section 1. Xpert cycle threshold will be analyzed as a continuous variable in addition to the pre-specified thresholds. Analyses will also consider transformations such as delta cycle threshold.

6.5 Statistical Methods

6.5.1 Primary Analysis

The primary analysis will estimate the lower bound of a one-sided 95% confidence interval of the difference in success rates between arms B and C. If the lower bound is greater than -7%, this will be evidence that the treatment-shortening arm is not inferior to the standard duration arm. Confidence intervals will be constructed using Wald intervals, with inverse weighting according to site-estimated variances, as a stratified analysis. Additional analyses of the primary endpoint will consider a non-stratified-based confidence interval of the difference.

6.5.2 Secondary Analyses

The difference (and 95% confidence interval) in treatment success rates between a combined A+B Arm (with Arm A participants selected to represent a true 6-month standard of care population) and a combined Arm A+C (with the remaining Arm A participants selected to represent a treatment shortening strategy arm, and no overlap in Arm A participants assigned to B and C). Logistic regression will be used to develop risk scores based on the various markers. Markers that are statistically significant univariately will be evaluated in multivariate models. ROC analyses will be conducted, when appropriate, using the non-parametric AUC estimator and the corresponding bootstrap variance estimators for hypothesis testing. Permutation methods and leave-one-out cross-validation will be implemented to control the family-wise error rate and adjust for over-fitting bias. When evaluating the relationship between time of specimen sampling with diagnostic ability (e.g., comparing early versus late Xpert cycle threshold as a predictor of relapse risk), time-dependent ROC curves will be estimated.

6.6 Sample Size Considerations

For this study, the sample size is calculated for Arms B and C, which are used for the primary endpoint. Because these are lower risk participants, we expect a treatment success

rate of 97%. Table 8 provides power calculations for sample sizes of 129 and 155 per group, adjusted for a 10% loss to follow-up rate. With true success rates of 97% in both arms, study power is greater than 90% with only 129 participants per group. However, to increase power to accommodate a scenario in which the true success rate in the four-month treatment arm is slightly lower than the six-month arm, a sample size of 155 per treatment arm was selected. We expect that approximately 50% of participants will be classified as higher risk and be placed into Arm A, giving a total study sample size of 620 participants. The proportion of participants considered lower risk by our treatment completion criteria will be monitored during the study. If it becomes expected that Arm A will complete enrollment (i.e., 310 participants) prior to Arms B/C, a corrective action plan will be developed to slow enrollment in Arm A so that Arms B/C enrollment can catch up.

Table 8. Power calculations for total sample sizes of 129 and 155 per group (arms B and C) for different success rates across and between treatment arms.

Success rate by study arm		Power for concluding NI with 7% margin, one-sided 5% type I error rate, and 10% loss to follow-up	
6-month tx	4-month tx	Sample size 129 per group	Sample size 155 per group
0.99	0.99	0.999	1
0.99	0.98	0.984	0.994
0.99	0.97	0.863	0.912
0.98	0.98	0.985	0.994
0.98	0.97	0.903	0.942
0.98	0.96	0.726	0.792
0.97	0.97	0.932	0.963
0.97	0.96	0.803	0.862
0.97	0.95	0.621	0.689
0.96	0.96	0.862	0.911
0.96	0.95	0.716	0.782
0.96	0.94	0.545	0.609
0.95	0.95	0.792	0.851
0.95	0.94	0.644	0.711
0.95	0.93	0.487	0.547

Table 9 provides sample sizes for differing non-inferiority margins.

Table 9: Sample Sizes for 80%, 85%, and 90% power with a one-sided type I error rate of 5% for various non-inferiority margins, assuming equal success rates in Arms B and C.

		Sample size (per arm)		
Treatment success rate	NI margin	80% Power	85% Power	90% Power
0.90	0.04	696	809	964
0.90	0.05	446	518	617
0.90	0.06	310	360	429
0.90	0.07	228	265	315
0.95	0.04	368	427	509
0.95	0.05	235	274	326
0.95	0.06	164	190	226
0.95	*0.07*	*120*	***140***	*167*
0.97	0.04	225	262	312
0.97	0.05	144	168	200
0.97	0.06	100	117	139
0.97	0.07	66	85	102

Note that the sample size was determined based on an assumed treatment success rate of 95%, a non-inferiority margin of 7%, with 85% power.

6.7 Interim analyses

6.7.1 Early stopping for inferiority of treatment shortening arm

Interim analyses will be performed for safety in Arms B and C, to evaluate whether the poor outcome rate is worse in the arm with earlier treatment completion. A Fisher's Exact test will be performed after 1/3 and 2/3 of participants have been followed for 72 weeks from study entry, using a Pocock boundary. The stopping boundary is derived from a test of inferiority (of the treatment shortening arm) that corresponds to a z-score of 2.178 (i.e., a two-sided p-value of 0.029).

Note that as of September 14, 2020 the Data Safety Monitoring Board recommended that randomization to Arm C should be discontinued following the interim analysis conducted after approximately 1/3 of participants were followed for 72 weeks.

6.7.2 Early stopping for study futility

The premise of this study is that imaging and Xpert markers can identify a subset of participants with high success rates. If the success rate is low in this subset, this will call into question the basis for this study. Accordingly, monitoring for a low success rate is proposed. When half of the participants have completed their week 72 follow-up, the treatment success rate will be evaluated and presented to the Data and Safety Monitoring Board (DSMB). If more than 16 (of 75) participants in the standard treatment arm (Arm B) have a relapse, a recommendation to stop the trial will be considered. Alternatively, a recommendation to stop randomization into Arm C will also be considered. In this scenario, participants who would be eligible for shortening

will be put into Arm B. Under this scenario the highest achievable success rate in the standard treatment arm is 90%. Success rates lower than 90% would be concerning given the eligibility criteria for randomization, which represents a subset of participants with a low probability of relapse. Conditional power will be included in the DSMB reports to give guidance about the likelihood of concluding non-inferiority if the study continues to full enrollment. A table of conditional power computations will be included using a range of non-inferiority margins (e.g., ranging from 6%-10%), since determination of an acceptable margin may depend on multiple factors.

7 PK Substudy for Sub-Breakpoint Minimum Inhibitory Concentrations (MIC) Comparison

TB patients are known to have widely variable serum PK values and these differences appear to affect treatment outcome [59-61]. Because a given patient's serum drug concentration achieved would clearly affect the clinical interpretation of a given MIC result, a model incorporating both parameters may predict outcomes better than either one alone.

To test the sub-breakpoint MIC concept, we would ideally identify those at highest risk of relapse, and those who have relapsed, to see if there were differences in sub-breakpoint MIC and/or PK/sub-breakpoint MIC at baseline, compared with those who did not relapse, as it is not practical to collect PK data on all participants. Based on preliminary data, we believe that participants who enter Arm A due to a residual bacterial burden at the end of treatment determined by a Week 16 Xpert Ct <28 are at a higher risk of relapse. We can therefore target this group of participants for inclusion into this PK substudy. Because not all these participants will ultimately relapse, we will have PK data on both participants who relapse and those who do not. Additionally, any participant determined to have a poor treatment outcome will be invited to participate in this substudy. Participants with poor treatment outcomes who agree to participate in the PK substudy will remain in the overall study until they complete the substudy.

7.1 Substudy Procedures

PK data will only be collected for INH and RIF. The study timeline is as follows. (Note, PK visits may occur between main study visits.)

- Week 16: Identification of participants who move to Arm A from Arm B/C due to Week 16 Xpert Ct<28 and are eligible for this substudy.

  After these participants are identified and enrolled, they may be scheduled for PK visit dates at any time, even if it is not at the dates recommended below. The timeline below may be used as a guide.

- Weeks 16-24: Site staff introduce substudy to participants who moved to Arm A from Arm B/C at week 16 due to Xpert Ct<28. Those willing to participate will provide informed consent and will be instructed to come to a subsequent substudy visit (may occur on or between overall study visits) without having taken their TB drug dose for the day.

- When the participant attends their first substudy visit, the time of TB medication intake (e.g. pill box opening times) for the two previous days are recorded. A trough blood (approximately 2 mL) is drawn, then the TB drug dose is taken. Blood (approximately 2 mL) is then drawn at approximately 1, 2, and 6 hours post-dose. Note that PK sampling cannot be done if the participant takes his/her TB drug dose before coming to the clinic and sampling will be deferred to the next study visit.
- Windows for the blood draws are +/- 10 min for the 1 and 2 hour draws and +/- 20 min for the 6 hour draw.
- Participants are instructed to come to their second substudy visit without having taken their TB drug dose for that day. The second substudy visit can occur on any day after the first substudy visit while the patient is still taking TB medicines.
- For every participant enrolled in the substudy who moved to Arm A at week 16, a control participant, who did not move to arm A at week 16 and who was randomized to Arm B will be enrolled and PK sampling performed as described above on 2 separate days.

- Weeks 16-72: Identification of those participants who have poor treatment outcomes. Participants who have poor treatment outcomes be referred to the local TB clinic to continue or restart treatment. Concurrently, site staff will introduce the substudy to these participants, and those willing to participate will provide informed consent. Substudy participants will be instructed to come for two PK sampling visits without having taken their TB drug dose for the day. Substudy participants restarting treatment should schedule their 1st visit at least 7 days after restarting treatment to allow time for steady-state kinetics to develop.

Participation in the substudy concludes after two days of PK sampling are completed.

Two days of PK sampling per participant are necessary to differentiate inter-patient variability from within patient inter-occasion variability. Inter-occasion variability is "random" variability due to factors such as food intake, concomitant medicines, co-morbidities, etc., that may affect absorption and that change over time. With two sets of data points at different times for each patient, the PK modelers will be able to differentiate signal from noise much more effectively.

7.2 **Inclusion criteria:**

1) Participant moved to Arm A due to Week 16 Xpert Ct <28 OR participants determined to have a poor treatment outcome
2) Willing to come for two study visits) without having taken that day's TB drug dose, then stay for at least 6 hours for blood draws at 0, 1, 2, and 6 hours after taking the TB drug dose.
3) Willing to have samples stored
4) Willing to sign substudy informed consent

7.3 **Exclusion criteria**

1) None

7.4 Control participants

Each participant transferred from Arm B/C to Arm A at week 16 due to Xpert<28 will be invited to participate in this substudy. Enrolled substudy participants who do not develop poor treatment outcomes will be treated as control participants.

Participants invited to participate in the substudy following determination of a poor treatment outcome will not have direct controls.

7.5 Substudy Remuneration

The substudy participant will be compensated $25 (350 rand; 165 yuan) for each day of PK sampling completed.

8 Lung Magnetic Resonance Imaging Scans to Individualize Treatment Duration for Pulmonary Tuberculosis Patients Substudy

MRI has been revolutionized in two directions over the last 40 years. Initial human MRI scanners used magnetic fields around 0.05-0.35 tesla (T). As hardware and software technology improved, higher magnetic field machines became available generating increasing image resolution but also higher expense due to the systems and shielding required to maintain high magnetic fields, with 7T systems now marketed. In contrast, more recent advances have emphasized a return to lower field MRI systems that are less expensive and potentially scalable in resource-limited settings. The software and hardware advances developed for high magnetic field systems are now also being applied to lower field systems to offset the decreased image resolution to an acceptable diagnostic level [62]. It is in this context that we hypothesize that baseline and week 4 MRI imaging biomarkers will stratify TB patients into higher risk (require longer treatment) and lower risk (cured with shorter treatment) cohorts in accordance with the PredictTB early treatment completion criteria.

Although PET/CT scanning is unlikely to be widely available globally soon, low-field MRI scanning is starting to bridge the gap between developed and developing country availability and thus may be relevant to future global TB treatment algorithms. This substudy anticipates the coming wider availability of MRI systems to synergize with the ongoing PredictTB clinical trial, which already conducts PET/CT scans, to add a small substudy to collect additional lung MRI scans at baseline and week 4. If the PredictTB trial is successful, blood, sputum, or urine-based biomarkers that correlate with the PET/CT signature will be sought but adding MRI scans provides an additional potential route to global applicability and scalability of the treatment shortening algorithm.

8.1 Substudy Procedures

MRI image data will only be collected for participants in the PredictTB study who enroll in this substudy.

- Screening visit: PredictTB main study ICF and MRI substudy ICF will be introduced and signed.
- Baseline enrollment visit: Participants who meet the main study inclusion/exclusion criteria will be enrolled. Enrolled main study participants who also meet the MRI substudy

inclusion/exclusion criteria will also be enrolled, with baseline and week 4 pulmonary MRI scans scheduled.

The following study procedures will be performed/obtained for this substudy:

1) Day 0: pulmonary MRI scan, within 7 days of treatment initiation (provided pregnancy test is negative)
2) Week 4: pulmonary MRI scan, taken 4 weeks after the date of the previous scan, with a -3/+7-day window (provided pregnancy test is negative)

8.2 Inclusion Criteria

1) Enrolled onto the PredictTB study.
2) Has a screening or baseline GeneXpert semiquantitative reading of medium or high to reduce the enrollment of participants with minimal disease in whom an MRI scan may be of less benefit.
3) Willing to sign the MRI substudy informed consent.

8.3 Exclusion Criteria

1) Unable to undergo MRI as determined by MRI safety screen (e.g., pregnancy, metal in body, claustrophobia) using the standard screen conducted by the MRI imaging facility.
2) Estimated GFR <60mL/min/1.73m^2 to reduce the risk of NSF.
3) Received a MRI scan with gadolinium-based contrast agent within the last 12 months.
4) Intolerance to macrocyclic gadolinium-based contrast agents.

Note that a pregnancy test will be repeated in females before the week 4 MRI scan and, if positive, the participant will be withdrawn from the substudy and that scan will not be done.

8.4 Statistical Methods

The primary statistical analysis (Aim 1) will be whether our baseline and week 4 MRI lesion quantitation parameters classify PredictTB patients into the PredictTB-defined high and low risk cohorts (final week 16 classification into Arm A (high risk) vs. Arms B or C (low risk)) with classification accuracy (p) greater than chance (>50%). If the true classification accuracy is 72%, the substudy would have a power of 90% to rule out the null hypothesis of p=0.5 with 60 participants (Table 10). Sensitivity (correct MRI classification amongst the high-risk subjects) and specificity (correct MRI classification amongst the low-risk subjects) will be estimated, along with 95% confidence intervals.

Power	Sample size	True p
90%	30	0.78
	40	0.75
	50	0.73

	60	0.72
80%	30	0.75
	40	0.72
	50	0.70
	60	0.69

Table 10: MRI Substudy Power Calculation

Approximately 40 participants will be enrolled in South Africa and 20 in China. The correlation between quantitative MRI and PET-CT variables will be tested and presented on Bland-Altman plots.

8.5 Substudy Remuneration

The substudy participant will be compensated 350 rand; 100 yuan for each MRI scan completed.

9 Adverse Events, Unanticipated Problems, Deviations, and Non-Compliance

9.1 Reporting Procedures

9.1.1 Assessment of Safety

9.1.1.1 AEs and other reportable events are defined in Policy 801: Reporting Research Events

9.1.2 Reporting Procedures

9.1.2.1 Unanticipated problems, non-compliance, and other reportable events will be reported to the NIH IRB according to Policy 801.

9.2 Investigator Assessment of Adverse Events

9.2.1 Grading Adverse Events for Severity

The severity of each AE will be determined using the DAIDS Table for Grading the Severity of Adult and Pediatric Adverse Events (Corrected Version 2.1 – July 2017). Any events that are not listed in this toxicity table will be graded by the local investigator as follows:

Grade 1 - Mild	Transient or mild discomfort; no limitation in activity; no medical intervention/therapy required

Grade 2 - Moderate	Moderate limitation in activity – some assistance may be needed; no or minimal medical intervention/ therapy required
Grade 3 - Severe	Marked limitation in activity; some assistance usually required; medical intervention/therapy required, hospitalizations possible
Grade 4 - Life-threatening	Extreme limitation in activity, significant assistance required; significant medical intervention/therapy required, hospitalization or hospice care probable

Grade 5 – Death

9.2.2 Assessing Adverse Events for Relationship to Study

Any AE that occurs in a participant will be assessed for its relationship to the study. A causal relationship means an intervention caused (or is reasonably likely to have caused) the AE. This usually implies a relationship in time between one or more intervention and the AE—for example, the AE occurred shortly after the participant received the drugs/study agents/intervention.

For all AEs, the clinician who examines and evaluates the participant will determine the AE's causality based upon the temporal relationship to administration of the intervention, the pharmacology of any applicable study agents, and his/her clinical judgment.

The following scale will be used to reflect the PI's judgment as to the relationship between the intervention and the AE:

Definitely Related: The AE is clearly related to one or more of the interventions – follows a reasonable temporal sequence from administration of one or more of the interventions, follows a known or expected response pattern to the one or more of the interventions that is confirmed by improvement on stopping and reappearance of the event in repeated exposure and that could not be reasonably explained by the known characteristics of the participant's clinical state.

Probably Related*:* The AE and administration of the interventions are reasonably related in time and/or follows a known pattern of response, and the AE is more likely explained by one or more of the interventions than other causes.

Possibly Related: AE follows a reasonable temporal sequence from administration of the interventions, follows a known or expected response pattern to the suspected intervention or interventions, but that could readily have been produced by a number of other factors.

***Unlikely Related**:* A potential relationship between one or more of the interventions and the AE could exist (i.e., the possibility cannot be excluded), but the AE is most likely explained by causes other than one or more of the interventions (e.g., could readily have been produced by the participant's clinical state or could have been due to environmental or other interventions)

Unrelated: AE is clearly not related to one or more of the interventions– another cause of the event is most plausible and/or a clinically plausible, temporal sequence is consistent with the onset of the event and the intervention administration and/or event is biologically implausible.

9.3 Documenting and Recording of Events

At each contact with the participant, information regarding adverse events will be elicited by appropriate questioning and examinations. All events, both expected/unexpected and related/unrelated will be recorded on a source document. Source documents will include: progress notes, laboratory reports, consult notes, phone call summaries, survey tools and data collection tools. Source documents will be reviewed in a timely manner by the research team. All adverse events that are identified will be recorded on the appropriate case report form (CRF) and in the study chart. The start date, stop date, severity of each reportable event, and the investigator's judgment of the AE's relationship and expectedness to the study will also be recorded on the CRF. In the event that a participant is withdrawn from the study due to an AE, it must be recorded on the CRF as such.

9.4 Adverse Event Treatment

Once an AE is known, staff at the study site should ensure that the participant receives prompt and appropriate care. Should a participant call a study clinician to report an AE, it will be determined at that time if an extra visit(s) will be scheduled, in addition to providing appropriate medical advice. All actions taken by the investigator after observing an AE should be documented, including increased monitoring of the participant, suspension of any treatment, etc. Additionally, all calls will be documented in the participant's study chart.

9.5 Expected Adverse Events

Intervention or potential cause of AE	Adverse effects
Blood drawing	**Common:** Discomfort
	Significant but rare: Hematoma, Infection, nerve damage, syncope
FDG-PET/CT	**Significant but rare:** Hematoma, thrombophlebitis, infection, nerve damage, syncope
Induced Sputum	Can have coughing, wheezing, and or bronchospasms

Table 11: Expected Adverse Events

9.6 Investigator Reporting Responsibilities to The NIH Institutional Review Board (IRB)

All reportable events will be reported by email, telephone, or fax by the participating sites to the following:

	Phone	Fax	Email	Address
Clifton E. Barry, 3rd, Ph.D.	+1-301-693-4665	+1-301-402-0993	**cbarry@niaid.nih.gov**; cc:vincentjp@nih.gov	Building 33, Room 2w20D, Bethesda, MD
Ray Chen, MD (NIAID Medically Accountable Investigator)	+1-301-443-5816	+1-301-480-5713	rchen@niaid.nih.gov;	Building 33, Room 2w20C, Bethesda, MD

Table 12: NIAID Contact information

9.7 Local Study Site Reporting

9.7.1 To Local IRB

The Local PIs have the responsibility to report AEs to their local IRB. The Local PIs also have the responsibility to report to the NIH team.

9.7.1.1 *China Team*

In the event of an unanticipated problem (UP)

1. The local principal investigator must initially report the UP and summary of the problem to the local IRB as soon as possible (usually within 24 ~ 48 hours) after awareness.

2. After the initial report, the local PI must provide a formal report to the IRB within 7 days of awareness including the investigator's judgment of harmfulness to the participant(s) and others.

9.7.1.2 *South African Team*

In the event of an unanticipated problem (UP)

1. The Local principal investigator must initially report the UP and summary of the problem to the local IRB according to the local IRB SOPs as soon as possible after awareness.

2. After the initial report, the Local PI must provide a formal report including any relevant information to the IRB as necessary, including the investigator's judgment of harmfulness to the subject(s) and others.

9.7.2 To NIH Study Team

The Local PI has the responsibility to report actual or suspected non-compliance, actual or suspected major deviations, actual or suspected Unanticipated Problems, new information that might affect the willingness of a subject to enroll or remain in the study, and suspension or termination of research activities within 7 days of awareness to allow the NIH PI time to submit to the NIH IRB.

Death of a research subject that is possibly, probably, or definitely related to the research must be reported within 24 hours of awareness.

9.8 Reporting SAEs to Health Authorities and IECs / IRBs

The site team will inform the relevant health authorities of any reportable SAEs according to the local regulatory requirements. The Investigators will be responsible for informing the IECs or IRBs that reviewed the trial protocol as per the relevant IEC/IRB's SOPs.

10 Data Management Plan - Data Collection, Sample Storage and Publication

10.1 Data Collection

Study data will be collected on standardized paper CRFs. The local study team will use only NIAID study team-approved CRFs. These forms are to be completed on an ongoing basis during the study. Any type of corrections to paper CRFs must be initialed and dated by the person making the correction. The PI is responsible for assuring that the data collected are complete, accurate, and recorded in a timely manner. The CRFs will be collected and placed into a participant-specific binder. Source documentation (the point of initial recording of a piece of data) should support the data collected on the CRF, and be signed and dated by the person recording and/or reviewing the data. Some CRFs may also be source documents. Source documents include all recordings of observations or notations of clinical activities and all reports and records necessary for the evaluation and reconstruction of the clinical trial. Source documents include, but are not limited to, the participant medical records, electronic chart records, laboratory reports, electrocardiogram (EKG) tracings, x-rays, radiologist reports, biopsy reports, ultrasound photographs, participant progress notes, pharmacy records and any other similar reports or records of procedures performed during the participant's participation in the protocol. Data for CRFs will be collected during participant visits by health care providers and abstracted from the medical record. Once the data are collected, it will be reviewed by the local site monitoring team or their contractors. Any data compiled for statistic or other manipulation will be handled in a database. The scientific results from this study will require various formats, depending on the data type. Locked copies of these files containing the results will be compiled by the research supervisor and made available to monitoring and regulatory agencies as necessary.

10.2 Data Management

Data entered onto the CRF will be transferred into the study database, which will be managed and monitored by the study team. Access to the database is password controlled

and will be limited to those with data entry and management responsibilities, as well as monitors. Records are protected by ownership control and a complete log of all activities within the system is recorded. All study related data will be maintained on servers located at the sites and NIH and image files will be stored and indexed in the system.

10.3 Data Storage

All essential documentation for all study participants including history and physical findings, laboratory data, and results of consultations are to be maintained by the investigators in a secure storage facility for a minimum of three years per NIH policies. These records are to be maintained in compliance with IRB/EC (Institutional Review Board/Ethics Committee), local and government requirements, whichever is longest. All records are to be kept confidential to the extent provided by federal, state, and local law.

It is the investigator's responsibility to retain copies of source documents until receipt of written notification to the contrary from NIH. No study document should be destroyed without prior written agreement between NIH and the Principal Investigator.

Should the investigator wish to assign the study records to another party and/or move them to another location, the investigator must provide written notification of such intent to NIH with the name of the person who will accept responsibility for the transferred records and/or their new location. NIH must be notified in writing and written NIH permission must be received by the site prior to destruction or relocation of research records.

10.4 Publication of Research Findings

Collaborating protocol team members will own the data generated by or resulting from this project, and they may arrange for publication of this original research (with consent of all study investigators) in a primary scientific journal, and for copyright by the journal unless the journal's copyright policy would preclude individuals from making or having made a single copy of any such article for their own use.

10.5 Sample Transfer and Storage

Sputum, blood, plasma, urine, and other tissue specimens will be collected at the sites and processed in the respective research labs.

Should sharing of specimens with institutions not on the protocol be desired, an amendment will be submitted to the NIH IRB and the sites' relevant IECs for review. Clinical information shared about the samples would similarly require prior IRB/IEC approval. The research use of stored, unlinked or unidentified samples may be exempt from the need for prospective IRB review and approval. Exemption requests will be submitted in writing to the NIH Office of Human Subjects Research, which is authorized to determine whether a research activity is exempt. Cultured bacterial isolates may be transferred, with the agreement of the PI, without additional IRB review only as anonymous strains with limited data including no personal identifiers.

A 20% loss of or destruction of samples will constitute a compromise of the scientific integrity of the data collected and will be reported to all IRBs. At the termination of the

NIAID study, any remaining samples will be retained or disposed of as determined by the study sites and their respective IRBs, according to relevant South African or Chinese laws.

11 Clinical Monitoring Structure

11.1 Site Monitoring Plan

The Office of Clinical Research Policy and Regulatory Operations (OCRPRO) of the Division of Clinical Research (DCR) of NIAID/NIH will contract monitors for this trial for source document verification and regulatory compliance. The OCRPRO monitoring team will discuss a detailed monitoring plan with the PI.

The study will be conducted in compliance with this protocol, International Conference on Harmonization of Good Clinical Practices (ICH GCP) and all applicable regulatory requirements. This study monitoring will be conducted according to the "NIAID Intramural Clinical Monitoring Guidelines." Monitors under contract to the NIAID/OCRPRO will visit designated clinical research sites to monitor all aspects of the study in accordance with the appropriate regulations and the approved protocol. The objectives of a monitoring visit will be as follows:

1) To verify the existence of signed informed consent documents and documentation of the ICF process for each monitored participant
2) To verify the prompt and accurate recording of all monitored data points, and prompt reporting of all SAEs
3) To compare abstracted information (CRFs, data pulls) with individual participant's records and source documents (participant's charts, laboratory analyses and test results, physicians' progress notes, nurses' notes, and any other relevant original participant information)
4) To help ensure investigators are in compliance with the protocol

The monitors will also inspect the clinical site's regulatory files to ensure that applicable regulatory requirements are being followed. During the monitoring visits the investigator, and/or designee, and other study personnel will be available to discuss the study. The investigator (and/or designee) will make study documents (e.g., consent forms, CRFs) and pertinent hospital or clinical records readily available for inspection by the local IRB/IEC, the FDA, the China FDA, the Ministry of Food and Drug Safety (MFDS), the South African Health Products Regulatory Authority (SAHPRA), the site monitors, any other relevant local health authorities, and the NIAID staff for confirmation of the study data.

A specific protocol monitoring plan will be discussed with the Principal Investigator and study staff prior to enrollment. The plan will outline the frequency of monitoring visits based on such factors as study enrollment, data collection status and regulatory obligations.

Any changes or additions to the protocol will be submitted to all necessary IRBs/IECs and relevant regulatory agencies for review. The written IRB approvals will be filed in the investigator's study binder, and a copy of the approvals will be forwarded to the monitoring team. Furthermore, essential documents will be collected in the study binders and will include the following:

1) IRB/EC approvals for the study protocol and all amendments
2) All source documents and laboratory records
3) CRF copies
4) Informed consent forms

11.2 Data and Safety Monitoring Plan

A Data and Safety Monitoring Board (DSMB) comprised of global TB experts will provide oversight on the study. The NIAID Intramural DSMB includes independent experts that do not have direct involvement in the conduct of the study and have no significant conflicts of interests as defined by NIAID policy. The board will review the study data to evaluate the safety, study progress, and conduct of the study. The DSMB will meet at least twice per year to evaluate safety, study conduct, and scientific validity and integrity of the trial. As part of this responsibility, DSMB members must be satisfied that the timeliness, completeness, and accuracy of the data submitted to them for review are sufficient for evaluation of the safety and welfare of study participants. The DSMB will also assess the performance of overall study operations and any other relevant issues, as necessary. The first DSMB review will occur after IRB approval and before the initiation of the study. Additional DSMB reviews will occur with the interim analyses. The DSMB may also convene additional reviews as necessary. All treatment-related adverse events grade 3 or higher and serious adverse event reports will be sent by the PI electronically to the DSMB members and the DSMB Executive Secretary before each meeting and within 7 days after knowledge of a relapse, or per DSMB guidelines. Serious adverse events determined to be possibly, probably, or definitely related to study medication or procedures will be reported to the DSMB at the same time they are reported to the IRB. Enrollment will continue unless the DSMB requests stopping enrollment to perform a more in-depth review. The PI will notify the board at the time pausing or halting criteria are met and obtain a recommendation concerning continuation, modification, or termination of the study. The PI will submit the written DSMB summary reports with recommendations to the IRB. The DSMB will also be available for other clinical advice as requested.

Items reviewed by the DSMB include but are not limited to:

1) SAE and AE line listings and SAE narratives
2) Demographic information on study participants
3) TB disease recurrences
4) Interim analysis of treatment success rates
5) Factors that might affect the study outcome or compromise the confidentiality of the trial data (protocol violations, unmasking, etc.)
6) Factors external to the study, such as scientific or therapeutic developments, that may adversely affect participant safety or the ethics of the study

Interim analyses will be conducted according to section 6.7. During this time, the DSMB will continue to review the safety data as scheduled.

During annual Continuing Reviews, the study team will report the number of DSMB meetings that occurred since the last continuing review and forward any reports based

upon these meetings to the NIH IRB. Reports/recommendations issued by the DSMB that necessitate any changes in the conduct of the study will be reported to the NIH IRB within 3 weeks of the PI's notification.

12 Human Subject Protection

This protocol must receive the approval of all sites' Review Boards prior to implementation. Each site may start implementing the protocol separately if that respective site and NIAID have approval, as long as local regulations are also met. The study will be conducted in accordance with the design and specific provisions of this IRB approved protocol, in accordance with the ethical principles that have their origin in the Declaration of Helsinki, and that are consistent with Good Clinical Practice (GCP) and any applicable regulatory requirement(s), as well as in accordance with the NIH policies. The PI will assure that no deviation from, or changes to the protocol will take place without prior agreement and documented approval from the IRB, except where necessary to eliminate an immediate hazard(s) to the trial participants. The PI will promptly report to the IRBs and to NIH IRB any changes in research activity and will promptly report to the IRBs all unanticipated problems involving risk to human participants, or others.

12.1 Rationale for Participant Selection

According to a World Health Organization 2015 global TB report, South Africa has the sixth highest rate of new tuberculosis (TB) cases in the world, the highest incidence of TB/HIV in the world, and the 10th highest number of drug-resistant TB. According to a WHO 2015 report, the prevalence of tuberculosis in South Africa is approximately 375,840 (696/100,000 population), but with an annual incidence of 450,000 persons per year. HIV/TB co-infected patients constitute about 2/3 of the annual TB incidence in South Africa and is typically associated with a higher burden of disease. As previously mentioned, HIV participants will be excluded in this study. Although this excludes a significant TB-infected population, our rationale in this initial study is twofold:

1) The PET noise from HIV even in absence of IRIS may substantially confound the ability to tease out drug effects on TB alone.
2) Immunologic markers may respond differently in HIV+ participants compared to HIV- participants.

The population treated by the HPCH was selected for study for several reasons: (1) China has one of the highest rates of TB infection in the world with an estimated number of 89 cases of active tuberculosis per 100,000 (as reported by the World Health Organization for 2015). Henan province has a population of ~100 million people, and therefore a large concentration of TB; (2) Existence of a well-managed, specialized hospital with highly-trained staff and experienced in conducting clinical trials.

12.2 Participation of Children and Other Vulnerable Participants

Only persons 18 years of age and older will be enrolled as this is the age of consent in both South Africa and China. It is rare for patients less than 18 years old to be hospitalized at HPCH so very few participants will be functionally excluded on this basis. In addition, it remains culturally unacceptable to expose the under-age or elderly to risk without benefit in China. Pregnant woman will be excluded because of the risks for PET/CT scanning.

12.3 Risks/Benefits Analysis including Considerations of Alternatives to Participation

12.3.1 Potential Benefits to Study Participants

Participants who stop drug early and are cured of their TB will not be exposed unnecessarily to 8 additional weeks of potentially toxic drugs. Those who receive 24 weeks of treatment receive an indirect benefit of having their TB managed closely and they therefore may have a better outcome than others who are not on the study.

12.3.2 Potential Risks to Study Participants

For those who stop drug early, there is a risk that they will not yet be cured of the TB bacteria and will relapse. We expect this risk to be small because those with the most severe disease will not stop treatment early. Only those who meet the early completion criteria will discontinue treatment early. However, if our hypothesis is incorrect and the early completion criteria do not correctly identify people who are cured after 16 weeks of treatment, their chance of relapse will be higher than if they had received a full 24 weeks of treatment. To note, risk of relapse is also related to medication adherence. Even those who would otherwise have been cured with 16 weeks of treatment may relapse if their medication adherence is poor. If relapse does occur, we will refer the participants to the local TB clinic to restart treatment. The study team will provide the TB clinic with any available drug resistance data for this participant.

There are minor risks related to blood drawing, including discomfort, hematoma, and rarely an infection. Sputum collection may also be uncomfortable and sputum induction can cause wheezing or a tightness in the airways. It is generally thought to be a safe procedure. There is a risk, although rare, that placing an IV for the FDG-PET may result in a hematoma, thrombophlebitis, infection, or nerve damage.

Rifampin is a strong inducer of the hepatic cytochrome P450 enzyme system, which is the system that metabolizes many other drugs and therefore may cause a decrease in these other drug levels. Use of these other drugs are not prohibited, however they should only be used together with caution as dose adjustments of the concomitant drug may be necessary. Drugs include but are not limited to: warfarin, saquinavir and other HIV protease inhibitors, birth control pills or other hormonal contraception, phenytoin, digoxin, clarithromycin, caspofungin, voriconazole, ketoconazole, itraconazole, diltiazem, verapamil, lorazepam, atorvastatin, rosiglitazone, and celecoxib.

There is also risk associated with radiation exposure from imaging. Radiation from CXR is negligible. CXR will not be done during the study unless Arm A fills up faster than expected. In this case, a screening CXR may be used to screen out participants who have CXR features consistent with not meeting inclusion/exclusion criteria or early treatment completion criteria. These participants will not be considered further for the study. Both PET and CT components of the scan will expose participants to ionizing radiation. We have considered ways to limit the amount of radiation exposure participants will receive. By restricting the CT scan to the chest as the area of interest rather than performing a whole body scan (which is

customary for PET/CT) we will reduce radiation exposure. We have also reduced the radiation dose for the 3rd and possible 4th scans, by decreasing the energy current. This will result in a lower resolution CT scan, but should be adequate for the more marked changes expected over a longer period. The vast majority of participants will undergo 3 PET/CT scans in a calendar year during the course of the study. A small group of participants who relapse will receive an extra scan, and therefore it is possible that this small group would receive four scans within one calendar year. The expected maximum effective dose a participant will receive from scans scheduled as part of the study over one calendar year is in the appendix for each country.

Participants involved in the MRI imaging substudy may experience nervousness and/or claustrophobia during the MRI. While generally safe, it is not known whether an MRI would harm a fetus. Pregnant women are excluded from scanning. Gadolinium-based contrast agents (GBCAs) are used with MRI to improve visualization of internal organs, blood vessels, and tissues. The use of GBCAs also carries some risk, including side effects such as allergic reactions to the contrast agent. To date, the only known adverse health effect related to gadolinium retention is a rare condition called nephrogenic systemic fibrosis (NSF) that occurs in a small subgroup of patients with pre-existing kidney failure. Therefore, participants with any evidence of renal insufficiency will be excluded from the MRI substudy. There is evidence that small amounts of gadolinium can stay in the body. It is not known how gadolinium deposition may affect the body, but so far, studies have not found harmful effects in patients who have healthy renal function. Only macrocyclic GBCAs, which cause less retention than linear GBCAs, will be administered for MRI scans.

12.4 Privacy and Confidentiality

All laboratory specimens, evaluation forms, reports, and other records that leave the site will be identified by coded number only to maintain participant confidentiality. All records will be kept locked. All computer entry and networking programs will be done with coded numbers only. Clinical information will not be released without written permission of the participant, except as necessary for monitoring by the contract monitors, IRBs, NIAID, Office for Human Research Protections (OHRP), or other regulatory agencies.

13 References

1. *Controlled clinical trial of five short-course (4-month) chemotherapy regimens in pulmonary tuberculosis. Second report of the 4th study. East African/British Medical Research Councils Study.* Am Rev Respir Dis, 1981. **123**(2): p. 165-70.
2. *Long-term follow-up of a clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis. Singapore Tuberculosis Service/British Medical Research Council.* Am Rev Respir Dis, 1986. **133**(5): p. 779-83.
3. Fox, W., *Whither short-course chemotherapy?* Br J Dis Chest, 1981. **75**(4): p. 331-57.
4. Johnson, J.L., et al., *Shortening treatment in adults with noncavitary tuberculosis and 2-month culture conversion.* Am J Respir Crit Care Med, 2009. **180**(6): p. 558-63.
5. Gillespie, S.H., et al., *Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis.* N Engl J Med, 2014. **371**(17): p. 1577-87.

6. Jindani, A., et al., *High-dose rifapentine with moxifloxacin for pulmonary tuberculosis.* N Engl J Med, 2014. **371**(17): p. 1599-608.
7. Merle, C.S., et al., *A four-month gatifloxacin-containing regimen for treating tuberculosis.* N Engl J Med, 2014. **371**(17): p. 1588-98.
8. Phillips, P.P., et al., *Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials.* BMC Med, 2016. **14**(1): p. 19.
9. Coleman, M.T., et al., *PET/CT imaging reveals a therapeutic response to oxazolidinones in macaques and humans with tuberculosis.* Sci Transl Med, 2014. **6**(265): p. 265ra167.
10. Chen, R.Y., et al., *PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis.* Sci Transl Med, 2014. **6**(265): p. 265ra166.
11. Wallis, R.S., et al., *Tuberculosis biomarkers discovery: developments, needs, and challenges.* Lancet Infect Dis, 2013. **13**(4): p. 362-72.
12. Wallis, R.S., et al., *Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice.* Lancet, 2010. **375**(9729): p. 1920-37.
13. Horne, D.J., et al., *Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis.* Lancet Infect Dis, 2010. **10**(6): p. 387-94.
14. Phillips, P.P., K. Fielding, and A.J. Nunn, *An evaluation of culture results during treatment for tuberculosis as surrogate endpoints for treatment failure and relapse.* PLoS One, 2013. **8**(5): p. e63840.
15. Benator, D., et al., *Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial.* Lancet, 2002. **360**(9332): p. 528-34.
16. Nunn, A.J., et al., *Results at 30 months of a randomised trial of two 8-month regimens for the treatment of tuberculosis.* Int J Tuberc Lung Dis, 2011. **15**(6): p. 741-5.
17. Fox, W. and I. Sutherland, *A five-year assessment of patients in a controlled trial of streptomycin, para-aminosalicylic acid, and streptomycin plus para-aminosalicylic acid, in pulmonary tuberculosis.* Q J Med, 1956. **25**(98): p. 221-43.
18. Coleman, M.T., et al., *Early Changes by (18)Fluorodeoxyglucose positron emission tomography coregistered with computed tomography predict outcome after Mycobacterium tuberculosis infection in cynomolgus macaques.* Infect Immun, 2014. **82**(6): p. 2400-4.
19. Lin, P.L., et al., *Radiologic responses in cynomolgous macaques for assessing tuberculosis chemotherapy regimens.* Antimicrob Agents Chemother, 2013.
20. Carroll, M.W., et al., *Efficacy and safety of metronidazole for pulmonary multidrug-resistant tuberculosis.* Antimicrob Agents Chemother, 2013. **57**(8): p. 3903-9.
21. Marx, F.M., et al., *The temporal dynamics of relapse and reinfection tuberculosis after successful treatment: a retrospective cohort study.* Clin Infect Dis, 2014. **58**(12): p. 1676-83.
22. Hesseling, A.C., et al., *Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients.* Int J Tuberc Lung Dis, 2010. **14**(5): p. 560-70.
23. Weiner, M., et al., *Evaluation of time to detection of Mycobacterium tuberculosis in broth culture as a determinant for end points in treatment trials.* J Clin Microbiol, 2010. **48**(12): p. 4370-6.
24. Boehme, C.C., et al., *Rapid molecular detection of tuberculosis and rifampin resistance.* N Engl J Med, 2010. **363**(11): p. 1005-15.

25. Blakemore, R., et al., *A multisite assessment of the quantitative capabilities of the Xpert MTB/RIF assay.* Am J Respir Crit Care Med, 2011. **184**(9): p. 1076-84.
26. Fennelly, K.P., *An eXpert AFB smear?* Clin Infect Dis, 2012. **54**(3): p. 389-91.
27. Theron, G., et al., *The use of an automated quantitative polymerase chain reaction (Xpert MTB/RIF) to predict the sputum smear status of tuberculosis patients.* Clin Infect Dis, 2012. **54**(3): p. 384-8.
28. Hanrahan, C.F., et al., *Xpert MTB/RIF as a measure of sputum bacillary burden. Variation by HIV status and immunosuppression.* Am J Respir Crit Care Med, 2014. **189**(11): p. 1426-34.
29. Friedrich, S.O., et al., *Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment.* Lancet Respir Med, 2013. **1**(6): p. 462-70.
30. Friedrich, S.O., et al., *Suitability of Xpert MTB/RIF and genotype MTBDRplus for patient selection for a tuberculosis clinical trial.* J Clin Microbiol, 2011. **49**(8): p. 2827-31.
31. Shenai, S., et al., *Bacterial Loads Measured by the Xpert MTB/RIF Assay as Markers of Culture Conversion and Bacteriological Cure in Pulmonary TB.* PLoS One, 2016. **11**(8): p. e0160062.
32. Brennan, A.T., et al., *The interplay between CD4 cell count, viral load suppression and duration of antiretroviral therapy on mortality in a resource-limited setting.* Trop Med Int Health, 2013. **18**(5): p. 619-31.
33. Kurbatova, E.V., et al., *Sputum culture conversion as a prognostic marker for end-of-treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis of data from two observational cohort studies.* Lancet Respir Med, 2015. **3**(3): p. 201-9.
34. Nuermberger, E.L., et al., *Moxifloxacin-containing regimens of reduced duration produce a stable cure in murine tuberculosis.* Am J Respir Crit Care Med, 2004. **170**(10): p. 1131-4.
35. Nuermberger, E.L., et al., *Moxifloxacin-containing regimen greatly reduces time to culture conversion in murine tuberculosis.* Am J Respir Crit Care Med, 2004. **169**(3): p. 421-6.
36. Wang, J.Y., et al., *Adding moxifloxacin is associated with a shorter time to culture conversion in pulmonary tuberculosis.* Int J Tuberc Lung Dis, 2010. **14**(1): p. 65-71.
37. Burman, W.J., et al., *Moxifloxacin versus ethambutol in the first 2 months of treatment for pulmonary tuberculosis.* Am J Respir Crit Care Med, 2006. **174**(3): p. 331-8.
38. Rustomjee, R., et al., *A Phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis.* Int J Tuberc Lung Dis, 2008. **12**(2): p. 128-38.
39. Conde, M.B., et al., *Moxifloxacin versus ethambutol in the initial treatment of tuberculosis: a double-blind, randomised, controlled phase II trial.* Lancet, 2009. **373**(9670): p. 1183-9.
40. Dorman, S.E., et al., *Substitution of moxifloxacin for isoniazid during intensive phase treatment of pulmonary tuberculosis.* Am J Respir Crit Care Med, 2009. **180**(3): p. 273-80.
41. Kruuner, A., M.D. Yates, and F.A. Drobniewski, *Evaluation of MGIT 960-based antimicrobial testing and determination of critical concentrations of first- and second-line antimicrobial drugs with drug-resistant clinical strains of Mycobacterium tuberculosis.* J Clin Microbiol, 2006. **44**(3): p. 811-8.
42. Laszlo, A., et al., *Conventional and radiometric drug susceptibility testing of Mycobacterium tuberculosis complex.* J Clin Microbiol, 1983. **18**(6): p. 1335-9.

43. Springer, B., et al., *Quantitative drug susceptibility testing of Mycobacterium tuberculosis by use of MGIT 960 and EpiCenter instrumentation.* J Clin Microbiol, 2009. **47**(6): p. 1773-80.
44. Lew, W., et al., *Initial drug resistance and tuberculosis treatment outcomes: systematic review and meta-analysis.* Ann Intern Med, 2008. **149**(2): p. 123-34.
45. Colangeli, R., et al., *Bacterial Factors That Predict Relapse after Tuberculosis Therapy.* N Engl J Med, 2018. **379**(9): p. 823-833.
46. Rizzi, E.B., et al., *Detection of Pulmonary tuberculosis: comparing MR imaging with HRCT.* BMC Infect Dis, 2011. **11**: p. 243.
47. Zeng, J., et al., *MRI evaluation of pulmonary lesions and lung tissue changes induced by tuberculosis.* Int J Infect Dis, 2019. **82**: p. 138-146.
48. Buzan, M.T.A., et al., *MRI as indicator of pulmonary tuberculosis activity: from morphological to molecular level assessment - a case report.* Rom J Morphol Embryol, 2017. **58**(1): p. 193-196.
49. Kurihara, Y., et al., *MRI of pulmonary nodules.* AJR Am J Roentgenol, 2014. **202**(3): p. W210-6.
50. Nunn, A.J., P.P. Phillips, and D.A. Mitchison, *Timing of relapse in short-course chemotherapy trials for tuberculosis.* Int J Tuberc Lung Dis, 2010. **14**(2): p. 241-2.
51. Johnson, J.L. and B.A. Thiel, *Time until relapse in tuberculosis treatment trials: implication for phase 3 trial design.* Am J Respir Crit Care Med, 2012. **186**(5): p. 464.
52. Moy, M.P., et al., *A new, simple method for estimating pleural effusion size on CT scans.* Chest, 2013. **143**(4): p. 1054-9.
53. Podewils, L.J., et al., *Patterns of treatment interruption among patients with multidrug-resistant TB (MDR TB) and association with interim and final treatment outcomes.* PLoS One, 2013. **8**(7): p. e70064.
54. *Guidance for Industry Pulmonary Tuberculosis: Developing Drugs for Treatment*, F. CDER, DHHS, Editor. 2013.
55. Guerra-Assuncao, J.A., et al., *Recurrence due to Relapse or Reinfection With Mycobacterium tuberculosis: A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence and Active Follow-up.* J Infect Dis, 2014.
56. Abel, L., et al., *Human genetics of tuberculosis: a long and winding road.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1645): p. 20130428.
57. FDA. *FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings.* 16 May 2018 1 October 2019]; Available from: https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-gadolinium-based-contrast-agents-gbcas-are-retained-body.
58. Mitchison, D.A., et al., *Quality control in tuberculosis bacteriology. 2. The origin of isolated positive cultures from the sputum of patients in four studies of short course chemotherapy in Africa.* Tubercle, 1980. **61**(3): p. 135-44.
59. Pasipanodya, J.G., et al., *Serum drug concentrations predictive of pulmonary tuberculosis outcomes.* J Infect Dis, 2013. **208**(9): p. 1464-73.
60. Pasipanodya, J.G., S. Srivastava, and T. Gumbo, *Meta-analysis of clinical studies supports the pharmacokinetic variability hypothesis for acquired drug resistance and failure of antituberculosis therapy.* Clin Infect Dis, 2012. **55**(2): p. 169-77.
61. Reynolds, J. and S.K. Heysell, *Understanding pharmacokinetics to improve tuberculosis treatment outcome.* Expert Opin Drug Metab Toxicol, 2014. **10**(6): p. 813-23.

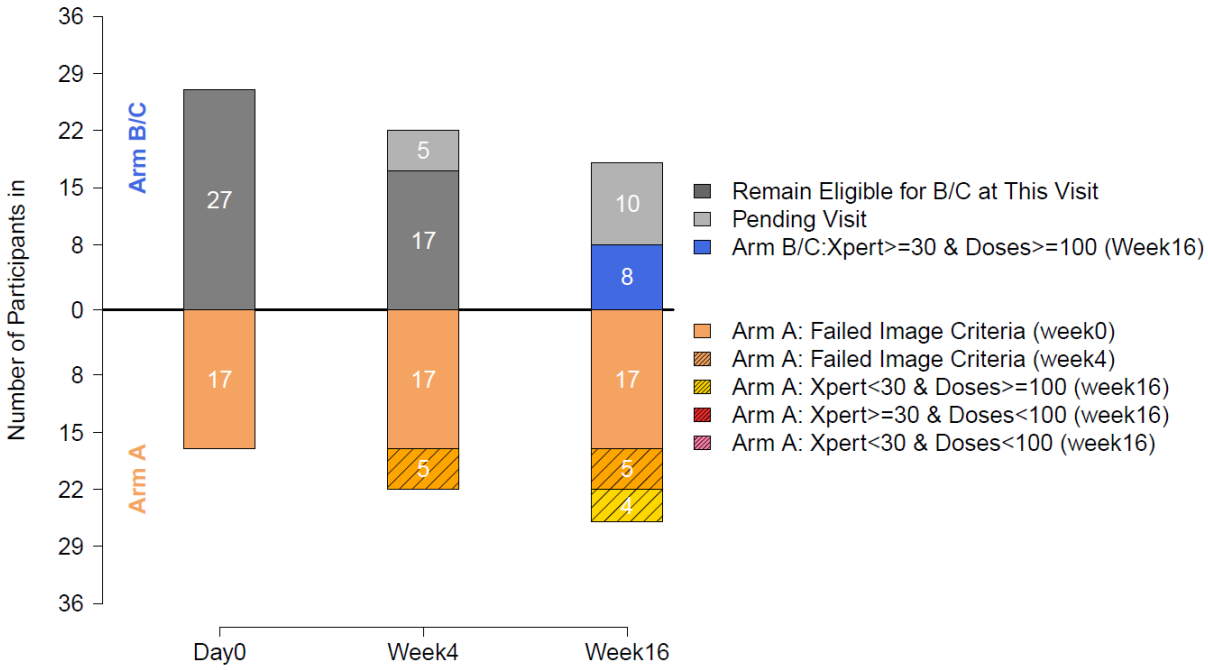
62. Marques, J.P., F.F.J. Simonis, and A.G. Webb, *Low-field MRI: An MR physics perspective.* J Magn Reson Imaging, 2019. **49**(6): p. 1528-1542.
63. USNRC, *Biological Effects of Radiation*, USNRC, Editor. 2004.

14 Appendix 1: Description of Changes to Early Treatment Completion Criteria (as of March 12, 2018)

Based on prior data as summarized in the background of the protocol (sec 1.3), an estimated 80-85% of all pulmonary drug-sensitive TB subjects are cured with 4 months of standard treatment. For the Predict TB study, we conservatively reduced this estimate to 50% of participants enrolled. The sample size calculated for the Predict TB trial is 310 participants to Arms B and C (155/arm). Based on the estimate that 50% of all participants would be cured at 4 months of treatment and therefore eligible to be randomized to Arms B and C, a total sample size of 620 was selected for the study with the understanding that the actual total would depend on the true proportion of participants randomized to Arms B and C (protocol section 3.2). If the proportion stratified to Arm A is <50%, the total sample size needed will be less than 620. If the proportion stratified to Arm A is >50%, more than 620 total participants will be needed. In addition to implications for resources and funding, an increase in total sample size raises questions about the relevance of study results. For example, if only 25% of participants are eligible for randomization, a successful study result will be much less generalizable and globally applicable because most TB patients would likely not meet such stringent requirements. Thus, we monitor carefully the proportion of participants to each arm.

Figure 1. describes the current arm distribution plot. At baseline imaging criteria evaluation, 61.4% (27/44) remain eligible for early treatment completion. At week 4, this proportion drops to 43.6% (17/39), with another 5 pending their week 4 PET/CT results. At week 16, the proportion eligible for early treatment completion falls to 23.5% (8/34), with another 10 still pending week 16 evaluation. Even if all 10 went to arms B and C, this would still only be 40.9% (18/44). We believe these data demonstrate that a significant arm imbalance exists in the study that threatens the overall scientific validity and that changes to the early treatment completion criteria are therefore justified. Changes to the criteria are needed early so that results remain valid.

Figure 1: Predict TB arm distribution plot based on current criteria (data as of March 7, 2018)



# of enrolled: N=61
# of on−study participants have not completed baseline visit: N=8
# of dropout: N=9
# of participants included in the plot: N=44(# of participants completed baseline visit minus dropout)

The current early treatment completion criteria are described below (taken from table 6 in the protocol):

Table 13: Early Treatment Completion Criteria: All early treatment criteria must be met for the participant to be eligible for randomization

Early Completion criteria:	Determined at Week 16 – unless known to have failed a radiographic criterion at baseline or week 4.
Radiographic criteria	Baseline PET/CT: • No total lung collapse of a single side, AND • No pleural effusion, AND • No single cavity air volume on CT scan >30 mL, AND • CT scan hard volume (-100 to +100 HU density) <200 mL, AND • PET total activity <1500 units Week 4 PET/CT: • All individual cavities decrease by >20% (unless cavity <2 mL), AND • CT scan hard volume does not increase by >10% unless the increase is <5 mL, AND • PET total activity does not increase by >30% unless the increase is <50 units

Bacterial load criterion	Week 16 Xpert cycle threshold ≥30
Adherence criterion	Minimum of 100 doses received by week 16

We propose making two changes to these criteria. The first is to change the week 16 Xpert MTB/RIF (Xpert) cycle threshold (Ct) cutpoint. The second is to change the baseline and week 4 radiographic criteria.

GeneXpert cycle threshold

We propose to change the Xpert Ct cutoff from Ct ≥30 to Ct ≥28. The initial criterion for Ct at week 16 was based on a cohort study in South Africa with MGIT culture, the only data available to us at the time. We adopted a stringent Ct value of 30 based upon analysis of these data for subjects to be randomized to Arms B and C. Xpert detects only DNA and does not determine the viability of detected bacilli. Note that LJ culture is the basis for outcome determination in Predict TB.

We recently received results from study TBTC-29, which collected Ct values and LJ culture. In evaluating the change, we consider the chance of missing an LJ+ result, as well as the sensitivity and specificity of various Ct cutpoints. In contrast to positive and negative predictive values, sensitivity and specificity do not depend on the underlying proportion of culture positive results, which varies over time and from study to study. That said, patient safety is a driving factor so we consider how many positive cultures might be missed for various cutpoints. This was defined as the probability of being LJ+ given a Xpert Ct value less than the threshold, i.e., P(LJ+ | Ct-). We assume what we consider are high proportions of LJ+ cultures (i.e., 10% and 5% at week 16 of treatment in a lower risk cohort of arm B/C) when making this decision. In contrast to the TBTC-29 study, which randomized all-comers and did not stratify participants by risk, the Predict TB study further excludes poorly adherent participants with too severe disease at baseline and not responding appropriately to treatment at one month. As a result, the expected LJ+ rates of 10% and 5% are thought to be on the high-side. Table 1 describes these proportions for the sensitivity and specificity estimates from TBTC-29. Based on these estimates, a Ct of 30 is expected to miss 2.1% of LJ+ results, while a threshold of 28 would miss 2.5%, assuming a 10% LJ+ rate. This translates to an increase in less than one participant being missed amongst those randomized to arm C. That is, if the background LJ+ rate is 10%, 3.3 (of 155 randomized to stop treatment at week 16) true LJ+ participants may be missed with Ct=30, and 3.9 may be missed with Ct=28. If the underlying LJ+ rate is 5%, this becomes 1.6 missed LJ+ participants with Ct=30 and 1.9 missed LJ+ participants with Ct=28. If the true underlying LJ+ rate is even lower, the difference between the two Ct values becomes even smaller.

Table 2: Sensitivity and specificity estimates from TBTC-29 for various Xpert Ct cutpoints, along with estimates of missed LJ+ and missed LJ- results for assumed (16 week) culture-positivity rates of 10% and 5%.

Xpert Ct Cutpoint	Sensitivity: P(Ct<c|LJ+)	Specificity: P(Ct>c|LJ-)	Chance of missed LJ+ P(LJ+|Ct-) with 10% LJ+ rate	Chance of missed LJ+ P(LJ+|Ct-) with 5% LJ+ rate	Chance of missed LJ+ P(LJ+|Ct-) with 2.5% LJ+ rate
Ct=31	0.93	0.43	0.018	0.008	0.004
Ct=30	0.91	0.46	0.021	0.010	0.005
Ct=29	0.89	0.49	0.024	0.012	0.006
Ct=28	0.88	0.52	0.025	0.012	0.006
Ct=27	0.86	0.55	0.028	0.013	0.006
Ct=26	0.84	0.60	0.029	0.014	0.007
Ct=25	0.79	0.66	0.034	0.016	0.008

Within Predict TB, 12 active participants have been eligible for Xpert Ct testing. (Note, the current protocol only allows for Xpert testing at week 16 amongst those not already allocated to arm A). Of the 12 participants with Xpert Ct results, 8 (67%) were negative on Xpert. All 8 participants with week 16 Xpert negative results were also LJ- on their most recent cultures. The remaining 4 had Ct values as described in Table 3.

Table 3: LJ culture results for participants with non-negative Xpert values at week 16.

Xpert	LJ results
18.2	TB+ up to week 2; other results NA
25.2	TB- at weeks 8 and 12; week 16 NA
28.4	TB- at weeks 4 and 8, other results NA
28.5	TB- at weeks 8, 12 & 16

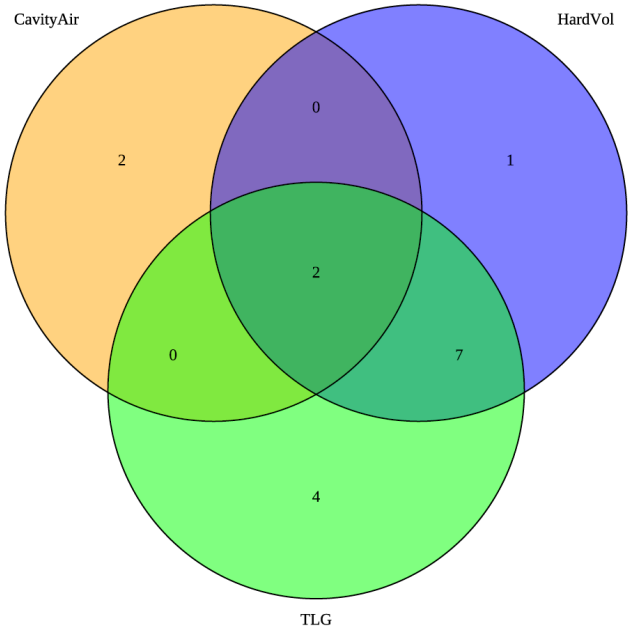
Thus changing the Xpert Ct cutpoint to 28 would have only retained an additional 2 participants in arms B and C. This change by itself would not be sufficient to correct the arm imbalance.

Radiographic criteria

To further correct the arm imbalance, we also propose to change the baseline and week 4 radiographic criteria. Our current early treatment completion criteria involve quantitating cavity air volume, CT hard volume, and PET total activity (also referred to as "total lesion glycolysis," or TLG). Prior studies (protocol sec 1.1) have validated that cavity on baseline CXR is a risk factor for treatment relapse. In our analyses of prior data, cavity size was also the strongest factor in predicting poor treatment outcome so we decided not to adjust this criterion. The data for CT hard volume and PET total activity as a risk factor for poor treatment outcomes, however, are weak. We selected these criteria based on our retrospective analysis of 92 pulmonary TB patients with baseline and week 4 PET/CT scans, as summarized in protocol sec 1.1. Figure 2 shows the

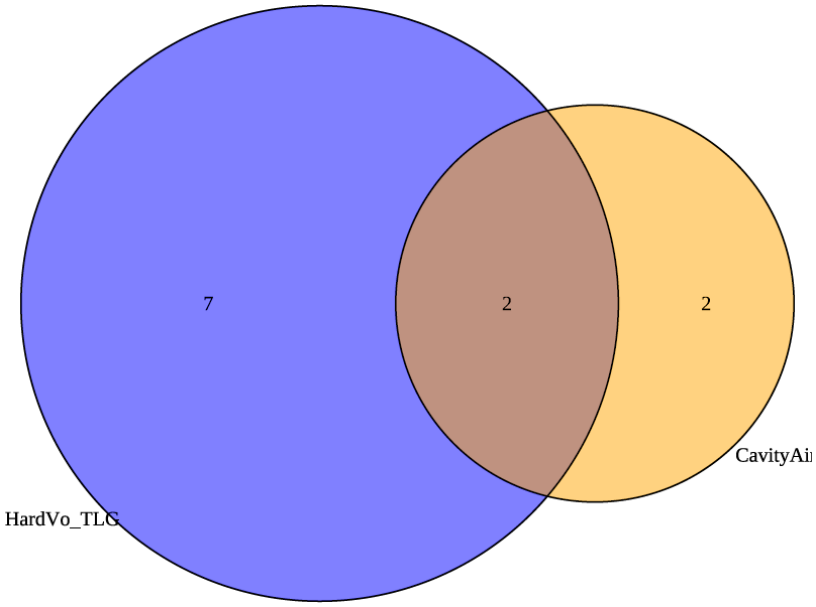
distribution of participants stratified to Arm A at baseline by the original radiology criteria. The numbers in the circles represent the number of participants that fall into arm A according to the defined criteria.

Figure 2 Venn diagram of the *original* baseline PET/CT criteria by which participants are stratified to Arm A.



The hard volume and total activity criteria are relatively well correlated in capturing participants, with only 5 participants moved to Arm A based on a single criterion, hard volume or PET total lesion glycolysis. Thus, instead of arbitrarily increasing the hard volume and PET total activity cutpoints, we decided to change the criteria from requiring both hard volume **AND** total activity to be below the thresholds to be considered **low risk,** to only requiring 1 criterion. That is, participants that have either hard volume **OR** total activity below the threshold at both baseline and week 4 will be considered low risk. The thresholds themselves will not change. Applying this change to the PET/CT criteria results in the following revised Venn diagram, which is the same as Figure 2 except for the 5 participants moved to Arm A based on hard volume or PET activity alone are no longer considered high risk.

Figure 3: Venn diagram of the *revised* baseline PET/CT criteria by which participants are stratified to Arm A



Proposed new early treatment completion criteria

The proposed revised early treatment completion criteria incorporating both Xpert Ct and radiographic criteria changes are:

Table 14: Proposed *Revised* Early Treatment Completion Criteria: All early treatment criteria must be met for the participant to be eligible for randomization (changes highlighted)

Early Completion criteria:	Determined at Week 16 – unless known to have failed a radiographic criterion at baseline or week 4.
Radiographic criteria	Baseline PET/CT: • No total lung collapse of a single side, AND • No pleural effusion, AND • No single cavity air volume on CT scan >30 mL, AND • CT scan hard volume (-100 to +100 HU density) <200 mL **OR** PET total activity <1500 units Week 4 PET/CT: • All individual cavities decrease by >20% (unless cavity <2 mL), AND • CT scan hard volume does not increase by >10% unless the increase is <5 mL **OR** PET total activity does not increase by >30% unless the increase is <50 units
Bacterial load criterion	Week 16 Xpert cycle threshold **≥28**
Adherence criterion	Minimum of 100 doses received by week 16

The original early treatment completion criteria, when applied to the data from an earlier study of 92 pulmonary TB patients from Cape Town, yielded the following stratification:

Table 5: Treatment outcome by *original* Predict TB early treatment completion risk stratification

	Treatment Outcome (Numbers of Participants)			
Risk Categorization	Cure	Failure	Programmatic Treatment Restart	Total
Low Risk	40	1	6	47
High Risk	33	7	5	45
Total	73	8	11	92

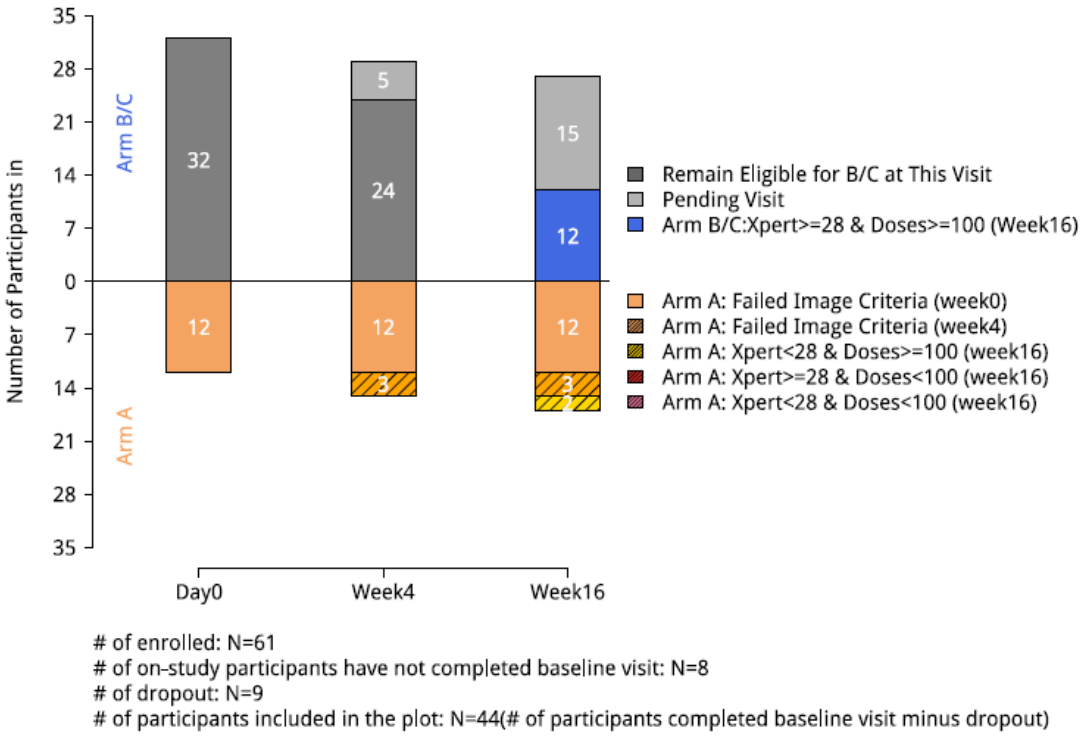
However, we know that our current Predict TB population is different from these earlier participants as more than half of these earlier participants were stratified as low risk (Arms B or C), in contrast with our current Predict TB participants. Applying the revised criteria to this earlier dataset results in the following new stratification:

Table 6: Treatment outcome by *revised* Predict TB early treatment outcome risk stratification

	Treatment Outcome (Numbers of Participants)			
Baseline Risk Categorization	Cure	Failure	Programmatic Treatment Restart	Total
Low Risk	55	2	8	65
High Risk	18	6	3	27
Total	73	8	11	92

Applying the revised criteria to the current active Predict TB participants results in the following arm distribution:

Figure 4: Predict TB arm distribution plot based on *revised* criteria (data as of March 7, 2018)



# of enrolled: N=61
# of on-study participants have not completed baseline visit: N=8
# of dropout: N=9
# of participants included in the plot: N=44(# of participants completed baseline visit minus dropout)

Using the revised baseline imaging criteria evaluation, 72.7% (32/44) remain eligible for early treatment completion. At week 4, this proportion drops to 61.5% (24/39) with another 5 pending their week 4 PET/CT results. At week 16, the proportion eligible for early treatment completion falls to 41.4% (12/29), with another 15 pending week 16 evaluation. Note that the additional 5 "pending" participants compared to the original criteria arm distribution plot at week 16 (Figure 1) are original Arm A participants who now moved to Arm B/C but did not get a week 16 Xpert Ct, which was not done in Arm A. The Xpert Ct evaluation of the 15 pending participants will determine the final arm stratification outcome, as there are currently no adherence issues. If 2/3 (10/15) achieve Ct ≥28, that will result in a 50/50 split. Currently, 12/14 (85.7%) participants had a Ct ≥28 at week 16. If this proportion holds, this proposed change will be sufficient to re-balance the arm distribution over time.

Finally, note that the currently randomized participants will not be affected by this change. Only participants before their week 16 randomization visit will be eligible to be stratified by these revised criteria, once all regulatory approvals have been obtained.

15 Appendix 2: Radiation Dosimetry

The FDG-PET/CT scans in South Africa and China will be performed on approved scanners using settings approved by the NIH Radiation Safety Committee to allow for a maximum of 4.5 rem per participant for 4 PET/CT scans and 1 CXR in a 12-month period. This radiation dose is

less than the maximal permissible annual research exposure of 5 rem/yr [63], however most participants will only receive 3 PET/CT scans. The NIH Radiation Safety Committee has reviewed the use of radiation in this research study and has approved this use as involving slightly greater than minimal risk and necessary to obtain the research information desired.

16 Changes to study procedures per Data Safety Monitoring Board recommendations

The NIAID Data Safety Monitoring Board conducted its 7th interim review of unblinded data on September 11, 2020. Following this review, the DSMB recommended halting randomization to Arms B and C. This recommendation is based on the result of the interim analysis in section 6.7.1 of the study protocol, when 1/3 of randomized participants have been followed for 72 weeks from study entry. The protocol stopping guideline for inferiority of the treatment shortening arm was met. The DSMB was also presented with conditional power calculations relating to the futility interim analysis; although the protocol-specified time for this analysis had not been reached, the results of the analysis were consistent with a decision to terminate randomization into Arms B and C, but to continue enrollment into Arms A and B at the discretion of the investigators. Pre-emptive re-treatment of participants who have completed the course of treatment in Arm C is not recommended and should be left to the discretion of the treating clinician.

17 NIH/SOUTH AFRICA Specific Appendices

In South Africa a saliva sample will be collected at both week 16 and week 24 visits for biomarkers irrespective of the PET/CT scan visit.

17.1 Remuneration in RSA

Study participants will receive the following compensation for each type of visit. If possible, the participants will be paid by SMS.

Screening:	150 Rand (approximately $12)
Each Study Visit:	150 Rand (approximately $12)
Each PET/CT:	350 Rand (approximately $25)
Extra sputum visit:	50 Rand (approximately $4); SATVI site: 150 Rand (approximately $12)
Each MRI scan:	350 Rand (approximately $25)

Travel expenses incurred will be paid according to the site's guidelines. Participants may receive further incentives, such as a phone card, as mentioned in section 5.2.

18 NIH/CHINA Specific Appendices

18.1 Blood volumes in China

Due to concerns about total blood volumes collected in China, the Chinese sites will not participate in all blood immunological marker analyses. Per section 5.7.2, China sites will NOT collect whole blood for FACS (1 mL) nor whole blood for PBMC isolation (30 mL). They will collect whole blood for serum (8 mL), blood for host mRNA (2.5 mL), and blood for host DNA (2 mL) for a total of approximately 12.5 mL whole blood at any one visit. Note

that the 2 mL blood for plasma drug levels (sec 5.7.3) occurs only at week 20, when blood biomarkers are not collected.

18.2 **Remuneration in China**

Study participants will receive the following compensation for each type of visit:

Screening:	50 Yuan (approximately $8)
Each Study Visit:	80 Yuan for visits D0-W12, 100 Yuan for W16 (plus 100 if they have completed every visit up to this time point), 20, 36, and 48, 300 Yuan for week 24, and 350 Yuan for week 72 (plus 350 Yuan if they have completed every visit up to this time point). At recurrence, participants will receive 100 Yuan.
Each PET/CT:	100 Yuan (approximately $15)
Each MRI scan:	100 Yuan (approximately $15)

Travel expenses incurred will be paid according to the site's guidelines. Participants may receive further incentives, such as a phone card, as mentioned in section 5.2.