**Title**
Model-based tools to improve treatment of tuberculosis patients



**Author**
Imperial, Marjorie Zamora

**Publication Date**
2020

Peer reviewed|Thesis/dissertation

Model-based tools to improve treatment of tuberculosis patients

by
Marjorie Zamora Imperial

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Rada Savic
7EC091D1336044E...

Rada Savic
Chair

DocuSigned by:
Payam Nahid
DocuSigned by:
Patrick Phillips
F39BACE611F44E6...

Payam Nahid

Patrick Phillips

Committee Members

**Acknowledgements**

First, I would like to thank all the study participants and volunteers, the staff at clinical sites, and all investigators for each of the clinical trials that are the fundamental foundation to this dissertation. No analysis would be possible without your contributions.

My deepest gratitude to my mentor and adviser, Dr. Rada Savic, for giving the me opportunity to do research and providing invaluable guidance throughout my graduate career. Her enthusiasm and intellectual curiosity have deeply inspired and motivated me. I feel honored to have had the opportunity to learn from such a well-respected leader in her field and am extremely grateful for what she has offered me. She has challenged me in many ways that has made me into a more confident researcher. I also thank her for her friendship and support at the personal level while I started a family during my graduate studies.

I would also like to offer my sincere appreciation for the insightful comments, constructive criticism, and continuous support at all stages of my research and career provided by the members of my dissertation committee, Drs. Payam Nahid and Patrick Phillips. They have acted as a second set of mentors and advisors for my research and have provided invaluable insights on the best way to present and frame my research projects and dissertation. Thank you to Rada, Payam and Patrick for trusting me with such impactful research and the opportunity and invitations to be part of the TB research community. I would also like to thank other members of my qualifying exam for their valuable expertise, Drs. Mark Segal, Susan Swindells and David Hermann. I would also like to thank Drs. Deanna Kroetz and Nadav Ahituv for their support scientifically and professionally.

I have had the honor of working with numerous collaborators. The work presented here and at other venues would not have been accomplished without their contributions. I would like to thank our collaborators at the Critical Path Institute (Klaus Romero, Bob Stafford, and Dan Hartley) for preparing the TBReFLECT database and answering all our data questions. I would

also like to thank the TBReFLECT steering committee and Drs. Gerry Davies, Katherine Fielding, Debra Hanna, David Hermann, Robert Wallis, John L. Johnson, Christian Lienhardt, Carl Mendel, James Neaton, and Andrew Vernon for their review of the manuscript published for Chapter 1. Thank you to Drs. Jerry Nedelman, Dan Everitt and Eugene Sun for their help on planning, executing, and publishing/presenting the work in Chapter 5. Thank you to Elly Trepman for editorial support on Chapter 5. I would also like to thank Dr. Christian Lienhardt for the invitation to a WHO workshop as a technical writer. I learned so much from joining experts in the field. Thank you to Drs. Susan Swindells and Kelly Dooley for always making me feel included and the invitations to dinner when I joined Rada at international conferences and workshops. I would also like to thank the Pharmacometrics group at Eli Lilly for the amazing internship opportunity. Thank you to Nieves, Gary, Johan, Mannie, Brian W., Evan, Brian M. and Lisa. I would also like to thank my mentors at Seattle Genetics and UW, Drs. Tae Han, Jash Unadkat, and Isabelle Ragueneau-Majlessi, for introducing me to clinical pharmacology and for giving me a chance and sparking such a strong interest to the field.

Completion of this dissertation could not have been accomplished without the support of the Savic Lab. I would like to thank all former and current lab members for the stimulating discussions, helping me get through many parts of my projects and the fun coffee breaks and lab outings, especially the lab scavenger hunt around the city and our trip to Angel Island. I also thank them for the gifts and the celebrations planned for me and my family. I would also like to thank the Clinical TB Team for being a sounding board every day and helping me work through some of the toughest challenges in our projects while also helping me realize that some issues are trivial. Thank you to Vincent for the many simulations, insightful questions and all his work on Chapter 3. Thank you to Emma and my PSPG lab mates for teaching me the valuable skills of being a mentor. I hope they were able to gain something from me because I learned so much from them. I

would also like to thank Craig for making sure my deliverables are met, keeping me accountable, and managing all the exciting projects in lab.

I would like to thank all of my PSPG classmates for their support and fun memories over the last five years. A special thank you to Emmalyn, Liz, and Kat for always being there to have quick or long chats over science, crafts, and life. I am grateful for their support through some of the toughest situations and always lending a helping hand when I needed it.

I would also like to thank all my friends back home. My time spent with them was always a breath of fresh air from the stressful times in graduate school. TBNS.

To my Imperial and Zamora family, thank you all for always encouraging and supporting me in all of my pursuits since I was young and inspiring me to always follow my dreams. Words cannot express how grateful and appreciative I am to my amazing parents for their love, prayers, and sacrifices for educating and preparing me for my future. I would like to thank my dad for always dropping me off and picking me up from school, from kindergarten at Camas to graduate school in SF, and my mom for always believing in me and wanting the best for me. Thank you to my brothers, Junior and Marwin, for always being there for me. Our bond is unbreakable and I thank them for the many laughs we have with our parents. Thank you to my in-laws and the Gutierrez family for allowing me to take over their dining room during this pandemic and babysitting DJ these last few months as I write this dissertation. I am forever grateful, gracias.

Finally, to my very caring, loving, and understanding husband, David, my sincere and heartfelt thanks. He has held the fort down throughout my time in graduate school. He has always been patient with me and a supportive sounding board while I ramble about lab and research ideas. His sacrifices as I complete this dissertation have not gone unnoticed. And lastly, my baby boy DJ is the light of my life and has given me the extra strength and motivation with his giddy smiles throughout the day. This dissertation is dedicated to him.

**Contributions**

Chapter 1 was modified from "A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis" as it was published in *Nature Medicine* in 2018. Christian Lienhardt had the initial idea for the meta-analysis. Marjorie Imperial and Rada Savic contributed to the study design. Debra Hanna and John. L. Johnson provided the data for analysis. Patrick Phillips and Katherine Fielding provided data management and statistical support. Marjorie Imperial executed the analysis. All authors discussed results and implications. Marjorie Imperial and Payam Nahid wrote the manuscript with contributions from Rada Savic and Patrick Phillips and all other authors.

Chapter 2 was modified from a manuscript in preparation "Predicting optimal treatment durations for tuberculosis patients: a risk stratification algorithm and clinical simulation tool". Marjorie Imperial and Rada Savic contributed to study design. Execution and analysis were carried out by Marjorie Imperial. Marjorie Imperial wrote the manuscript with contributions from Payam Nahid, Patrick Phillips, and Rada Savic.

Chapter 3 was modified from a manuscript in preparation "Development and application of an Integrated Biomarker Clinical Endpoint Tool for Late Stage Tuberculosis Regimen Development and Innovative Clinical Trial Designs". Marjorie Imperial, Patrick Phillips, and Rada Savic contributed to study design. Execution and analyses were carried out by Marjorie Imperial and Vincent Chang. Marjorie Imperial wrote the manuscript with contributions from Patrick Phillips and Rada Savic.

Chapter 4 was modified from a manuscript in preparation "Identification of novel proteomic signatures as predictors of tuberculosis treatment response". Marjorie Imperial and Rada Savic contributed to study design. Execution and analyses were carried out by Marjorie Imperial. Proteomic assays were performed at Meso Scale Diagnostics, LLC. by George Sigal and

colleagues. Marjorie Imperial wrote the manuscript with contributions from Payam Nahid and Rada Savic.

Chapter 5 was modified from a manuscript in preparation “Linezolid dosing strategies to minimize adverse events with high-dose, long-term treatment for extensively drug-resistant tuberculosis”. Marjorie Imperial, Jerry Nedelman, and Rada Savic contributed to study design. Execution and analyses were carried out by Marjorie Imperial. Marjorie Imperial wrote the manuscript with contributions from Jerry Nedelman and Rada Savic.

# Model-based tools to improve treatment of tuberculosis patients

Marjorie Zamora Imperial

**Abstract**

Tuberculosis (TB) infects 10 million people each year and kills more than any other infectious disease. All current approaches to TB treatment are based on a one-size-fits-all approach, which leads to undertreatment of patients with severe forms of disease and entails unnecessarily long treatment with potential toxicities for many patients in whom the disease is less severe. Shorter, efficacious, and better-tolerated oral regimens for TB are needed. Unfortunately, all recent Phase 3 clinical trials aimed to shorten treatment duration from 6 months to 4 months for drug susceptible TB failed. The aims of this dissertation were to quantitively characterize treatment response in TB patients and develop model-based tools that provide informed recommendations on optimal treatment regimens and strategies that: i.) maximize durable cure in all patients, ii.) maximize success of late stage regimen development, and iii.) minimize safety concerns associated with a highly potent, but toxic, high-dose linezolid-containing regimen.

In a patient-level pooled analysis of all recent Phase 3 clinical trials evaluating shorter treatments for drug susceptible TB , survival analysis identified risk factors of treatment outcomes. Based on these risk factors, a risk stratification algorithm and clinical simulation tool were developed to provide more individualized predictions of optimal treatment regimens to achieve high cure rates in TB patients.

TB regimen development is plagued with many challenges, the most serious being the inability to identify optimal regimens early and efficiently. To facilitate decisions on novel TB regimens that move forward through the development process, an integrated model was developed to describe the translational link between Phase 2 intermediate biomarkers, treatment characteristics, and patient risk factors to Phase 3 clinical outcomes. We provide clinical trial simulation tools to

design innovative clinical trial designs that permit evidence-based decisions on moving the best regimens forward in late stage clinical development .

The TB regimen development process is also challenged by the lack of reliable, quantitative, non-culture-based biomarkers that inform individual level and trial level treatment response. We showed that leveraging longitudinal sputum culture results and drug exposure and applying advanced nonlinear mixed effect modeling with machine learning approaches offer insights into the response dynamics following anti-TB treatment. We identified candidate proteomic signatures that can potentially predict treatment response.

Lastly, using modeling and simulation approaches, we quantified the pharmacokinetic-toxicodynamic relationship of a high-dose linezolid-containing regimen for extensively drug resistant TB. We provide practical data-driven recommendations about linezolid dosing adjustments to optimize therapeutic effects and minimize adverse events.

The quantitative, model-based tools developed in this dissertation contributes to providing evidence-based recommendations on optimal treatment strategies for current and novel TB regimens.

**Table of Contents**

## List of Figures

## List of Tables

**List of Abbreviations**

| | |
|---|---|
| ART | antiretroviral therapy |
| ATS | American Thoracic Society |
| AUC | area under the concentration time curve over 24 hours |
| AUROC | area under the receiver operating characteristic curve |
| BID | twice daily (Latin bis in die) |
| BMI | body mass index |
| BPaL | Bedaquiline, pretomanid, linezolid |
| CART | classification and regression tree |
| CDC | Center for Disease Control and Prevention |
| CDISC | Clinical Data Interchange Standard Consortium |
| CI | confidence interval |
| $C_{max}$ | maximum concentration |
| $C_{min}$ | minimum concentration |
| CTB2 | Consortium for Tuberculosis Biomarkers |
| DMID | Division of Microbiology and Infectious Disease |
| DS-TB | drug susceptible tuberculosis |
| ECL | electrochemiluminescence |
| fAUC | free area under the concentration time curve over 24 hours |
| FDA | Food and Drug Administration |
| Hb | hemoglobin |
| HIV | human immunodeficiency virus |
| HR | hazard ratio |
| IC50 | half maximal inhibitory concentration |
| IDSA | Infectious Diseases Society of America |
| LJ | Lowenstein-Jensen |
| LLOQ | Lower limit of quantification |
| MDR-TB | multi-drug resistant tuberculosis |
| MGIT | Mycobacteria Growth Indicator Tube |
| MIC | minimum inhibitory concentration |
| MIRU | mycobacterial interspersed repetitive unit |
| MITT | modified intent-to-treat |
| MSD | Meso Scale Diagnostics, LLC |
| NIAID | National Institute of Allergy and Infectious Diseases |
| NIH | National Institute of Health |
| NONMEM | Nonlinear Mixed Effect Modeling |
| OFV | objective function value |
| PET-CT | Positron Emission Tomography-Computed Technology |
| PI | prediction interval |

| | |
|---|---|
| PK | pharmacokinetics |
| QD | once daily (Latin quaque die) |
| ROC | receiver operating characteristic curve |
| RSE | relative standard error |
| Rx | treatment |
| TB | tuberculosis |
| TB-PACTS | Platform for Aggregation of Clinical Tuberculosis Studies |
| TBTC | Tuberculosis Trials Consortium |
| TI/NR | treatment intolerant or non-responsive |
| TMIC | time above minimum inhibitory concentration |
| VPC | visual predictive check |
| WHO | World Health Organization |
| XDR-TB | extensively drug resistant tuberculosis |

# Chapter 1: A patient-level pooled analysis of treatment-shortening regimens for drug susceptible pulmonary tuberculosis*


## Abstract

Tuberculosis (TB) kills more people than any other infectious disease. Three pivotal trials testing 4-month regimens failed to meet non-inferiority margins; however, approximately four-fifths of participants were cured. Through a pooled analysis of patient-level data with external validation, we identify populations eligible for 4-month treatment, define phenotypes that are hard to treat and evaluate the impact of adherence and dosing strategy on outcomes. In 3,405 participants included in analyses, baseline smear grade of 3+ relative to <2+, HIV seropositivity and adherence of ≤90% were significant risk factors for unfavorable outcome. Four-month regimens were non-inferior in participants with minimal disease defined by <2+ sputum smear grade or non-cavitary disease. A hard-to-treat phenotype, defined by high smear grades and cavitation, may require durations >6 months to cure all. Regimen duration can be selected in order to improve outcomes, providing a stratified medicine approach as an alternative to the 'one-size-fits-all' treatment currently used worldwide.


* Modified from the publication: Imperial, MZ, *et.al* A patient-pooled analysis of treatment-shortening regimens for drug susceptible pulmonary tuberculosis. *Nature Medicine* 2018; **24:**1708-15.

## Introduction

Three recent international randomized Phase 3 trials evaluating 4-month fluoroquinolone-containing regimens in adults with pulmonary, drug-susceptible tuberculosis (TB) failed to achieve non-inferiority as compared to the standard 6-month control regimen (OFLOTUB[1], ClinicalTrials.gov number, NCT00216385; REMoxTB[2], NCT00864383; RIFAQUIN[3], ISRCTN number, 44153044). These trials evaluated later-generation fluoroquinolones (gatifloxacin and moxifloxacin), as single substitutions for ethambutol or isoniazid in multi-drug regimens with the objective of shortening treatment duration from six to four months. In each of the three trials, the 4-month regimen did not satisfy the criteria for non-inferiority. However, the experimental four-month regimens did cure approximately four-fifths of the participants, suggesting that a large proportion of global TB cases could be successfully treated with shorter duration.[1-3]

Since the introduction of highly effective rifampin-based regimens in the 1970s and 1980s, the treatment of TB has been a "one-size-fits-all" paradigm, with a 6-month regimen comprised of four drugs (isoniazid, rifampin, pyrazinamide and ethambutol) used for all patients with drug-susceptible pulmonary TB.[4,5] Regimen administration is coupled with various adherence interventions at the programmatic level, including directly observed therapy, to ensure regimen intake.[4] In programs, the one-size-fits-all paradigm leads to undertreatment of patients with severe forms of disease, and entails unnecessarily long treatment with potential toxicities for many patients in whom there is a lower disease burden, which in turn may result in increased rates of loss to follow-up.[6] In clinical trials, one-size-fits-all experimental regimens have been consistently inadequate to cure the hardest-to-treat TB patients indicating that treatment duration is a critical determinant for cure.[7] Moreover, even for the standard 6-month regimen, based on the recent trials 5-8% of patients fail treatment or relapse, and 15-20% experience composite unfavorable outcomes.[1-3,8] TB is not a uniform clinical entity, and presents with wide variation in severity of

disease at the time of diagnosis. Yet, current TB regimen development efforts are aimed at using new drugs with increased potency to identify shorter treatments for all patients, regardless of severity of disease. This approach places otherwise efficacious drugs and regimens at risk of being abandoned, consequently impeding the identification of new TB regimens that are curative if used with greater precision.

In this pooled analysis of individual patient datasets from these high-quality, contemporary trials, we sought to identify characteristics of those patients who were cured with 4-month regimens and conversely those with hard-to-treat phenotypes of TB who might require longer treatment durations. We evaluated both baseline characteristics as well as on-treatment markers of risk, including dosing frequency and adherence, for their ability to stratify the study population into easy- or hard-to-treat phenotypes of TB.

## Methods

### Study design

This study utilized individual patient data from four recent, international, randomized Phase 3 trials (OFLOTUB[1], REMoxTB[2], RIFAQUIN[3] and DMID 01-009[9]) that compared 4-month regimens to standard 6-month WHO and ATS/CDC/IDSA endorsed regimens for drug-susceptible pulmonary TB.[4,5] The OFLOTUB trial compared an experimental 4-month gatifloxacin-based regimen to a 6-month standard regimen.[1] The REMoxTB trial compared two experimental 4-month moxifloxacin-based regimens to a 6-month standard regimen.[2] The RIFAQUIN trial compared experimental 4-month or 6-month moxifloxacin- and high-dose rifapentine-based intermittent regimens to a 6-month standard regimen.[3] A fourth independent TB treatment-shortening trial sponsored by the National Institute of Allergy and Infectious Diseases (NIAID) and conducted by the NIH-funded Tuberculosis Research Unit compared a 4-month

standard regimen (with no fluoroquinolone) to a 6-month standard regimen in adults with non-cavitary disease and 2-month culture negative status (DMID 01-009, NCT00130247).[9] The pooled analyses are focused on data from participants receiving the 4-month experimental regimens and 6-month standard regimens, and do not include the once-weekly (in continuation phase) fluoroquinolone 6-month experimental regimen in the RIFAQUIN trial. The three trials that compared four fluoroquinolone-based tuberculosis regimens to a 6-month standard regimen provided data for identifying markers and models for risk stratification, while the DMID 01-009 trial data were used for external validation. We defined the experimental group as all study participants allocated to any of the 4-month experimental regimens and the control group as all study participants allocated to the 6-month standard regimen. The protocol for each study was reviewed and approved by ethics committees and regulatory committees described in the original publications and all patients provided written informed consent.[1-3,9]

**Data acquisition, management and harmonization**

Integrated and standardized individual level data in each of the trials were obtained through the Platform for Aggregation of Clinical TB Studies (TB-PACTS; https://c-path.org/programs/tb-pacts/). Data sharing was directed by comprehensive Data Contribution Agreements with sponsors. Before data were pooled, we compared trial protocols, case report forms, and data dictionaries to harmonize databases. Data queries were resolved through direct consultations with each trial team and Critical Path data managers. After pooling data, data inputs were checked for missing or duplicated values, for consistency and plausibility. Final dataset specification is available in Appendix Table A.1 and Appendix Table A.2, and access to original databases is available through TB-PACTS. Data from the DMID 01-009 were obtained directly from the sponsor.

**Efficacy outcomes**

The primary efficacy endpoint of the pooled analysis was time to an unfavorable outcome for a maximum of 24 months after start of treatment (participants in the OFLOTUB study were followed until 24 months after start of treatment and RIFAQUIN and REMoxTB for 18 months), as defined according to each trial protocol and described in the original publications. Trial-specific definitions of unfavorable outcome were broadly similar but included some differences, which are outlined in Appendix Table A.1. For example, reinfections confirmed by mycobacterial interspersed repetitive unit (MIRU) typing were excluded from the composite definition of unfavorable outcome in the primary analysis of the REMoxTB and RIFAQUIN trials, whereas they were included in the composite definition of unfavorable outcome in the primary analysis of the OFLOTUB trial. The secondary efficacy outcome was the non-parametric Kaplan-Meier estimate of unfavorable outcome at 24 months after start of treatment.

**Baseline predictors**

The primary analysis set included baseline predictors, which were missing in no more than 10% of participants: age, race, body mass index (BMI), sex, presence of cavitation on chest radiograph and smear grade (Appendix Table A.2). Weight was also considered for inclusion in the primary analysis but ultimately was not included due to its moderate correlation with body mass index (BMI, Spearman coefficient 0.74). No major covariate imputation was done, with two exceptions: (1) black race was assigned for all participants in the OFLOTUB trial, in which race information was not available, given that all OFLOTUB sites were in Africa and similar demographic characteristics were observed in other studies at their African sites (majority black); (2) median height for females and males of available data was used for 291 participants with missing height to calculate BMI, defined as the weight in kilograms divided by the squared height in meters (additional details available in Appendix Table A.2). Smear grading was specific for

each microscopy method, each study, and, in the RIFAQUIN trial, each study center; described in study protocols and lab manuals.[1-3,9] RIFAQUIN and OFLOTUB trials reported smear grade using a negative, 1+, 2+, and 3+ system, while REMoxTB and the validation study reported smear grade using a 1+, 2+, 3+ and 4+ system. A conversion chart available in the REMoxTB trial lab manual was used to synchronize all smear data to the same grading scale.[2] Additional patient characteristics (smoking, cough grade, other radiographic measures) were considered but not included in the primary analysis due to large proportions of missing data (>10%, Appendix Table A.2).

**On-treatment predictors**

On-treatment culture time-point universally applied in all trials was month-two culture status on Lowenstein-Jensen (LJ) solid medium or in liquid medium using the Mycobacteria Growth Indicator Tube (MGIT) system. Culture positivity on either media was used for analyses, with preference for solid culture if available. Univariate Cox proportional hazard analysis for merged MGIT and LJ culture data (as described above), MGIT data only, and LJ data only showed similar results in each treatment group (Appendix Table A.3). Treatment adherence was calculated as the number of days that doses were taken divided by the prescribed number of days. For participants with an unfavorable event during the treatment phase, the adherence calculation was adjusted for duration completed, e.g. full adherence was assigned for study participants who took all doses up to time of event, if the event appeared during treatment.

Individuals with missing data between the predefined sets of predictors were excluded from the multivariate analysis. There were no major correlations between predefined set of baseline and on-treatment predictors.

**Statistical analysis**

All analyses were conducted using modified intent-to-treat (MITT) and per-protocol populations, with the former used for primary analysis (per-protocol results available in published manuscript). Definitions for analysis populations are provided in the clinical trial protocols.[1-3,9]

To identify risk factors of time to unfavorable outcomes, we performed multivariate Cox proportional hazards analysis. Hazard ratios with 95% Wald confidence intervals were reported. Analyses were conducted separately for the experimental and control regimens as the hard-to-treat phenotypes may be different for different treatment durations. All multivariate analyses were adjusted for study country. The proportional hazard assumption was tested using Schoenfeld residuals, with a $p < 0.05$ for non-proportionality. Model selection for multivariate Cox analysis started with a full model (included all predefined predictors) that was followed by a backward stepwise approach ($p > 0.05$ to remove) then a forward stepwise approach to test predictors that were removed in the backward step ($p < 0.01$ to include). Predictors were included using linear relationships. Non-inferiority analyses were performed in study participant subgroups, according to identified risk factors in the multivariate Cox analysis. The test for interaction for each subgroup was performed prior to non-inferiority sub-group tests.[10] The absolute difference in proportion of unfavorable outcomes was calculated using inverse probability study-weighted Kaplan-Meier estimates at 24 months after start of treatment to include maximal patient-years of follow up and retain maximal data.[11] Non-inferiority was assessed using the upper bound of the two-sided 90% confidence interval, determined by bootstrapping 500 samples, and a non-inferiority margin of 6 percentage points, which was used in all the parent trials.[1-3]

Further analyses were performed to assess impact of 7/7 (REMoxTB and RIFAQUIN) and 6/7 weekly (OFLOTUB) dosing strategies on outcomes. First, we compared Kaplan-Meier estimates for 7/7 and 6/7 weekly dosing strategies in study participants who completed their

prescribed treatment. Second, we performed separate Cox proportional hazards analyses for trials with different weekly dosing strategies and assessed total number of days that the drugs were taken (total doses) and treatment duration (time between first and last dose dates) as predictors of treatment outcomes. To allow for pragmatic interpretation, hazard ratios were reported for total doses of 156 to 181 (on average 6/7 doses per week) and 112 to 155 (on average 5/7 doses per week) relative to 182 (on average 7/7 doses per week) for the REMoxTB and RIFAQUIN analysis (7/7 weekly dosing strategies for 26 weeks). For the OFLOTUB analysis (6/7 weekly dosing strategy for 24 weeks), hazard ratios were reported for total doses of 112 to 143 (on average 5/7 doses per week) relative to 144 (on average 6/7 doses per week). We have used an arbitrarily lower cutoff of 112 total doses as it coincides with 4 months of treatment on 7/7 dosing strategy and most of the data were clustered above this cutoff point. We have performed sensitivity analysis with cutoffs of at least 130 (exact number of doses if participant took 5/7 doses for 26 weeks) for the REMoxTB and RIFAQUIN analysis and 120 (exact number of doses if participant took 5/7 doses for 24 weeks) for the OFLOTUB analysis. Each analysis was adjusted for study country.

All data management, analyses and visualization were performed using R Statistical Software (version 3.4.3, https://www.r-project.org/).

**Data availability**

The standardized data for the OFLOTUB (ClinicalTrials.gov number, NCT00216385) , REMoxTB (NCT00864383), and RIFAQUIN (ISRCTN number, 44153044) trials that support the findings of this study are publicly available to qualified researchers through the Platform for Aggregation of Clinical TB Studies (TB-PACTS, https://c-path.org/programs/tb-pacts/). The DMID 01-009 (NCT00130247) data that support the findings of this study are available from the Tuberculosis Research Unit at Cape Western Reserve University but restrictions apply to the availability of these data, which were used under agreement for the current study. Data are however

available from the authors upon reasonable request and with permission from the Tuberculosis Research Unit.

## Results

### Study participants

A total of 3411 study participants treated for drug susceptible tuberculosis with one of four fluoroquinolone-containing 4-month regimens (n=2001) or the standard 6-month regimen (n=1404) were included in the modified intention-to-treat (MITT) analyses of the OFLOTUB[1], REMoxTB[2], and RIFAQUIN[3] trials; six participants were excluded from the current analyses due to inability to verify treatment allocation in source databases. The external validation data set (DMID 01-009[9]) includes 193 study participants treated with a 4-month experimental regimen (no fluoroquinolone) and 193 study participants treated with the standard 6-month regimen (Figure 1.1). Baseline characteristics of participants did not differ across the experimental and control groups within analysis datasets with exception of race and Senegal country (both $p < 0.001$, Table 1.1); 12% of the participants were HIV-infected.

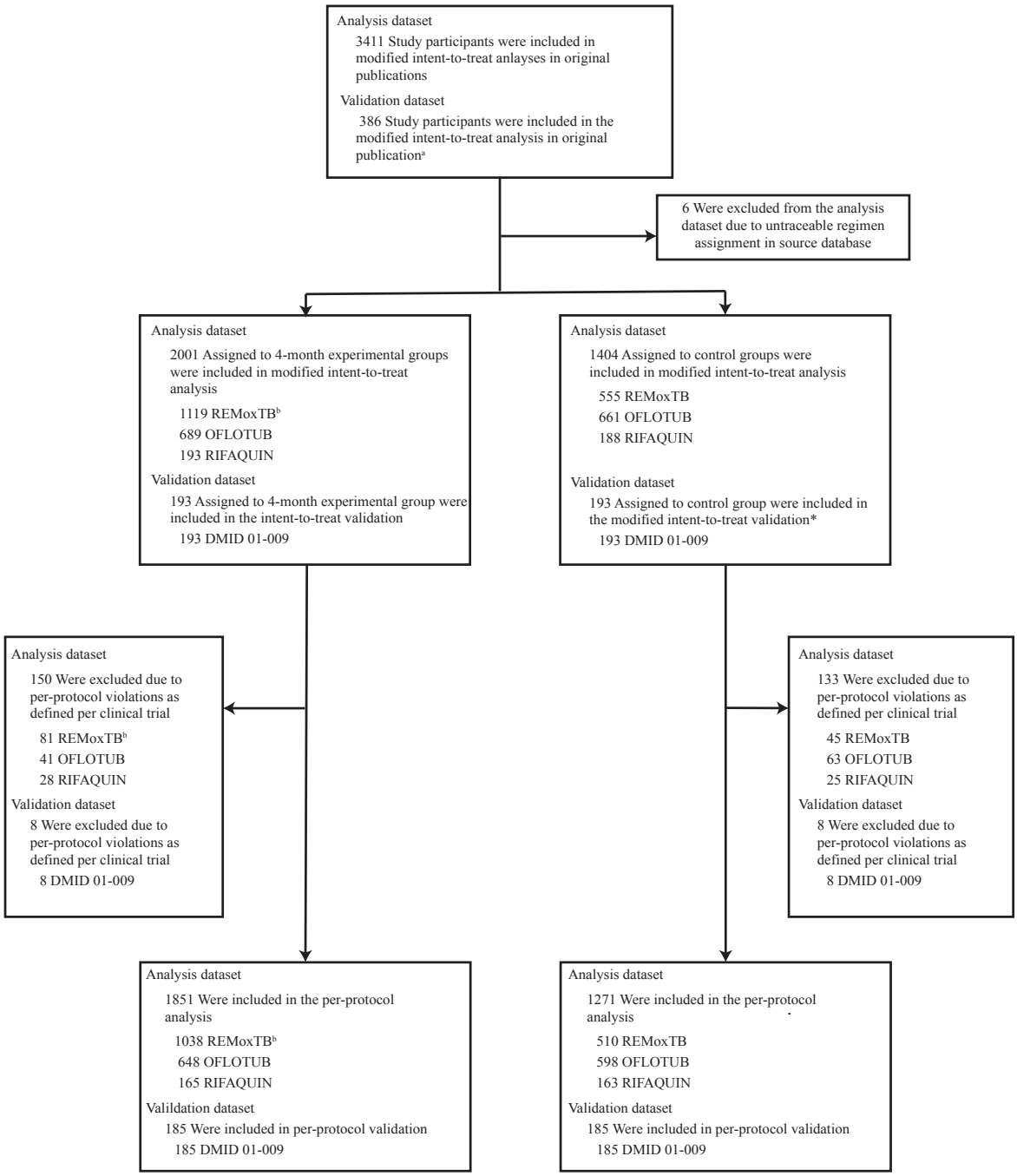



**Figure 1.1** Analysis and validation populations. Individual patient data from three trials were pooled for analysis. The original results were published in ref. 1 (OFLOTUB), ref. 2 (REMoxTB), and ref. 3 (RIFAQUIN). Data from a fourth trial, DMID 01-009, were used for external validation and previously published in ref. 9. The modified intent-to-treat population was used for the analysis. aFor the validation dataset, the time to event analysis population in the original publication was used. bREMoxTB consisted of two 4-month experimental groups.

**Table 1.1** Baseline characteristics of study participants in modified intent-to-treat analysis.

| | **Analysis Dataset (OFLOTUB, REMoxTB, RIFAQUIN)** | | **Validation Dataset (DMID 01-009)** | |
|---|---|---|---|---|
| | **Experimental Group** | **Control Group** | **Experimental Group** | **Control Group** |
| **Characteristic** | **(N = 2001)** | **(N = 1404)** | **(N = 193)** | **(N=193)** |
| **Country - no. (%)** | | | | |
| Benin | 122 (6) | 108 (8) | - | - |
| Botswana | 11 (<1) | 12 (<1) | - | - |
| China | 12 (<1) | 8 (<1) | - | - |
| Guinea | 191 (10) | 184 (13) | - | - |
| India | 228 (11) | 114 (8) | - | - |
| Kenya | 165 (8) | 122 (9) | - | - |
| Malaysia | 43 (2) | 20 (1) | - | - |
| Senegal | 129 (6) | 138 (10) | - | - |
| South Africa | 811 (41) | 516 (37) | - | - |
| Tanzania | 122 (6) | 67 (5) | - | - |
| Thailand | 65 (3) | 34 (2) | - | - |
| Zambia | 35 (2) | 21 (1) | - | - |
| Zimbabwe | 67 (3) | 60 (4) | - | - |
| Brazil | - | - | 67 (35) | 68 (35) |
| Philippines | - | - | 46 (24) | 46 (24) |
| Uganda | - | - | 80 (41) | 79 (41) |
| **Female sex - no. (%)** | 592 (30) | 415 (30) | 76 (39) | 76 (39) |
| **Race - no. (%)[a]** | | | | |
| Black or African American | 1326 (66) | 1066 (76) | - | - |
| Asian | 349 (17) | 178 (13) | - | - |
| Other | 326 (16) | 160 (11) | - | - |
| **Age- yrs[b]** | | | | |
| Median | 30 | 29 | 29 | 27 |
| Interquartile range | 24-39 | 24-38 | 23-38 | 22-36 |
| **Weight- kg** | | | | |
| Median | 52 | 52 | 54 | 55 |
| Interquartile range | 46-58 | 47-58 | 49-62 | 49-61 |
| **Body mass index[c]** | | | | |
| Median | 18.4 | 18.3 | 20.3 | 19.5 |
| Interquartile range | 16.9-20.2 | 16.9-20.1 | 18.7-22.2 | 18.5-22.0 |
| **HIV positivity - no. (%)[d]** | 248 (12) | 220 (16) | 0 (0) | 0 (0) |
| **CD4 cell count [e]** | | | | |
| Median | 363 | 317 | - | - |
| Interquartile range | 265-493 | 241-444 | - | - |
| ≤ 300 - no. (%) | 74 | 81 | - | - |
| > 300 - no (%) | 135 | 99 | - | - |
| **Cavitation- no. (%)[f]** | 1247 (62) | 847 (60) | 0 (0) | 0 (0) |
| **Smear- no. (%)[g]** | | | | |
| Negative | 151 (8) | 85 (6) | 85 (44) | 85 (44) |
| 1+ | 332 (17) | 232 (17) | 26 (14) | 30 (15) |
| 2+ | 503 (25) | 404 (29) | 32 (17) | 36 (18) |
| 3+ | 988 (49) | 667 (48) | 50 (26) | 42 (22) |

[a]Race was missing for all OFLOTUB study participants, black race was assigned to all study participants given all OFLOTUB sites were in Africa.
[b]Age was missing for 5 study participants.
[c]Body mass index was defined as the weight in kilograms divided by the squared height in meters. Height was missing for 291 study participants, median height for females and males were used to calculate body mass index.
[d]Human immunodeficiency virus (HIV) status was missing for 9 study participants.
[e]CD4 cell count cutoff was variable across trials (described in Supplementary Table 2). CD4 cell counts summary statistics was based only on study participants co-infected with HIV, but were missing for 79 HIV co-infected study participants.
[f]Cavitation status was missing for 200 study participants.
[g]Smear grade was based on clinical trial defined grading, but readjusted so all data was on the same scale. Smear grade was missing for 43 study participants.

**Primary outcome analysis**

Multivariate Cox analysis of baseline risk factors for unfavorable outcomes included 3154/3405 (93%) participants with no missing baseline covariates; 1843/2001 (92%) participants were allocated to one of the 4-month experimental regimens and 1311/1404 (93%) participants were allocated to the control regimens. In participants assigned to 4-month experimental regimens, baseline smear 3+ relative to smear negative or 1+ grade and HIV seropositivity were the two major baseline clinical risk factors for unfavorable outcomes with an adjusted hazard ratio of 1.4 (95% confidence interval [CI],1.1-1.9) and 1.4 (95% CI 1.1-1.9), respectively, adjusted also for age and sex. Higher risk was shown in older participants (adjusted hazard ratio [HR], 1.1 per 10 years increase; 95% CI, 1.0-1.2) and male sex (HR, 1.6; 95% CI 1.3-2.1) study participants. After inclusion of on-treatment culture and adherence as risk factors, 1668/2001 (83%) experimental arm participants were available for analysis. Non-adherence was the most significant risk factor for unfavorable outcome with adjusted hazard ratios of 5.7 (95% CI, 3.3-9.9) for participants who missed 10% or more prescribed doses and 1.4 (95% CI, 1.0-1.9) for participants who missed less than 10% of prescribed doses relative to participants who completed treatment without any missed doses. Month-2 culture positivity was significantly associated with unfavorable outcome (HR, 2.2 (95% CI, 1.7-2.9)). After adjustment for on-treatment factors, lower body mass index (BMI,

representative of malnutrition) was a risk factor for unfavorable outcome (HR, 1.4 per 5 kg/$m^2$ decrease; 95% CI 1.1-1.7) (Figure 1.2).

In the 1311/1404 (93%) participants allocated to the 6-month control regimen, HIV seropositivity was the most significant baseline risk factor for unfavorable outcomes with an adjusted hazard ratio of 2.3 (95% CI, 1.6-3.3). Older age (HR, 1.3 per 10 years increase, 95% CI, 1.1-1.4), male sex (HR, 1.5; 95%CI, 1.1-2.1), and lower BMI (HR, 1.3 per 5 kg/$m^2$ decrease; 95% CI, 1.0-1.7) at study entry had higher risk of unfavorable outcomes. 1186/1404 (84%) control arm participants contributed data both for baseline and on-treatment risk factors. Non-adherence was the most significant on-treatment risk factor for unfavorable outcomes with adjusted hazard ratio of 5.9 (95% CI, 3.3-10.5) for participants who missed 10% or more and 2.4 (95% CI, 1.6-3.6) for participants who missed less than 10% of prescribed doses relative to participants who completed treatment without any missed doses. On-treatment culture positivity was also identified as a significant risk factor for unfavorable outcomes (month-2 HR, 1.8; 95% CI, 1.3-2.7). After adjustment for on-treatment factors, HIV positivity (HR, 3.1; 95% CI, 2.0-4.6), male sex (HR, 1.5; 95% CI, 1.0-2.4), and lower BMI (HR, 1.5 per 5 kg/$m^2$ decrease; 95% CI, 1.0-2.0) remained as factors associated with high risk (Figure 1.3). In the per-protocol analysis, results were similar in the experimental and control groups when compared to the primary modified intent-to-treat analysis (data not shown).

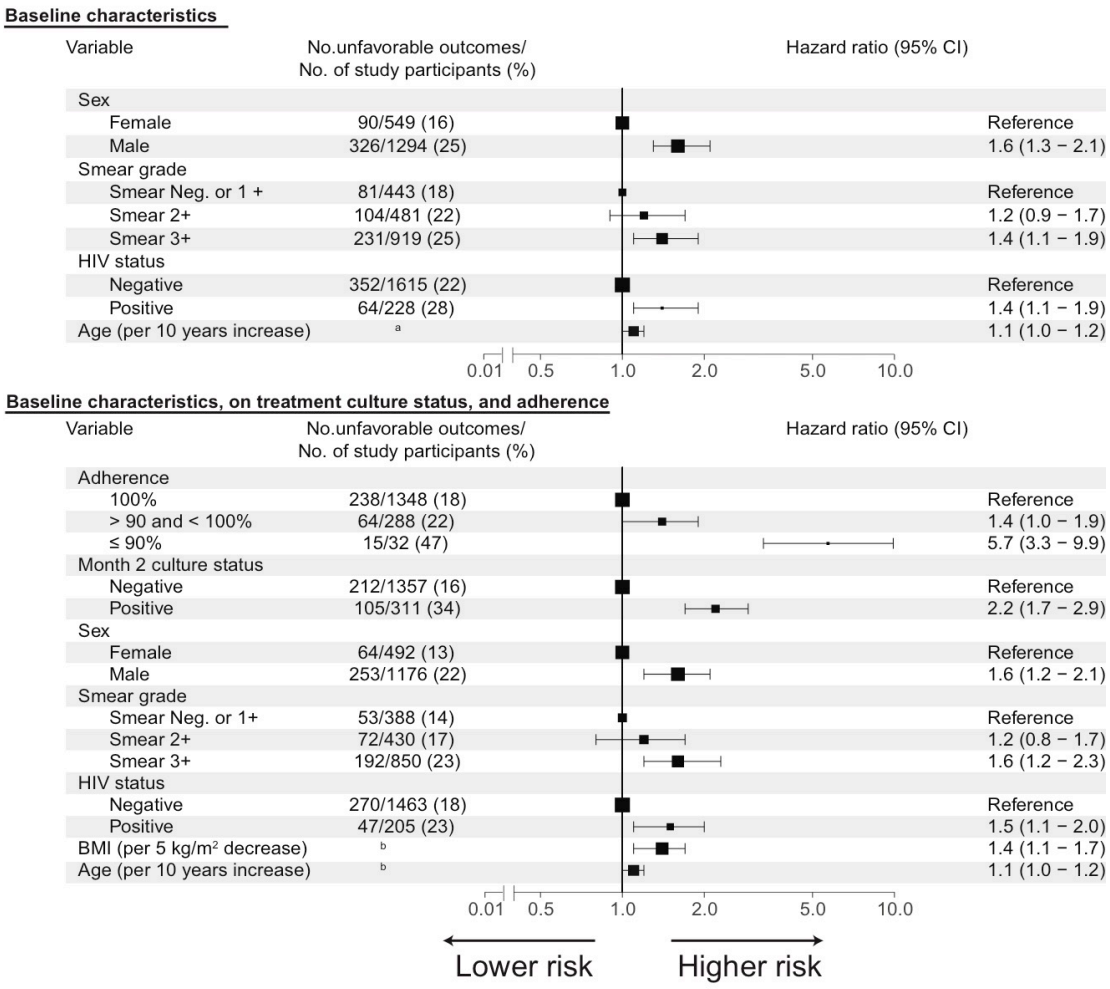



**Figure 1.2** Multivariate Cox regression analysis for experimental group with baseline predictors (top) and baseline and on-treatment predictors (bottom). [a]Age <30 years, 179/916 (20%) unfavorable outcomes and age ≥30 years, 237/927 (26%) unfavorable outcomes. [b]Age <30 years, 136/830 (16%) unfavorable outcomes and age ≥30 years, 181/838 (22%) unfavorable outcomes; BMI ≥17 kg/m$^2$, 226/1,247 (18%) unfavorable outcomes and BMI <17 kg/m$^2$, 91/421 (22%) unfavorable outcomes.

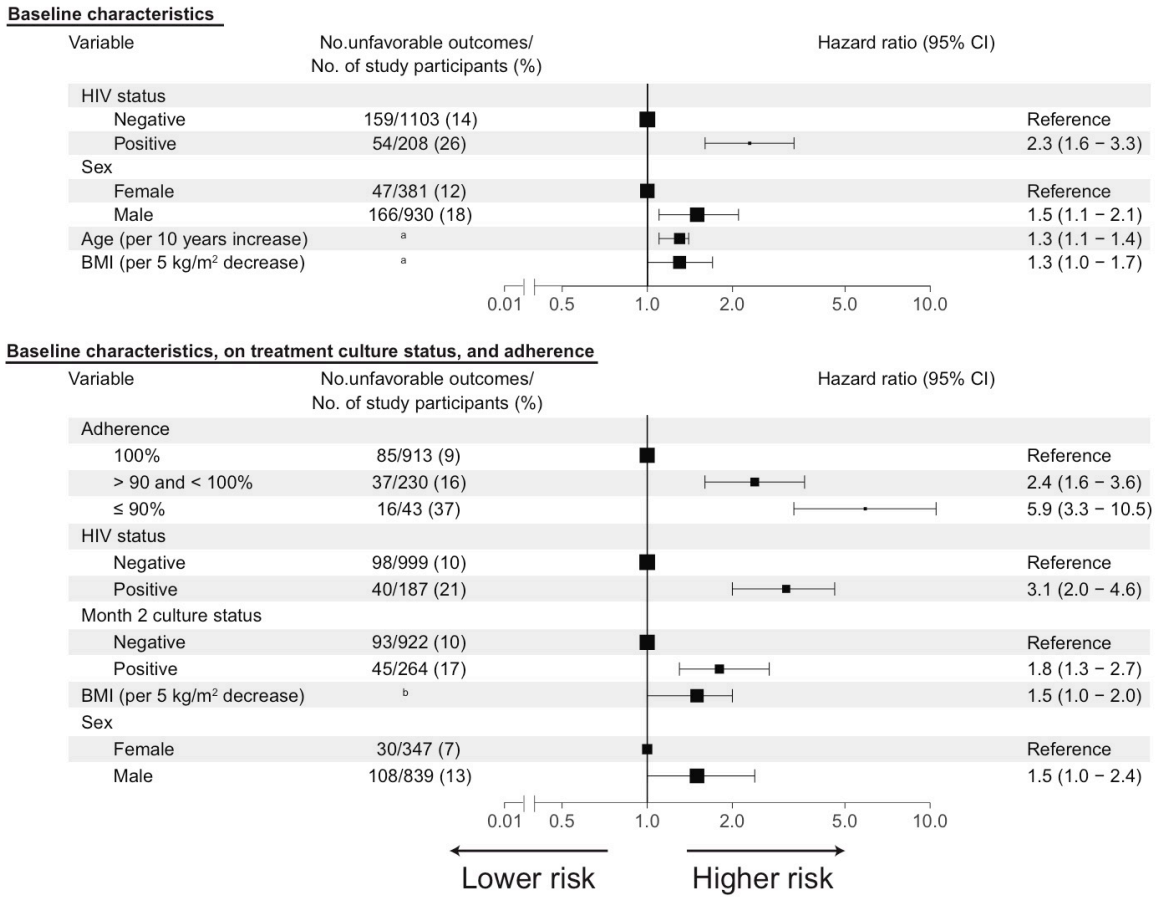



**Figure 1.3** Multivariate Cox regression analysis for control group with baseline predictors (top) and baseline and on-treatment predictors (bottom). [a]Age <30 years, 92/657 (14%) unfavorable outcomes and age ≥30 years, 121/654 (19%) unfavorable outcomes; BMI ≥17 kg/m$^2$, 156/989 (16%) unfavorable outcomes and BMI <17 kg/m$^2$, 57/322 (18%) unfavorable outcomes. [b]BMI ≥17 kg/m$^2$, 102/901 (11%) unfavorable outcomes and BMI <17 kg/m$^2$, 36/285 (13%) unfavorable outcomes.

## Noninferiority test

The proportion of unfavorable outcomes at 24 months for study participants with a baseline smear negative or 1+ grade was similar in experimental and control regimens, indicating noninferiority (difference in study adjusted Kaplan-Meier estimate of unfavorable outcome, 2.6; 90% CI, -0.4 to 5.6; P=0.05 for interaction). Additionally, study participants with non-cavitary disease had a similar proportion of unfavorable outcomes between experimental and control regimens (difference in study adjusted Kaplan-Meier estimate of unfavorable outcome, 3.1; 90% CI, 0.9 to 5.4; P=0.06 for interaction). In an easy-to-treat phenotype of TB consisting of patients

with 1+ or negative smear or non-cavitary disease that comprised 47% (1591/3405) of the study population, the 4 month regimens were noninferior to the 6-month control regimen (Figure 1.4). In a hard-to-treat phenotype of TB consisting of patients with 3+ smear and cavitary disease that comprised 34% (1162/3405) of the study population, the 4-month regimens were clearly inferior.

**External validation**

Using an independent data set available from the DMID 01-009 trial in patients with non-cavitary disease, the patient population eligible for a 4-month rifampin-containing regimen was validated, confirming that for study participants with low-to-moderate smear grade, a standard regimen shortened to 4 months was noninferior to standard 6-month regimen. We confirmed that the driver of high rates of unfavorable outcomes in the 4-month DMID 01-009 regimen was due to study participants with high smear grade (Figure 1.4).

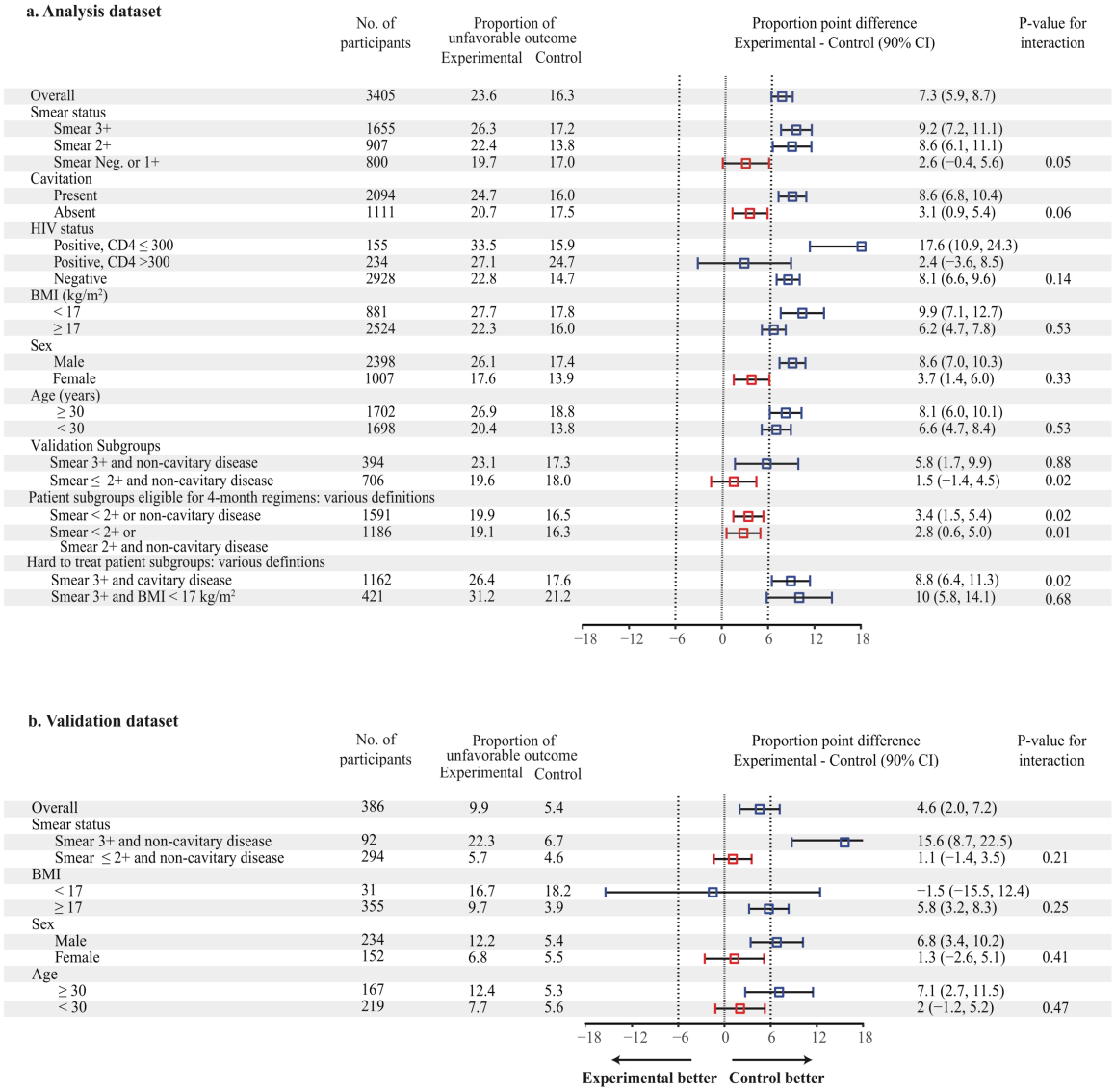



**Figure 1.4** Difference in proportion of unfavorable outcomes between the experimental group and control group, overall and according to subgroups. a. Non-inferiority tests based on analysis dataset. b. Validation of non-inferiority tests in panel a based on an independent validation dataset. The 90% confidence interval of the difference in proportion of unfavorable outcomes were determined by bootstrapping 500 samples. Red squares denote experimental subgroups that were non-inferior to the control subgroups and blue squares denote subgroups that did not show non-inferiority. Study participants in the validation dataset were HIV-uninfected adults with non-cavitary disease and 2-month culture negative status.

**Impact of dosing frequency**

Kaplan-Meier estimates show that study participants who fully adhered to a 6/7 weekly dosing treatment had a higher probability of unfavorable outcome than those who adhered to and completed a 7/7 weekly dosing treatment (HR 2.7, 95% CI 1.1-6.7, after adjustment for treatment duration and country) (Figure 1.5).

To assess the impact of partial adherence on standard of care under a 7/7 or 6/7 dosing strategy, 1285 participants who completed at minimum 4 months of treatment (112 total doses) were included in the Cox regression analysis. This analysis set included 687 participants who were prescribed treatment with a 7/7 weekly dosing strategy for 26 weeks (REMoxTB and RIFAQUIN trials) and 598 participants prescribed under a 6/7 weekly dosing strategy for 24 weeks (OFLOTUB trial). On a 7/7 weekly dosing strategy for 26 weeks, participants who took 156 to 181 total doses (corresponding to an average of 6 doses per week or missing up to 14% pills) or 112 to 155 total doses (corresponding to an average of 5 doses per week or missing 14-33% pills) had significantly higher risk of unfavorable outcomes relative to those who took all 182 prescribed doses (7 doses per week), with hazard ratios of 2.4 (95% CI, 1.3-4.3) and 28.9 (95% CI, 10.5-80.0), respectively, adjusted for treatment duration and country (Figure 1.5). Similarly, participants receiving 112 to 143 doses (average of 5 doses per week) had a higher risk of unfavorable outcomes relative to those who took the complete 144 prescribed doses (6 doses per week) for 24 weeks, with hazard ratio of 2.4 (95% CI 1.2 to 4.8), adjusted for treatment duration and country (Figure 1.5).

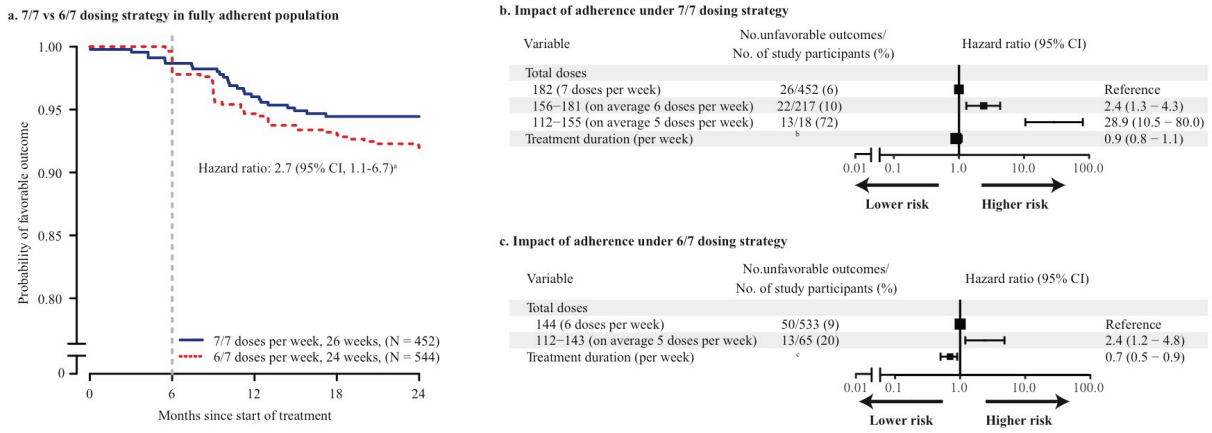



**Figure 1.5** Analysis of 7/7 and 6/7 dosing strategies and impact of adherence in the control group. a. Kaplan–Meier estimates for fully adherent study participants (n = 996) after treatment with 7/7 or 6/7 dosing strategies. b. Multivariate analysis with total number of doses taken for study participants who took at least 4 months of treatment under 7/7 dosing strategies for 26 weeks (REMoxTB and RIFAQUIN trials), after adjustment for country and treatment duration. c. Multivariate analysis with total number of doses taken for study participants who took at least 4 months of treatment under 6/7 dosing strategies for 24 weeks (OFLOTUB trial), after adjustment for country and treatment duration. Effect sizes for country are available in Supplementary Table 7. In b and c, the HRs with 95% Wald CI are reported. aHazard ratio with 95% Wald CI for 6/7 dosing strategy relative to 7/7 dosing strategy for fully adherent population after adjustment for country and treatment duration. bTreatment duration <182 days, 21/110 (19%) unfavorable outcomes and treatment duration ≥182 days, 40/577 (7%) unfavorable outcomes. cTreatment duration <169 days, 21/155 (14%) unfavorable outcomes and treatment duration ≥169 days, 42/443 (9%) unfavorable outcomes.

## Discussion

In this individual patient pooled analysis of recent phase 3 treatment shortening trials, we have shown that adult patients with minimal disease, as defined by low smear grade or the absence of cavitation were at lower baseline risk for unfavorable outcomes; in this population the experimental 4-month regimens are effective. Patients with either of these low-risk characteristics, which define an easy-to-treat phenotype of TB, comprised 47% (1591/3405) of the total study population. Conversely, we have shown that a smear grade of 3+ and the presence of cavitation on chest radiograph at baseline define a hard-to-treat phenotype, comprising 34% (1162/3405) of the population, and this group may require longer durations of treatment than the current standard 6-month regimen to achieve the highest cure rates feasible. In our analyses, other baseline characteristics associated with unfavorable outcomes included being HIV-infected and having a

lower BMI at study entry. Male sex was consistently and independently linked with poor likelihood of cure in both control and experimental regimens. The etiology for this association is not clear, particularly given the association persists after adjusting for severity of disease and adherence. Our definitions of TB phenotypes were validated in an independent trial dataset of patients with non-cavitary disease. Whereas this trial was stopped early due to higher rates of unfavorable outcomes in the experimental 4-month regimen, we confirmed that a 4-month regimen would be effective for patients with negative, 1+ or 2+ smears in non-cavitary disease at baseline. We also confirmed that patients with high smear grade (smear 3+) at baseline were more likely to fail treatment regardless of receiving 4-month or 6-month regimens, as compared to those with lower smear grades at baseline. Given the established importance of cavitation in disease prognosis and treatment response[5,12,13], we included this characteristic in the analyses of non-inferiority for various sub-groups, despite the fact that cavitation was not a significant variable in the multivariate analysis and only marginally significant in the univariate analysis. In analyses limited to the trials providing detailed chest radiograph readout data, specifically OFLOTUB and RIFAQUIN, we confirmed that cavity size, bilateral disease and disease extent measured by zone scores were all significant risk factors for unfavorable outcome (Appendix Figure A.1 and Appendix Figure A.2), confirming that disease severity determined by chest radiograph remains an important tool for the definition of hard-to-treat phenotypes and prediction of treatment outcome. Overall, we show that the combination of smear grading and cavitary status adequately define easy-to-treat and hard-to-treat groups, however, we also show subgroups that allow for stratification when chest radiographic information is not available.

In this study, we also found that across both experimental and standard control regimens, minimal non-adherence and missed doses were associated with significantly increased risk for unfavorable outcome. Missing as few as 1 in 10 doses of a regimen was associated with a five-

fold increase in risk. Missed doses had a stronger association with poor outcome than failure to achieve culture conversion at 2 months. Consistent with our analyses of non-adherence, dosing frequencies of less than 7 of 7 days increase the chances of unfavorable outcome, even if participants were fully adherent (Figure 1.5). Current U.S. TB treatment guidelines state, on the basis of clinical experience and program practicality, that 5-days-a-week drug administration is an acceptable alternative to 7-days-a-week administration, and that either approach may be considered as meeting the definition of "daily" dosing.[5] Our findings suggest otherwise and provide data-driven evidence to support using 7 of 7 dosing.[14,15] The finding that the current rifampin-based regimen used worldwide has "low forgiveness" for non-adherence or missed doses has important implications for TB care as well as for future design and conduct of clinical trials. A regimen with excellent efficacy under rigorous clinical trial settings that is otherwise unforgiving of missed doses will fail in the field. New and improved adherence interventions for TB have been introduced to facilitate treatment completion, however, such tools can be limited by issues of scale-up, generalizability and cost.[16,17] A more durable and patient-centered solution is the targeted development of regimens comprised of drugs with long half-lives and steady pharmacokinetic profiles that will accommodate less than perfect adherence patterns in the field, without penalty to the efficacy of the regimen. Our findings in this regard highlight the critical value of additionally conducting pragmatic clinical trials that assess the effectiveness and robustness of regimens under programmatic conditions.

In this study, we found that the 4-month fluoroquinolone containing regimens met the margin for non-inferiority in participants with a baseline smear negative or 1+ grade or non-cavitary disease. Conversely, we found that a hard-to-treat phenotype of TB defined by high smear grades and cavitation on baseline chest radiograph was associated with unfavorable outcomes. Randomized trials conducted by the British Medical Research Council largely in the pre-HIV era

have previously illustrated that the majority of patients do not need six months of standard therapy.[18,19] Our analyses support this position and suggest that the current "one-size-fits-all" model of care leads to under-treatment of patients with severe forms of disease, and unnecessarily long treatment (with unjustified risk of drug toxicity) for many patients with less extensive disease. We believe our results provide justification to evaluate a stratified approach to TB therapeutics. Using baseline markers to determine the optimal stratum for a given patient, with decisions for treatment extension further enhanced by use of on-treatment measures of adherence and clinical, microbiologic and radiographic markers, the feasibility of achieving cure in all patients with TB, rather than a majority, is enhanced. Pursuit of the highest possible cure rates in TB is an important public health priority, and perhaps more important than treatment shortening, as suggested by recent modelling work that shows increases in treatment efficacy will have the greatest impact on reducing mortality and burden of disease worldwide.[20] The tools necessary for using stratified medicine approaches to TB care at the programme level are already in use in many settings, including HIV testing, CD4 cell counts for HIV-positive patients, chest radiography, smear microscopy, and scales for measuring height and weight for calculation of BMI. Future trials that test stratified medicine approaches to TB care should also evaluate newer tools (e.g., GeneXpert cycle threshold), which in turn would allow for algorithms for selecting duration to be further refined, offering additional characteristics and options for determining risk. Nonetheless, some patients will have limited access to these diagnostics and in such settings, either a simpler stratification algorithm can be developed (for example, smear grade and BMI, as shown in ) or the currently used "one-size-fits-all" approach may still remain the most practical and implementable option.

Our study has limitations. Data sharing principles are supported in the TB therapeutics field, however, data collection was not standardized across the included trials.[21,22] Future protocols

should use minimum data set standards, compliant with CDISC standards (https://www.cdisc.org), to allow robust pooled analyses in the future. Second, chest radiograph interpretation was not uniform, as such we could not analyze size and number of cavities in all three studies (Appendix Figure A.1 and Appendix Figure A.2). Third, very limited pharmacokinetic data were available, hampering our ability to explore dosing, drug-exposure and outcome relationships. We advocate for the inclusion of population pharmacokinetics in phase 3 trials to address the variability in responses across geographic regions and populations. Our comparison of 6/7 with 7/7 dosing was a comparison between trials rather than within trials, and therefore may be confounded by other study differences. Finally, only 12% of participants had HIV-coinfection and many were not on effective ART regimens, thus, caution should be used in generalizing our findings to immunocompromised populations. Strengths of our analyses include the inclusion of large data sets from four international registration-quality phase 3 trials conducted across diverse human populations in high TB burden settings in South America, sub-Saharan Africa and Asia, performance of microbiologic assays by quality-controlled laboratories, and the careful recording of study treatment under direct observation.

In sum, our validated analyses of individual patient data from contemporary randomized clinical trials provide three major findings. First, we show that low smear grades at baseline or the absence of cavitation identifies a population at low risk for recurrence in whom 4-month rifampin containing regimens may be effective. Conversely, high sputum smear grade at baseline in conjunction with the presence of cavities defines a hard-to-treat phenotype that may require longer durations of treatment than the current standard of care, in order to achieve high cure rates. There is also a third phenotype made up of the remaining patients for whom treatment shortening may also be possible. Second, we show that minor degrees of non-adherence or missed doses significantly increase the risk for poor outcomes. Third, we show that simple baseline and on-

treatment markers could be used to select treatment duration with greater precision, providing a programmatically viable alternative to the "one-size-fits-all" paradigm used worldwide. Our results indicate that stratified medicine principles should be further evaluated in clinical trials of TB therapeutics.

## References


1. Merle CS, Fielding K, Sow OB, et al. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med* 2014; **371**: 1588-98.
2. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med* 2014; **371**: 1577-87.
3. Jindani A, Harrison TS, Nunn AJ, et al. High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N Engl J Med* 2014; **371**: 1599-608.
4. Organization GWH. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. Licence: CC BY-NC_SA 30 IGO; 2017.
5. Nahid P, Dorman SE, Alipanah N, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clinical Infectious Diseases* 2016; **63**: e147-e95.
6. Jo K-W, Yoo J-W, Hong Y, et al. Risk factors for 1-year relapse of pulmonary tuberculosis treated with a 6-month daily regimen. *Respiratory Medicine* 2014; **108**: 654-9.
7. Benator D, Bhattacharya M, Bozeman L, 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 (London, England)* 2002; **360**: 528-34.
8. Goldberg S. TBTC Study 31: Rifapentine-containing Tuberculosis Treatment Shortening Regimens (S31/A5349).
9. Johnson JL, Hadad DJ, Dietze R, et al. Shortening Treatment in Adults with Noncavitary Tuberculosis and 2-Month Culture Conversion. *Am J Respir Crit Care Med* 2009; **180**: 558-63.

10. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med* 2007; **357**: 2189-94.

11. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine* 2005; **24**: 3089-110.

12. Savic RM, Weiner M, MacKenzie WR, et al. Defining the optimal dose of rifapentine for pulmonary tuberculosis: Exposure-response relations from two phase II clinical trials. *Clinical pharmacology and therapeutics* 2017; **102**: 321-31.

13. Alipanah N, Cattamanchi A, Menzies R, Hopewell PC, Chaisson RE, Nahid P. Treatment of non-cavitary pulmonary tuberculosis with shortened fluoroquinolone-based regimens: a meta-analysis. *The International Journal of Tuberculosis and Lung Disease* 2016; **20**: 1522-8.

14. Vernon AA, Iademarco MF. In the Treatment of Tuberculosis, You Get What You Pay for…. *American Journal of Respiratory and Critical Care Medicine* 2004; **170**: 1040-2.

15. Chang KC, Leung CC, Yew WW, Ho SC, Tam CM. A Nested Case–Control Study on Treatment-related Risk Factors for Early Relapse of Tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 2004; **170**: 1124-30.

16. Ngwatu BK, Nsengiyumva NP, Oxlade O, et al. The impact of digital health technologies on tuberculosis treatment: a systematic review on behalf of the Collaborative group on the impact of digital technologies on TB. *Eur Respir J* 2018; **51**.

17. DiStefano MJ, Schmidt H. mHealth for Tuberculosis Treatment Adherence: A Framework to Guide Ethical Planning, Implementation, and Evaluation. *Global Health: Science and Practice* 2016; **4**: 211-21.

18. Fox W. Whither short-course chemotherapy? *British journal of diseases of the chest* 1981; **75**: 331-57.

19. Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 1999; **3**: S231-79.

20. Kendall EA, Shrestha S, Cohen T, et al. Priority-Setting for Novel Drug Regimens to Treat Tuberculosis: An Epidemiologic Model. *PLOS Medicine* 2017; **14**: e1002202.

21. CDISC-Tuberculosis Therapeutic Area User Guide v2.0. 2016.

22. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors. *New England Journal of Medicine* 2016; **374**: 384-6.

## Chapter 2: Predicting optimal treatment durations for tuberculosis patients: a risk stratification algorithm and clinical simulation tool


### Abstract

**Background** No evidence-based tools exist to guide decisions on the optimal treatment duration for tuberculosis. We developed (1) survival models to predict individual risk of unfavorable outcomes, (2) a quantitative risk stratification algorithm that stratifies individuals into risk groups, and (3) a clinical tool that predicts optimal treatment duration for each patient receiving rifampin containing regimens.

**Methods** Data from four Phase 3 trials, each evaluating treatment duration shortening from 6 months to 4 months, were obtained from a public repository. Parametric survival models were used to describe time to unfavorable outcomes. Regimen, baseline, and on-treatment characteristics were evaluated as predictors of outcomes. Exact regression coefficients of significant predictors were used to calculate individual risk scores and a target cure rate of 93% was used to predict optimal treatment duration.

**Results** A six-item risk score (HIV status, smear grade, sex, cavitation, BMI and month 2 culture status) successfully grouped patients into low (794/3405, 23%), moderate (1624/3405, 48%), and high (987/3405, 29%) risk, requiring treatment durations of 4, 6 and greater than 6 months, respectively, to reach 93% target cure rates. Of the 393/3405 patients who had TB-related unfavorable events, 49% (194/393) were identified as high risk, 38% (151/393) as moderate risk and 12% (48/393) as low risk.

**Conclusions** Our results show that stratified medicine approaches, where duration is selected with precision based on patient risk, is feasible and safely achieves high cure rates in tuberculosis

patients. A validated, interactive, evidence-based tool for selection of optimal treatment durations is provided.

## Introduction

Innovation in and simplification of tuberculosis (TB) therapy is desperately needed. There is an increase in the number of new and repurposed compounds undergoing evaluation as part of novel treatment regimens. Current TB drug development programs are focused on identifying shorter TB treatment regimens that maximize treatment completion without compromising on overall cure rates.[1,2] However, numerous translational gaps hinder the developmental pathway and improved tools and approaches are necessary to accelerate the identification of optimal treatment regimens for TB patients.[1,3]

Whereas current practice guidelines highlight individual risk factors that based on post-hoc analyses suggest an extension in treatment duration may be warranted (bacterial burden, extent of cavitary disease, culture positivity at 8 weeks, etc.), there have been no tools developed that indicate likelihood of achieving cure based on an integrated suite of baseline and on-treatment risk factors.[4,5] Such tools could provide for individualized prediction of optimal treatment regimens for TB patients. Moreover, there have been no tools to date that could estimate likelihood of durable cure when treatment is shortened to durations less than 6 months.

As short and ultrashort duration regimens, such as those evaluated by TB Trials Consortium (TBTC) Study 31/A5349[6] and TRUNCATE TB[7], are now being evaluated in the treatment of TB patients, stakeholders are increasingly seeking to integrate innovative clinical trial approaches and tools into their decision making to facilitate early and effective deployment of the best regimens. In this study, we leveraged data from three large contemporary Phase 3 trials to develop a data-driven framework and clinical tools that can be used to recommend optimal treatment durations for stratified patient subgroups with the aim to provide equal likelihood of durable cure in all patients. Specifically, we developed: (1) parametric survival models to predict individual risk of unfavorable outcomes, (2) a quantitative risk stratification algorithm that defines

stratification groups using individual patient risk scores, and (3) a clinical simulation tool that predicts optimal treatment duration of rifampin-containing regimens based on individual patient risk and target cure rates. We externally validated our models and tools using data from a fourth Phase 3 trial. Our tools can be used to alert clinicians to patients with high risk of unfavorable outcome who may need longer treatment, to identify patients that may be able to be treated with shorter durations, and finally to inform a priori decisions regarding optimal durations for new regimens being considered for phase 3 clinical trials.

## Methods

### Study design and data collection

Individual participant data (n=3405) from three recent international, randomized phase III trials (OFLOTUB[8], REMoxTB[9], and RIFAQUIN[10]) that compared 4-month fluoroquinolone containing regimens to the standard 6-month regimen was used for model development. A fourth trial, DMID 01-009[11] (n = 386), was used for external validation and tested a 4-month standard regimen (no fluoroquinolone). Additional information on study design for these trials are available in the original publications and described in Chapter 1 Methods Section.

### Efficacy outcomes

The primary efficacy endpoint was time to an unfavorable outcome for a maximum of 18 months after start of treatment. Participants who were not followed for at least 18 months were censored at their last available timepoint. Due to the composite definitions used to label unfavorable outcomes, we developed two separate models for: a.) time to TB-related outcomes and b.) time to non-TB-related outcomes. TB-related outcomes included treatment failures, deaths due to TB, relapse, and exogenous reinfection (for the OFLOTUB study only). Non-TB-related outcomes included dropouts, withdrawal of consent, lost to follow up, adverse events, other deaths,

and inadequate treatment. For each of the models, time was censored at time of alternative outcome (i.e. when modelling time to TB-related outcomes, non-TB-related outcomes were censored at time of event). This approach requires independent censoring, meaning that we are assuming censoring does not change the probability of the event of interest for each model.[12]

**Predictors of efficacy outcomes**

Variables collected in all studies and missing in less than 10% of participants were tested as potential predictors of risk of unfavorable outcomes. Covariate search was performed in a stepwise manner. First, exposure and regimen composition factors were tested. Due to limited pharmacokinetic data collected in each trial (83% of population without pharmacokinetic data), exposure of drugs was described by: i.) treatment duration, defined as the duration of days the participant was on treatment, ii.) number of treatment days, defined as the total number of treatment days drugs were administered, and iii.) cumulative rifampin dose, defined as number of treatment days taken multiplied by individual rifampin daily dose. Actual treatment duration, number of treatment days, and cumulative rifampin dose were all determined based on complete dosing histories that were recorded during the trial with distributions shown in Figure 2.1. Number of treatment days was adjusted for the intermittent schedule used with high dose rifapentine in the RIFAQUIN 4-month experimental arm during the continuation phase. The 900 mg rifapentine twice weekly administration (total of 1800 mg weekly) was translated to 3 treatment days per week if both doses were administered (1800 mg per week/600 mg per day). Isoniazid and moxifloxacin inclusion were tested as regimen composition predictors of unfavorable outcome.

Second, baseline characteristics and month 2 culture conversion were tested as potential predictors of risk of unfavorable outcomes. The baseline analysis set included age, race, BMI, sex, presence of cavitation on chest radiograph and smear grade. Major imputations for baseline characteristics and more information on data specification and harmonization is available and has

been reported in our previous publication (Chapter 1).[4] For missing categorical data, the mode value for the entire population was assigned. For missing continuous data, the median value for the entire population was assigned. Finally, the effect of study and region of clinic site (African vs. non-African) were tested and the final model was adjusted if found significant.

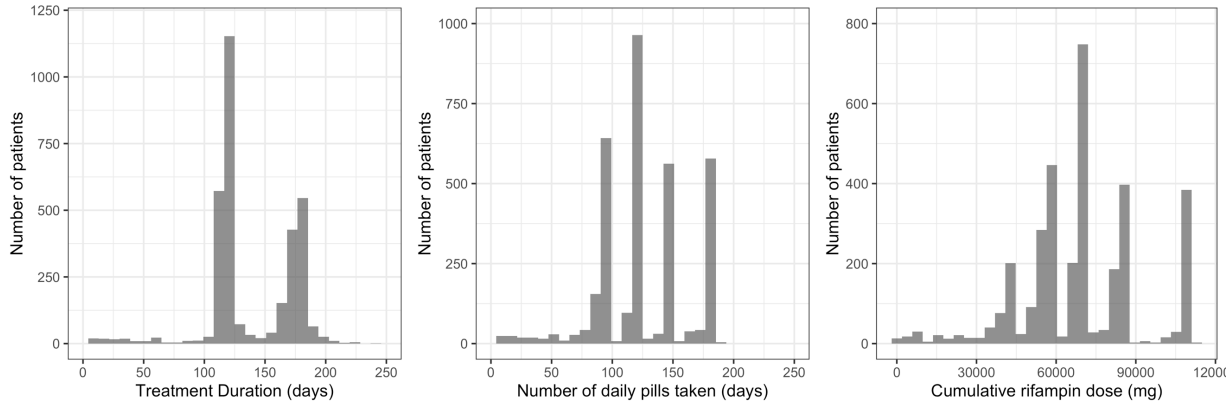



**Figure 2.1** Distribution of treatment duration (days), number of treatment days and cumulative rifampin dose (mg) for patients in the model development dataset.

**Model development and evaluation**

Two separate parametric survival models were developed to describe time to TB-related unfavorable outcomes and time to non-TB-related unfavorable outcomes. To describe time to event for each model, a parametric survival function was used, according to the following survival equation:

$$S_t = e^{-\int_0^t h(t)d(t)}$$

where survival at time $t$, $S_t$, is a function of the cumulative hazard risk from time 0 to time $t$. The hazard, $h(t)$, describes the instantaneous risk of having an event at time $t$. The hazard distribution of each of the models were evaluated by testing different functions that best described the risk of having an event including the exponential, Gompertz, Weibull, and surge functions.

Identification of predictors were tested on all model parameters of the best fit hazard risk function (i.e. baseline hazard risk and shape parameters) for each TB-related unfavorable outcome model and non-TB-related unfavorable outcome model. An automated procedure with Stepwise Covariate Model software (PsN, SourceForge, Slashdot Media, San Francisco, CA, USA) was used. This method involves stepwise testing of linear and nonlinear relationships with forward inclusion (change in objective function value, ΔOFV, of 6.63, $p < 0.05$ to include for 1 degree of freedom) and backwards exclusion (ΔOFV of 10.83, $p < 0.01$ to remove for a degree of freedom). In case of categorical covariates, ΔOFV at the respective p-values may be different depending on the degrees of freedom. Linear, sigmoidal EMAX, and sigmoidal EMAX with hill coefficients relationships were assessed when testing exposure predictors on hazard function parameters. Linear relationships were tested for all other predictors. The final models contained covariates that met the predefined statistical criteria.

Model development was performed using the nonlinear mixed effect approach available in the NONMEM program (version 7.4). The Laplacian estimation method was employed for the time to event variable. The model building procedure was guided by likelihood testing and internal validation techniques, including Kaplan-Meier visual predictive checks to assess calibration and area under the receiver operating characteristic curve (AUROC) for discrimination. Kaplan-Meier visual predictive checks involved performing 800 simulations with the analysis datasets and final model, then comparing the observed Kaplan-Meier curve with 95% prediction interval of the simulated data. AUROC for discrimination was determined by calculating the c-index at 6-months post end of treatment. After final model development, external validation with the DMID 01-009 dataset was also evaluated using the same calibration and discrimination techniques.

**Derivation of individual risk scores and optimal treatment durations**

Exact regression coefficients for significant baseline and on-treatment (month 2 culture status) predictors of TB-related unfavorable outcomes in the final model were used to derive a continuous risk score , $Risk\ Score_i$, for each individual, $i$. The individual risk scores, which take into account all significant baseline and on-treatment predictors, and the final predictive model for time to TB-related outcome were used to calculate the number of treatment days required to reach a specified target cure rate, $CURE_{target}$. The optimal treatment duration for each individual, $TRT\_DURATION_i$, can then be determined based the number of treatment days required to reach a target cure rate and the weekly dosing schedule. Optimal treatment duration calculations in this manuscript are based on a 7/7 weekly dosing schedule and full adherence. Full derivation of $Risk\ Score_i$ and $TRT\_DURATION_i$ are described in the Chapter 2 Supplemental Results.

**Validation of risk stratification algorithm and treatment duration**

The risk stratification algorithm to define low, moderate, and high risk patients were based on the predicted optimal treatment duration of the control regimen (isoniazid, rifampin, pyrazinamide, and ethambutol) required to reach less than or equal to 7% TB-related outcomes ($CURE_{target}$ = 0.93) at 18 months follow-up from the start of treatment. Low risk group was defined as patients requiring less than or equal to 18 weeks of treatment, moderate risk group requiring 19 to 24 weeks of treatment, and high risk group requiring more than 24 weeks of treatment. To validate the calibration of our risk stratification algorithm, we first calculated the individual risk scores for all individuals in the derivation cohort (REMoxTB, RIFAQUIN and OFLOTUB) and grouped patients into low, moderate, and high risk groups. Then, we compared the observed Kaplan Meier rates of TB-related unfavorable outcome after treatment of 4-month and 6-month regimens for each risk group. For good calibration, we expected that the low risk group would be treated successfully (Kaplan Meier rates at or above 93%) by the 4 or 6-month

regimens, moderate risk groups treated successfully only by the 6-month regimen, and the high risk group not treated successfully with neither the 4- or 6-month regimens. Additionally, we calculated the hazard ratio based on Cox regression analysis to compare success rates in the 4-month and 6-month regimens for each risk group. For external validation of the risk stratification algorithm, we used the independent dataset available from DMID 01-009. The same statistical methods used for interval validation was also used to externally validate the calibration of our risk stratification algorithm.

**Clinical simulation tool**

An interactive clinical simulation application based on the final parametric survival model for TB-related unfavorable outcomes was developed using the Shiny package in the R programming language (version 1.3.2). The tool allows evaluation by simulation of different treatment strategies in various population groups based on predictors included in the final models. All simulations of virtual populations are based on the available model development and external validation populations used in this analysis. This means that the uncertainty of simulated populations reflects the distribution of patients available in the development and external validation population and the global TB burden included in these clinical trials.

**Results**

**Data characteristics**

The model development population included 3405 patients with drug susceptible TB. As described previously, baseline characteristics did not differ between experimental and control groups (Table 2.1).[4] In the 4-month experimental group, 1257/2001 (63%) of patients were treated with a regimen that included isoniazid. The median number of treatment days was 114 in the 4-month experimental group and 169 in the 6-month control group (Table 2.1). Month 2 culture

conversion rates were higher in the 4-month experimental group than 6-month control group (Table 2.1, P = 0.01). Of the 3405 patients, 656 patients had an unfavorable outcome during follow-up (up to 18 months), with shorter time to unfavorable outcome when treated with 4-month experimental regimens [univariate hazard ratio [HR]: 1.6 (95% confidence interval [CI]:1.3-1.8)]. Of the 656 patients with unfavorable outcomes, 393 had a TB-related outcome, with shorter time to TB-related outcome when treated with 4-month experimental regimens [HR: 2.5 (95% CI: 2.0-3.1)], and 263 had a non-TB-related outcome (145 in the 4-month experimental group and 118 in the 6-month experimental group), with no evidence of difference in time to non-TB-related outcome among 4-month and 6-month regimens [HR: 0.87 (95% CI: 0.68-1.1)] (Figure 2.2).

**Table 2.1** Baseline, on-treatment and regimen characteristics of study participants included in the model development population**.**

| | **4-month experimental group** | **6-month control group** |
|---|---|---|
| **Characteristic** | **(N = 2001)** | **(N = 1404)** |
| **Site Region** | | |
| Sub-Saharan Africa | 1653 (83) | 1228 (88) |
| India | 228 (11) | 114 (8) |
| Asia | 120 (6) | 62 (4) |
| **Female sex - no. (%)** | 592 (30) | 415 (30) |
| **Age- yrs**[a] | | |
| Median | 30 | 29 |
| Interquartile range | 24-39 | 24-38 |
| Range | 16-81 | 17-77 |
| **Weight- kg** | | |
| Median | 52 | 52 |
| Interquartile range | 46-58 | 47-58 |
| Range | 35-98 | 35-137 |
| **Body mass index**[b] | | |
| Median | 18.4 | 18.3 |
| Interquartile range | 16.9-20.2 | 16.9-20.1 |
| Range | 12.0-40.7 | 12.1-50.9 |
| **HIV positivity - no. (%)**[c] | 248 (12) | 220 (16) |
| **Cavitation- no. (%)**[d] | 1247 (62) | 847 (60) |
| **Smear- no. (%)**[e] | | |
| Negative or 1+ | 483 (24) | 317 (23) |
| 2+ | 503 (25) | 404 (29) |
| 3+ | 988 (49) | 667 (48) |

**Table 2.1 continued.**

| Characteristic | 4-month experimental group (N = 2001) | 6-month control group (N = 1404) |
|---|---|---|
| **Regimen Composition** | | |
| Isoniazid | 1257 (63) | 1404 (100) |
| Rifapentine | 193 (10) | 0 (0) |
| Moxifloxacin | 1312 (66) | 0 (0) |
| Gatifloxacin | 689 (34) | 0 (0) |
| **Treatment duration (days)[f]** | | |
| Median | 119 | 175 |
| Interquartile range | 114-119 | 169-182 |
| Range | 2-202 | 4-239 |
| **Number of treatment days[g]** | | |
| Median | 114 | 144 |
| Interquartile range | 96-119 | 144-182 |
| Range | 1-120 | 1-189 |
| **Cumulative rifampin dose (mg)[h]** | | |
| Median | 57600 | 86400 |
| Interquartile range | 51600 - 71400 | 79200 - 108600 |
| Range | 450 -72000 | 450-113400 |
| **Number of daily doses taken[i]** | | |
| Median | 114 | 144 |
| Interquartile range | 96-119 | 144-182 |
| Range | 1-120 | 1-189 |
| Month 2 culture positivity[i] | 336 (17) | 285 (20) |

[a]Age was missing for 5 study participants.
[b]Body mass index was defined as the weight in kilograms divided by the squared height in meters. Height was missing for 291 study participants, median height for females and males were used to calculate body mass index.
[c]Human immunodeficiency virus (HIV) status was missing for 9 study participants.
[d]Cavitation status was missing for 200 study participants.
[e]Smear grade was based on clinical trial defined grading, but readjusted so all data was on the same scale. Smear grade was missing for 43 study participants.
[f]Treatment duration, defined as the number of days the participant was on treatment, was missing for 106 study participants.
[g]Number of treatment days, defined as the total number of treatment days drugs were administered, was missing for 38 study participants.
[h]Cumulative rifampin dose, defined as number of treatment days multiplied by individual rifampin daily dose, was missing for 38 study participants.
[i]Month 2 culture was missing for 308 study participants.

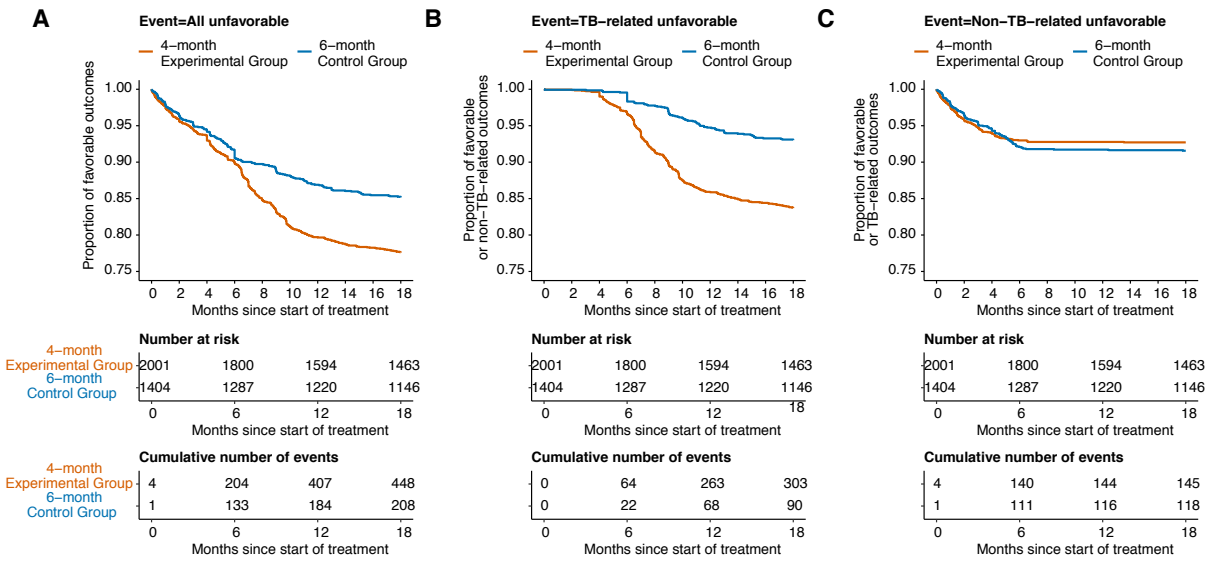



**Figure 2.2** Kaplan-Meier estimates of time to unfavorable outcomes, TB-related outcomes and non-TB-related outcomes. (A) All unfavorable outcomes as defined according to each trial protocol. (B) TB-related outcomes including treatment failures, deaths due to TB, relapse, and reinfection (OFLOTUB study only). (C) Non-TB-related outcomes including dropouts, withdrawal of consent, lost to follow up, adverse events, other deaths, and inadequate treatment. Each plot was stratified by assigned regimen duration.

## Model development and evaluation

The hazard risk for TB-related outcomes was best described with a surge function:

$$h(x_i, t) = \lambda(x_i) t^{\beta} exp(-\alpha t)$$

where $\lambda(x_i)$ is the baseline hazard risk dependent on the covariate values, $x_i$, for each individual, $i$; $\alpha$ and $\beta$ are the surge function shape parameters; and $t$ is time in months from start of treatment. Decreased number of treatment days and exclusion of isoniazid increased hazard risk of TB-related outcomes (29% (relative standard error, RSE = 9) increase per 28 day decrease in number of treatment days, 32% (48) increase for exclusion of isoniazid, Table 2.2). Baseline factors that increased hazard risk included HIV co-infection (86% (RSE = 29) increase), higher smear grade (68% (36) increase for smear 3+ relative to smear 1+ or negative), male sex (64% (32) increase), cavitary disease (26% (57) increase), and lower BMI (18% (41) increase per 5 kg/m$^2$ decrease). Inclusion of month 2 culture status improved discrimination with an increase in AUROC from 0.69 (95% CI, 0.66-0.72) to 0.72 (0.69-0.75) (Table 2.2). Calibration of the final predictive model

was good, as shown in the visual predictive checks (Figure 2.3 and Appendix Figure A.3 and Appendix Figure **A.4**)

To model probability of non-TB-related outcomes, a Gompertz function was used to describe the hazard risk:

$$h(x_i, t) = \lambda(x_i)\exp(-\alpha t)$$

where $\lambda(x_i)$ is the baseline hazard risk dependent on the covariate values, $x_i$, for each individual, $i$; $\alpha$ is the Gompertz function shape parameter; and $t$ is time in months from start of treatment.

Increasing age was the sole factor that increased the hazard risk of non-TB-related outcomes (23% (RSE = 29) increase per 10 year increase; Table 2.2, Figure 2.3 and Appendix Figure A.5). Because the final model for non-TB-related outcomes were independent of treatment specific factors, derivation and prediction of subsequent risk scores and optimal treatment durations were based solely on the final model for TB-related outcomes.

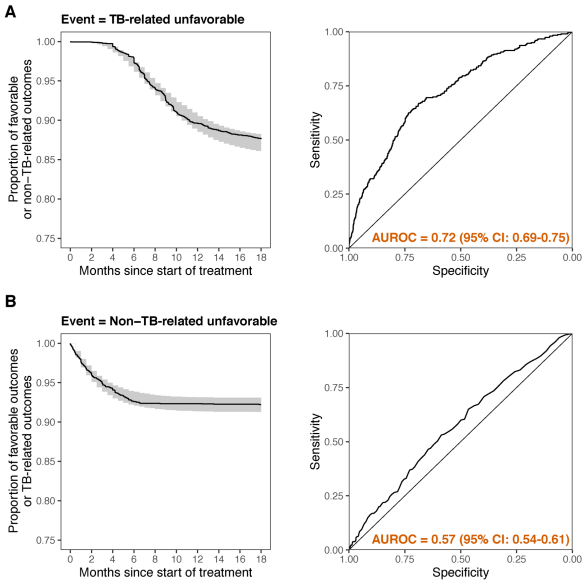



**Figure 2.3** Kaplan-Meier visual predictive checks and receiver operating characteristic curves. (A) Visual predictive check (left) and ROC curve (right) for final model describing time to TB-related outcome. (B) Visual predictive check (left) and ROC curve (right) for final model describing time to non-TB-related outcome.

**Table 2.2** Estimated parameters for models describing TB-related outcomes and non-TB-related outcomes. [a]

| | **Model 1A**[e] **Event** = TB-related outcomes **Predictors** = Treatment days and baseline characteristics | **Model 1B**[e] **Event** = TB-related outcomes **Predictors** = Treatment days, baseline characteristics, and culture | **Model 2** **Event** = Non-TB-related outcomes **Predictors** = Baseline characteristics |
|---|---|---|---|
| **AUROC (95% confidence interval)** | **0.69 (0.66-0.72)** | **0.72 (0.69-0.75)** | **0.57 (0.54-0.61)** |
| **Parameter Description** | **Estimate (RSE[d])** | **Estimate (RSE[d])** | **Estimate (RSE[d])** |
| **Baseline hazard[b], $TV\lambda$** | $10^{-4.0}$ (11) | $10^{-4.1}$ (11) | 0.03 (8) |
| **Shape parameter[b], $\alpha$** | 0.52 (24) | 0.52 (24) | 0.38 (6) |
| **Shape parameter 2[b], $\beta$** | 3.9 (26) | 3.9 (27) | - |
| **Covariate effects [c]: Percent increase in baseline hazard** | | | |
| Per 28 day decrease in number of treatment days | 28 (10) | 29 (9) | - |
| For month 2 culture positivity | - | 145 (19) | - |
| For HIV co-infection | 90 (28) | 86 (29) | - |
| For smear 3+ relative to smear negative or 1+ | 86 (31) | 68 (36) | - |
| For smear 2+ relative to smear negative or 1+ | 23 (91) | 18 (110) | - |
| For male sex | 72 (30) | 64 (32) | - |
| For cavitary disease at baseline | 38 (43) | 26 (57) | - |
| For exclusion of isoniazid in regimen | 30 (51) | 32 (48) | - |
| Per 5 kg/m$^2$ decrease in BMI | 14 (56) | 18 (41) | - |
| Per 10 year increase in age | - | - | 23 (29) |

[a]Model 1A describes TB-related outcomes and included number of treatment days and baseline characteristics as predictors. Model 1B describes TB-related outcomes and included number of treatment days, baseline characteristics, and culture at week 8 as predictors. Model 2 describes non-TB-related outcomes and included baseline characteristics as predictors.

[b]Hazard of TB-related outcomes was described with the surge function, $h(t)_{TB} = \lambda(x_i)t^{\beta}exp(-\alpha t)$,and hazard of non-TB-related outcomes was described with the Gompertz function, $h(t)_{Non-TB} = \lambda(x_i)exp(-\alpha t)$. $\lambda(x_i)$ are the baseline hazard risk parameters and are dependent on the typical value, $TV\lambda$, and covariate vector, $x_i$, for individual $i$, for respective models. $\alpha$ and $\beta$ are shape parameters for respective models.

[c]Covariate effects added using linear relationships. For continuous covariates, the following relationship was used: $P(x) = TVP\big(1 + \theta(COV - COV_{median})\big)$, where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect centered around the covariate median value ($COV_{median}$) and $COV$ is the individual covariate value. For binary covariates, the following relationship was used: $P(x) = TVP\big(1 + \theta(COV)\big)$ where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect for the individual covariate value $COV$ (value of either 0 for reference or 1 for test group). Increased effect (positive covariate effect) on baseline hazard refers to increased hazard risk of poor outcomes in this model.

[d]RSE = relative standard error standardized to parameter estimate (typical value or median).

[e]Adjusted for region of clinic site (African vs. non-African)

**Risk stratification algorithm and optimal treatment durations**

Optimal treatment duration was predicted based on a six-item hazard risk score: month 2 culture, HIV status, baseline smear grade, sex, baseline cavitation, and baseline BMI. The derivations and final formulas to calculate individual risk scores and optimal treatment durations are presented in Chapter 2 Supplemental Results. Based on the predicted optimal treatment durations for each individual to reach a 93% target cure rate, 794/3405 (23%) patients in the model development population were assigned to a low risk group (≤ 18 weeks required for 93% cure rate) with risk scores ranging from 0 to 1.67, 1624/3405 (48%) patients were assigned to a moderate risk group (19-24 weeks required for 93% cure rate) with risk scores ranging from 1.68 to 3.20, and 987/3405 (29%) patients were assigned to a high risk group (> 24 weeks required for 93% cure rate) with risk scores ranging from 3.21 to 14.73. The distribution of individual risk scores and predicted optimal treatment durations in the model development population are shown in Figure 2.4A and Figure 2.4B, respectively. Figure 2.4C illustrates the distribution of different risk factors across the three risk strata. Patients with individual risk factors are still distributed among low, moderate, and high risk groups showing that risk group assignment is dependent on a patient's combination of risk factors, rather than a single variable.

The performance of the risk stratification algorithm is presented in observed Kaplan-Meier rates shown in Figure 2.5 and Appendix Figure A.4. Of the 393 patients who had TB-related outcomes, 49% (194/393) were defined as high risk according to our risk stratification algorithm identified, 38% (151/393) as moderate risk and 12% (48/393) as low risk. Patients in the low risk group treated with either a 4- or 6-month regimen had similar risk of TB-related outcomes [HR: 1.7 (95% CI: 0.9-3.1)], with cure rates above or approximately at the 93% target cure rate threshold. In the moderate risk group, only patients treated with a 6-month regimen resulted in cure rates above 93%, with 4-month regimens leading to significantly more risk of TB-related

outcomes than the 6-month regimens [HR: 3.4 (95% CI: 2.2-5.2)]. Finally, in the high risk group, cure rates after treatment with a 4- or 6-month regimen were below the 93% threshold, with the 4-month regimens leading to significantly higher risk than the 6-month regimen [HR: 2.5 (95% CI: 1.8-3.4)]. No interaction between regimens and risk groups were identified, suggesting that risk of TB-related outcomes increases in higher risk groups independent of treatment duration (P value for interaction = 0.4).

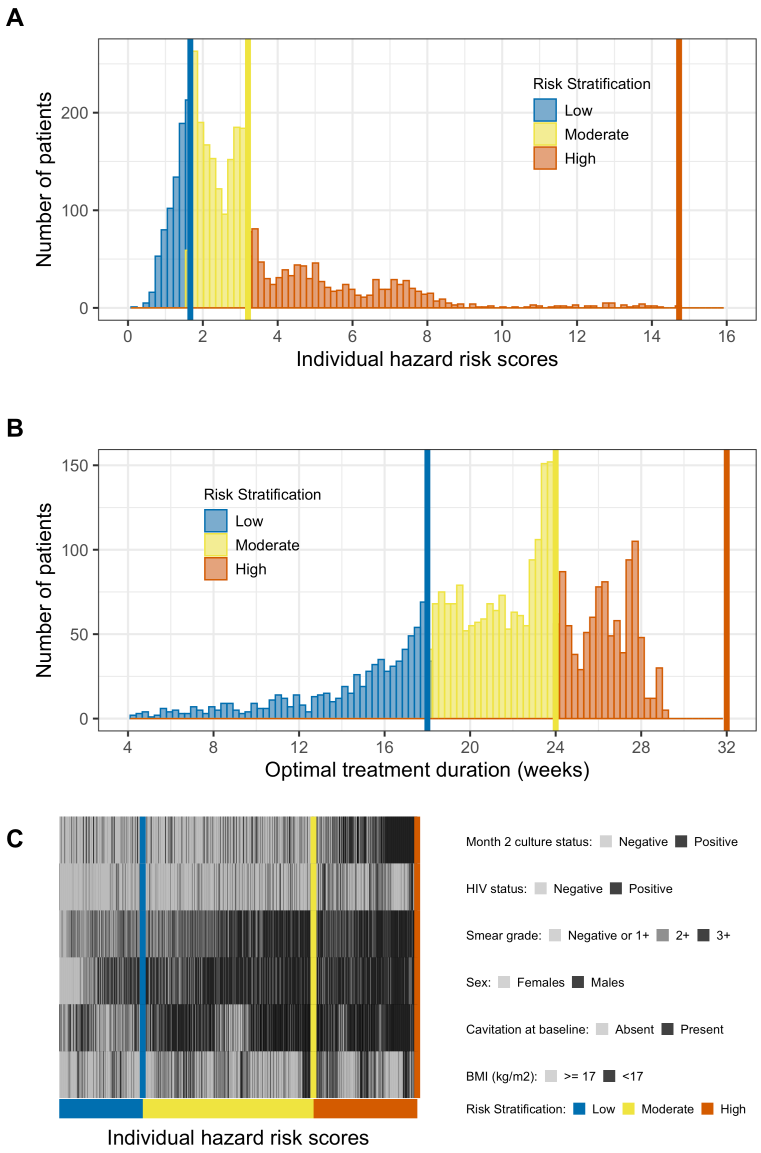



**Figure 2.4** Distribution of individual risk scores, optimal treatment durations for target cure of 93% and risk factors in the model development population. (A) Distribution of individual risk scores stratified by low, moderate, and high-risk group. (B) Distribution of individual optimal treatment durations for target cure rate of 93% at 18 months since start of treatment stratified by low, moderate, and high-risk groups. (C) Heat map distribution of identified risk factors among low, moderate, and high-risk groups. All individuals are arranged on the x-axis from lowest risk score to highest risk score and each column in each row (risk factor) represents a single individual. Low risk group was defined as patients requiring less than or equal to 18 weeks of treatment, moderate risk group requiring 19 to 24 weeks of treatment, and high-risk group requiring more than 24 weeks of treatment for a target cure rate of 93%.

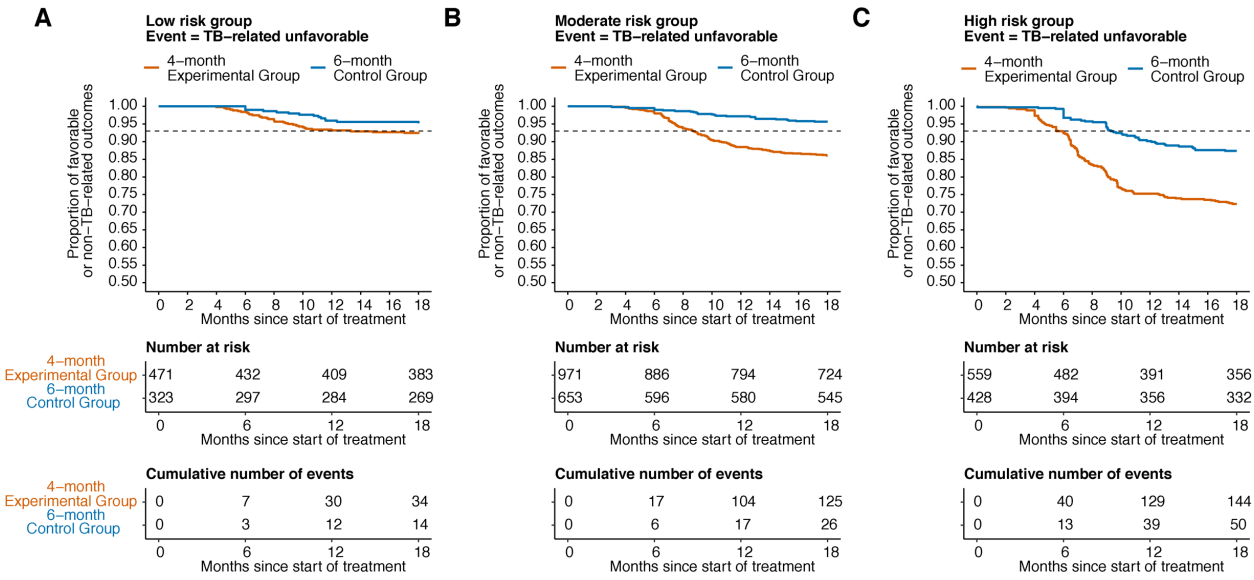



**Figure 2.5** Kaplan-Meier estimates to validate calibration of risk stratification algorithm using model development population for (A) low risk group stratified by regimen duration, (B) moderate risk group stratified by regimen duration, (C) high risk group stratified by regimen duration. Dashed line shows target cure rate of 93% at 18 months since start of treatment.

## External validation

The final TB-related outcome model and stratification algorithm were externally validated using an independent dataset available from the DMID 01-009 trial that includes 386 patients with non-cavitary disease at baseline and culture conversion at month 2. This external dataset represents a sub-population of primarily lower risk patients with 266/386 (69%) patients in the low risk group, 116/386 (30%) in the moderate risk group, and 4/386 (1%) in the high risk group (Figure 2.6). The final TB-related outcome model had similar model discrimination and calibration with the external dataset as compared to the model development dataset (AUROC of 0.78 (95% CI, 0.65-0.90)), Figure 2.7). The observed Kaplan-Meier estimates of TB-related outcomes confirm that patients in the low risk group can be treated with a 4-month regimen and patients in the medium risk group require at least 6 months of treatment to reach 93% target cure rates. No TB-related outcomes were reported in the four patients categorized in the high risk group (Figure 2.8).

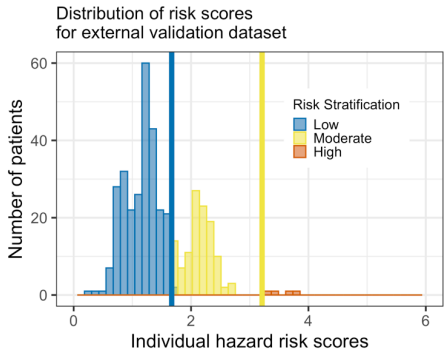



**Figure 2.6** Distribution of risk scores for external validation population. Risk groups are defined by target cure rate of 93% at 18 months since start of treatment.

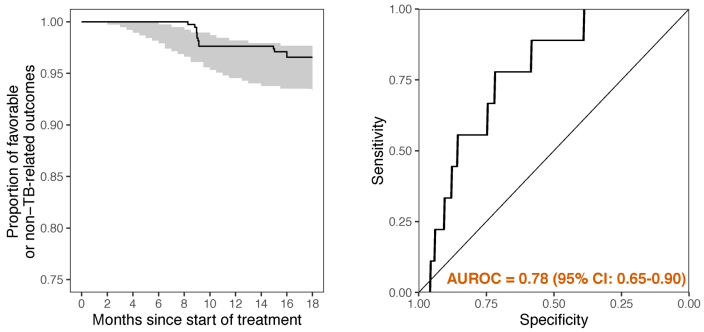



**Figure 2.7** Kaplan-Meier visual predictive checks and receiver operating characteristic curves for external validation dataset. Visual predictive check (left) and ROC curve (right) for final model describing time to tuberculosis related outcome.

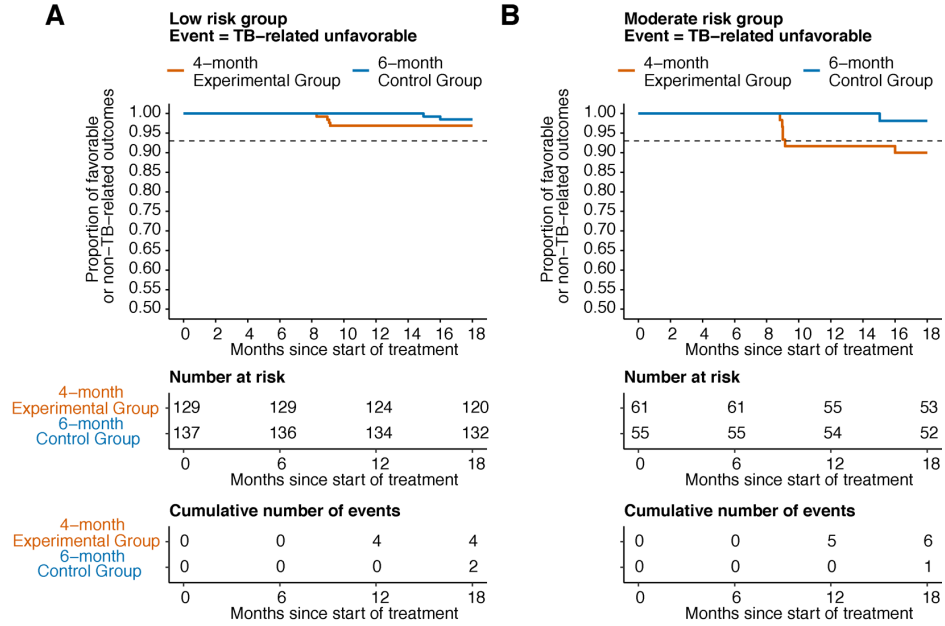



**Figure 2.8** Kaplan-Meier estimates to validate calibration of risk stratification algorithm using external validation population for (A) low risk group stratified by regimen duration and (B) medium risk group stratified by regimen duration. Only 3 individuals in the 4-month experimental group and 1 individual in the 6-month control group, none of which had a Tuberculosis related unfavorable outcome, were categorized as high risk in the external validation dataset so Kaplan-Meier graph is not shown. Dashed line shows target cure rate of 93% at 18 months since start of treatment.

## Clinical simulation tool

We developed a clinical simulation tool based on the final model for TB-related outcomes to provide recommended treatment interventions in stratified groups with the aim to maximize TB clinical cure rates. This interactive tool can generate critical knowledge essential for regimen optimization in a clinical or research setting by highlighting those subgroups of patients who are at higher risk of unfavorable outcomes and may require treatment adjustments. Input parameters for the application include arguments about regimen of interest (composition, duration, and dosing frequency), baseline characteristics, month 2 culture conversion, and patient adherence. Users can assess predicted proportions of patients without TB-related outcomes following 8 to 36 weeks of treatment, patient risk scores (at baseline and when on-treatment data becomes available) to stratify patients, and optimal treatment durations of standard rifampin dose regimens for a target cure rate in specified subgroups. A snapshot of the application is shown in Figure 2.9.

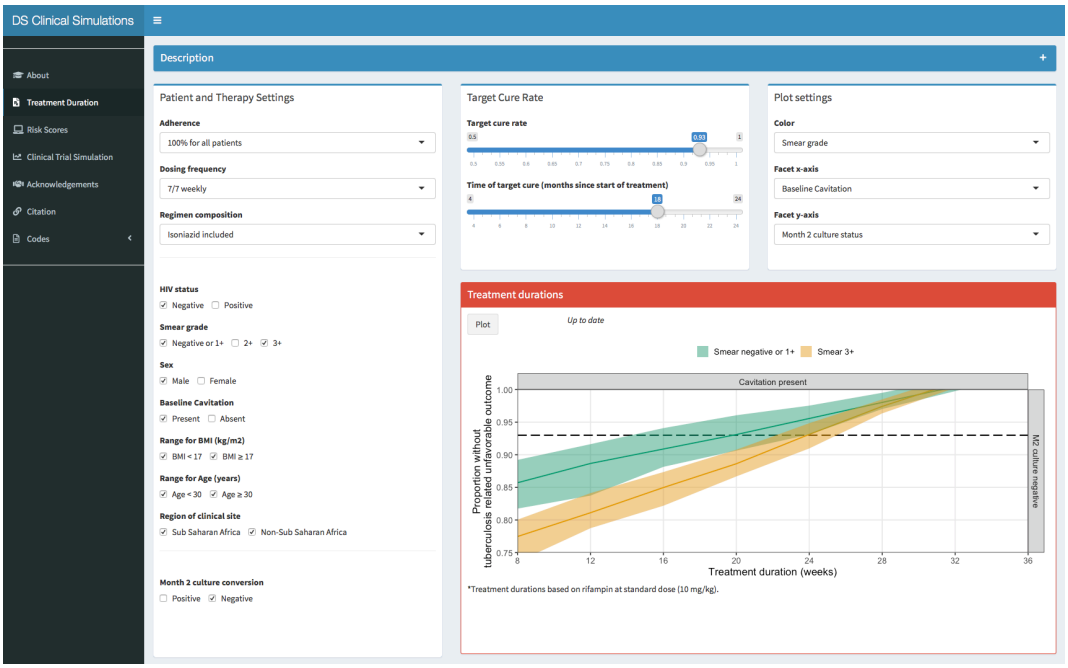



**Figure 2.9** Interactive clinical simulation tool. A page in the clinical simulation tool that displays the predicted proportions of patients without TB-related unfavorable outcomes following 8 to 36 weeks of treatment. In this tab, the user can investigate the optimal treatment duration for various stratified populations.

## Discussion

In this study, we developed and validated a clinical tool that successfully stratified TB patients into low (794/3405, 23%), moderate (1624/3405, 48%), and high (987/3405, 29%) risk groups. In conjunction, through risk stratification, we are able to predict the optimal treatment durations of standard rifampin dose (10mg/kg) regimens for each stratum, where low risk patients can be treated with a 4-month regimen, moderate with a 6-month regimen, and high likely requiring regimens exceeding 6 months without compromising on cure rates. Based on our results, we have developed interactive clinical tools that provide quantitative evidence-informed recommendations on optimal treatment interventions in stratified populations.

Our risk stratification algorithm uses six pragmatic markers of risk that are routinely collected in the clinic: HIV status, baseline bacterial burden, sex, cavitation, BMI, and month 2 culture status. The risk stratification algorithm successfully grouped low risk patients eligible for 4-month standard rifampin dose regimens, moderate risk patients requiring at least 6 months, and high risk patients with suboptimal relapse rates with 6-month standard rifampin dose regimens. Because our external validation dataset only included non-cavitary patients with month 2 culture conversion treated with 4- or 6-month durations, it was only capable of validating the optimal durations for low and medium risk groups. For high risk patients, it was not validated whether standard rifampin dose regimens exceeding six months will result in better treatment outcomes. However, we learned that this high risk group indeed require more effective regimens to reach target cure rates and are likely the cause of unsuccessful shortening of TB treatments when using one-size-fits-all regimens. For example, observed cure rates for this group treated with the 6-month control was 88% (378/428) compared to the low and moderate risk group of 96% (936/976) (Figure 2.5). Possible alternative interventions that can be tested to improve efficacy in these patients include increasing daily rifamycin doses or substituting drugs for those with better lesion

penetration properties and/or more potent bactericidal or sterilizing activity.[5,13,14] These potentially more effective regimens may also allow for ultrashort treatments for low and moderate risk groups.

Presently, only two separate studies have investigated the relationship between treatment duration and rates of relapse. In one study, a meta-regression model was developed based on published historical data to predict rates of relapse using treatment duration and month 2 culture conversion rates as predictors.[15] The model was capable of predicting the expected rates of relapse in the 4-month experimental regimens from the REMoxTB and RIFAQUIN trials, and found that the tested regimens can indeed shorten TB treatment, but only to 5 months. This model predicted wide confidence and prediction intervals suggesting that other important factors were likely unaccounted for in the model making it difficult to make appropriate treatment recommendations in individuals or patient subgroups, particularly in high risk groups who are the main drivers of relapse. In a second study, a translation pharmacokinetic-pharmacodynamic model derived from preclinical mice data was used to predict the results of a number of clinical trials with reasonable success, including the observed 4-month experimental and 6-month control regimens tested in the REMoxTB and RIFAQUIN study.[16] Although, this translational model is valuable in informing regimen optimization and predict outcomes of future and on-going trials, particularly in early stages of regimen development, it also does not take into account established major high risk factors that may influence cure rates. Our model now provides an evidence-based tool developed from individual patient data from four large randomized controlled trials and accounts for established patient and disease specific risk factors. We are now capable of quantitatively predicting, with good precision, rates of TB-related outcomes and provide recommendations on optimal treatment durations in stratified groups to maximize cure.

We developed separate parametric survival models for TB-related and non-TB-related outcomes to determine whether different risk factors affect each outcome. Indeed, treatment and

disease specific risk factors only affected TB-related outcomes and proposed treatment interventions would only improve relapse, treatment failures, and TB-related deaths. Still, non-TB-related outcomes are undesirable because of risk of disease transmission and emergence of drug resistant strains.[17] To improve non-TB-related outcomes, possible interventions include well-informed patients of potential risk of inadequate treatment, use of all efforts (i.e. phone calls, home visits) by clinicians and researchers to contact participants that miss routine visits, and closely monitoring adverse events.

Our study has limitations. First, with modest model discrimination, our risk score still has room for improvement if more quantitative and sensitive measures of disease burden and severity are used. Smear grade to measure baseline bacterial burden and month 2 culture as a putative surrogate biomarker of clinical endpoint both have been proven to have low sensitivity.[18-20] Promising new techniques showing association with baseline and longitudinal bacterial burden include cycle threshold in Gene Xpert assays and quantification of sputum or urine lipoarabinomannan levels.[21-26] Our model can be continually revised to account for more information as additional data become available and reliable relationships between current and promising new markers become established and routinely collected. Second, all tested regimens in the studies included in our analysis used rifampin-based regimens at standard-suboptimal doses. Predicted optimal treatment durations will likely be underestimated when high dose rifamycin regimens are considered, potentially allowing even shorter regimens to be tested. As clinical data on high-dose rifamycin regimens become available this model can be revised. Caution is also advised if generalizing our findings to regimens of other compositions as predictors of relapse may be different. Third, although the proposed six risk factors are routinely collected at baseline or during treatment (month 2 culture), missing values may arise. Our model can predict a risk score for patients with missing values by using a simplified model (without missing factor) or imputing

the missing value based on observed patient distributions. Strengths of our analyses include the inclusion of four large datasets from phase 3 trials conducted across diverse populations in high TB-burden settings; the predictive model is evidence-based and fully parametric with minimal assumptions about the shape of hazard risk; the stratification algorithm is based on pragmatic and routinely available makers; an interactive clinical tool has been developed to handle complex calculations; and our predictive model had similar discrimination and calibration in the model development and validation cohort and that of other predictive models of risk of relapse.

In conclusion, we have developed and validated a risk stratification algorithm capable of categorizing patients into low, medium, and high risk groups. Additionally, an interactive clinical simulation tool has been developed to predict the optimal treatment duration for standard-dose rifampin regimens in stratified populations. Our validated interactive clinical tools allow for more informed decision making and could be used to accelerate regimen development and decision making while maximizing expected cure rates. Finally, our tool can also greatly benefit tuberculosis patients who can now be assigned optimal treatment duration based on their phenotype.

**Supplemental results: Full derivation of individual risk scores and optimal treatment durations**

The surge hazard function, $h(x_i, t)$, best described the instantaneous risk of TB-related outcomes:

$$h(x_i, t) = \lambda(x_i) t^{\beta} exp(-\alpha t)$$

where $\lambda(x_i)$ is the baseline hazard risk dependent on the covariate values, $x_i$, for each individual, $i$; $\alpha$ and $\beta$ are the surge function shape parameters; and $t$ is time in months from start of treatment. The cumulative hazard from 0 to time, t, can be analytically derived using the following:

$$Cumhaz(x_i, t) = \int_0^t h(x_i, t) d(t) = \lambda(x_i) \left[ \frac{t^{\beta} (\alpha t)^{-\beta} [\Gamma(\beta + 1) - \Gamma(\beta + 1, \alpha t)]}{\alpha} \right]$$

$$= \lambda(x_i) g(t, \alpha, \beta)$$

where $\Gamma(a)$ is the gamma function and $\Gamma(a, b)$ is the incomplete gamma function. The second term in the cumulative hazard can be grouped, $g(t, \alpha, \beta)$ and is independent of $x_i$.

Survival (proportion not experiencing a TB-related outcome or proportion cured), $S$, can then be determined by the cumulative hazard with the following:

$$S = \exp(-Cumhaz(x_i, t)) = \exp(-\lambda(x_i) g(t, \alpha, \beta))$$

$\lambda(x_i)$ is defined as the product of the typical value (median of population) of the baseline hazard, $TV\lambda$, and linear relationships with individual covariate values, $x_i$:

$$\lambda(x_i) = TV\lambda f(x_i)$$

$$= TV\lambda \Big[ (1 + EFF_{M2CUL})(1 + EFF_{HIV})(1 + EFF_{Smear})(1 + EFF_{Sex}) \Big( 1 + EFF_{Region} \Big) (1 + EFF_{CAV}) (1 + EFF_{BMI} \left( \frac{18.37}{5} + \frac{BMI_i}{5} \right)) (1 + EFF_{INH}) (1 + EFF_{Rx\ Days} \left( \frac{119}{28} - \frac{No.\ of\ treatment\ days}{28} \right)) \Big]$$

where the effect parameters, $EFF_z$, for each covariate, $z$, are described in Table 2.3, and $BMI_i$ is the invidual's BMI value (kg/m$^2$) and $No. of\ treatment\ days$ is the total number of days drugs are administered .

**Table 2.3** Covariate effect parameters to calculate individual risk scores.

| **Parameter** | **Value** |
|---|---|
| $\boldsymbol{EFF_{M2CUL}}$ | |
| Month 2 culture negative | 0 |
| Month 2 culture positive | 1.45 |
| $\boldsymbol{EFF_{HIV}}$ | |
| Negative | 0 |
| Positive | 0.86 |
| $\boldsymbol{EFF_{Smear}}$ | |
| Negative or 1+ | 0 |
| 2+ | 0.18 |
| 3+ | 0.68 |
| $\boldsymbol{EFF_{Sex}}$ | |
| Female | 0 |
| Male | 0.64 |
| $\boldsymbol{EFF_{Region}}$ | |
| Sub Saharan African | 0 |
| Non-Sub Saharan African | 0.54 |
| $\boldsymbol{EFF_{CAV}}$ | |
| Negative | 0 |
| Positive | 0.26 |
| $\boldsymbol{EFF_{BMI}}$ **($\boldsymbol{BMI_i}$= individual patient's BMI value)** | 0.18 |
| $\boldsymbol{EFF_{INH}}$ | |
| Included | 0 |
| Excluded | 0.26 |
| $\boldsymbol{EFF_{Rx\ Days}}$ | 0.29 |

Individual risk scores, $Risk\ Score_i$, were defined as the product of the $TV\lambda$ and the six identified baseline and on-treatment risk factors (excluding regimen characteristics) and adjusted for region of clinical site. A scaling factor of $10^4$ was added. The following formula was used:

$$Risk\ Score_i = 10^4 TV\lambda(1 + EFF_{M2CUL})(1 + EFF_{HIV})(1 + EFF_{Smear})(1 + EFF_{Sex})\big(1 + EFF_{Region}\big)(1 + EFF_{CAV})(1 + EFF_{BMI}\left(\frac{18.37}{5} + \frac{BMI_i}{5}\right))$$

The baseline hazard risk can then be redefined as:

$$\lambda(x_i) = \frac{Risk\ Score_i}{10^4}(1 + EFF_{INH})(1 + EFF_{Rx\ Days}\left(\frac{119}{28} - \frac{No.of\ treatment\ days}{28}\right))$$

And survival can be redefined as:

$$-\ln(S) = \lambda(x_i)g(t,\alpha,\beta)$$

$$= \frac{Risk\ Score_i}{10^4}(1 + EFF_{INH})(1$$

$$+ EFF_{Pill\ count}\left(\frac{119}{28} - \frac{No.of\ treatment\ days}{28}\right))g(t,\alpha,\beta)$$

Predicted optimal treatment durations are based on the standard regimen that includes isoniazid, rifampin, pyrazinamide, and ethambutol, therefore, $EFF_{INH}$= 0 and $EFF_{Pill\ count}$ = 0.29 (Table 2.3). Rearranging and solving for optimal $No.of\ treatment\ days$ required for a target cure rate, $S = CURE_{target}$, we have:

$$No.of\ treatment\ days = 119 - \left[\frac{28\left(\frac{-\ln(CURE_{target})}{(Risk\ Score_i/10^4)\ (g(t,\alpha,\beta))} - 1\right)}{EFF_{Pill\ count}}\right]$$

Predicted optimal treatment duration (weeks), $TRT\_DURATION_i$, for a 7/7 weekly dosing schedule is then defined as:

$$TRT\_DURATION_i(weeks) = No.of\ treatment\ days/Treatment\ days\ per\ week =$$

$$\frac{1}{7}\left[119 - \left[\frac{28\left(\frac{-\ln(CURE_{target})}{(Risk\ Score_i/10^4)(g(t,\alpha,\beta))} - 1\right)}{EFF_{Pill\ count}}\right]\right]$$

## References


1. Lienhardt C, Nahid P. Advances in clinical trial design for development of new TB treatments: A call for innovation. *PLOS Medicine* 2019; **16**: e1002769.
2. WHO End TB strategy. Target regimen profiles for TB treatment Candidates: rifampicin-susceptible, rifampicin-resistant and pan-TB treatment regimens WHO Library Cataloguing-in-Publication Data Target regimen profiles for TB treatment: candidates: rifampicin-susceptible, rifam. 2016.
3. Ginsberg AM, Spigelman M. Challenges in tuberculosis drug research and development: Commentary. *Nature Medicine* 2007; **13**: 290-4.
4. Imperial MZ, Nahid P, Phillips PPJ, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nature Medicine* 2018; **24**: 1708-15.
5. Strydom N, Gupta SV, Fox WS, et al. Tuberculosis drugs' distribution and emergence of resistance in patient's lung lesions: A mechanistic model and tool for regimen and dose optimization. *PLOS Medicine* 2019; **16**: e1002773.
6. Dorman SE, Nahid P, Kurbatova EV, et al. High-dose rifapentine with or without moxifloxacin for shortening treatment of pulmonary tuberculosis: Study protocol for TBTC study 31/ACTG A5349 phase 3 clinical trial. *Contemporary Clinical Trials* 2020; **90**.
7. Papineni P, Phillips P, Lu Q, Cheung YB, Nunn A, Paton N. TRUNCATE-TB: an innovative trial design for drug-sensitive tuberculosis. *International Journal of Infectious Diseases* 2016; **45**: 404.
8. Merle CS, Fielding K, Sow OB, et al. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med* 2014; **371**: 1588-98.

9. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med* 2014; **371**: 1577-87.

10. Jindani A, Harrison TS, Nunn AJ, et al. High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N Engl J Med* 2014; **371**: 1599-608.

11. Johnson JL, Hadad DJ, Dietze R, et al. Shortening Treatment in Adults with Noncavitary Tuberculosis and 2-Month Culture Conversion. *Am J Respir Crit Care Med* 2009; **180**: 558-63.

12. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016; **133**: 601-9.

13. Dorman SE, Savic RM, Goldberg S, et al. Daily rifapentine for treatment of pulmonary tuberculosis: A randomized, dose-ranging trial. *American Journal of Respiratory and Critical Care Medicine* 2015; **191**: 333-43.

14. Velásquez GE, Brooks MB, Coit JM, et al. Efficacy and Safety of High-Dose Rifampin in Pulmonary Tuberculosis. A Randomized Controlled Trial. *American Journal of Respiratory and Critical Care Medicine* 2018; **198**: 657-66.

15. Wallis RS, Peppard T, Hermann D. Month 2 Culture Status and Treatment Duration as Predictors of Recurrence in Pulmonary Tuberculosis: Model Validation and Update. *PLOS ONE* 2015; **10**: e0125403.

16. Bartelink I, Zhang N, Keizer R, et al. New Paradigm for Translational Modeling to Predict Long-term Tuberculosis Treatment Response. *Clinical and Translational Science* 2017; **10**: 366-79.

17. WHO | Global tuberculosis report 2018. WHO: World Health Organization; 2019.

18. Rockwood N, du Bruyn E, Morris T, Wilkinson RJ. Assessment of treatment response in tuberculosis. *Expert Review of Respiratory Medicine* 2016; **10**: 643-54.

19. Phillips PPJ, Mendel CM, Burger DA, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Medicine* 2016; **14**: 19.

20. Horne DJ, Royce SE, Gooze L, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. *The Lancet Infectious Diseases* 2010; **10**: 387-94.

21. Wallis RS, Kim P, Cole S, et al. Tuberculosis biomarkers discovery: developments, needs, and challenges. *The Lancet Infectious Diseases* 2013; **13**: 362-72.

22. Blakemore R, Nabeta P, Davidow AL, et al. A Multisite Assessment of the Quantitative Capabilities of the Xpert MTB/RIF Assay. *American Journal of Respiratory and Critical Care Medicine* 2011; **184**: 1076-84.

23. Hanrahan CF, Theron G, Bassett J, et al. Xpert MTB/RIF as a Measure of Sputum Bacillary Burden. Variation by HIV Status and Immunosuppression. *American Journal of Respiratory and Critical Care Medicine* 2014; **189**: 1426-34.

24. Shenai S, Ronacher K, Malherbe 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**: e0160062.

25. Kawasaki M, Echiverri C, Raymond L, et al. Lipoarabinomannan in sputum to detect bacterial load and treatment response in patients with pulmonary tuberculosis: Analytic validation and evaluation in two cohorts. *PLOS Medicine* 2019; **16**: e1002780.

26. Friedrich SO, Rachow A, Saathoff E, et al. Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. *The Lancet Respiratory Medicine* 2013; **1**: 462-70.

## Chapter 3: Development and Application of an Integrated Biomarker Clinical Endpoint Tool for Late Stage Tuberculosis Regimen Development and Innovative Clinical Trial Designs


### Abstract

**Background** TB regimen development is plagued with many challenges, the most serious being the inability to identify optimal regimens early and efficiently. We developed integrated models that described the translational link between sputum-based Phase 2B intermediate biomarkers (time to culture conversion, culture conversion status after 8 weeks of treatment, and slope of time to detection profiles from MGIT system after 8 weeks of treatment), patient risk factors, and treatment characteristics to Phase 3 clinical outcomes at an individual level. We applied our tool to recommend minimum Phase 2B treatment efficacy targets and to design innovative late stage trials that maximize success of TB regimen development with the most promising regimens.

**Methods** Data from four Phase 3 trials that compared six novel regimens to the 6-month standard of care for treatment of drug susceptible TB was used to develop and evaluate integrated parametric models. Simulations were performed to assess cure rates of novel regimens with culture conversion hazard ratios of 2 and 3 for treatment durations between 2 and 6 months. Cure rates for adjuvant immunotherapeutic strategies were also assessed. Phase 3 trial designs with the standard one-duration-fits-all approach or stratified medicine approaches were evaluated to maximize trial success. The simulation process involved individual level predictions of Phase 2B outcomes that were sequentially used as input for individual level predictions of Phase 3 outcomes.

**Results** Time to culture conversion better predict Phase 3 outcomes compared to culture conversion status at 8 weeks or slope of time to detection at 8 weeks [AUROC=0.72 (95% CI: 0.68-0.77) vs. 0.67 (0.62-0.71) vs. 0.66 (0.62-0.71)]. Potent regimens with culture conversion

hazard ratios of 3 or more are required to shorten treatment durations to 4 months using a one-duration-fits-all approach. With these potent regimens, a stratified medicine approach may allow for low risk populations to be treated with ultrashort 2-month durations, but high risk populations will still require durations of at least 5.5 months to maximize cure rates. Adjuvant therapies have the potential to reduce relapse by 50% with culture conversion hazard ratios above 1.5. Phase 3 designs with innovative stratified medicine approaches or adjuvant therapeutic strategies have potential to introduce a paradigm shift to the TB regimen development process that include superiority tests.

**Conclusions** We provide a clinical trial simulation tool that can be used to design optimal Phase 2 and Phase 3 trials that permit informed decisions about moving the best regimens forward in the TB regimen development process.

## Introduction

Shorter, efficacious, and better-tolerated oral regimens for tuberculosis (TB) are needed to meet the ambitious goals of the WHO End TB Strategy.[1] The TB research community has identified several new chemical entities with promising preclinical data and are now evaluating novel regimen compositions in the clinical phases.[2] However, the developmental process of new TB regimens remains slow and costly and is plagued by many challenges, the most serious being the inability to identify optimal regimens early and efficiently.[3] Phase 3 trials that recently failed to shorten TB treatment duration from 6 to 4 months using fluoroquinolones has led to a thorough review on how decisions are made to move novel regimens through the developmental process.[4-6]

Late stage TB regimen development rely on sputum-based culture conversion in Phase 2B trials as an intermediate biomarker to predict Phase 3 clinical endpoints of treatment failure and relapse up to 24 months following treatment completion. Predominately, Phase 2B studies have relied on a dichotomous measure of culture conversion after 8 weeks of treatment due to its simplicity and abundance of historical data showing its modest association with long term outcomes.[7,8] New approaches to Phase 2B designs are now being adapted where more quantitative measures are used for early treatment response over the first 8 weeks of treatment with more intensive sampling of sputum at earlier timepoints, such as time to culture conversion or slope of longitudinal quantitative cultures (i.e. time to detection from the Mycobacteria Growth Indicator Tube, MGIT, system; or colony-forming units from the Lowenstein-Jensen, LJ, medium). These endpoints permit smaller trials and are thought to be more reliable by capturing an element of time on treatment.[9,10] However, the translational link between these intermediate biomarkers, treatment duration and Phase 3 clinical outcomes is not fully understood and the minimum treatment effect required in Phase 2B trials (i.e. minimum hazard ratio for time to culture conversion) for novel

regimens to shorten TB treatment is unknown making it difficult to decide which regimens to advance to late stage clinical development.

Epidemiologic models strongly suggest that novel regimens to end the TB epidemic require high cure rates, even more so than treatment shortening. In one study, increasing efficacy from 94% to 99% was predicted to have the greatest impact on reducing mortality, transmission, and burden of disease.[11] Additionally, only a modest population-level effect on 10-year transmission and mortality was predicted for a 4-month regimen with non-inferior cure rates to a 6-month standard of care.[12] Therefore, unless TB treatment is radically shortened to 2 months or less, 4-month regimens that are non-inferior to the 6-month standard regimen will be inadequate to reach the all goals outlined in the WHO End TB Strategy. Rather, superior regimens are required. However, novel regimens are unlikely to achieve higher cure rates than the 93% observed in standard regimens.[4-6] Even if possible, very large sample sizes would be required for Phase 3 trials that are already expensive and time consuming.[13-15] Therefore, innovation to curing TB is required. Stratified medicine approaches where duration is based on patient risk factors or adjuvant immunotherapeutic strategies with host-directed therapies or therapeutic vaccines administered with standard regimens are promising innovative approaches to increase efficacy of treatment regimens.[16,17] However, the optimal clinical trial design of incorporating these innovative approaches have not been evaluated. In this study, we developed integrated parametric models that describe the translational link between sputum-based Phase 2B intermediate biomarkers (time to culture conversion, culture conversion status after 8 weeks of treatment, and slope of time to detection profiles from the MGIT system after 8 weeks of treatment), patient risk factors, and treatment characteristics (composition and duration) to Phase 3 clinical outcomes at an individual level. Then, we applied our tool to recommend minimum Phase 2B treatment efficacy targets and to design innovative late stage trials that maximize success of TB regimen development with the

most promising regimens. Our goal was to provide a clinical trial simulation tool to design optimal Phase 2 and Phase 3 trials that permit informed decisions about moving the best regimens forward to Phase 3 clinical trials.

## Methods

### Data

Individual participant data (n = 4003) from four Phase 3 trials (OFLOTUB, REMoxTB, RIFAQUIN, and DMID 01-009)[4-6,18] that compared six 4- or 6-month novel regimens to the 6-month standard regimen for treatment of drug susceptible TB was used to develop integrated parametric models. The OFLOTUB trial tested a 4-month gatifloxacin-containing regimen, REMoxTB tested two 4-month moxifloxacin-containing regimens, RIFAQUIN tested a 4-month and a 6-month moxifloxacin- and high-dose-rifapentine-containing intermittent regimen, and DMID 01-009 tested a 4-month regimen with standard drugs. Additional information is available in the original publication. [4-6,18]

OFLOTUB, RIFAQUIN, and DMID 01-009 collected sputum samples every one or two months during treatment. REMoxTB serially collected sputum samples weekly to 8 weeks and monthly to end of treatment. In each trial, participants were followed for up to one or two years post treatment with sputum samples collected monthly, 3-monthly, or 6-monthly. For analysis in this study, participants were censored at 18 months after start of treatment.

### Integrated model development

Model development involved two parts: 1. integrated model development to predict Phase 3 clinical outcomes and 2. parametric model development for Phase 2B intermediate biomarkers. In part 1, treatment characteristics, observed individual level Phase 2B outcomes, and patient phenotypes were tested as predictors of Phase 3 clinical outcomes. In part 2, the best Phase 2B

outcome that predicted Phase 3 clinical outcomes from part 1 was modeled with treatment characteristics and patient phenotypes tested as predictors.

***Integrated model development for Phase 3 clinical outcomes***

The Phase 3 clinical outcome used for parametric survival modeling was time to a TB-related outcome; a composite endpoint of treatment failure at the end of treatment, relapse during follow up for 18 months from start of treatment, deaths due to TB, and exogenous reinfection (for the OFLOTUB study only). The definitions of each outcome were taken from the original publications. [4-6,18] Defaults and deaths from a non-TB cause were classified as a missing endpoint and censored in our analysis. Hereafter, a patient who experienced a TB-related outcome is referred to as experiencing a poor outcome and a patient who does not experience a TB-related outcome is referred to have been cured. Cure rates are defined as the proportion of cured patients.

Phase 2B outcomes investigated to be predictors of poor outcomes were culture status after 8 weeks of treatment, time to culture conversion up to 6 months after start of treatment, and slope of time to detection profiles from the MGIT system after 8 weeks of treatment. Time to culture conversion was defined as the time from start of treatment to the first of two negative sputum cultures at different visits without an intervening positive culture, irrespective of whether there were subsequent positive cultures. Slope of time to detection profiles from the MGIT system after 8 weeks of treatment was defined as the difference in time to detection at 8 weeks vs baseline (pretreatment) in days divided by treatment time in days; the slope had units of detection days per treatment days.

Separate analysis was performed based on available data from the five trials. Data from all four trials were pooled to assess the universally available culture status after 8 weeks of treatment as a predictor of poor outcomes. Culture status on LJ solid medium was used because it was available in all four trials. Serially collected sputum samples from the REMoxTB trial was further

analyzed separately to assess culture status after 8 weeks of treatment, time to culture conversion up to 6 months, and slope of time to detection profiles after 8 weeks of treatment all on MGIT liquid medium. In the REMoxTB trial, MGIT liquid cultures and LJ solid cultures were available but MGIT data was used because it is believed to be more sensitive and current trials are more commonly beginning to shift to the use of liquid media.

Predictors of poor outcomes were tested in the following manner: 1) treatment characteristics (treatment duration defined as total number of treatment days drugs were administered and regimen composition) were tested independently and significant factors ($P < 0.01$) were retained; 2) Phase 2B outcomes were added independently and the single most significant factor with the highest area under the receiver operating characteristic curve (AUROC) was retained; and 3) patient risk factors were tested as additional predictors and significant factors were retained using a stepwise covariate search ($p < 0.05$ to include and $p < 0.01$ to remove).

***Parametric model development for Phase 2B intermediate biomarkers***

Time to culture conversion was described with a parametric survival model. Regimen composition and patient risk factors were tested as predictors of time to culture conversion using a stepwise covariate search. The final model was adjusted for geographical region of clinical site (Africa, India, East Asia, South America).

Model building of Phase 2B and Phase 3 outcomes was guided by Kaplan-Meier visual predictive checks to assess calibration and AUROC for discrimination. Model development was performed using the nonlinear mixed effect approach available in the NONMEM program (version 7.4).

**Clinical trial simulations**

The final models that describe Phase 2B and Phase 3 outcomes were integrated to perform clinical trial simulations. The simulation process involved individual level predictions of Phase 2B

outcomes that were sequentially used as input for individual level predictions of Phase 3 outcomes (Figure 3.1). Input parameters for clinical trial simulations include:

i. sample size per treatment arm,
ii. recruitment/enrollment strategies (enriched for low risk patients, enriched for high risk patients, or distributions that reflect current Phase III trials),
iii. sampling intervals and study duration for Phase 2B trials or sample intervals and follow up duration for Phase 3 trials,
iv. individual level patient characteristics,
v. expected improvement in Phase 2B outcomes for novel experimental regimens relative to standard regimen (culture conversion hazard ratio adjusted for risk factors),
vi. expected improvement in poor outcomes after adjusting for improvement in Phase 2B outcomes (percent decrease in relapse after adjusting for culture conversion improvements),
vii. treatment duration with 7 of 7 daily dosing, and
viii. patient adherence (full adherence assigned in this study).

Hereafter, improvements in Phase 2B outcomes described by culture conversion hazard ratios and improvements in poor outcomes described by decrease in relapse are referred to as adjusted improvements after accounting for other identified risk factors in the final models. Low, moderate, and high risk patients were defined based on risk stratification algorithms described in Chapter 2.

Simulations were performed to assess cure rates of novel regimens with culture conversation hazard ratios of 1, 2, and 3 and decreases in poor outcomes between 0 to 33% for treatment durations of 2 to 6 months. A target cure rate of 93%, the observed cure rate in 6-month standard regimen data, was used to compare regimens and inform minimum culture conversion targets required to shorten treatment duration. Simulations were also performed to predict cure

rates for adjuvant immunotherapies (with host-directed therapies or therapeutic vaccines) that are given in combination with the 6-month standard regimen with culture conversion hazard ratios of 1, 2, and 3. A prolonged duration effects between 1 to 2 weeks was also included to describe the prolonged immunological response from adjuvant immunotherapies. For adjuvant immunotherapies, a target of 50% reduction in poor outcomes compared to the standard regimen alone was used. The effect of enriched recruitment strategies, uncertainty in effect size of Phase 2B outcomes, and uncertainty in the translational link between Phase 2B and Phase 3 outcomes on predicted cure rates was assessed.

Via simulation study, we evaluated operating characteristics of Phase 3 trial designs for novel regimens. The performance of Phase 3 trials was characterized by cure rates of novel regimens at 18 months after start of treatment, power to detect non-inferior results relative to the 6-month standard regimen with a non-inferior margin of 6%,[4-6] and power to detect superior results relative to 6-month standard regimen. Performance was compared between one-duration-fits-all and stratified medicine approaches for novel regimens with culture conversion hazard ratios of 2 and 3. For one-duration-fits-all approaches, a 4-month treatment duration was used. For stratified medicine approaches, low risk patients were treated with 2 months, moderate risk patient with 4 months, and high risk patients with 6 months. Designs were also characterized by i.) sample sizes of 250 and 500 patients per arm and ii.) recruitment/enrollment strategies that are consistent with current Phase 3 trial populations, are enriched for 80% low risk patients, or are enriched for 80% high risk patients.

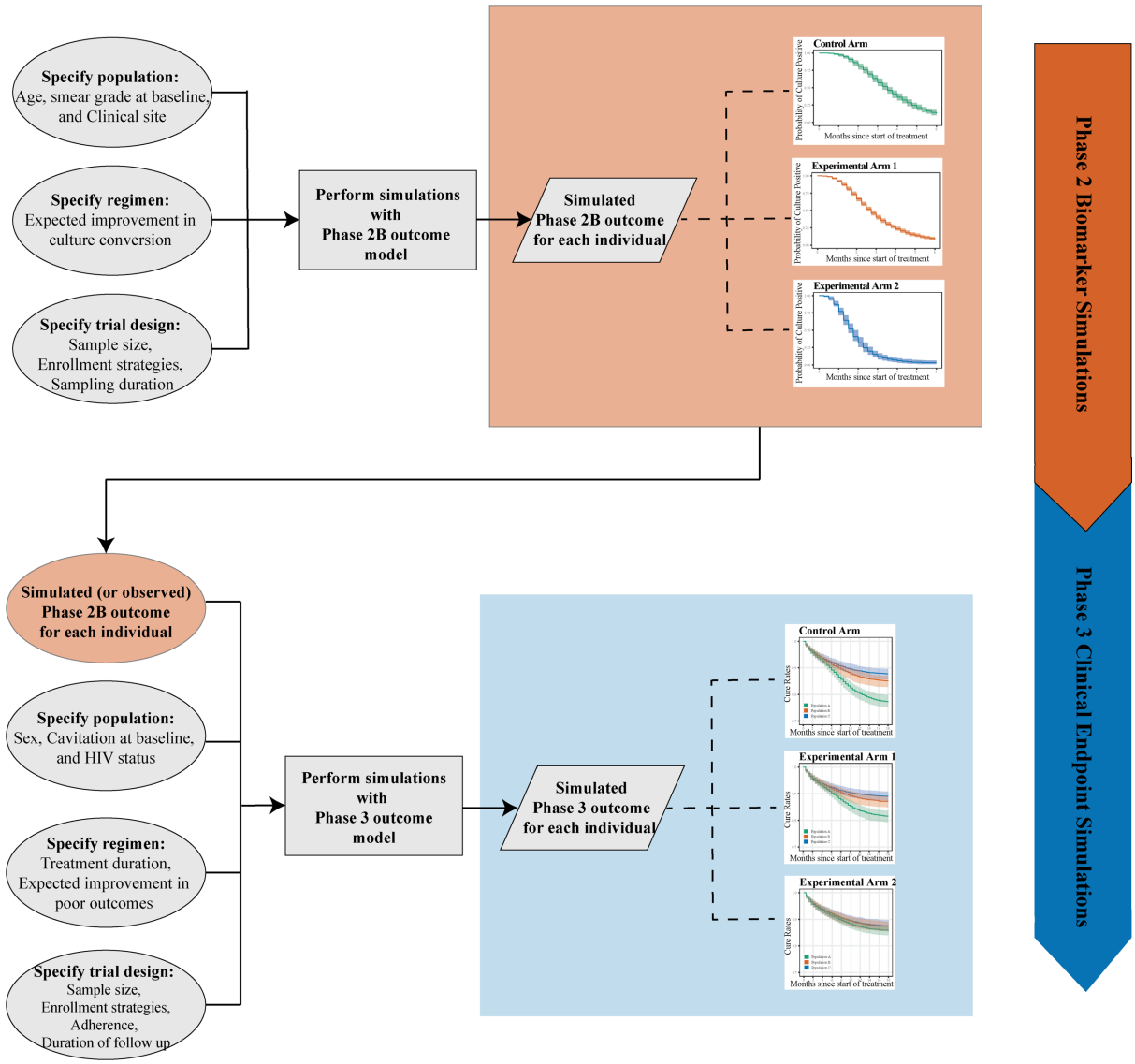



**Figure 3.1** Clinical trial simulation workflow. Clinical trial simulations involved individual level predictions of Phase 2B outcomes using Model 8 described in Table 3.4 that were sequentially used as input for individual level predictions of Phase 3 outcomes using Model 7 described in Table 3.3. For Phase 2B outcomes, individual level patient characteristics, regimen characteristics, and trial design characteristics were specified for model simulations. For Phase 3 outcomes, simulated Phase 2B outcomes (or observed values can be used if available), individual patient characteristics, regimen characteristics, and trial design characteristics were specified for model simulations. Output for each simulation are individual level Phase 2B or Phase 3 outcomes.

## Results

### Data characteristics

Data was pooled from four Phase 3 trials (N = 4003) that compared six novel experimental regimens (one 6-month regimen and five 4-month regimens) to the 6-month standard regimen. In solid medium, 643 of 4003 pooled study participants (16%) had positive culture after 8 weeks of treatment. For REMoxTB participants, 764 of 1674 study participants (46%) had positive culture after 8 weeks of treatment, median time to culture conversion was 56 days and median slope of time to detection was 3.3 detection days per treatment day, all in liquid medium (Table 3.1).

**Table 3.1** Data characteristics

| **Characteristic** | **N = 4003** |
|---|---|
| **Treatment regimen N (%)[a]** | |
| **6-month regimens** | |
| 2HRZE/4HR (standard regimen) | 1597 (40) |
| 2EMRZ/4$P_1M_1$ | 212 (5) |
| **4-month regimens** | |
| 2HRZE/2HR | 193 (5) |
| 2HRZG/2HRG | 689 (17) |
| 2HRZM/2MHR | 568 (14) |
| 2EMRZ/2MR | 551 (14) |
| 2EMRZ/2$P_2M_2$ | 193 (5) |
| **Geographical region of clinical site** | |
| Africa | 3228 (81) |
| India | 342 (9) |
| East Asia | 274 (7) |
| South America | 159 (4) |
| **Female sex - no. (%)** | 1234 (31) |
| **Age- yrs[b]** | |
| Median | 29 |
| Interquartile range | 24-38 |
| Range | 16-81 |
| **Weight- kg** | |
| Median | 52 |
| Interquartile range | 47-58 |
| Range | 32-137 |
| **Body mass index[c]** | |
| Median | 18.5 |
| Interquartile range | 16.0-20.4 |
| Range | 12.0-50.9 |
| **HIV positivity - no. (%)[d]** | 517 (13) |

**Table 3.1 continued.**

| Characteristic | N = 4003 |
|---|---|
| **Cavitation- no. (%)[e]** | 2219 (55) |
| **Smear- no. (%)[f]** | |
| Negative or 1+ | 1118 (28) |
| 2+ | 1009 (25) |
| 3+ | 1833 (45) |
| **Culture positivity after 8 weeks in solid medium[g]** | 643 (16) |
| **Culture positivity after 8 weeks liquid medium (REMoxTB only)[h] 196** | 764 (46) |
| **Time to culture conversion (days) (REMoxTB only)[h]** | |
| Median | 56 |
| Interquartile range | 42-90 |
| Range | 1-168 |
| **Slope of time detection profiles after 8 weeks (detection days/treatment days) (REMoxTB only)[h]** | |
| Median | 3.3 |
| Interquartile range | 1.5-4.7 |
| Range | -4.9-5.4 |

[a]H = isoniazid, Z = pyrazinamide, E = ethambutol, R = rifampin, P = rifapentine, M =moxifloxacin, G=gatifloxacin. Subscripts on P and M define rifapentine and moxifloxacin weekly dosing; 1, once weekly rifapentine (1200 mg) and moxifloxacin (400 mg); 2, twice weekly rifapentine (900 mg) and moxifloxacin (400 mg). All other drugs were administered daily according to published guidelines. First set of drugs represent 2-month intensive phase regimen and second set of drugs represent 4-month or 2-month continuation phase regimen.
[b]Age was missing in 5 participants (<1%).
[c]Body mass index was defined as the weight in kilograms divided by the squared height in meters.
[d]HIV status was missing in 9 study participants (<1%).
[e]Cavitation status was missing for 215 study participants (5%).
[f]Smear grade was missing for 43 study participants (1%).
[g] Culture status in Lowenstein-Jensen solid medium. Culture status after 8 weeks of treatment was missing for 340 study participants (8%).
[h] Culture data from Mycobacterial Growth Indicator Tube (MGIT) liquid medium used in separate analysis with REMoxTB trial data only. Based on N=1674. Culture status after 8 weeks of treatment was missing for 196 study participants (12%). Time to culture conversion was missing for 142 study participants (8%). Slope of time to detection after 8 weeks of treatment was missing for 196 study participants (12%).

## Integrated model development

Culture status after 8 weeks of treatment had modest AUROC for predicting poor outcomes in the pooled data from four Phase 3 trials (0.67 (95% confidence interval, CI, 0.63-0.70); Table 3.2) and in the REMoxTB data only (0.67 (0.62-0.71) (Table 3.3). Comparing more quantitative Phase 2B outcomes, time to culture conversion had higher AUROC for predicting poor outcomes than the slope of time to detection profiles after 8 weeks of treatment (0.72 (0.68-0.77) vs. 0.66 (0.62-0.71), Table 3.3). Integrating time to culture conversion, treatment duration and identified

patient risk factors (HIV status, sex, and cavitation at baseline) further improved the prediction of poor outcomes (Table 3.3 and Figure 3.2).

A time to culture conversion model was developed to link to the Phase 3 clinical outcome model. Patient risk factors that prolonged risk of culture positivity were higher smear grade at baseline (60% (relative standard error, RSE=9%) decrease in culture conversion hazard for smear 3+ relative to smear 1+ or negative) and older age (8% (RSE = 21) decrease in culture conversion hazard per 10 year increase in age). The final model that predicts time to culture conversion had an AUROC of 0.73 (0.70-0.75) (Table 3.4). Calibration of the final integrated models were good, as shown in the visual predictive checks (Figure 3.2 and Figure 3.3).

**Table 3.2** Model parameter estimates for time to poor outcome in data pooled from four trials (DMID 01-009, OFLOTUB, REMoxTB, RIFAQUIN).

| | **Model 1** Treatment Duration | **Model 2** Treatment duration and culture status |
|---|---|---|
| **AUROC (95% confidence interval)** | **0.60 (0.58-0.64)** | **0.67 (0.63-0.70)** |
| **Parameter Description** | **Estimate (RSE[c])** | **Estimate (RSE[c])** |
| **Baseline hazard[a], $TV\lambda$** | $10^{-3.6}$ (13) | $10^{-3.7}$ (12) |
| **Shape parameter[a], $\alpha$** | 0.52 (24) | 0.52 (24) |
| **Shape parameter 2[a], $\beta$** | 3.9 (26) | 3.9 (26) |
| **Linear effect on baseline hazard (%) [b]** | | |
| Per 28 day decrease in treatment duration | 27 (10) | 30 (8) |
| For culture positivity after 8 weeks in solid medium | - | 190 (16) |

[a]Hazard of poor outcomes was described with the surge function, $h(t) = \lambda(x)t^{\beta}exp(-\alpha t)$. $\lambda(x)$ is the baseline hazard risk and is dependent on the typical value, $TV\lambda$, and covariate vector, $x$. $\beta$ and $\alpha$ are shape parameters.

[b]Covariate effects added using linear relationships. For continuous covariates, the following relationship was used: $P(x) = TVP\left(1 + \theta(COV - COV_{median})\right)$, where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect centered around the covariate median value ($COV_{median}$) and $COV$ is the individual covariate value. For binary covariates, the following relationship was used: $P(x) = TVP\left(1 + \theta(COV)\right)$ where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect for the individual covariate value $COV$ (value of either 0 for reference or 1 for test group). Increased effect (positive covariate effect) on baseline hazard refers to increased hazard risk of poor outcomes in this model.

[c]RSE = relative standard error standardized to parameter estimate (typical value or median), units in %

-Parameter not included in model.

**Table 3.3** Model parameter estimates for time to poor outcome in the REMoxTB data only

| | **Model 3** Treatment Duration | **Model 4** Treatment duration and culture status | **Model 5** Treatment duration and slope of time to detection | **Model 6** Treatment duration and time to culture conversion | **Model 7** Treatment duration, time to culture conversion and patient risk factors |
|---|---|---|---|---|---|
| **AUROC (95% confidence interval)** | **0.61 (0.58-0.65)** | **0.67 (0.62-0.71)** | **0.66 (0.62-0.71)** | **0.72 (0.68-0.77)** | **0.74 (0.70-0.79)** |
| **Parameter Description** | **Estimate(RSE[d])** | **Estimate (RSE[d])** | **Estimate(RSE[d])** | **Estimate (RSE[d])** | **Estimate (RSE[d])** |
| **Baseline hazard[a], $TV\lambda$** | $10^{-2.8}$ (16) | $10^{-3.0}$ (15) | $10^{-2.8}$ (16) | $10^{-3.0}$ (15) | $10^{-3.4}$ (13) |
| **Shape parameter[a], $\alpha$** | 0.34 (40) | 0.34 (40) | 0.34 (40) | 0.34 (40) | 0.34 (40) |
| **Shape parameter 2[a], $\beta$** | 2.4 (44) | 2.4 (44) | 2.5 (0.44) | 2.4 (44) | 2.5 (44) |
| **Linear effect on baseline hazard (%) [b]** | | | | | |
| Per 28 day decrease in treatment duration | 29 (11) | 30 (10) | 30 (10) | 30 (13) | 30 (9) |
| For culture positivity after 8 weeks | - | 127 (26) | - | - | - |
| Per detection day decrease in slope of time to detection after 8 weeks | - | - | 22 (22) | - | - |
| For HIV co-infection | - | - | - | - | 129 (39) |
| For male sex | - | - | - | - | 84 (39) |
| For cavitary disease at baseline | - | - | - | - | 75 (41) |
| **Exponential effect on baseline hazard[c]** | | | | | |
| For 28 day increase in time to culture conversion | - | - | - | 0.31 (13) | 0.29 (15) |

[a]Hazard of poor outcomes was described with the surge function, $h(t) = \lambda(x)t^{\beta}exp(-\alpha t)$. $\lambda(x)$ is the baseline hazard risk and is dependent on the typical value, $TV\lambda$, and covariate vector, $x$. $\beta$ and $\alpha$ are shape parameters.

[b]Covariate effects added using linear relationships. For continuous covariates, the following relationship was used: $P(x) = TVP\left(1 + \theta(COV - COV_{median})\right)$, where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect centered around the covariate median value ($COV_{median}$) and $COV$ is the individual covariate value. For binary covariates, the following relationship was used: $P(x) = TVP\left(1 + \theta(COV)\right)$ where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect for the individual covariate value $COV$ (value of either 0 for reference or 1 for test group). Increased effect (positive covariate effect) on baseline hazard refers to increased hazard risk of poor outcomes in this model.

[c] Covariate effects added using exponential relationships. For continuous covariates, the following relationship was used: $P(x) = TVP\left(1 + e^{\theta(COV - COV_{median})}\right)$, where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect centered around the covariate median value ($COV_{median}$) and $COV$ is the individual covariate value. Increased effect (positive covariate effect) on baseline hazard refers to increased hazard risk of poor outcomes in this model.

[d]RSE = relative standard error standardized to parameter estimate (typical value or median), units in %

-Parameter not included in model.

**Table 3.4** Model parameter estimates for time to culture conversion. [a]

| | **Model 8**<br>Patient risk factors |
|---|---|
| **AUROC (95% confidence interval)** | **0.73 (0.70-0.75)** |
| **Parameter Description** | **Estimate (RSE[d])** |
| **Baseline hazard[b], $TV\lambda$** | 1.0 (14) |
| **Shape parameter[b], $\alpha$** | 0.93 (10) |
| **Shape parameter 2[b], $\beta$** | 2.2 (7) |
| **Linear effect on baseline hazard (%)[c]** | |
| For smear 3+ relative to smear negative or 1+ | -60 (9) |
| For smear 2+ relative to smear negative or 1+ | -18 (71) |
| For 10-year increase in age | -8 (21) |
| **Linear effect on shape parameter[c]** | |
| For smear 3+ relative to smear negative or 1+ | -30 (17) |

[a] The final model was adjusted for geographical region of clinical site (Africa, India, East Asia, South America).

[b]Hazard of poor outcomes was described with the surge function, $h(t) = \lambda(x)t^{\beta}exp(-\alpha(x)t)$. $\lambda(x)$ is the baseline hazard risk and is dependent on the typical value, $TV\lambda$, and covariate vector, $x$. $\alpha(x)$ is the shape parameter and is dependent on the typical value, $TV\alpha$, and covariate vector, $x$. $\beta$ is the second shape parameter.

[c]Covariate effects added using linear relationships. For continuous covariates, the following relationship was used: $P(x) = TVP\big(1 + \theta(COV - COV_{median})\big)$, where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect centered around the covariate median value ($COV_{median}$) and $COV$ is the individual covariate value. For binary covariates, the following relationship was used: $P(x) = TVP\big(1 + \theta(COV)\big)$ where $TVP$ is the typical value for parameter $P$, $\theta$ is the reported covariate effect for the individual covariate value $COV$ (value of either 0 for reference or 1 for test group). Decreased effect (negative covariate effect) on baseline hazard and shape parameters refers to prolonged culture conversion in this model.

[d]RSE = relative standard error standardized to parameter estimate (typical value or median), units in %

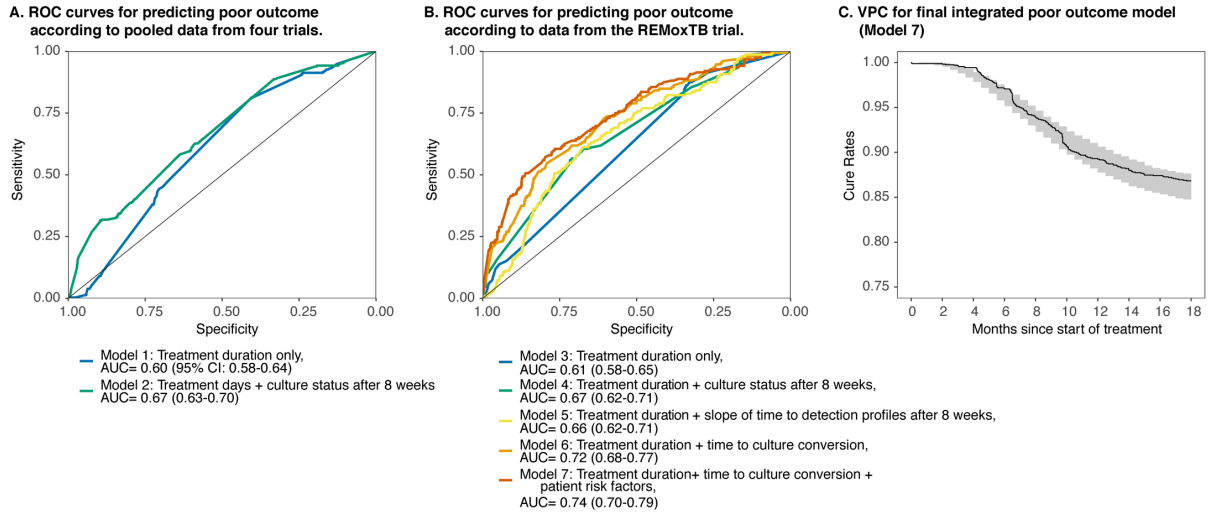



**Figure 3.2** Receiver operating characteristic curves and Kaplan Meier visual predictive checks for time to poor outcome models. (A) ROC curves for predicting time to poor outcome according to pooled data from four Phase 3 trials (DMID 01-009, OFLOTUB, REMoxTB, RIFAQUIN). (B) ROC curves for predicting time to poor outcome according to data from REMoxTB trial only. (C) VPC for final integrated poor outcome model (Model 7). Model 1 and 3 included treatment duration only, Model 2 and 4 included treatment duration and culture status after 8 weeks of treatment (solid medium for Model 2 and liquid medium for Model 4), Model 5 included treatment duration and slope of time to detection profiles after 8 weeks of treatment, Model 6 included treatment duration and time to culture conversion, and Model 7 included treatment duration, time to culture conversion, and patient risk factors.

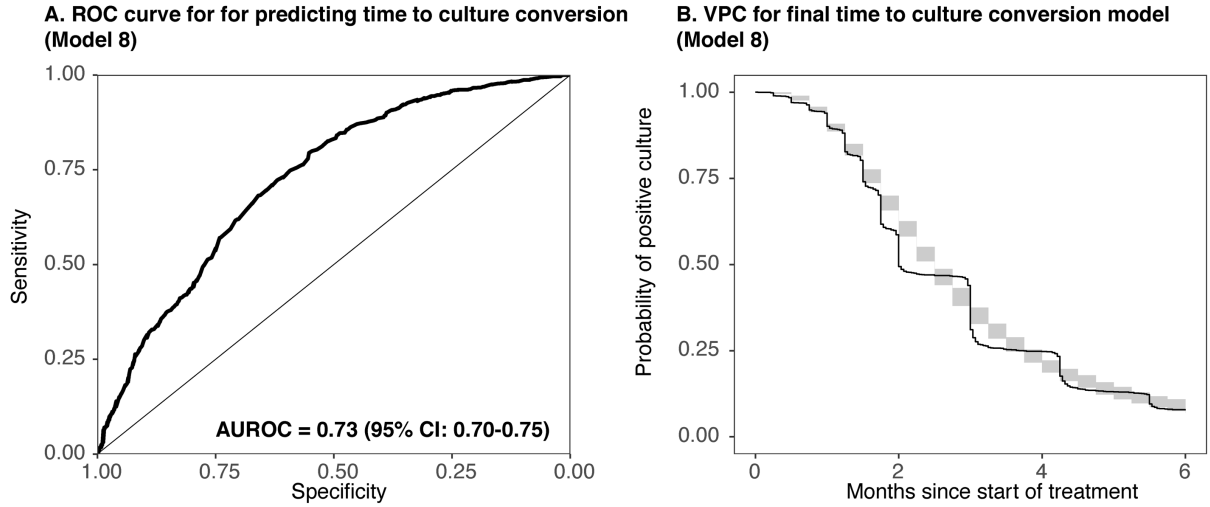



**Figure 3.3** Receiver operating characteristic curves and Kaplan-Meier visual predictive checks for final time to culture conversion model. (A) ROC curve for predicting time to culture conversion and (B) VPC for final time to culture conversion model

**Phase 2B treatment efficacy targets**

Via simulation study, the final integrated models were used to inform treatment durations required for novel regimens with culture conversion hazard ratios of 1, 2, and 3 relative to standard regimen. For a one-duration-fits-all approach, as would be expected, treatment durations of 6 months are required for regimens with a culture conversion hazard ratio of 1 (i.e. the same as the standard regimen) to reach target cure rates of 93%. More potent regimens with culture conversion hazard ratios of 2 and 3 were predicted to require treatment durations of 4.5 to 5.5 and 4 to 5.5 months to reach target cure rates, respectively (Figure 3.4A). Stratifying patients by risk group showed potential for the low risk subgroup to be treated for as low as 2 months with highly potent regimens (i.e. hazard ratio of 3). However, the moderate risk subgroup requires durations of at least 3.5 months to reach target cure rates and the high risk subgroup requires at least 5.5 months (Figure 3.4B). For adjuvant immunotherapies that are co-administered with the standard 6-month regimen, a minimum culture conversion hazard ratio of 1.5 is required to reach a 50% reduction in poor outcomes compared to the standard regimen alone (Figure 3.5).

The effect of enrollment/recruitment strategies, uncertainty in culture conversion hazard ratios, and uncertainty in the translational link between Phase 2B to Phase 3 outcome on predicted cure rates was assessed for a 4-month regimen with a culture conversion hazard ratio of 2. Recruitment strategies had the largest effect on predicted cure rates. Compared to recruitment reflecting recent Phase 3 trials, enrichment for 80% low risk patients led to 4.2 percentage points above the true cure rate and enrichment for 80% high risk patients led to 4.8 percentage points below the true cure rate. Uncertainty in culture conversion hazard ratios also had large effects on predicted cure rates. If the estimated culture conversion hazard ratio from a Phase 2B trial is at the $5^{th}$ percentile of the uncertainty distribution, the estimated cure rate can be as low as 3.3 points

below the expected true cure rate. The uncertainty in the translational link between Phase 2B to Phase 3 outcomes had marginal effects on predicted cure rates (Figure 3.6).

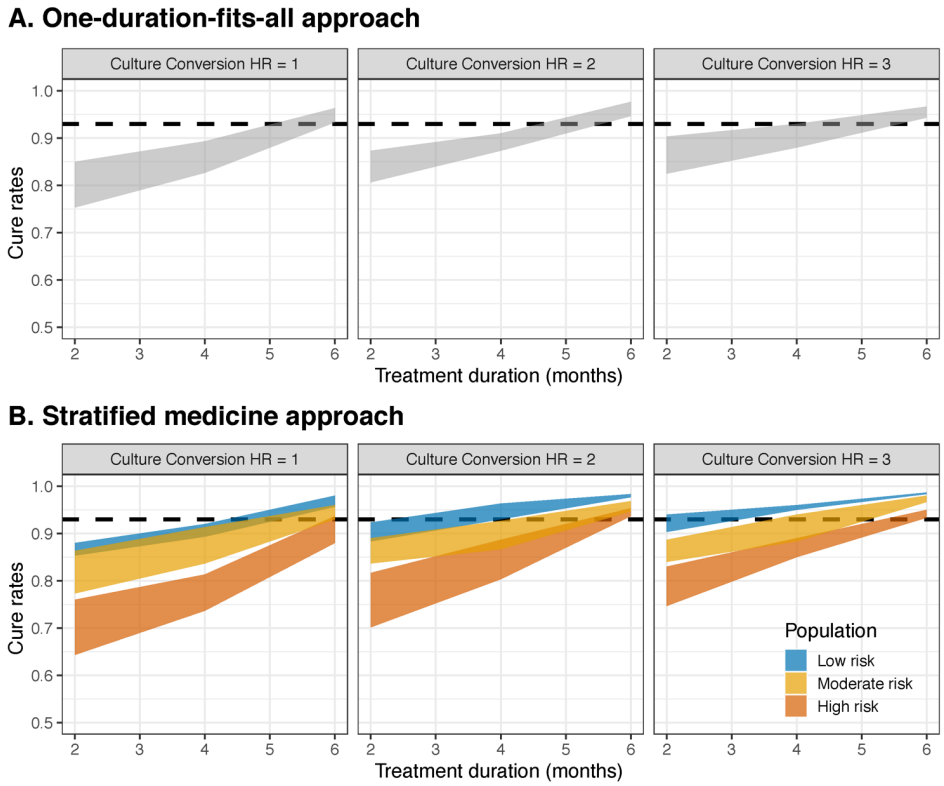



**Figure 3.4** Optimal treatment durations for novel regimens. Cure rates for novel regimens with Phase 2B culture conversion hazard ratios (HR) of 1 (left), 2 (middle), and 3 (right) administered for 2 to 6 month durations. (A) Simulations for one-duration-fits-all approach. (B) Simulations by risk group for a stratified medicine approach. Shaded areas represent regimens with 0 to 33% decrease in relapse after adjusting for culture conversion improvements.

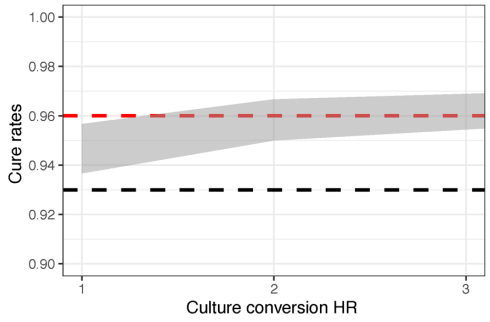



**Figure 3.5** Optimal culture conversion hazard ratios for adjuvant immunotherapeutic strategies. Cure rates for adjuvant immunotherapies co-administered with the standard 6-month regimen for Phase 2B culture conversion hazard ratios (HR) of 1 to 3 are shown. Shaded area represents adjuvant therapies with prolonged duration effects of 1 to 2 weeks.

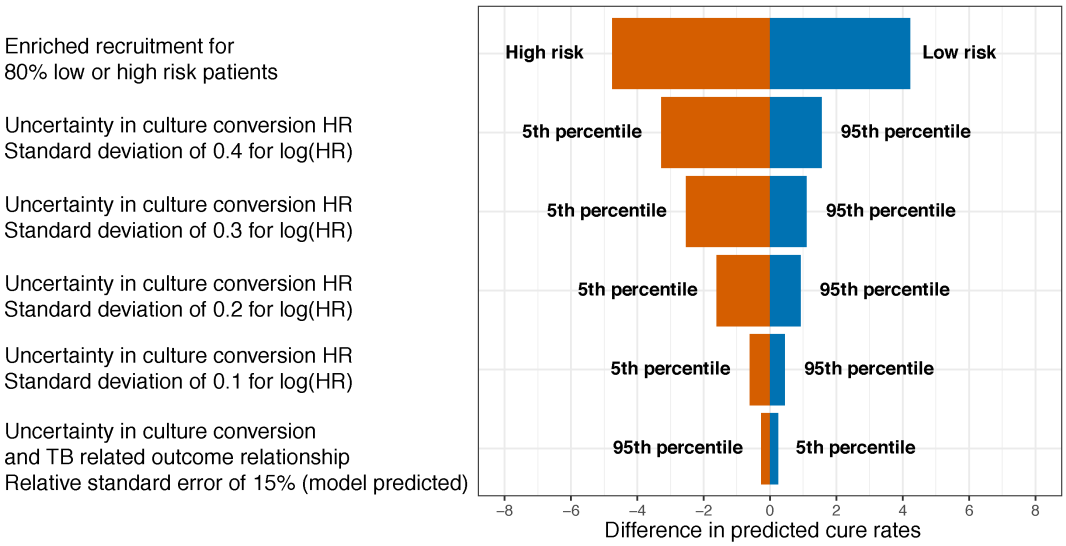



**Figure 3.6** Effect of enrollment/recruitment strategies, uncertainty in culture conversion hazard ratio, and uncertainty in the model parameter that links Phase 2B to Phase 3 outcomes on predicted cure rates for a novel 4-month regimen with a culture conversion hazard ratio (HR) of 2 relative to the 6-month standard regimen. Reference recruitment reflected distributions observed in recent Phase 3 trials. Predicted cure rates for enriched enrollment/recruitment strategies with 80% low risk patients or 80% high risk patients, the 5th and 95th percentile of culture conversion HR uncertainty with standard deviations of 0.1, 0.2, 0.3, and 0.4 for log(HR), and the 5th and 95th percentile of the estimated model parameter uncertainty that links Phase 2B to Phase 3 outcomes (median = 0.29, relative standard error of 15%, see Table 3.3 Model 7)were compared to true cure rates.

### Phase 3 clinical trial simulations

The clinical trial simulation tool was used to identify Phase 3 clinical trial designs and novel regimens that maximize trial success and cure rates with current one-duration-fits-all and innovative stratified medicine approaches. Simulations showed that one-duration-fits-all superiority trial designs for novel 4-month regimens with culture conversion hazard ratios of 2 or 3 are unlikely to be successful, with powers below 10% (Figure 3.7A and Figure 3.7B, bottom panels). However, stratified medicine designs with a highly potent regimen that has a culture conversion hazard ratio of 3 is predicted to have 80% power to show superiority if the trial has 500 participants per arm and is enriched for high risk patients (Figure 3.7D, bottom panel). For the same potent regimen, one-duration-fits-all and stratified medicine trial designs with 500 participants per arm had at least 90% power to show non-inferiority (Figure 3.7B and Figure **3.7**D, middle panels). Comparing enrollment/recruitment strategies, one-duration-fits-all non-inferiority

designs with enrichment for high risk patients had decreased power compared to enrichment for low risk patients (Figure 3.7A and Figure 3.7B, middle panels). However, for stratified medicine non-inferiority trial designs, enrichment for high risk patients had increased power compared to enrichment for low risk patients (Figure 3.7C and Figure 3.7D, middle panels).

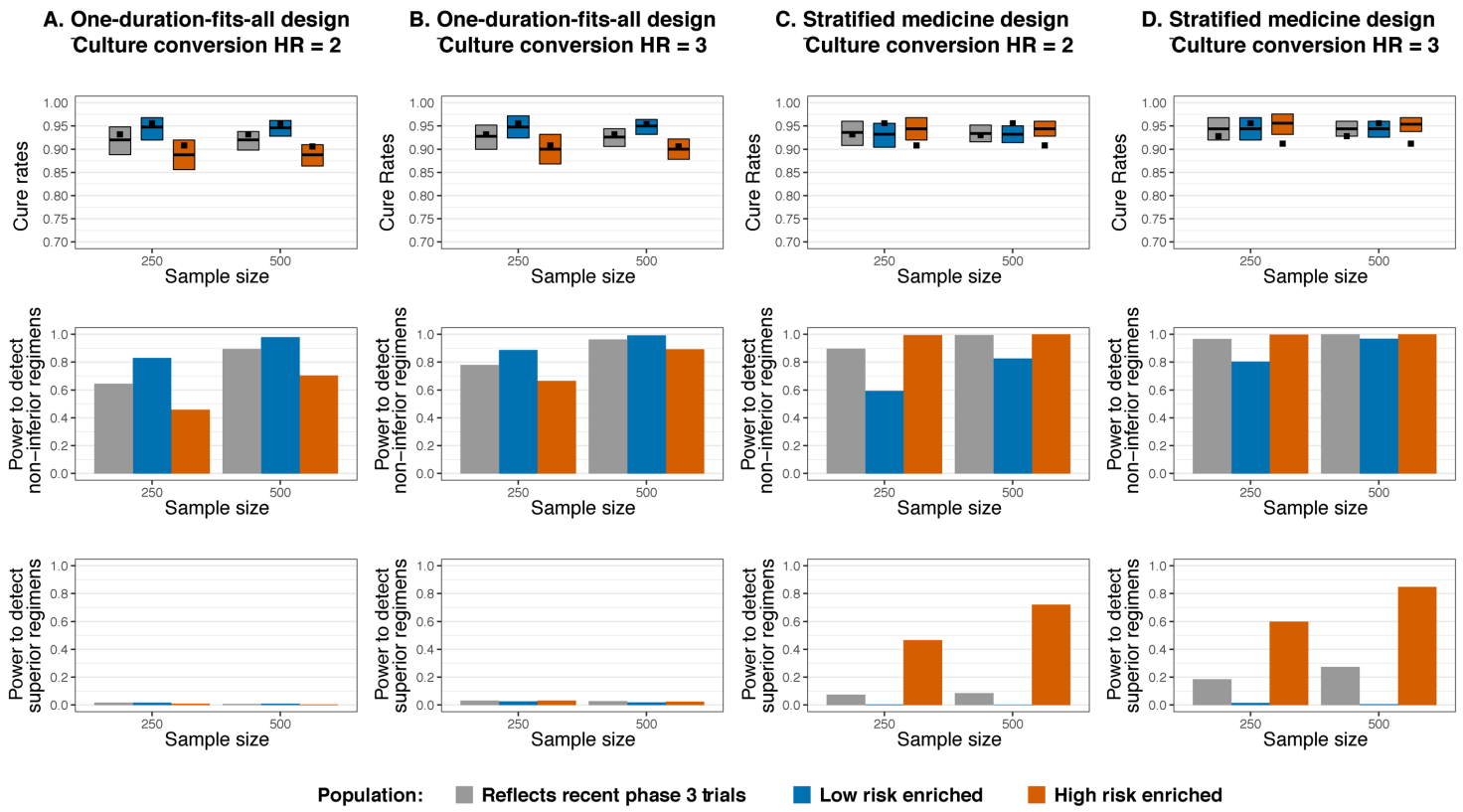



**Figure 3.7** Clinical trial simulations for Phase 3 trials**.** 5000 simulations were performed for clinical trial designs with combinations of sample sizes of 250 or 500; and enrollment/recruitment strategies that reflect recent Phase 3 trials, are enriched for low risk patients (80% low,10% moderate, and 10% high risk patients), or enriched for high risk patients (10% low, 10% moderate, and 80% high risk patients). One-duration-fits-all approach with a treatment duration of 4 months were assessed for novel regimens with (A) culture conversion hazard ratios (HR) of 2 and (B) culture conversion HR of 3. Stratified medicine approach with treatment durations of 2 months for low risk patients, 4 months for moderate risk patients, and 6 months for high risk patients were assessed for novel regimens with (C) culture conversion HR of 2 and (D) culture conversion HR of 3. Each regimen was assumed to have a 33% decrease in relapse after adjusting for culture conversion improvements. Top panel shows cure rates from 5000 simulations. Crossbar shows the median and 90% prediction interval for experimental arm and black point shows the median of the 6-month standard of care control arm with consistent sample sizes and randomizations of low, moderate, and high risk groups for experimental and control arms. Middle panel shows power to detect noninferior outcomes between experimental arm and 6-month standard of care control arm. Noninferior margin of 6% was used. Bottom panel shows power to detect superior outcomes between experimental arm and 6-month standard of care control arm from 5000 simulations.

## Discussion

In this study, we have developed a clinical trial simulation tool that integrates Phase 2B outcomes, treatment duration, and patient risk factors to predict Phase 3 clinical outcomes. Applying the tool, we predicted that highly potent regimens with culture conversion hazard ratios of 3 or higher are required to reach target cure rates. With highly potent regimens, a stratified medicine approach may allow for low risk subgroups to be treated with ultrashort 2-month durations but high risk subgroups may still require durations of at least 5.5 months to maximize cure rates. Adjuvant immunotherapeutic strategies that prolonged immune response when co-administered with the 6-month standard regimen may also have the potential to increase cure rates with culture conversion hazard ratios above 1.5. We found that clinical trial designs with these innovative approaches have potential to introduce a paradigm shift to the TB regimen development process that include superiority tests.

Our data shows that time to culture conversion, a summary measure of the longitudinal profile of culture results over time, had higher AUROC for predicting long-term poor outcomes compared to culture status after 8 weeks of treatment or slope of time to detection profiles after 8 weeks of treatment (Table 3.3 and Figure 3.2). However, the ability of time to culture conversion to discriminate between favorable and poor outcomes is still modest with a maximum AUROC of 0.74 after adjusting for patient risk factors and treatment duration. This level of discrimination agree with previously published analysis with REMoxTB data using a logistic regression approach compared to the parametric survival modeling we performed to predict Phase 3 clinical outcomes.[19] We also found that the uncertainty in the translational link between Phase 2B and Phase 3 outcomes had marginal effects on predicted cure rates further highlighting the low sensitivity of culture-based intermediate markers to predict long-term poor outcomes (Figure 3.6). However, this may also be explained, in part, by the low-unvarying culture conversion hazard

ratios for the two regimens tested in the REMoxTB trial (1.17 for both novel fluoroquinolone-containing regimens compared to standard regimen)[5] that were included in the final integrated model. Sputum samples were not collected serially at earlier timepoints than 1 or 2 months in the OFLOTUB, RIFAQUIN, and DMID 01-009 trial, therefore, not included in the final integrated model. As more longitudinal culture and long-term outcome data become available from Phase 3 clinical trials that test novel potent regimens (i.e. SimpliciTB, NCT03338621[20] and TBTC Study 31, NCT02410772[15]), our clinical trial simulation tool can be validated to confirm the translational links identified.

Although potentially imperfect, the integrated relationship between intermediate culture-based biomarkers, treatment duration, patient risk factors and clinical outcomes, was used to inform the minimum efficacy targets required in Phase 2B trials to maximize cure rates and success in late stage TB regimen development. Our simulations suggest that culture conversion hazard ratios of 3 or more are required to maximize cure rates for shorter durations than 6 months. In a Phase 2B study, a very promising regimen that includes bedaquiline, pretomanid, moxifloxacin, and pyrazinamide (BPaMZ) was shown to have a culture conversion hazard ratio of 3.3 in patients with rifampin-resistant TB.[10] In the current Phase 3 trial (SimpliciTB), the BPaMZ regimen is being tested for 4 months in drug-susceptible and 6 months in drug-resistant patients.[20] Prospectively, our clinical trial simulation tool predicts that a regimen with a culture conversion hazard ratio of 3, similar to the BPaMZ regimen, can potentially shorten drug susceptible TB treatments to 4 to 5.5 months using a one-duration-fits-all approach. However, low risk subgroups may only require 2 months of the BPaMZ regimen and therefore may be over treated with a 4-month duration; potentially exposing patients to unnecessarily long and toxic treatments. In contrast, we predict that high risk subgroups may be undertreated requiring durations longer than

the 4 months currently investigated. These predictions will need to be confirmed when longitudinal culture data and long-term outcome data become available from the clinical trial.

Noninferiority has been the design of choice for all recent Phase 3 trial for drug susceptible TB because the very high success rates with the standard regimen making it unlikely to find an experimental regimen showing superior outcomes without exceptionally large sample sizes. In addition, some loss of efficacy because of shorter treatments in exchange for more effectivity in programmatic terms, such as improved adherence, improved patient management, and reduced exposure to toxic drugs, is deemed beneficial to the TB community.[4-6,13-15,20-22] Unfortunately, all recent Phase 3 trials with non-inferiority designs testing 4-month experimental regimens failed.[4-6] In previous analysis, high risk subgroups (i.e. patients with high bacterial burden and cavitary disease) were shown to require regimens longer than the standard 6 months and are likely the reason for failed non-inferiority trials (Chapter 1 and 2).[16] In this study, we found that innovate stratified medicine approaches, where duration is selected based on patient phenotypes, not only improves success of non-inferiority trials but also introduces potential to design and conduct unprecedented superiority trials in TB regimen development with increased cure rates. However, success rates of Phase 3 trials using a stratified medicine approach was dependent on enrollment strategies. In TB, the high risk subgroup is a curable population with the most potential for improved efficacy compared to the low risk subgroup that already have high cure rates even with shorter standard regimens.[16] Therefore, designs with stratified medicine approaches have more power to detect non-inferiority or superiority between experimental and standard regimens when enrollment is enriched for high-risk patients and designs with one-duration-fits-all approaches have more power to detect non-inferiority between experimental and standard regimens when enrollment is enriched for low-risk patients; as high risk patients will likely still be under treated. Enrichment in confirmatory Phase 3 trials is limited and unlikely to be acceptable by regulators;

as evidence of proven efficacy and safety in all patients is required for approval. Innovative approaches that include stratified medicine and strata level objectives may be an alternative approach to show efficacy and safety in all patients. For example, strata level objectives may involve showing superiority for high risk subgroups treated with regimens 6 months or longer and non-inferiority for low risk groups treated with shorter regimens. This approach may allow for maximized cure rates all patients and reduced toxicity concerns in at least low risk subgroups but, with treatment durations based on patient phenotypes, improvement of effectivity in programmatic terms will need to be assessed.

Adjuvant immunotherapeutic strategies, where host-directed therapies or therapeutic vaccines are administered with standard regimens, are another attractive approach to increase TB cure rates.[17] However, efficacy of adjuvant immunotherapies in patients is difficult to quantify because of the lack of clinical data, as clinical trials are still in the design and development stages.[23,24] In our tool, we added a prolonged duration effect from the standard regimen to quantitatively describe the long-lasting immune response because of immunotherapies. We predicted that adjuvant therapies with long-lasting immune response require culture conversion hazard ratios of 1.5 compared to the standard regimen alone to reduce poor outcomes by 50%. Prospective Phase 3 trials that test adjuvant immunotherapies to treat TB will need to confirm these predictions and will help better quantify the prolonged duration effect.

In conclusion, we provide a clinical trial simulation tool that can be used to design optimal clinical trials that permit informed decisions about moving the most promising regimens forward in the TB regimen development process. The tool integrates the most studied and broadly acceptable intermediate biomarkers assessed in Phase 2B trials, treatment duration, and patient risk factors to predict Phase 3 clinical outcomes. Model simulations provided Phase 2B efficacy targets for novel regimens and optimal trial designs for innovative stratified medicine approaches

and adjuvant immunotherapeutic strategies. We have developed a framework that can be extended and adapted to provide more precise predictions of cure rates and further facilitate decisions when more quantitative and sensitive biomarkers are discovered and more data from Phase 3 trials that testing novel regimens become available.

## References


1. WHO End TB Strategy. 2015.
2. Pipeline | Working Group for New TB Drugs. https://www.newtbdrugs.org/pipeline/clinical (accessed April 29, 2020.
3. Lienhardt C, Nahid P. Advances in clinical trial design for development of new TB treatments: A call for innovation. *PLOS Medicine* 2019; **16**: e1002769.
4. Merle CS, Fielding K, Sow OB, et al. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med* 2014; **371**: 1588-98.
5. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med* 2014; **371**: 1577-87.
6. Jindani A, Harrison TS, Nunn AJ, et al. High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N Engl J Med* 2014; **371**: 1599-608.
7. Phillips PPJ, Fielding K, Nunn AJ. An Evaluation of Culture Results during Treatment for Tuberculosis as Surrogate Endpoints for Treatment Failure and Relapse. *PLoS ONE* 2013; **8**: e63840.
8. Wallis RS, Peppard T, Hermann D. Month 2 Culture Status and Treatment Duration as Predictors of Recurrence in Pulmonary Tuberculosis: Model Validation and Update. *PLOS ONE* 2015; **10**: e0125403.
9. Dawson R, Diacon AH, Everitt D, et al. Efficiency and safety of the combination of moxifloxacin, pretomanid (PA-824), and pyrazinamide during the first 8 weeks of antituberculosis treatment: A phase 2b, open-label, partly randomised trial in patients with drug-susceptible or drug-resistant pul. *The Lancet* 2015; **385**: 1738-47.
10. Tweed CD, Dawson R, Burger DA, et al. Bedaquiline, moxifloxacin, pretomanid, and pyrazinamide during the first 8 weeks of treatment of patients with drug-susceptible or

drug-resistant pulmonary tuberculosis: a multicentre, open-label, partially randomised, phase 2b trial. *The Lancet Respiratory Medicine* 2019; **7**: 1048-58.

11. Kendall EA, Shrestha S, Cohen T, et al. Priority-Setting for Novel Drug Regimens to Treat Tuberculosis: An Epidemiologic Model. *PLOS Medicine* 2017; **14**: e1002202.
12. Fofana MO, Knight GM, Gomez GB, White RG, Dowdy DW. Population-level impact of shorter-course regimens for tuberculosis: A model-based analysis. *PLoS ONE* 2014; **9**.
13. Nunn AJ, Phillips PPJ, Gillespie SH. Design issues in pivotal drug trials for drug sensitive tuberculosis (TB). *Tuberculosis* 2008; **88**: S85-S92.
14. Merle CSC, Sismanidis C, Sow OB, et al. A pivotal registration phase III, multicenter, randomized tuberculosis controlled trial: Design issues and lessons learnt from the Gatifloxacin for TB (OFLOTUB) project. *Trials* 2012; **13**: 61.
15. Dorman SE, Nahid P, Kurbatova EV, et al. High-dose rifapentine with or without moxifloxacin for shortening treatment of pulmonary tuberculosis: Study protocol for TBTC study 31/ACTG A5349 phase 3 clinical trial. *Contemporary Clinical Trials* 2020; **90**.
16. Imperial MZ, Nahid P, Phillips PPJ, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nature Medicine* 2018; **24**: 1708-15.
17. Young C, Walzl G, Du Plessis N. Therapeutic host-directed strategies to improve outcome in tuberculosis. *Mucosal Immunology* 2020; **13**: 190-204.
18. Johnson JL, Hadad DJ, Dietze R, et al. Shortening Treatment in Adults with Noncavitary Tuberculosis and 2-Month Culture Conversion. *Am J Respir Crit Care Med* 2009; **180**: 558-63.

19. Phillips PPJ, Mendel CM, Burger DA, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Medicine* 2016; **14**: 19.

20. Trial to Evaluate the Efficacy, Safety and Tolerability of BPaMZ in Drug-Sensitive (DS-TB) Adult Patients and Drug-Resistant (DR-TB) Adult Patients - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03338621 (accessed April 30, 2020.

21. Tweed C, Wills GH, Crook AM, et al. A Partially-Randomised Phase 3 Trial of Pretomanid, Moxifloxacin and Pyrazinamide in Combination for the Treatment of Drug-Susceptible and Drug-Resistant Pulmonary Tuberculosis. *SSRN Electronic Journal* 2020.

22. Papineni P, Phillips P, Lu Q, Cheung YB, Nunn A, Paton N. TRUNCATE-TB: an innovative trial design for drug-sensitive tuberculosis. *International Journal of Infectious Diseases* 2016; **45**: 404.

23. Coler RN BS, Pine SO, Orr MT, Reese V, Windish HP, Davis C, Kahn M, Baldwin SL, Reed SG. Therapeutic immunization against Mycobacterium tuberculosis is an effective adjunct to antibiotic treatment. . *J Infect Dis* 2013; **207**(8): 1242-52.

24. Larsen SE BS, Orr MT, Reese VA, Pecor T, Granger B, Dubois Cauwelaert N, Podell BK, Coler RN. Enhanced Anti-Mycobacterium tuberculosis Immunity over Time with Combined Drug and Immunotherapy Treatment. *Vaccines (Basel)* 2016; **6**(2).

## Chapter 4: Identification of novel proteomic signatures as predictors of tuberculosis treatment response


### Abstract

**Background** Traditionally, sputum-based culture status after 8 weeks of treatment is used to monitor tuberculosis (TB) treatment response. Robust, quantitative, non-culture-based biomarkers of treatment outcomes represent an advance for individual monitoring of TB patients. Here, we leverage serially-collected sputum culture results and drug exposure in clinical trial participants to identify proteomic signatures that predict response to TB treatment.

**Methods** Blood samples collected at enrollment and week 8 of treatment from 538 participants from two Phase 2B trials were screened for 70 markers of infection, inflammation, and metabolism. Biomarker assays were quantified using a multiplexed electrochemiluminescence assay. Nonlinear mixed effect modeling and classification and regression tree analysis was used to integrate proteomic, drug exposure, and clinical risk factor data for prediction of longitudinal treatment response profiles from the Mycobacteria Growth Indicator Tube assay. An independent dataset of TB patients with confirmed durable cure 12 months post-treatment completion was used to validate identified proteomic signatures.

**Results** We identified SAA1, CD40L, and RANTES as candidate proteomic biomarkers with early, dynamic treatment response profiles. By integrating the proteomic signatures with drug exposure, and clinical risk factors we predicted week 8 and week 12 culture conversion status with an AUROC of 0.82 (95% CI: 0.76-0.87) and 0.85 (0.78-0.92), respectively. Fast and moderate responders, as predicted by proteomic biomarker levels, had shorter time to culture conversion compared to slow responders (3.1, 95% CI 1.9-5.2, and 2.0, 1.2-3.4, respectively). In the independent dataset of cured participants, 80% (24/30) were predicted to be fast to moderate

responders. Longitudinal proteomic levels suggest treatment response may be predicted with levels after 2 weeks of treatment.

**Conclusions** In a rigorously followed cohort of diverse TB patients enrolled in two Phase 2B clinical trials, we identified and validated proteomic signatures that predict early treatment response dynamics. Further investigation is warranted with novel non-rifampin-based regimens, more long-term outcomes and biomarker levels at earlier timepoints.

## Introduction

Establishing the efficacy of newer regimens has been a major challenge in tuberculosis (TB) drug development because of the lack of reliable surrogate markers.[1] Sputum-based markers, predominately culture status after 8 weeks of treatment or time to culture conversion, are the most frequently used intermediate markers to quantify early drug response and predict clinical outcomes of treatment failure and relapse up to 24 months of follow up. Unfortunately, despite their historical use as a measure of bacterial burden, they have shown to have poor surrogate properties at the individual and trial level with low sensitivity and modest specificity.[2-4] New biomarkers and technologies are needed to both monitor treatment response in individual patients and to support innovation in TB clinical trials.

Discovery of robust, quantitative, non-culture-based biomarkers of anti-TB treatment outcomes that can be measured early in treatment would represent a major advance for the TB therapeutics field.[1,5,6] Blood-based markers are promising as they are easy to collect, available throughout treatment, quantitative, thereby providing opportunity to improve predictive power, and can lead to development of cost-effective point-of-care assays that don't require sophisticated mycobacteriology laboratory infrastructure. A number of studies have identified several blood-based host biomarkers that change during TB treatment or are associated with culture status after 8 weeks of treatment.[7-18] However, most of these studies are single center studies, assess single markers, use convenience samples, have modest sample sizes or rely on case-control or observational designs. In seeking to overcome such limitations, 70 different markers of infection, inflammation, and metabolism were previously investigated in over 300 participants enrolled in an international Phase 2B clinical trial conducted across nine countries (TB Trial Consortium Study 29, NCT00694629)[19]. A subset of biomarkers were shown to be associated with disease severity and bacterial burden at baseline and strongly modulated by tuberculosis treatment with

greater modulation in early culture converters than late culture converters.[19-21] Although the rigorous analyses identified promising novel biomarkers that could be useful for evaluating disease severity and treatment monitoring, these biomarkers alone or in combination with clinical risk factors only had modest predictive power to discriminate individual treatment status dichotomously using culture status after 8 weeks of treatment (area under the receiver operating characteristic curve, AUROC, 0.66).[20] Longitudinal information from sputum culture results in conjunction with population-based pharmacokinetics (PK) and drug exposure were not fully leveraged in the previous analyses, warranting further investigation into the response dynamics to better understand treatment response heterogeneity between individuals.

Here, we aimed to improve predictive power of proteomic biomarkers by maximizing the use of the rich, longitudinal time to detection data collected from the Mycobacteria Growth Indicator Tube (BACTEC MGIT 960) assay throughout treatment to describe individual treatment response. In addition, we pooled additional data from participants enrolled in a second clinical trial that tested higher rifamycin doses (TB Trial Consortium Study 29X, NCT01043575)[22]. High interindividual variability of rifamycin PK (more than 4-fold variation of exposure for a given dose) have been highly predictive of individual culture conversion,[22,23] therefore was integrated with the search for proteomic biomarkers, along with other identified clinical risk factors (i.e. HIV status, smear grade, and cavitation), to predict of individual treatment response. Nonlinear mixed effect methodology was used to model longitudinal time to detection profiles and empirical Bayesian estimates were obtained from the final model to describe patient-specific treatment response parameters. Machine learning approaches with classification and regression tree (CART) analysis were then used to examine the role of 70 biomarkers as predictors of the patient-specific treatment response parameters.[24] Our proposed candidate biomarkers of treatment response were then validated as predictors of long-term outcomes from an independent dataset.

## Methods

### Data

We pooled data from adults, 18 years and older, with pulmonary tuberculosis enrolled in two randomized Phase 2 clinical trials (TB Trial Consortium Studies 29 and 29X) that compared efficacy and safety of rifampin to rifapentine during the 8-week intensive phase of treatment.[19,22] These trials were conducted in Brazil, Hong Kong, Kenya, Peru, South Africa, Spain, Vietnam, Uganda, and the United States. In Study 29, participants were randomized to receive rifapentine (10 kg/mg/dose) or rifampin (10 mg/kg/dose) 5 days per week for 8 weeks. In Study 29X, participants received rifapentine (10, 15, or 20 mg/kg/dose) or rifampin (10 mg/kg/dose) daily for 8 weeks. Isoniazid, pyrazinamide, and ethambutol were also administered during the 8-week intensive phase and rifampin and isoniazid were administered during the 4-month continuation phase in accordance to published guidelines.[25,26] Sputum specimens for culture were collected at baseline (enrollment) for pretreatment levels and weeks 2, 4, 5, 6, 8, 12 and monthly thereafter during treatment for Studies 29 and 29X participants. Study design and results for both trials have been previously published.[19,22]

In Studies 29 and 29X, 546 protocol correct participants (319 from Study 29 and 227 from Study 29X) had paired baseline (pretreatment) and week 8 blood samples for biomarker testing. Eight participants had missing baseline smear grade or chest radiograph data and were excluded from the analysis. In total, 538 participants irrespective of dose were included in the analysis of proteomic biomarkers that account for drug only (rifampin vs rifapentine) to predict treatment response. This analysis would inform if regimen composition and proteomic biomarkers are adequate for predictions. For 335 participants treated with rifapentine, a second analysis was performed by integrating proteomic biomarkers, drug exposure and clinical risk factors to predict treatment response. Rifampin PK data was not collected in the studies therefore not included in

the second analysis. Comparing the results of these two analyses would inform the level of drug exposure and clinical data necessary to maximize predictive value of proteomic biomarkers. A population pharmacokinetics-pharmacodynamic model for the Studies 29 and 29X dataset have been previously developed and were used in this analysis for post-hoc estimates of individual rifapentine PK parameters (area under the concentration time curve, AUC).[23]

An independent dataset for external evaluation of potential biomarkers was obtained from the multi-network partnered Consortium for TB Biomarkers (CTB2; https://www.tbbiorepository.org funded through U.S. Food and Drug Administration, U.S. National Institutes of Allergy and Infectious Diseases and Bill & Melinda Gates Foundation) and included 21 cured participants from the Phase 3 REMoxTB trial (NCT00864383) and 9 cured participants enrolled and managed through the National Tuberculosis Program at the same sites at which REMoxTB was conducted.[27] Cured was defined as confirmed durable cure 12 months post-treatment completion. In the REMoxTB trial, adults with pulmonary TB received the standard 6-month regimen or one of two experimental 4-month regimens that replaced isoniazid or ethambutol with moxifloxacin. Adults with pulmonary TB from the National Tuberculosis Program are presumed to have received the standard 6-month regimen. All participants in the independent dataset received standard rifampin doses. Blood for proteomic analysis was collected at baseline and weeks 2, 4, 8, 17, 26, and 52 and sputum was collected at baseline and weeks 2, 4, 8, 17, 26, and 52 for REMoxTB participants and at baseline, and weeks 4, 8, 26, and 52 for National Tuberculosis Program participants.

**Longitudinal treatment response variables**

Treatment response for each individual was characterized using the longitudinal sputum culture results from the MGIT assay to describe 1.) time to detection profiles which reflect bacterial burden over time, 2.) time to first negative culture status defined as time from start of treatment to

the first culture conversion from positive to negative, irrespective of whether there were any subsequent positive results, and 3.) time to stable culture conversation defined as time from start of treatment to the first of two negative cultures at different visits without an intervening positive result. Time to detection profiles were used to identify proteomic biomarkers of treatment response that grouped participants into slow, moderate, and fast responders. Time to first culture negative status and time to stable culture conversion were used to compare responder groups in Cox regression analysis, with the former being the primary marker because it is less impacted by study design (i.e. how often sputum is collected) and better aligns with time to detection profiles as it describes the treatment time in an individual's profile when time to detection becomes censored such that culture becomes negative.

Nonlinear mixed effect modeling was used to describe time to detection profiles with the following empirical logistic model:

$$t_{detection} = BASE + \frac{MAX - BASE}{1 + e^{-\alpha(T_{treatment} - T_{50})}}$$

where $t_{detection}$ is the time to bacterial detection in the MGIT assay at treatment time, $T_{treatment}$; $BASE$ is baseline time to bacterial detection; MAX is the model derived time to bacterial detection maximum value (artifact of modeling due to assay censoring at 44 days); $\alpha$ is the slope of the profile; and $T_{50}$ is the treatment time to reach 50% of MAX. $t_{detection}$, $BASE$, and $MAX$ are in the units of days and represent assay times; $T_{treatment}$ and $T_{50}$are in the units of weeks and represent treatment times; and $\alpha$ is in the units 1/week. A publication for this model in currently in progress. Figure 4.1 shows the effect of covariates on each of the model parameters while holding all other parameters constant. Covariates on $T_{50}$ has the most pronounced effect on the treatment time when time to detection profiles reach censored values of 44 days such that culture becomes negative. In

this study, we aimed to identify proteomic biomarkers that describe empirical Bayesian estimates of unexplained interindividual variability of $T_{50}$. The $T_{50}$ parameter is described by the following:

$$T_{50} = TVT_{50}(x) * \exp(\eta)$$

where $TVT_{50}(x)$ is the typical value of $T_{50}$ for individuals with covariates, $x$, and $\eta$ is the empirical Bayesian estimate of unexplained interindividual variability of $T_{50}$ with a mean of zero and variance of $\sigma_{\eta}^2$, which is estimated in population modeling. Hereafter, the empirical Bayesian estimates, $\eta$, of unexplained inter-individual variability of $T_{50}$ is referred to as an individual's response time. A positive value for response time increases $T_{50}$ and represents a slower response (longer time on treatment to convert culture) for an individual relative to a typical patient with the same covariates, while a negative value for response time decreases $T_{50}$ and represents a faster response (shorter time on treatment to convert culture) for an individual relative to a typical patient with the same covariates. A zero value for response time represents an individual with no variation compared to the typical patient with the same covariates. Drug or drug exposure and clinical risk factors of smear grade, geographic site (Africa vs non-Africa) and cough before treatment (productive vs. nonproductive) were covariates included for determination response time.

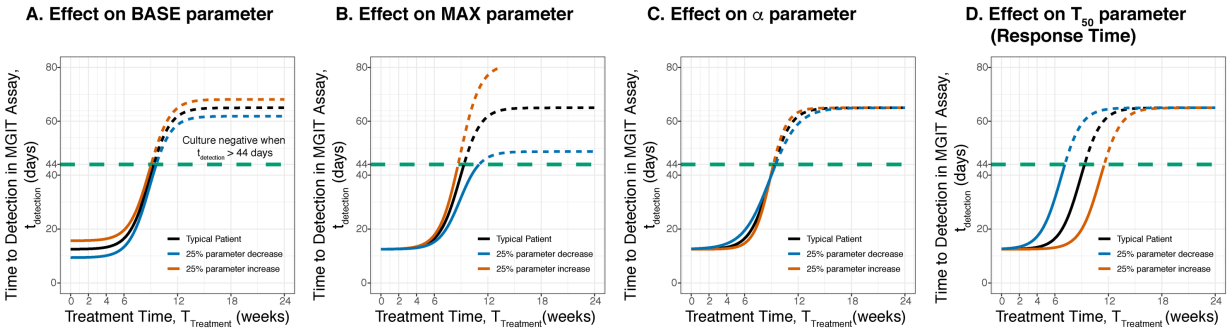



**Figure 4.1** Covariate effects on time to detection parameters. Covariate effects are shown on the (A) *BASE* parameter, (B) *MAX* parameter, (C) $\alpha$ parameter, and (D) $T_{50}$ parameter. Black curves represent a typical patient, red curves represent a 25% increase in parameter values due to a covariate, and blue curves represent a 25% decrease in parameter values due to a covariate. $t_{detection}$, time to bacterial detection in the MGIT assay; $T_{treatment}$, treatment time; $BASE$, baseline time to bacterial detection; MAX, model derived maximum time to bacterial detection; $\alpha$, slope of the profile; and $T_{50}$, treatment time to reach 50% of MAX.

**Proteomic methods**

Proteomic assays have been described in a precious publication.[20] Briefly, a total of 70 biomarkers were measured in 14 multiplexed assay panels using sandwich immunoassay format (proteins) or a competitive immunoassay format (neopterin), using electrochemiluminescence (ECL) detection.[28] The ECL assays employed consumables and instrumentation from Meso Scale Diagnostics, LLC (MSD). The assay components for each panel included an MSD MULTI-ARRAY® 96- well plate having an array of capture antibodies in each well, a set of labeled detection antibodies for each analyte in the panel (labeled with the MSD SULFO-TAG™ ECL label), a combined calibration standard containing a mixture of the target analytes, an assay diluent and a detection antibody diluent. For the neopterin competitive assay, a labeled neopterin analog was used in the place of the labeled detection antibody. In total, 14 biomarker panel assays were tested, six MSD commercial kits and eight custom assay panels that were newly developed for analysis of this data.

MSD received 500 μL of each sample at their core facility (Gaithersburg, MD) where assays were conducted with investigators and technicians blinded to participant data. Each sample was tested in duplicate with each of the 70 assays. Concentrations were reported as the average value of the duplicate measurements; values below the limits of detection were assigned a concentration equal to the limits of detection. Coefficient of variations were determined for the biomarker levels measured in the control samples run on each assay plate; the median control coefficient of variation (and in quantitation range) across the different assays was 10% (9%–13%).

Serum levels were quantified for 69 proteins and one metabolic marker. As described previously, 8 of 70 markers that provided levels greater than twice the limit of detection for <25% of the samples were excluded from further analysis.[20] In addition, one marker, IL12P40, was

excluded from analyses in this study because it was not available from the panel of assays for Study 29X participants.

**Statistical methods**

Statistical analysis was conducted using the R statistical programming language (version 3.4). We performed CART analysis to examine proteomic biomarkers that predict response time after adjusting for drug or drug exposure and clinical risk factors. Baseline, week 8 and ratio of week 8 to baseline proteomic biomarker levels (61 biomarkers, 183 variables total) were tested simultaneously as predictors of response time. CART uses nonparametric techniques that examine both linear and nonlinear interactions simultaneously in the whole dataset and creates a hierarchy of predictors, starting with the most predictive to the least predictive.[24] The CART algorithm model works by repeatedly splitting predictor variables into multiple nodes, so that the outcomes in each final node is as homogenous as possible. The split cutoff for each predictor is defined so that the residual mean squared error is minimized across the samples that fall within the sub partition. Tree pruning, where decisions trees are simplified to remove splits that do not significantly improve overall quality of the model, as described by the root mean squared error, was performed by automatically assessing 10 values of the complexity parameter in CART analysis using the caret and rpart packages in R.[29,30] The complexity parameter imposes penalties for having too many splits and the best model with the optimal complexity parameter value was chosen that minimized the 10-repeated 10-fold cross validation root mean squared error. A random selection of 70% of participants in Studies 29 and 29X was used to train the model, while the remaining 30% was used to test the model. Following CART analysis, the final decision tree was used to define slow, moderate, and fast responders using the identified proteomic signatures. Cox regression analysis was performed to assess adjusted hazard ratios between slow, moderate, and fast responders for time to first negative culture conversion and time to stable culture conversion.

Hazard ratios were adjusted for drug or drug exposure and clinical factors. Using the linear predictor from the Cox regression results, area under the receiver operating characteristics curve (AUROC) analysis was applied to assess discrimination of culture conversion after 8 and 12 weeks of treatment.

**Funding**

Written informed consent was obtained from all study participants for collection of serum for TB-related research. In addition, the institutional review board at University of California, San Francisco approved this ancillary study to assess putative biomarkers of treatment response (approval #12-10360). The substudy was supported through a Partnerships for Biodefense RO1, funded by the NIH, NIAID (R01AI104589).

**Results**

**Data**

Table 4.1 shows demographic and clinical data of 538 participants included in the train and test datasets and 30 participants included in the independent dataset. In the train and test datasets, 203 participants were treated with the standard regimen (rifampin at 10 mg/kg) and 335 were treated with rifapentine at 10 mg/kg, 15 mg/kg or 20 mg/kg as substitution for standard-dose rifampin during the 8-week intensive phase. In the independent dataset, 17/30 (57%) participants received the 6-month standard regimen and 13/30 (43%) received a 4-month experimental regimen. Of 538 participants in the train and test dataset, 53 (10%) were HIV co-infected. No HIV co-infected participants were included in the independent dataset. At the time of modeling, smear grade, cavitation and other clinical factors were not available for the independent.

**Table 4.1** Regimen and participant characteristics

| | Training Dataset | Testing Dataset | Independent Dataset |
|---|---|---|---|
| **N** | **379** | **159** | **30** |
| **Treatment regimen N (%)[a]** | | | |
| 2HRZE/4HR | 143 (38) | 60 (38) | 17 (57) |
| $2HP_{10}ZE/4HR$ | 158 (42) | 68 (43) | - |
| $2HP_{15}ZE/4HR$ | 40 (11) | 22 (14) | - |
| $2HP_{20}ZE/4HR$ | 38 (10) | 9 (6) | - |
| 2HRZM/2MHR | - | - | 3 (10) |
| 2EMRZ/2MR | - | - | 10 (33) |
| **Enrolled at African Site N (%)[b]** | 214 (57) | 96 (60) | 20 (67) |
| **Female N (%)** | 123 (33) | 46 (29) | 12 (40) |
| **Age (years)** Median [interquartile range] | 32 (24-44) | 30 (25-44) | 28 (23-35) |
| **Weight at baseline (kg)** Median [interquartile range] | 55 (50-61) | 54 (49-62) | 52 (47-60) |
| **BMI at baseline ($kg/m^2$)** Median [interquartile range] | 20 (18-22) | 19 (18-21) | 19 (17-22) |
| **HIV positive N (%)** | 39 (10) | 14 (9) | 0 (0) |
| **History of smoking N (%)** | 161 (43) | 61 (38) | Not available |
| **Smear grade at baseline** | | | |
| Negative | 24 (6) | 13 (8) | Not available |
| 1+ | 119 (31) | 49 (31) | Not available |
| 3+ | 91 (24) | 34 (21) | Not available |
| 4+ | 145 (38) | 63 (40) | Not available |
| **Cavitation at baseline** | 268 (71) | 113 (71) | Not available |
| **Cough before treatment** | | | |
| Productive | 345 (91) | 140 (88) | Not available |
| Nonproductive | 22 (6) | 14 (9) | Not available |
| No cough | 12 (3) | 5 (3) | Not available |
| **Rifapentine AUC** Median [interquartile range][c] | 264 (174-417) | 250 (157-369) | Not available |

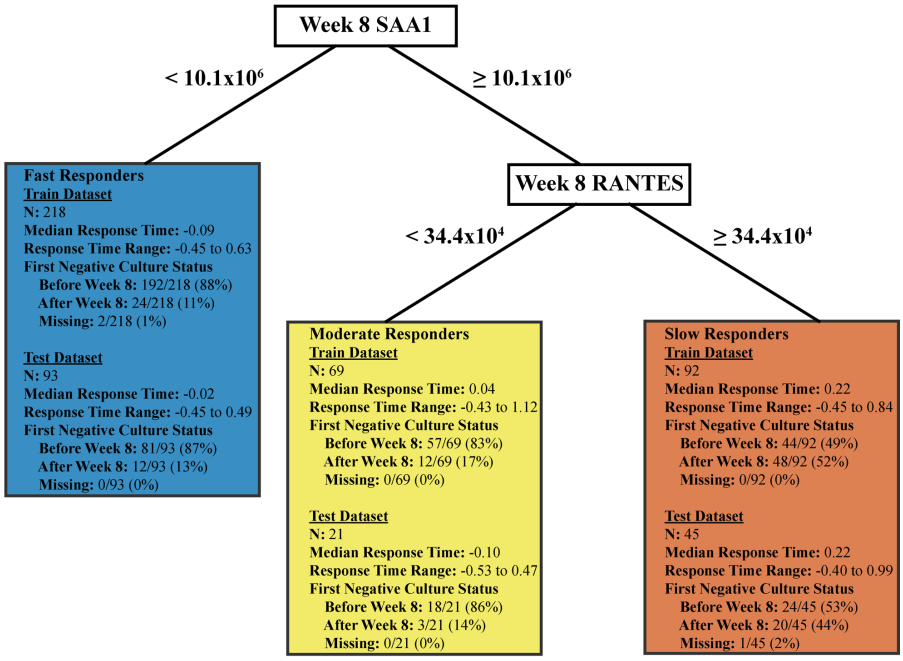
[a]H, isoniazid; Z, pyrazinamide; E, ethambutol; R, rifampin; P, rifapentine;M, moxifloxacin at 400 mg/dose. Subscripts on P define rifapentine dose (10 mg/kg/dose, 15 mg/kg/dose, or 20 mg/kg/dose). All other drugs were administered according to published guidelines. First set of drugs represent drugs included in 2-month intensive phase and second set of drugs represent drugs included in 4-month or 2-month continuation phase of regimen.
[c]Rifapentine area under the curve estimated from previously published population pharmacokinetic model for Studies 29 and 29X . [1]
HIV, human immunodeficiency virus; BMI, body mass index; AUC, area under concentration time curve

## Proteomic signature after adjusting for rifamycin drug

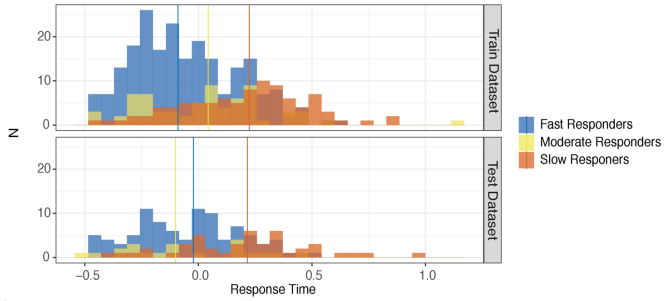
Week 8 SAA1 and RANTES levels were the top proteomic biomarkers that predict response time after adjusting for rifamycin drug (rifampin vs rifapentine) (Figure 4.2A). The distribution of response times and longitudinal time to detection profiles by response group as defined by the week 8 SAA1 and RANTES proteomic signature are shown in Figure 4.2B and

Figure 4.2C. Time to first negative culture status was not available for three participants. Among those participants who are fast responders, 88% (273/311) culture converted before week 8 and 12% (36/311) after week 8; among those who are moderate responders, 83% (75/90) culture converted before week 8 and 17% (15/90) after week 8; and among those who are slow responders, 50% (68/137) culture converted before week 8 and 50% (68/137) after week 8 (Figure 4.2A). The adjusted hazard ratios for time to first culture negative status of fast and moderate responders compared to slow responders were 2.6 (95% CI, 2.1-3.2) and 1.9 (1.4-2.5), respectively, for the combined train and test dataset (Table 4.2 and Figure 4.2D). The AUROC for this proteomic signature to predict week 8 and week 12 culture conversion status was 0.77 (95% CI, 0.73-0.82) and 0.77 (0.71-0.84), respectively, for the combined train and test dataset; an improvement when compared to models without proteomic biomarkers (AUROC $\leq$ 0.60, Table 4.3). Similar hazard ratios and AUROC were observed for predicting time to stable culture conversion (Appendix Table A.4 and Appendix Table A.5).

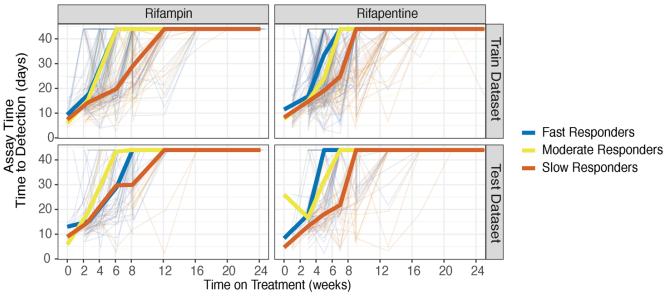
**A. Decision Tree**




**B. Distribution of Response Times**




**C. Time to Detection Profiles**




**D. Time to First Negative Culture Status**

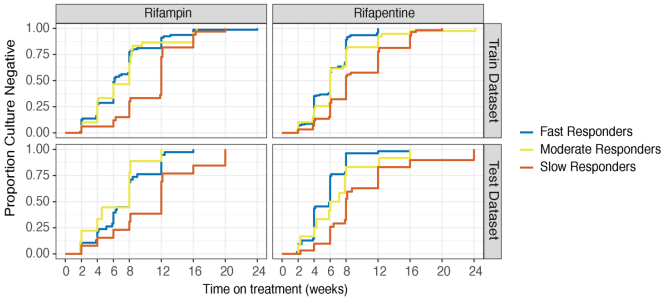


**Figure 4.2** Proteomic signature that predicts response time after adjusting for drug (rifampin vs rifapentine). (A) Decision tree that predicts fast, moderate, and slow responders based on week 8 SAA1 and RANTES levels. (B) Distribution of response time (unexplained interindividual variability of $T_{50}$ ) stratified by fast, moderate, and slow responders as defined by proteomic signature. Vertical lines represent the median response time for each group. (C) Individual time to detection profiles (thin lines) and median time to detection profiles (thick lines) color coded by fast, moderate, and slow responders as defined by proteomic signature and stratified by drug (rifampin vs rifapentine) and dataset. (D) Time to first culture negative status color coded by fast, moderate, and slow responders as defined by proteomic signature and stratified by drug (rifampin vs rifapentine) and dataset.

**Table 4.2** Adjusted hazard ratios for time to first negative culture status. P value for interaction between response groups and drug or drug exposure not significant at a P = 0.01 level.

| | **Train Dataset** | **Test Dataset** | **Train + Test Dataset** | **Independent Dataset** |
|---|---|---|---|---|
| **Proteomic signature adjusted for drug (rifampin vs rifapentine)** | | | | |
| **N** | 379[a] | 158[a] | 538[a] | 30 |
| **First Negative Culture Hazard Ratio Median (95% CI)** | | | | |
| Moderate responders relative to slow responders | 1.8 (1.3-2.5) | 2.3 (1.3-3.9) | 1.9 (1.4-2.5) | 1.7 (0.5-5.1) |
| Fast responders relative to slow responders | 2.5 (1.9-3.2) | 2.9 (1.9-4.3) | 2.6 (2.1-3.2) | 2.3 (0.8-6.5) |
| **Proteomic signature adjusted for drug exposure and clinical risk factors** | | | | |
| **N** | 236[a] | 99[a] | 335[a] | - |
| **First Negative Culture Hazard Ratio Median (95% CI)** | | | | |
| Moderate responders relative to slow responders | 2.0 (1.0-3.9) | 1.7 (0.7-4.1) | 2.0 (1.2-3.4) | - |
| Fast responders relative to slow responders | 3.6 (1.9-6.7) | 2.3 (0.9-5.7) | 3.1 (1.9-5.2) | - |
| [a]Two participants in train dataset and one participant in test dataset missing time to first negative culture status | | | | |

**Table 4.3** AUROC values to predict time to first negative culture status. (95% confidence interval)

| | | **AUROC at 8 weeks** | | | **AUROC at 12 weeks** | | |
|---|---|---|---|---|---|---|---|
| **Adjusted for:** | **Proteomics Included?** | **Train Dataset** | **Test Dataset** | **Train + Test Dataset** | **Train Dataset** | **Test Dataset** | **Train + Test Dataset** |
| Drug (Rifampin vs Rifapentine) | No | 0.59 (0.53-0.65) | 0.60 (0.51-0.70) | 0.60 (0.55-0.65) | 0.58 (0.49-0.68) | 0.52 (0.35-0.67) | 0.56 (0.48-0.64) |
| Drug (Rifampin vs Rifapentine) | Yes | 0.76 (0.71-0.83) | 0.77 (0.68-0.85) | 0.77 (0.73-0.82) | 0.78 (0.72-0.85) | 0.75 (0.60-0.91) | 0.77 (0.71-0.84) |
| Rifapentine AUC and clinical risk factors | No | 0.74 (0.68-0.83) | 0.79 (0.69-0.89) | 0.74 (0.67-0.81) | 0.85 (0.78-0.93) | 0.82 (0.70-0.94) | 0.83 (0.77-0.89) |
| None | Yes | 0.72 (0.64-0.81) | 0.65 (0.52-0.79) | 0.71 (0.64-0.78) | 0.80 (0.70-0.92) | 0.65 (0.52-0.79) | 0.81 (0.74-0.88) |
| Rifapentine AUC and clinical risk factors | Yes | 0.81 (0.76-0.90) | 0.83 (0.73-0.91) | 0.82 (0.76-0.87) | 0.92 (0.88-0.97) | 0.78 (0.62-0.94) | 0.85 (0.78-0.92) |

In the independent dataset of 30 cured participants, 16 (53%) were categorized as fast responders, 8 (27%) as moderate responders, and 6 (20%) as slow responders. Among those participants who are fast responders, 81% (13/16) culture converted before week 8 and 19% (3/16) converted after week 8; among those who are moderate responders, 63% (5/8) culture converted before week 8 and 38% (3/8) after week 8; and among those who are slow responders, 50% (3/3) culture converted before week 8 and 50% (3/3) after week 8 (Figure 4.3A) Although the confidence interval was wide, as expected in this small independent dataset, the adjusted hazard ratio point estimates for first culture negative status between responder groups were consistent with the train and test datasets (Table 4.2).

The longitudinal profiles of SAA1 and RANTES levels was available for up to 52 weeks after start of treatment for the 30 cured participants in the independent dataset. Fast responders were qualitatively shown to have the steepest decline in SAA1 levels compared to moderate and slow responders that can already be observed after only 2 weeks of treatment (Figure 4.3B). Furthermore, at 2 weeks fast responders had a median level of SAA1 below the week 8 split threshold identified in the CART analysis above (Week 8 SAA1 levels = $10.1 \times 10^6$). The median RANTES level for the slowest responders did not decline during the first 8 weeks of treatments but the median RANTE level for moderate responders declined after only 4 weeks of treatment before beginning to plateau (Figure 4.3C). Data from participants that failed treatment were not included in this analysis but data acquisition is on-going and analysis will be updated when data becomes available.

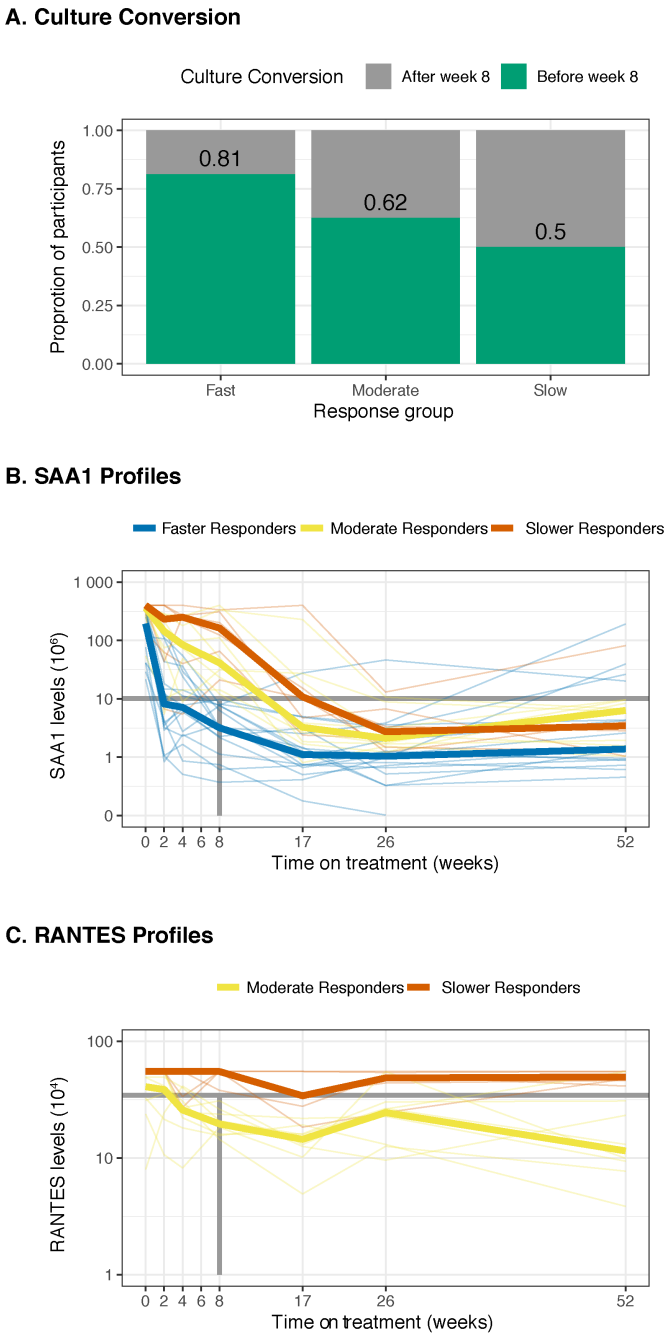
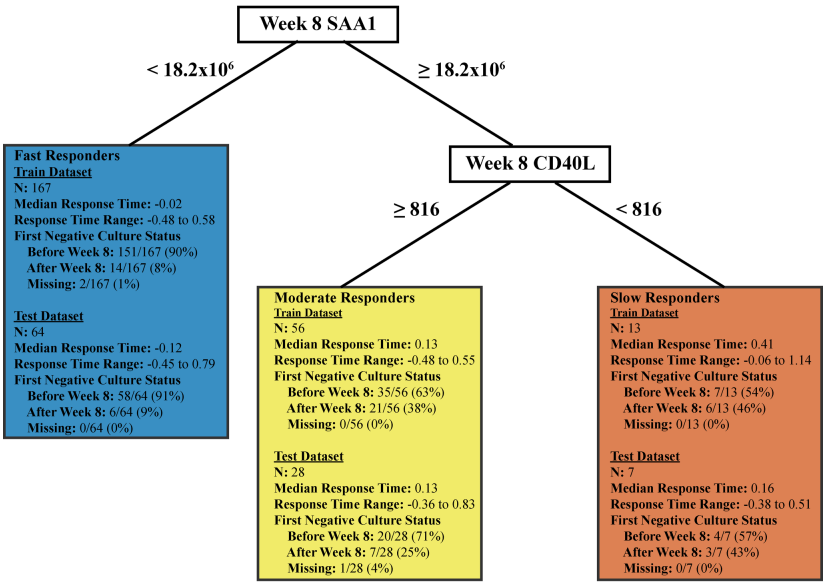



**Figure 4.3** Validation of drug adjusted proteomic signature in independent dataset. (A) Proportion of participants with culture conversion before or after week 8 for fast, moderate, and slow responders as defined by week 8 SAA1 and RANTES proteomic signature. (B) SAA1 profiles color coded by fast, moderate, and slow responders as defined proteomic signature. (C) RANTES profiles color coded by moderate, and slow responders as defined by proteomic signature.

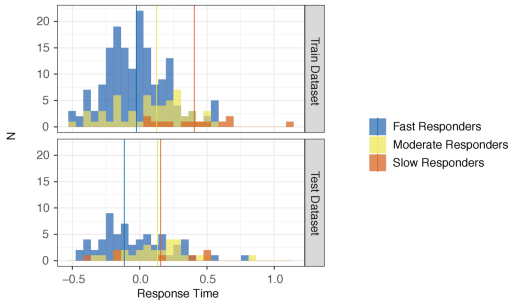
**Proteomic signature after adjusting for rifapentine exposure and clinical risk factors**

Week 8 SAA1 and CD40L levels were the top proteomic biomarkers that predict response time after adjusting for rifapentine drug exposure and clinical risk factors (Figure 4.4A). The distribution of response times and time to detection profiles by response group as defined by the week 8 SAA1 and CD40L proteomic signature are shown in Figure 4.4B and Figure 4.4C. Participants with rifapentine exposure below the median had slower response compared to participants with rifapentine exposure above the median in the same response groups (solid vs dashed lines, Figure 4.4C). Time to first negative culture status was not available for three participants. Among those participants who are fast responders, 90% (189/231) culture converted before week 8 and 9% (20/231) after week 8; among those who are moderate responders, 65% (55/84) culture converted before week 8 and 33% (28/84) after week 8; and among those who are slow responders, 55% (11/20) culture converted before week 8 and 45% (9/20) after week 8 (Figure 4.4A). The adjusted hazard ratios for time to first culture negative status of fast and moderate responders compared to slow responders were 3.1 (95% CI, 1.9-5.2) and 2.0 (1.2-3.4), respectively, for the combined train and test dataset (Table 4.2 and Figure 4.4D). The integration of rifapentine exposure, clinical risk factors, and proteomic biomarkers had an AUROC to predict week 8 and week 12 culture conversion of 0.82 (95% CI: 0.76-0.87) and 0.85 (0.78-0.92), respectively, for the combined train and test dataset; higher than models including drug exposure and clinical factors or proteomic biomarkers only (Table 4.3). Similar hazard ratios and AUROC were observed for predicting time to stable culture conversion (Appendix Table A.4 and Appendix).

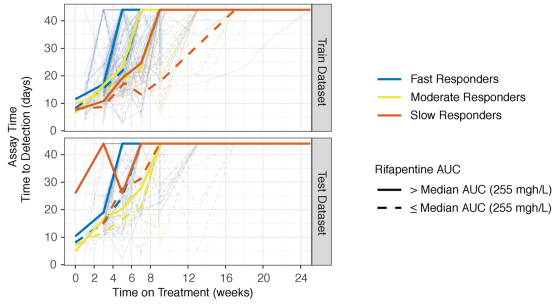
**A. Decision Tree**




**B. Distribution of Response Times**




**C. Time to Detection Profiles**




**D. Time to First Negative Culture Status**

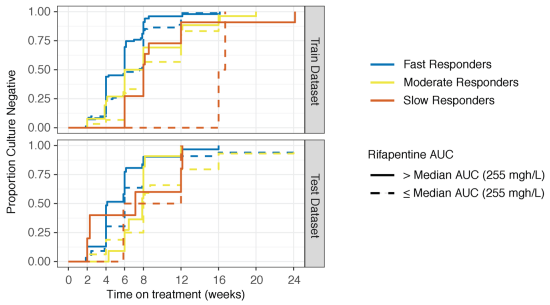


**Figure 4.4** Proteomic signature that predict response time after adjusting for rifapentine exposure and clinical risk factors. (A) Decision tree that predicts fast, moderate, and slow responders based on week 8 SAA1 and CD40L levels (B) Distribution of response time (unexplained interindividual variability of $T_{50}$), stratified by fast, moderate, and slow responders as defined by proteomic signature. Vertical lines represent the median response time for each group. (C) Individual time to detection profiles (thin lines) and median time to detection profiles (thick lines) color coded by fast, moderate, and slow responders as defined by proteomic signature. (D) Time to first culture negative status color coded by fast, moderate, and slow responders as defined by proteomic signature. Solid lines represent subpopulation with rifapentine exposure above the median (AUC > 255 mgh/L) and dashed lines represent subpopulation with rifapentine at or below the median (AUC ≤ 255 mgh/L).

## Discussion

In this study, we have shown that leveraging longitudinal sputum culture results and drug exposure and applying advanced nonlinear mixed effect modeling with machine learning approaches offer insights into the early response dynamics following anti-TB treatment and help identify candidate proteomic signatures that can predict treatment response. We identified SAA1, RANTES, and CD40L as candidate proteomic biomarkers that can predict time to detection profiles. By integrating proteomic signatures, drug exposure, and clinical risk factors we predicted week 8 and week 12 culture conversion status with high AUROC (0.82, 95% CI: 0.76-0.87, and 0.85, 0.78-0.92, respectively). Week 8 proteomic levels were the top predictors of response time compared to baseline or ratio of week 8 to baseline levels. Although only baseline and week 8 proteomic levels were available in the train and test datasets, earlier proteomic data in the independent dataset suggest that levels after only 2 weeks of treatment may have potential to predict treatment response.

The proteomic signatures identified in this study are predictors of treatment response as defined by response time in time to detection profiles from the MGIT assay. This is not the same as showing an effect on clinical outcome, namely treatment failure or relapse. However, these time to detection profiles are used to define intermediate biomarkers in current Phase 2 trials, such as culture status after 8 weeks of treatment and time to culture conversion, to predict clinical outcomes. We also showed that categorizing participants into slow, moderate, and fast responders

defined by proteomic signatures are risk factors of time to first negative culture status and time to stable culture conversion. Therefore, we anticipate that treatment response as predicted by the proposed proteomic signatures are correlated with clinical outcomes. Furthermore, we have validated our proteomic signatures in cured participants and plan to validate them in failed and relapse participants when data becomes available.

The top predictor of response time in the two proteomic signatures we identified was week 8 levels of SAA1, an acute phase protein mainly expressed in the lung and produced in response to infection, tissue injury, and malignancy.[31] SAA1 and CRP, a similar acute phase protein, have been associated with baseline severity of TB disease and bacterial burden and have shown significant decreases during anti-TB treatment.[11,17,18,20,32] In this study, we showed that lower levels of SAA1 at week 8 was associated with faster response times; consistent with anticipated improvements of acute phase protein response to TB treatment. In addition, week 8 levels of RANTES or CD40L were capable to differentiate participants with the slowest response times. In TB, RANTES has a protective role in *M. tuberculosis* infection by forming granulomas, limiting pathogen growth, and preventing lung tissue damage [33] Interestingly, the relationship between RANTES levels and treatment response have differed among ethnic populations (i.e. African, Eurasian, and Asian) and comorbidities (i.e. HIV co-infection) and RANTES polymorphisms have been associated with increased risk of TB.[34-39] CD40-CD40L pathway has been identified as a critical mechanism for generation of antigen specific T-helper-17 cells that produce IL-17 responses which have emerged as an important role for protecting and controlling immunity to TB.[40] Although these markers have been associated with treatment response in TB patients, there use alone or in combination with clinical risk factors have not shown sufficient predictive value for clinical use.[20] Instead, in this study, we found that integrating serially-collected culture data,

drug exposure parameters, clinical risk factors and proteomic biomarkers substantially improved the predictions of early treatment response.

Based on variable importance, week 8 proteomic levels were primarily the top predictors of response times and contributes the largest model improvements by reducing root mean squared errors (Appendix Table A.6). This may suggest that monitoring when biomarker levels are within normal levels may indicate when patients are in a recovery phase. In this study, participants without TB (i.e. healthy controls) were not included to assess normal ranges of biomarker levels. However, recent studies that have monitored longitudinal profiles of inflammatory markers show treated and survived TB patients approach normal ranges observed in healthy controls.[10,41-46] Blood-based biomarkers would provide even more value if levels before week 8 can predict treatment outcomes. In the independent dataset with proteomic levels measured at weeks 2 and 4, participants predicted to be fast responders already had median proteomic levels below the week 8 split thresholds identified in the CART analysis (Figure 4.3). Similar decreasing patterns of inflammatory markers in TB patients have been shown during the first 2 weeks of treatment with levels plateauing to normal ranges by 8 weeks.[10,41-46] Blood-based biomarkers that can reliability predict treatment response within the first 2 weeks of treatment would greatly increase efficiency of clinical care and regimen development and provide major advantages over sputum-based approaches.

In the independent dataset, 80% of participants were predicted to be fast to moderate responders but 20% were still predicted to be slow responders despite having been cured. In a recent study that used Positron Emission Tomography-Computed Tomography (PET-CT) scans to monitor treatment response in patients, the heterogenous nature of disease pathology and prognosis during treatment was illustrated and similar PET-CT patterns of active disease at diagnosis were identified in bacteriologically cured patients at the end of treatment.[47] Interestingly, many cured

individuals developed new lesions during treatment, and 68% of cured individuals had persistent pulmonary inflammation even one year after the completion of therapy. This may explain, in part, the small portion of cured participants predicted to be slow responders and the relatively high levels of inflammatory markers in a small portion of cured participants (5/30 above SAA1 split threshold) through week 52 (at least 6 months of follow-up) (Figure 4.3). Participants who failed treatment or relapsed are not yet part of the independent dataset, but we hypothesize that a larger proportion of these participants will have sustained high levels of proteomic levels and will be categorized as slow responders.

There were limitations in our study. First, the two proteomic signatures identified in this study were not uniform potentially implying that signatures may predict differently for each regimen or may depend on available drug exposure and clinical data. PK data was not available for participants who administered standard-dose rifampin limiting our second analysis to a subset of the data including only participants who administered rifapentine and making it difficult to confirm if the same proteomic signature can predict response after accounting for rifampin exposure. Varying signatures for each regimen would imply poor trial level surrogacy but may still be a valuable tool for more individualized treatments. The utility of these proteomic signatures as individual- and trial-level surrogates should be further evaluated as part of robust clinical trials that test rifamycin- and non-rifamycin-based regimens and collect long term outcomes and PK samples. Third, the parent trials for training and testing the model did not have samples from earlier timepoints. Based on the longitudinal proteomic data in the independent dataset, we may have missed important changes in proteomic levels that have stabilized before week 8. Fourth, our population only consist of 10% HIV co-infected individuals so separate analysis for HIV co-infected individuals was not plausible. The primary strength of our study lies in the integration within randomized clinical trials, which provided the infrastructure to conduct measurements

rigorously. PK, clinical, radiographic, and bacteriological data were collected using standardized forms and methods, acid-fast bacilli smears and cultures were performed at a single quality-controlled microbiological laboratory, and all biomarker assays including those for the independent dataset were done at a single quality-controlled immunology laboratory.

In conclusion, across 70 host biomarkers evaluated in the largest cohort of diverse TB patients enrolled in clinical trials, we identified proteomic signatures that can predict treatment response. We found that proteomic levels after 8 weeks of treatment predict treatment response but levels after 2 weeks of treatment may be sufficient. Accounting for individual drug exposure proved to be valuable in improving the predictive value of proteomic biomarkers and since drug exposure and proteomic blood-based biomarkers are easily measurable and available throughout treatment, together they have potential to project long term relapse-free outcomes and inform individualized dosing or treatment strategies to maximize cure for all.

**References**


1. Wallis RS, Kim P, Cole S, et al. Tuberculosis biomarkers discovery: developments, needs, and challenges. *The Lancet Infectious Diseases* 2013; **13**: 362-72.
2. Horne DJ, Royce SE, Gooze L, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. *The Lancet Infectious Diseases* 2010; **10**: 387-94.
3. Phillips PPJ, Fielding K, Nunn AJ. An Evaluation of Culture Results during Treatment for Tuberculosis as Surrogate Endpoints for Treatment Failure and Relapse. *PLoS ONE* 2013; **8**: e63840.
4. Phillips PPJ, Mendel CM, Burger DA, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Medicine* 2016; **14**: 19.
5. Walzl G, Ronacher K, Djoba Siawaya JF, Dockrell HM. Biomarkers for TB treatment response: Challenges and future strategies. *Journal of Infection* 2008; **57**: 103-9.
6. Nahid P, Saukkonen J, Mac Kenzie WR, et al. Tuberculosis biomarker and surrogate endpoint research roadmap. *American Journal of Respiratory and Critical Care Medicine* 2011; **184**: 972-9.
7. Andrade BB, Pavan Kumar N, Mayer-Barber KD, et al. Plasma Heme Oxygenase-1 Levels Distinguish Latent or Successfully Treated Human Tuberculosis from Active Disease. *PLoS ONE* 2013; **8**: e62618.
8. Azzurri A, Sow OY, Amedei A, et al. IFN-γ-inducible protein 10 and pentraxin 3 plasma levels are tools for monitoring inflammation and disease activity in Mycobacterium tuberculosis infection. *Microbes and Infection* 2005; **7**: 1-8.

9. Coussens AK, Wilkinson RJ, Hanifa Y, et al. Vitamin D accelerates resolution of inflammatory responses during tuberculosis treatment. *Proceedings of the National Academy of Sciences of the United States of America* 2012; **109**: 15449-54.

10. Djoba Siawaya JF, Beyers N, Van Helden P, Walzl G. Differential cytokine secretion and early treatment response in patients with pulmonary tuberculosis. *Clinical and Experimental Immunology* 2009; **156**: 69-77.

11. Djoba Siawaya JF, Bapela NB, Ronacher K, et al. Immune parameters as markers of tuberculosis extent of disease and early prediction of anti-tuberculosis chemotherapy response. *Journal of Infection* 2008; **56**: 340-7.

12. Huang CT, Lee LN, Ho CC, et al. High serum levels of procalcitonin and soluble TREM-1 correlated with poor prognosis in pulmonary tuberculosis. *Journal of Infection* 2014; **68**: 440-7.

13. Jayakumar A, Vittinghoff E, Segal MR, et al. Serum biomarkers of treatment response within a randomized clinical trial for pulmonary tuberculosis. *Tuberculosis* 2015; **95**: 415-20.

14. Lee JH, Chang JH. Changes of plasma interleukin-1 receptor antagonist, interleukin-8 and other serologic markers during chemotherapy in patients with active pulmonary tuberculosis. *The Korean journal of internal medicine* 2003; **18**: 138-45.

15. Mihret A, Bekele Y, Bobosha K, et al. Plasma cytokines and chemokines differentiate between active disease and non-active tuberculosis infection. *Journal of Infection* 2013; **66**: 357-65.

16. Ostrowski SR, Ravn P, Hoyer-Hansen G, Ullum H, Andersen AB. Elevated levels of soluble urokinase receptor in serum from mycobacteria infected patients: Still looking for

a marker of treatment efficacy. *Scandinavian Journal of Infectious Diseases* 2006; **38**: 1028-32.

17. De Groote MA, Nahid P, Jarlsberg L, et al. Elucidating Novel Serum Biomarkers Associated with Pulmonary Tuberculosis Treatment. *PLoS ONE* 2013; **8**.
18. Kedia K, Wendler JP, Baker ES, et al. Application of multiplexed ion mobility spectrometry towards the identification of host protein signatures of treatment effect in pulmonary tuberculosis. *Tuberculosis* 2018; **112**: 52-61.
19. Dorman SE, Goldberg S, Stout JE, et al. Substitution of rifapentine for rifampin during intensive phase treatment of pulmonary tuberculosis: study 29 of the tuberculosis trials consortium. *The Journal of infectious diseases* 2012; **206**: 1030-40.
20. Sigal GB, Segal MR, Mathew A, et al. Biomarkers of Tuberculosis Severity and Treatment Effect: A Directed Screen of 70 Host Markers in a Randomized Clinical Trial. *EBioMedicine* 2017; **25**: 112-21.
21. Babu S. Biomarkers for Treatment Monitoring in Tuberculosis: A New Hope. *EBioMedicine* 2017; **26**: 13-4.
22. Dorman SE, Savic RM, Goldberg S, et al. Daily rifapentine for treatment of pulmonary tuberculosis: A randomized, dose-ranging trial. *American Journal of Respiratory and Critical Care Medicine* 2015; **191**: 333-43.
23. Savic RM, Weiner M, MacKenzie WR, et al. Defining the optimal dose of rifapentine for pulmonary tuberculosis: Exposure-response relations from two phase II clinical trials. *Clinical pharmacology and therapeutics* 2017; **102**: 321-31.
24. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Classification and Regression Trees: CRC Press; 1984: 1-358.

25. Blumberg HM, Burman WJ, Chaisson RE, et al. American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. *American journal of respiratory and critical care medicine* 2003; **167**: 603-62.

26. Nahid P, Dorman SE, Alipanah N, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clinical Infectious Diseases* 2016; **63**: e147-e95.

27. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med* 2014; **371**: 1577-87.

28. Debad JD, Morris JC, Lynch V, Magnus P, Bard AJ. Dibenzotetraphenylperiflanthene: Synthesis, Photophysical Properties, and Electrogenerated Chemiluminescence. *J Am Chem Soc* 1996; **118**(10): 2374-9.

29. Kuhn M. Package 'caret', Classification and Regression Training. 2020.

30. Therneau T. Survival Package R CRAN.

31. Sung HJ, Ahn JM, Yoon YH, et al. Identification and validation of SAA as a potential lung cancer biomarker and its involvement in metastatic pathogenesis of lung cancer. *Journal of Proteome Research* 2011; **10**: 1383-95.

32. Nahid P, Bliven-Sizemore E, Jarlsberg LG, et al. Aptamer-based proteomic signature of intensive phase treatment response in pulmonary tuberculosis. *Tuberculosis (Edinburgh, Scotland)* 2014; **94**: 187-96.

33. Vesosky B, Rottinghaus EK, Stromberg P, Turner J, Beamer G. CCL5 participates in early protection against Mycobacterium tuberculosis *Journal of Leukocyte Biology* 2010; **87**: 1153-65.

34. Lee MR, Tsai CJ, Wang WJ, et al. Plasma biomarkers can predict treatment response in tuberculosis patients: A prospective observational study. *Medicine (United States)* 2015; **94**: e1628.

35. Coussens AK, Wilkinson RJ, Nikolayevskyy V, et al. Ethnic Variation in Inflammatory Profile in Tuberculosis. *PLoS Pathogens* 2013; **9**.

36. Wolday D, Tegbaru B, Kassu A, et al. Expression of Chemokine Receptors CCR5 and CXCR4 on CD4+ T Cells and Plasma Chemokine Levels During Treatment of Active Tuberculosis in HIV-1-Coinfected Patients. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2005; **39**: 265-71.

37. Kouhpayeh HR, Taheri M, Baziboroon M, Naderi M, Bahari G, Hashemi M. CCL5 rs2107538 Polymorphism Increased the Risk of Tuberculosis in a Sample of Iranian Population. *Prague medical report* 2016; **117**: 90-7.

38. Hu L, Zhang K, Yao L, Wang J. Chemokine (C-C motif) ligand 5 -28C>G is significantly associated with an increased risk of tuberculosis: a meta-analysis. *International journal of clinical and experimental medicine* 2015; **8**: 13211-8.

39. Areeshi MY, Mandal RK, Panda AK, Haque S. A Meta-Analysis of the Association between the CC Chemokine Ligand 5 (CCL5) -403 G>A Gene Polymorphism and Tuberculosis Susceptibility. *PLoS ONE* 2013; **8**: e72139.

40. Sia JK, Bizzell E, Madan-Lala R, Rengarajan J. Engaging the CD40-CD40L pathway augments T-helper cell responses and improves control of Mycobacterium tuberculosis infection. *PLOS Pathogens* 2017; **13**: e1006530.

41. Wilson D, Moosa M-YS, Cohen T, Cudahy P, Aldous C, Maartens G. Evaluation of Tuberculosis Treatment Response With Serial C-Reactive Protein Measurements. *Open forum infectious diseases* 2018; **5**: ofy253.

42. Cudahy PGT, Warren JL, Cohen T, Wilson D. Trends in C-reactive protein, D-dimer, and fibrinogen during therapy for HIV-associated multidrug-resistant tuberculosis. *American Journal of Tropical Medicine and Hygiene* 2018; **99**: 1336-41.

43. Zhao Y, Yang X, Zhang X, et al. IP-10 and RANTES as biomarkers for pulmonary tuberculosis diagnosis and monitoring. *Tuberculosis* 2018; **111**: 45-53.

44. Chowdhury IH, Ahmed AM, Choudhuri S, et al. Alteration of serum inflammatory cytokines in active pulmonary tuberculosis following anti-tuberculosis drug therapy. *Molecular Immunology* 2014; **62**: 159-68.

45. Wilson D, Nachega J, Morroni C, Chaisson R, Maartens G. Diagnosing smear-negative tuberculosis using case definitions and treatment response in HIV-infected adults. *International Journal of Tuberculosis and Lung Disease* 2006; **10**: 31-8.

46. Lawn SD, Obeng J, Acheampong JW, Griffin GE. Resolution of the acute-phase response in West African patients receiving treatment for pulmonary tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2000; **4**: 340-4.

47. Malherbe ST, Shenai S, Ronacher K, et al. Persisting positron emission tomography lesion activity and Mycobacterium tuberculosis mRNA after tuberculosis cure. *Nature Medicine* 2016; **22**: 1094-100.

## Chapter 5: Linezolid dosing strategies to minimize adverse events with high-dose, long-term treatment for extensively drug-resistant tuberculosis


### Abstract

**Background** High-dose linezolid given with bedaquiline and pretomanid has the potential to transform treatment of extensively drug resistant tuberculosis, but limited information is available about linezolid dosing to optimize treatment and minimize toxicity. We evaluated linezolid clinical trial data to provide model-based dosing recommendations to minimize toxicity.

**Methods** Pharmacokinetic-toxicodynamic modeling and simulations were performed with longitudinal data from 104 participants in the Nix-TB trial (NCT02333799) (treatment duration, 6 months; follow-up, 24 months after treatment completion). All participants initially administered a linezolid dosage of 1200 mg once daily or 600 mg twice daily. Dose adjustments were allowed per discretion of the investigator to manage linezolid toxicity. Linezolid pharmacokinetic profiles that accounted for individual dosing histories were predicted from the pharmacokinetic model and used as inputs to drive models for hemoglobin levels, platelet counts, and peripheral neuropathy scores. Final models were used to simulate and compare pharmacokinetics and safety outcomes following daily doses of 1200 mg linezolid as well as alternative dosing strategies.

**Results** Linezolid pharmacokinetics had nonlinear elimination, with AUC 2.7-fold greater at 1200 mg once daily than 600 mg once daily. Toxicity profiles were similar between 1200 mg once daily vs 600 mg twice daily. Linezolid pharmacokinetics predicted anemia, thrombocytopenia, and peripheral neuropathy ($p<0.001$). Regarding severe anemia, simulations indicated that: i) the median time to onset of severe anemia was 9 (90% CI: 9-11) weeks; ii) hemoglobin levels at pretreatment and week 4 could be used to predict severe anemia (AUROC=0.91 (0.88-0.93); iii) a greater than 10% decrease in hemoglobin at week 4 would have maximum sensitivity (82%) and

specificity (84%) for prediction; iv) a dose reduction to 600 mg triggered by this simple marker would prevent 63% (49-74) cases and v) recovery time to normal levels was 6 weeks (6-8) after dose adjustment. Severe peripheral neuropathy mostly occurred after 4 to 6 months of treatment and reversed within 6 months after linezolid dosage reductions or discontinuation. Less than 1% of severe thrombocytopenia was predicted at linezolid dosages of 1200 mg daily.

**Conclusions** Simple monitoring and biomarker-guided dose-adjustment strategies have great potential to anticipate and avoid toxicities associated with the use of high dose linezolid-containing regimens in patients with extensively drug resistant tuberculosis.

## Introduction

The WHO estimated that in 2017 there were approximately 500,000 cases of multi-drug resistant tuberculosis (MDR-TB), including 9% of the people who had extensively drug resistant TB (XDR-TB). Treatment success rates in XDR-TB patients is low, at around 34%, and it has been urged that new drugs and regimens are desperately needed to improve cure rates.[1]

Linezolid is one of the most potent repurposed agents for the treatment of drug resistant TB that has been shown to be efficacious against MDR and XDR-TB when added to failing regimens. In a recent meta-analysis with 12,030 MDR-TB patients, treatment with linezolid significantly reduced mortality and increased treatment success compared to failure or relapse.[2] Based on this evidence, a rapid communication published in 2018 from the WHO prioritized its use for MDR-TB.[3] In the Phase 3 Nix-TB trial (NCT02333799) that evaluated high-dose linezolid (1200 mg daily) in combination with bedaquiline and pretomanid (BPaL) for 6 months to treat XDR and treatment intolerant or non-responsive (TI/NR) MDR-TB, 102 of 109 (94%) of enrolled participants completed treatment (all surviving participants), and 98 (90%) had culture negative status at 6 months post treatment, exceeding the 34% success rate currently reported by WHO.[1,4,5]

Based on the efficacy and safety data from the Nix-TB trial, the FDA approved BPaL for treatment of XDR-TB and TI/NR MDR-TB.[4,5] However, the use of linezolid is controversial because of concerns over safety and tolerability. In the Nix-TB trial, adverse events including peripheral neuropathy (81% of participants), anemia (37%), and thrombocytopenia (6%) resulted in substantial frequencies of participants who required linezolid discontinuation (28%), interruption (46%), or dose reduction (39%). The mean maximum interruption of linezolid administration because of an adverse event was 38 days. Nevertheless, the hematological and neuropathy adverse events were generally reversible leading to the approved recommended

starting dose and duration for linezolid in the combination regimen of 1200 mg daily for 26 weeks, with dose adjustments for any known linezolid adverse reactions.[4,5]

Another review of linezolid treatment for MDR-TB and XDR-TB showed adverse events in 55% of patients at all doses investigated (range, 300 to 1200 mg daily).[6] Although linezolid trough levels have been associated with linezolid-related adverse events in XDR-TB patients, 40% of patients with trough levels < 2 mg/L develop an adverse event.[7] Despite the narrow therapeutic range, linezolid is among the most effective drugs for treatment of MDR-TB and XDR-TB.[4,5] Therefore, it is important to define optimal dosages, treatment durations, and best clinical practices for the use of linezolid in TB. However, limited information is available about the optimal time and extent of linezolid dosage adjustments required to optimize efficacy and minimize adverse events.

In this study, we quantitatively characterized the relationship between linezolid dose, plasma concentrations and the time course of all major toxicities (anemia, thrombocytopenia, and peripheral neuropathy) using longitudinal long-term data from the Nix-TB trial. Then, we provided model-based recommendations about optimal dosing to minimize adverse events.

## Methods

### Study design

Data from the Nix-TB clinical trial that evaluated the BPaL regimen after 6-months of treatment in participants with pulmonary XDR-TB or TI/NI MDR-TB was used in this study. The Nix-TB study enrolled participants presenting with one of the following three pulmonary TB conditions: i. XDR-TB with documented resistance to isoniazid, rifamycins, a fluoroquinolone and an injectable, ii. MDR-TB with documented non-response to treatment with best available regimen for 6 months or more prior to enrollment who in the opinion of the investigator have been adherent

to treatment and will be adherent to study regimen (TI MDR-TB), or iii. MDR-TB with documented intolerance to p-aminosalicylic acid, ethionamide, aminoglycosides or fluoroquinolones (NR MDR-TB). All pulmonary TB conditions must have documented culture positive results for *Mycobacterium Tuberculosis* within 3 months prior to screening. Males or nonpregnant females of at least 14 years of age and 30 kg body weight were included in the study. Participants were required to provide consent to HIV testing. Additional inclusion and exclusion criteria details are available in the original study protocol.[4]

Participants were treated for 6 months with bedaquiline at 400 mg once daily (QD) for 2 weeks then 200 mg three times per week, pretomanid at 200 mg QD, and linezolid 1200 mg daily at 600 mg twice daily (BID; according to initial protocol) or 1200 mg QD (according to amended protocol) schedules. The protocol was amended to require all participants to begin treatment with linezolid at the 1200 mg QD dosage because toxicity was thought to be caused by mitochondrial toxicity and may be lessened by lowering the time the linezolid concentration is greater than the threshold for mitochondrial toxicity.[4] For suspected linezolid-induced toxicities, linezolid was dose reduced, interrupted, or discontinued while participants remained on bedaquiline and pretomanid. Time, duration, and dose reductions of all dose adjustments for each individual were recorded and used for our analysis. Pretreatment levels were defined as measurements made before the first dose of BPaL taken. All participants provided pre-dosing pharmacokinetic (PK) samples at week 2, 8, and 16 and a subset of participants provided intensive PK samples at week 16 with samples collected at pre-dose and 0.5, 1, 2, 4, 8, 12, 12.5, 13, 14, 16, 20 and 24 hours after dosing. PK for bedaquiline, bedaquiline metabolite M2, and pretomanid were also available but not used for the current analysis.

Complete blood counts including hemoglobin level and platelet count were performed at pretreatment, weekly up to week 16 and at week 20 and 26 of treatment, and peripheral neuropathy

screening was performed at pretreatment, weeks 4, 8, 12, 16, 20, and 26 during treatment and months 3, 6, 12, and 24 after completion of treatment. Subject elicited symptoms (i. interference with walking or sleeping, ii. pain, aching or burning in feet, iii. "pins and needles" in feet or legs, iv. numbness (lack of feeling) in feet or legs), perception of vibration (left and right foot), and deep tendon reflexes (left and right foot) were collected at each peripheral neuropathy screen.

The investigators were responsible for defining and eliciting adverse events by observing the participants and recording adverse events observed by themselves or reported by the participants during the trial. Severity of adverse events were made per the Division of Microbiology and Infectious Disease (DMID) Adult Toxicity table.[8]

Of the 109 enrolled patients, 88 patients were included in the model development process. An additional 16 patients who were still in the study at time of model development were used for external validation of the final population models. Five patients had unverifiable dosing histories, hence were not included.

Additional information on study design and data collection is available in the original publication.[4]

**Population pharmacokinetic-toxicodynamic modeling**

Population PK modeling started with simple one or two compartment distribution models with linear elimination and first order absorption. More complex models with respect to absorption and elimination (nonlinear elimination) were then tested to identify the best model for our population. Between subject variability was modeled exponentially and residual variability was modeled with a proportional error model.

It was critical to investigate drug dosing and schedule information as potential predictors of toxicity response. Therefore, the focus in the PK-toxicodynamic modeling was on investigating the potential for dynamic linezolid concentration-time profiles, which take into account individual

dose reductions and dose interruptions, as the drug exposure variable to predict hemoglobin level, platelet count, and peripheral neuropathy scores. A sequential approach was used to develop PK-toxicodynamic models. First, empirical Bayesian estimates of PK parameters were used to estimate individual concentration-time profiles taking into account dose adjustments. Then, instantaneous linezolid concentrations were input for the toxicity models to drive response as a covariate with linear, Emax, or sigmodal Emax relationships.

Hemoglobin levels and platelet counts were characterized with delayed response models. Mechanistic, semi-mechanistic, and more empirical indirect models were tested separately for each hemoglobin level and platelet count model. The effect of instantaneous linezolid concentrations was included by inhibiting the first-order proliferative rate of progenitor cells (mechanistic or semi-mechanistic models) or more empirically the zero-order synthesis rate of response in an indirect model. The most mechanistic model that supported the data for each hematological response was used. Between subject variability was modeled exponentially in each model. Additive and/or proportional error models were explored to explain residual unexplained variability.

As several participants had increased hemoglobin levels at the end of BPaL treatment compared to pretreatment (73/88, 83%), we modeled hemoglobin levels obtained from participants of other trials without linezolid (OFLOTUB (NCT00216385)[9], REMoxTB (NCT00864383)[10], and RIFAQUIN (ISRCTN 44153044)[11]) to describe the resolution of anemia upon starting anti-TB treatment (data not shown). Empirical models (i.e., linear, exponential, logistic growth models) were tested to fit the independent dataset. The model that described hemoglobin profiles in the absence of linezolid was included in the Nix-TB hemoglobin model as an additional synthesis rate of response (Figure 5.1).

For peripheral neuropathy modeling, the maximum score of four subject-elicited symptom questions was used. The four symptom questions were based on: i.) interference with walking, ii.) pain, aching, or burning in feet or legs, iii.) pins and needles in feet or legs, and iv.) numbness in feet or legs. Participants were asked to rank each symptom from 0 (normal) to 10 (worst/severe) for up to 24 months post treatment. The maximum scores were grouped into four categories for modeling: normal (maximum score 0), minimal (1-3), modest (4-7) and severe (8-10). A proportional odds model was used to describe the proportion of scores over time, with a distributional effect compartment to account for the generally delayed onset of neuropathy. Additional information on proportional odds modeling used in this analysis is available in Appendix Text A.1.

Covariates (age, sex, weight, BMI and HIV status) were tested as predictors of model parameters that describe PK, hemoglobin levels, platelet counts, and peripheral neuropathy scores using an automated stepwise covariate modeling procedure. This method involves stepwise testing of linear and nonlinear relationships in a forward inclusion ($p < 0.01$) and backward exclusion ($p < 0.001$) procedure. Model evaluation was performed throughout the model development process by evaluating goodness-of-fit plots, visual predictive checks, and bootstrap estimates and confidence intervals. Additional information on the model structures, differential equations, and final model estimates are available in Figure 5.1, Appendix Text A.1, and Appendix Table A.7 - Appendix Table A.13.

**Model-based simulations**

The final PK-toxicodynamic models were used to perform 1000 simulations for 500 patients over 6 months for hematological toxicities and 24 months (6 months of treatment and 18 months of recovery) for peripheral neuropathy. Simulations assessed steady-state PK (area under the concentration-time curve over 24 hours, AUC, maximum concentrations, $C_{max}$, and minimum concentrations, $C_{min}$); probability of ≥ grade 3 hematological toxicities according to the DMID Adult Toxicity table (severe anemia, hemoglobin < 8 g/dL; severe thrombocytopenia, platelets < $50x10^9$ platelets/L)[8]; probability of peripheral neuropathy graded scores; management and reversibility of toxicities; and attainment of efficacy targets. Normal levels were based on DMID table for hematological toxicities (hemoglobin level ≥ 10.6 g/dL; platelet count, ≥ $100 \times 10^9$/L).

We considered linezolid dosages of 600 mg or 1200 mg total daily (BID or QD) for 6 months. Strategies to manage toxicities were investigated with linezolid dosage reductions to 600 mg or 300 mg total daily or discontinuations. Anemia management strategies involved monitoring weekly hemoglobin and steady-state linezolid trough concentrations. To assess reversal of peripheral neuropathy, the first occurrence of a modest or severe score was considered an event, and a dose adjustment was made at the time of the event via simulations. Following the dose adjustment, the time of reversal was defined as the time of the first of two consecutive minimal or normal scores. The distribution of the time from the event to reversed neuropathy was determined.

Although efficacy outcomes are not evaluated in this report, to allow some judgment of the potential impact of linezolid dosage adjustments, two PK-based pharmacodynamic targets for linezolid were assessed using free-drug concentrations in the simulations: i) ratio of the free area under the concentration-time curve over 24 hours (fAUC) to minimum inhibitory concentration (MIC) (fAUC/MIC) > 100, and ii) time above MIC (TMIC).[12,13] MIC of 0.5 mg/L and protein binding fraction of 31% were used.[14-17]

**Statistical analysis**

Separate Cox regression analysis was performed to identify predictors associated with the time to first occurrence of an investigator-reported peripheral neuropathy, anemia, and thrombocytopenia adverse event. Participant characteristics (HIV status, BMI, weight, age, and sex) and linezolid exposure (steady state AUC, $C_{max}$, or $C_{min}$ and observed trough after 2 weeks of treatment) were assessed as predictors for each investigator-reported adverse event. Hemoglobin levels at pretreatment, week 1, 2, and 4 were assessed as predictors of investigator-reported anemia adverse events. Area under the receiver operating characteristic curve (AUROC) was used to assess model discrimination. Data management, statistical analysis and visualization were performed using the R software (v3.4.1). Pharmacokinetic and safety data were characterized based on a population nonlinear mixed effects modelling approach using the software NONMEM 7.4.1.

## Results

### Participant and data characteristics

There were no significant differences in participant characteristics or pretreatment hemoglobin levels, platelet counts, or peripheral neuropathy scores between participants who initially had 600 mg BID vs 1200 QD linezolid dosages. Approximately half of the participants (50/104) were HIV co-infected (Table 5.1). Median levels of hemoglobin levels increased from 12.1 (range: 7.4-16.1) at pretreatment to 13.5 (9.5-19.4) at the end of treatment (Table 1, $P < 0.001$), and platelet counts decreased from 354 (137-1083) at pretreatment to 262 (116-840) at the end of treatment (Table 5.1).

### Population model development

Linezolid pharmacokinetics was best described by a two-compartment disposition model with nonlinear (Michaelis-Menten) elimination. No covariates were included in the final model to describe variability in PK. Hemoglobin levels were best described with an indirect model (Figure 5.1). Instantaneous linezolid concentrations that account for individual dosing histories were shown to inhibit the synthesis rate of hemoglobin with a maximum inhibition of 100% and a half-maximal inhibitory concentration (IC50) of 7.7 mg/L. A logistic growth model that empirically described the hemoglobin levels without linezolid was included in the Nix-TB hemoglobin model as an additional synthesis rate of response. Platelet counts were best described with a semi-mechanistic hematological toxicity model (Figure 5.1). Instantaneous linezolid concentrations were shown to inhibit the proliferation rate for the stem-cell-like platelet-progenitor cell compartment with 8% maximum inhibition and IC50 of 4.1 mg/L. Peripheral neuropathy scores were described by a proportional odds model with probabilities driven by linezolid concentrations in an effect compartment to describe the 2- to 3-month delay in onset. Model details are available in Figure 5.1, Appendix Text A.1 and Appendix Table A.7 - Appendix Table A.13. Each model

described its respective data reasonably well, as shown by the visual predictive checks with model-development data and external-validation data (Figure 5.2).

**Table 5.1** Participant characteristics and summary of data available for linezolid dosing model development and external validation in extensively drug-resistant tuberculosis.

| | **Model development** | | **External Validation** |
|---|---|---|---|
| **Initial Dose** | **600 BID** | **1200 QD** | **1200 QD** |
| **Participant characteristics** | | | |
| N | 42 | 46 | 16 |
| Males [N (%)] | 23 (55) | 23 (50) | 7 (43) |
| Age (y) [median (range)] | 31 (18-55) | 36 (21-60) | 36 (17-48) |
| Weight (kg) [median (range)] | 59 (29-112) | 54 (33-89) | 54 (32-106) |
| BMI ($kg/m^2$) [median (range)] | 19.8 (12.4-41.1) | 19.7 (13.6-36.1) | 18.9 (15.1-38.9) |
| HIV positive [N (%)] | 18 (42) | 25 (54) | 7 (43) |
| Creatinine clearance (ml/min) [med (range)][a] | 102 (46-167) | 104 (42-180) | 107 (43-179) |
| **Dose adjustments** | | | |
| Participants with dose adjustment | 38 (90) | 40 (87) | 10 (90) |
| Time of first dose adjustment (days) [med (range)] | 61.5 (18-174) | 75.5 (9-170) | 60.5 (13-153) |
| **Pharmacokinetic data** | | | |
| Participants in PK substudy [N (%)] | 16 (38) | 4 (8) | 5 (31) |
| Total evaluable PK samples [n] | 243 | 154 | 100 |
| **Hemoglobin data** | | | |
| No. of Hb samples per participant [med (range)] | 19 (5-24) | 19 (10-24) | 20 (17-24) |
| Total evaluable Hb samples [n] | 773 | 835 | 319 |
| Pretreatment Hb (g/dL) [median (range)] | 12.4 (8.5-16.1) | 11.8 (8.7-15.6) | 12.1 (7.4-13.9) |
| End of treatment Hb (g/dL) [median (range)][b] | 13.6 (9.8-19.4) | 13.7 (9.5-17.0) | 12.8 (11.2-16.8) |
| **Platelet data** | | | |
| No. of platelet samples per participant [med (range)] | 19 (5-24) | 19 (9-24) | 19 (16-24) |
| Total evaluable platelet samples [n] | 761 | 816 | 315 |
| Pretreatment Platelet ($10^9$ platelets/L) [median (range)] | 354 (137-1045) | 348 (188-1083) | 436 (139-730) |
| End of treatment Platelet (g/dL) [median (range)][c] | 254 (116-840) | 262 (175-478) | 312 (167-409) |
| **Peripheral Neuropathy data** | | | |
| No. of neuropathy scores per participant [median (range)] | 10 (2-14) | 8 (3-11) | 10 (10-13) |
| Total evaluable neuropathy scores [n] | 418 | 382 | 170 |
| No. of participants without neuropathy at pretreatment [N (%)][d] | 32 (76) | 31 (67) | 16 (100) |
| No. of participants with minimal neuropathy at pretreatment [N (%)][d] | 5 (12) | 5 (11) | 0 (0) |
| No. of participants with modest neuropathy at pretreatment [N (%)][d] | 4 (10) | 10 (22) | 0 (0) |
| No. of participants with severe neuropathy at pretreatment [N (%)][d] | 1 (2) | 0 (0) | 0 (0) |

Hb, hemoglobin; BMI, body mass index, HIV: human immunodeficiency virus
[a]Calculated with Cockcroft-Gault equation using serum creatine levels and ideal body weight. Creatinine clearance missing for two participants in the external validation dataset.
[b]End of treatment hemoglobin levels missing for 7 participants in model development dataset.
[c]End of treatment platelet counts missing for 8 participants in model development dataset.
[d]Based on maximum of four subject elicited symptom questions.

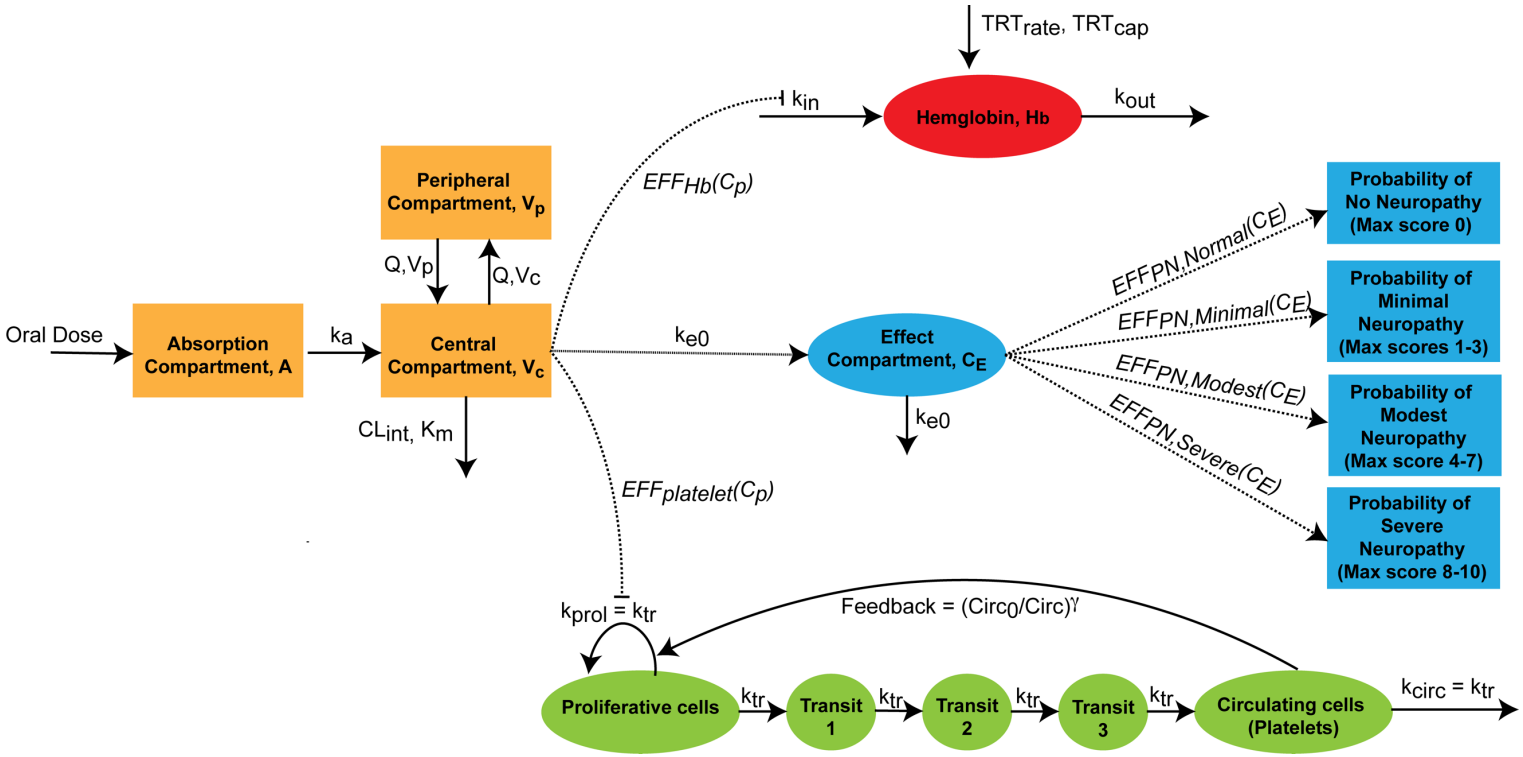



**Figure 5.1** Structural pharmacokinetics-toxicodynamic model for linezolid concentrations, hemoglobin levels, platelet counts and peripheral neuropathy symptom scores. Pharmacokinetic model (orange): A, absorption compartment; Vc, apparent central volume of distribution; Vp, apparent peripheral volume of distribution; Q, apparent inter-compartment clearance; ka, absorption rate constant; Km, Michaelis-Menten constant; and CLint, Michaelis-Menten intrinsic clearance; VMAX = CLint*Km, maximum rate of elimination (saturated elimination); Cp, linezolid concentration in central compartment. Hemoglobin level model (red): Hb, hemoglobin compartment; kout, first order hemoglobin elimination rate constant; kin, zero-order hemoglobin synthesis rate constant, which equals the product of pretreatment hemoglobin, BASEHb, and kout; TRTrate, rate of increase in hemoglobin, and TRTcap, maximum hemoglobin level, for logistic growth function describing improved hemoglobin levels upon initiating anti-TB treatment; EFFHb(Cp), linezolid concentration-response relationship (i.e., sigmoidal EMAX relationship). Peripheral neuropathy model (blue): CE, effect compartment linezolid concentrations; ke0, first order effect compartment rate constant; EFFy(CE), linezolid effective concentration-response relationship (i.e., sigmoidal EMAX relationship). Platelet count model (green): Prol, proliferative/progenitor cells compartment; Transit x, transit compartments representing cell maturation process; Circ, circulation compartment; kprol, first-order proliferation rate constant, assumed equal to ktr; ktr, first-order transit rate constant; kcirc, first-order rate of elimination from circulation compartment; Circo, baseline platelet level; and $\gamma$ homeostatic feedback parameter.

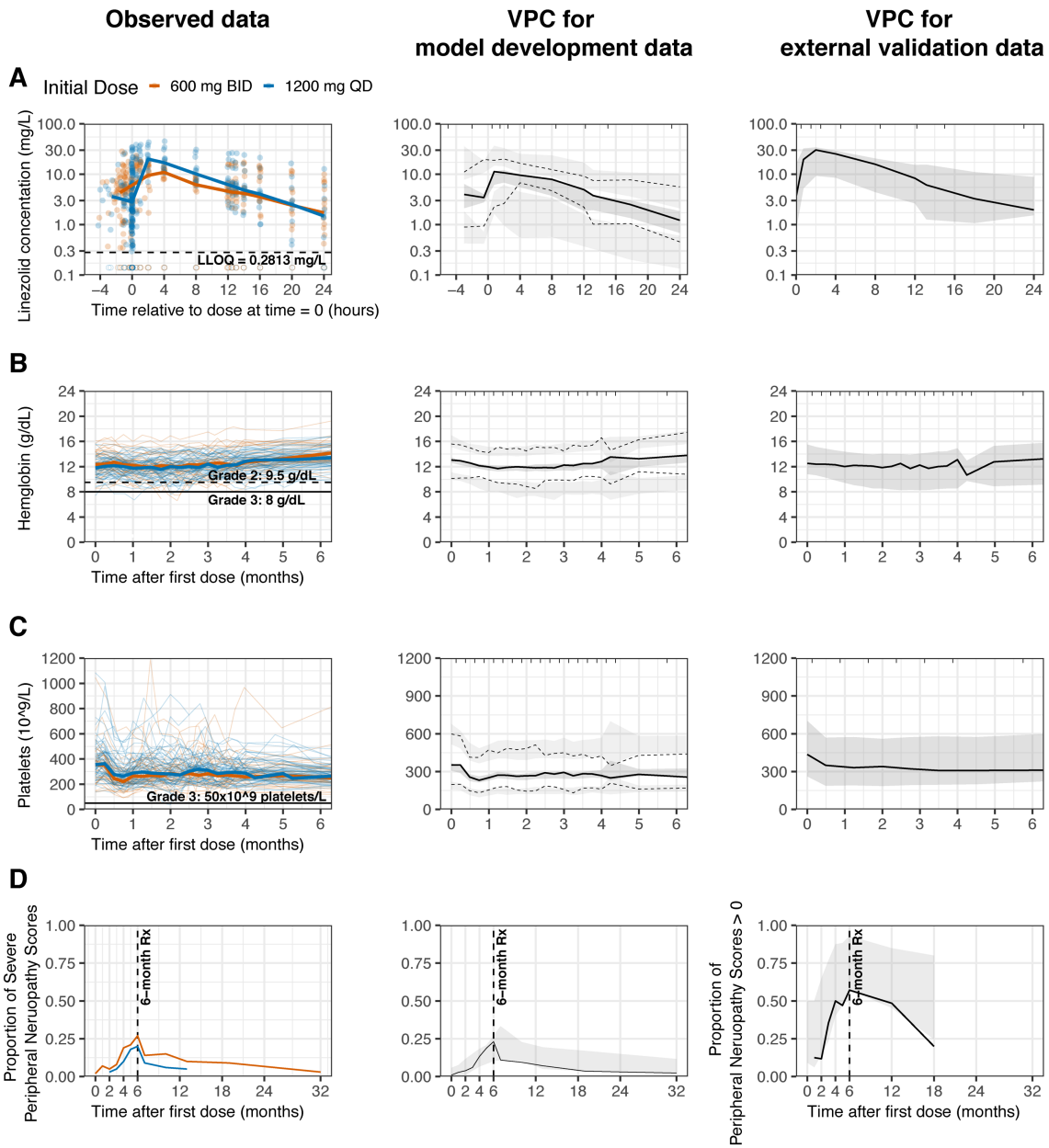



**Figure 5.2** Linezolid pharmacokinetic model and pharmacokinetic-toxicodynamic models: observed data and visual predictive checks. Left panel shows the observed data (points or thin solid line) and median (proportion for (D)) of the observed data (thick solid line) stratified by initial dosage (red = 600 mg BID, blue = 1200 mg QD). Middle panel shows the visual predictive checks for model development data. For (A), (B), and (C), the solid black line is the median and dashed black lines show the 5$^{th}$ and 95$^{th}$ percentiles of the observed data. The shaded areas are the 95% confidence intervals of the 95$^{th}$, median, and 5$^{th}$ percentiles of the model predicted simulations, respectively. For (D) the solid black line is the observed proportion and the shaded area is the 95% prediction interval of model predicted simulations. Right panel shows the visual predictive checks for external validation data. Solid black line is the median (proportion for (D)) of the observed data and the shaded area is the 95% prediction interval of model predicted simulations. (A) pharmacokinetic model, (B) pharmacokinetic-toxicodynamic model for hemoglobin levels, (C) pharmacokinetic-toxicodynamic model for platelet counts, and (D) pharmacokinetic-toxicodynamic model for peripheral neuropathy severe symptom scores (based on maximum of four symptom scores). Additional predictive checks for peripheral neuropathy normal, minimal, and modest symptom scores available in Appendix Table A.14. LLOQ, lower limit of quantification; Rx, treatment

In each toxicity model, predicted instantaneous linezolid concentrations that accounted for individual dosing histories and interindividual variation had better predictions of longitudinal toxicity profiles than models that included observed week 2 trough levels (Table 5.2). Participant characteristics did not predict the potency of linezolid exposure or pretreatment toxicity levels (final models described in Appendix Table A.7 - Appendix Table A.13).

**Table 5.2** Linezolid drug exposure as a predictor of linezolid-induced toxicities in participants with extensively drug-resistant tuberculosis.

| | **Anemia[a]** | **Thrombocytopenia[a]** | **Peripheral Neuropathy[a]** | **Hemoglobin[b]** | **Platelets[b]** | **Peripheral Neuropathy Scores[b]** |
|---|---|---|---|---|---|---|
| Instantaneous linezolid concentrations | Not tested | Not tested | Not tested | **P << 0.001** | **P <<< 0.001** | **P << 0.001** |
| Week 2 linezolid troughs | P = 0.2 | P = 0.2 | P =0.9 | P = 0.4 | **P << 0.001** | P = 0.9 |
| P values for inclusion of covariates as predictors of time to first occurrence of investigator-reported clinical event in Cox regression analysis[a] or as predictors of longitudinal hemoglobin levels, platelet counts, and peripheral neuropathy scores in the PK-toxicodynamic models[b]. | | | | | | |

**Effect of linezolid dosages on toxicity and efficacy**

Simulations were performed to predict PK and toxicity profiles following 1200 and 600 mg total daily doses. The simulated AUC at steady state was similar between BID vs QD dosing at the same total daily doses. Higher maximum linezolid concentrations and lower minimum concentrations were observed with QD than BID dosages. The AUC at linezolid 1200 mg QD was 2.7 (90% CI: 2.6-2.8) times greater than 600 mg QD because of nonlinear clearance in the model (Figure 5.3 and Table 5.3). In simulations, ≥ 99% of participants treated with 1200 mg total daily met the target of fAUC/MIC > 100, but 600 mg total daily resulted in 19% (600 mg QD, median fAUC/MIC: 134) or 26% (300 mg BID, median fAUC/MIC: 124) of participants below this target (Table 5.3).

Toxicity profiles were not substantially different between BID and QD dosing at the same total daily dose. However, simulations showed higher frequencies of severe anemia and severe

peripheral neuropathy scores but no differences in platelet toxicity with total daily linezolid of 1200 mg vs 600 mg (Figure 5.3 and Table 5.3).

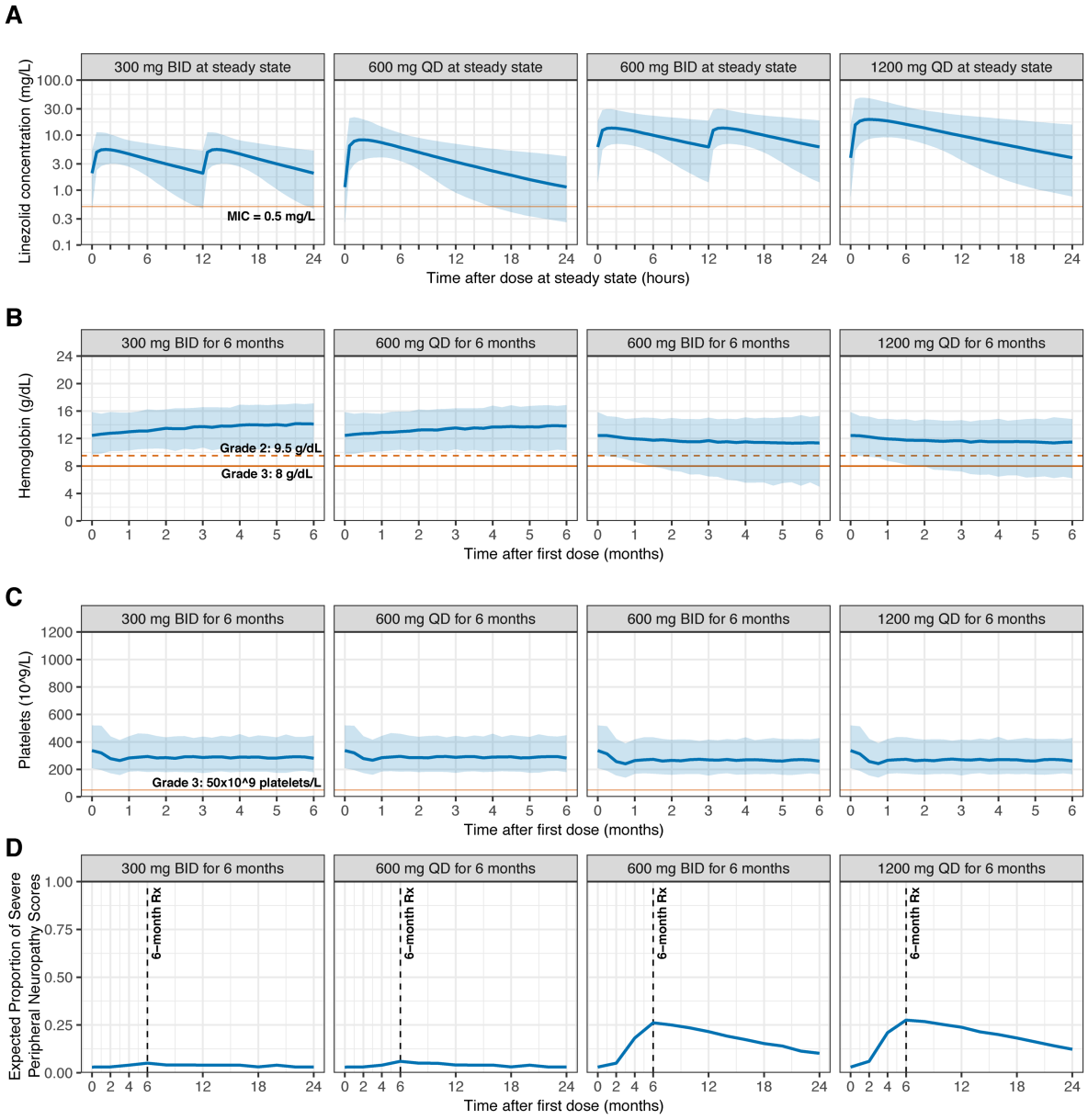



**Figure 5.3** Simulated pharmacokinetic and toxicity profiles after linezolid total daily doses of 600 mg or 1200 mg. (A) Simulated pharmacokinetic profiles. Solid blue lines represent the typical patient (median of simulations) and the shaded areas the 90% prediction interval. Red solid line represents an MIC of 0.5 mg/L. (B) Simulated hemoglobin profiles. Solid blue lines represent the typical patient and the shaded areas the 90% prediction intervals. Red dashed line represents a DMID definition of Grade 2 toxicity (hemoglobin levels < 9.5 mg/L) and red solid line represents a DMID definition of Grade 3 toxicity (hemoglobin levels < 8.0 mg/L). (C) Simulated platelet profiles. Solid blue lines represent the typical patient (median of simulations) and the shaded areas the 90% prediction intervals. Red solid line represents the DMID definition of Grade 3 toxicity (platelet levels < 50x109 platelets/L). (D) Simulated expected proportions of severe peripheral neuropathy maximum scores. Simulated proportions of normal, minimal, and modest score available in Appendix Figure A.6.

**Table 5.3** Simulated pharmacokinetic, efficacy, and toxicity parameters after total linezolid daily doses of 600 mg or 1200 mg and proposed dosage adjustments for management of anemia toxicity in participants with extensively drug-resistant tuberculosis.

| | **600 mg total daily dose for 6 months** | | **1200 mg total daily dose for 6 months** | | **Initial 1200 mg QD then management strategy for anemia at week 4** | | |
|---|---|---|---|---|---|---|---|
| | **300 mg BID** | **600 mg QD** | **600 mg BID** | **1200 mg QD** | **Reduced to 600 mg QD** | **Continued with 1200 mg QD** | **Combined[c]** |
| **Pharmacokinetics - Pharmacodynamics** | | | | | | | |
| $C_{max}$ (mg/L)[a] | 6 (3-12) | 9 (4-22) | 14 (7-31) | 21 (10-51) | 8 (4-20) | 21 (10-49) | 17 (5-46) |
| $C_{min}$ (mg/L)[a] | 2.0 (0.5-5.3) | 1.1 (0.3-4.1) | 6.1 (1.4-18.0) | 3.9 (0.8-15.0) | 1.1 (0.3-4.1) | 4.4 (0.8-15.6) | 3.0 (0.5-12.9) |
| AUC (mg/L*h)[a] | 90 (53-164) | 97 (56-177) | 236 (125-520) | 262 (138-561) | 91 (52-169) | 268 (139-628) | 228 (68-564) |
| TMIC over 24 hours[a] | 24 (22-24) | 24 (16-24) | 24 (24-24) | 24 (24-24) | 24 (15-24) | 24 (24-24) | 24 (19-24) |
| % of patients with fAUC/MIC > 100[b] | 74 (71-78) | 81 (78-84) | 99 (99-100) | 100 (100-100) | 77 (74-80) | 100 (100-100) | 94 (92-96) |
| **Anemia** | | | | | | | |
| % of patients with Grade 3 or greater toxicity (Hb < 8 g/dL)[b] | < 1 (0-1) | 1 (0-2) | 20 (17-23) | 16 (14-19) | 11 (9-14) | 4 (3-6) | 6 (4-8) |
| % of patients with early dose reduction[b] | 8 (6-10) | 9 (7-11) | 27 (24-30) | 29 (26-32) | 100 | 0 | 29 (26-32) |
| **Platelets** | | | | | | | |
| % of Grade 3 or greater toxicity (Platelets < $50x10^9$ Platelets/L)[b] | < 1 (0-<1) | < 1 (0-<1) | < 1 (0-<1) | < 1 (0-<1) | - | - | - |

**Table 5.3 continued.**

| | 600 mg total daily dose for 6 months | | 1200 mg total daily dose for 6 months | | Initial 1200 mg QD then management strategy for anemia at week 4 | | |
|---|---|---|---|---|---|---|---|
| | **300 mg BID** | **600 mg QD** | **600 mg BID** | **1200 mg QD** | **Reduced to 600 mg QD** | **Continued with 1200 mg QD** | **Combined** |
| **Peripheral Neuropathy** | | | | | | | |
| % of patients without neuropathy at 6 months[b] | 65 (62-67) | 62 (59-65) | 26 (24-29) | 24 (22-27) | - | - | - |
| % of patients with minimal neuropathy at 6 months[b] | 16 (13-18) | 16 (14-19) | 17 (15-20) | 17 (15-20) | - | - | - |
| % of patients with modest neuropathy at 6 months[b] | 14 (12-17) | 16 (13-18) | 30 (27-33) | 31 (27-34) | - | - | - |
| % of patients with severe neuropathy at 6 months[b] | 5 (4-7) | 6 (5-8) | 26 (24-29) | 27 (25-31) | - | - | - |

AUC, area under the concentration-time curve over 24 hours; $C_{max}$, maximum concentration; $C_{min}$, minimum concentration; fAUC/MIC, ratio of the free AUC curve to minimum inhibitory concentration; TMIC, time above MIC; Hb, hemoglobin

[a]Median (90% prediction interval)

[b]Median (90% confidence interval)

[c]Parameters for combined population with reduced dosage to 600 mg QD and continued dosage of 1200 mg QD.

-Synergistic modeling not performed

**Management of anemia**

At least 1 investigator-reported anemia adverse event was reported in 35 of the 88 model development participants (40%) (median time to first anemia event, 8 weeks (90% CI: 7-10)). Hemoglobin levels after 4 weeks of linezolid treatment had higher AUROC for predicting reported anemia adverse events than hemoglobin levels at earlier timepoints (pretreatment, 1 week, or 2 weeks), linezolid exposure parameters (steady state AUC, $C_{max}$, or $C_{min}$ and observed trough after 2 weeks of treatment) or participant characteristics (Figure 5.4A and Appendix Table A.15). In simulations with the hemoglobin toxicity model, the median time to onset of severe anemia was 9 weeks (90% CI: 9-11), and the AUROC for predicting model-simulated severe anemia was also higher for hemoglobin levels after 4 weeks than linezolid trough concentrations (Figure 5.4B). The threshold of 10% decrease in hemoglobin levels at 4 weeks vs pretreatment had the highest combination of sensitivity and specificity in identifying severe anemia (both > 0.80) (Figure 5.4C). With this threshold as a trigger for linezolid dosage reduction from 1200 to 600 mg QD to minimize severe anemia, simulations showed that the frequency of severe anemia events could potentially be decreased by a median of 63% from 16% to 6% (Figure 5.5 and Table 5.3). In the patients that meet the trigger and dose is reduced to 600 mg QD, the median recovery to pretreatment hemoglobin levels was 9 weeks (90% CI: 9-12) and to normal levels was 6 weeks (90% CI: 6-8). Decreasing the linezolid dosage from 1200 to 600 mg QD did not substantially affect efficacy targets (Table 5.3). Dosage adjustments to 300 mg once daily or discontinuation yielded similar results (data not shown).

**Management of thrombocytopenia**

Five of 88 model development participants (6%) had an investigator-reported thrombocytopenia adverse event, with only one being severe (1%). Simulations with the platelet toxicity model also showed that the frequency of severe thrombocytopenia was less than 1% for

linezolid dosages of 600 mg or 1200 mg total daily. Management strategies for severe thrombocytopenia were not further investigated because of the low frequency in the clinical trial and model simulations.

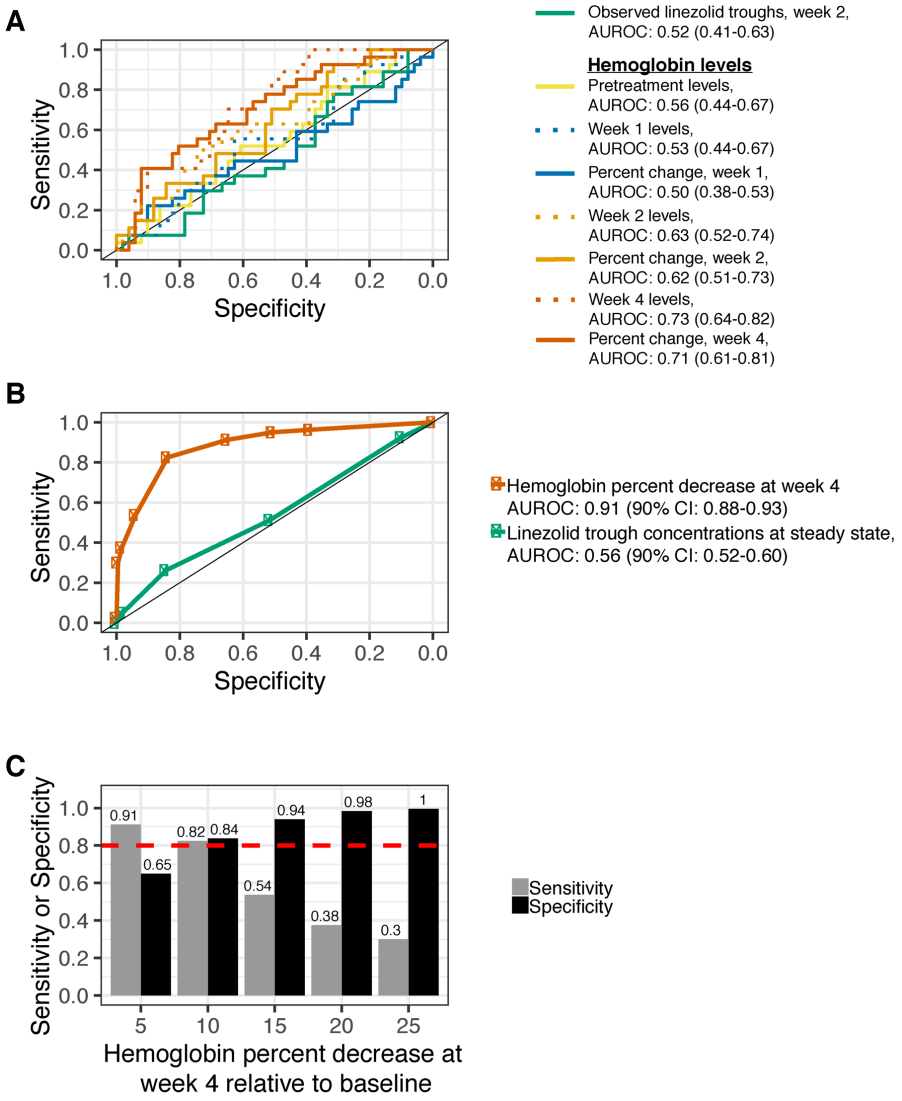



**Figure 5.4** Predictors of anemia associated with linezolid treatment (BPaL regimen). (A) ROC curves for univariate models that predict investigator-reported anemia adverse events. (B) ROC curves for simulated prediction of severe anemia (as defined by DMID Grade 3 or higher toxicity; hemoglobin levels < 8 g/dL) using the hemoglobin population model (C) Simulated sensitivity and specificity rates to predict severe anemia for management strategies using various hemoglobin percent change thresholds at week 4.

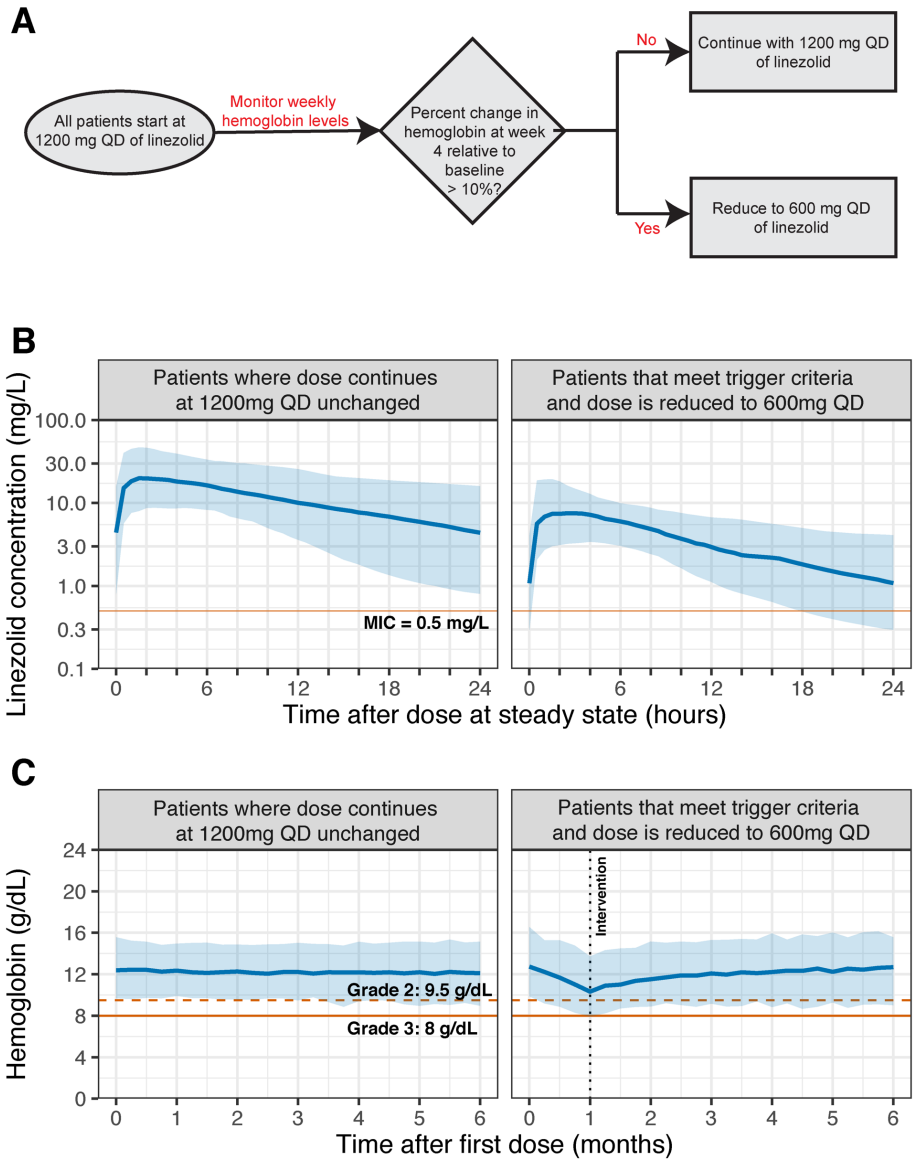



**Figure 5.5** Proposed management strategy to predict and minimize severe anemia associated with linezolid treatment (BPaL regimen) for extensively drug-resistant tuberculosis. (A) Decision tree algorithm. (B) Simulated steady state pharmacokinetic profiles after implementing anemia toxicity management strategies for initial linezolid 1200 mg QD dosage. (C) Simulated hemoglobin level profiles after implementing anemia toxicity management strategy for initial linezolid 1200 mg QD dosage. Solid blue lines represent typical participant (median of simulations); shaded areas represent 90% prediction interval.

**Management of peripheral neuropathy**

The distribution of maximum scores from 104 participants (88 model development + 16 external validation) shows that some pretreatment peripheral neuropathy was present, the frequency of peripheral neuropathy increased from 3 to 6 months after beginning BPaL treatment, and the majority of participants who have completed two years post treatment follow-up reversed or improved (Figure 5.6A). Observed data and simulations showed that peripheral neuropathy had reversed in most participants within 2 to 6 months after linezolid dosage reductions to 600 mg or 300 mg daily or discontinuations, but improvement continued for up to 1 year or more (Figure 5.6C-Figure 5.6E). In simulations, greater decreases in linezolid dosages were associated with slightly faster times to reversal of neuropathy, but linezolid discontinuation did not provide a substantial advantage over dosage reduction. Most severe neuropathy occurred toward the end of treatment (4 to 6 months after treatment initiation), but severe neuropathy persisted in 3 of 85 participants (4%) 1 year and 1 of 43 participants (2%) 2 years after treatment completion (based on available data at time of analysis) (Figure 5.6A). In the three participants with severe neuropathy at 1 year after treatment completion, peripheral neuropathy was no longer severe at 2 years after treatment completion (Figure 5.7). However, one participant had increasing maximum neuropathy scores even at the end of the 2 year follow up with a severe score in interference with walking and moderate score in numbness to the feet and legs (Figure 5.8).

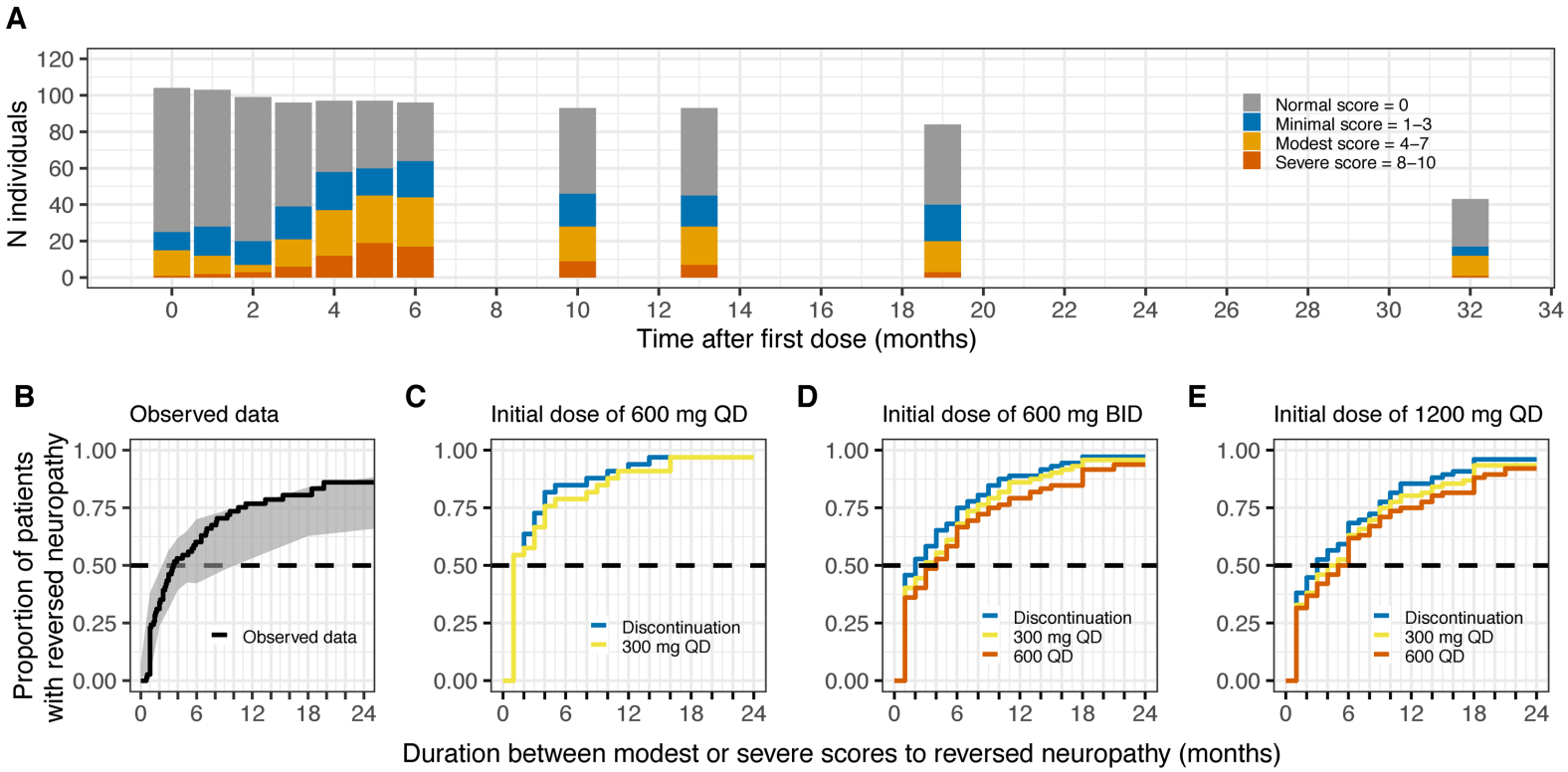



**Figure 5.6** Reversibility of peripheral neuropathy associated with linezolid treatment (BPaL regimen) for extensively drug-resistant tuberculosis. (A) Distribution of maximum subject-elicited peripheral neuropathy symptom scores over time. (B) Proportion of participants with peripheral neuropathy reversed based on observed data. Shaded area, 95% prediction interval of model predicted simulation that take into account dosing histories. (C), (D), and (E) Proportion of participants with peripheral neuropathy reversed after dose adjustments based on model-simulated data. (C) Initial linezolid dosage 600 mg QD. (D) Initial linezolid dosage 600 mg BID. (E) Initial linezolid dosage 1200 mg QD. Reduced dosage: red, 600 mg QD; yellow, 300 mg QD; blue, linezolid discontinued. The point at which each curve crosses the dashed black line is the time from dosage reduction to reversal of neuropathy in 50% of participants.

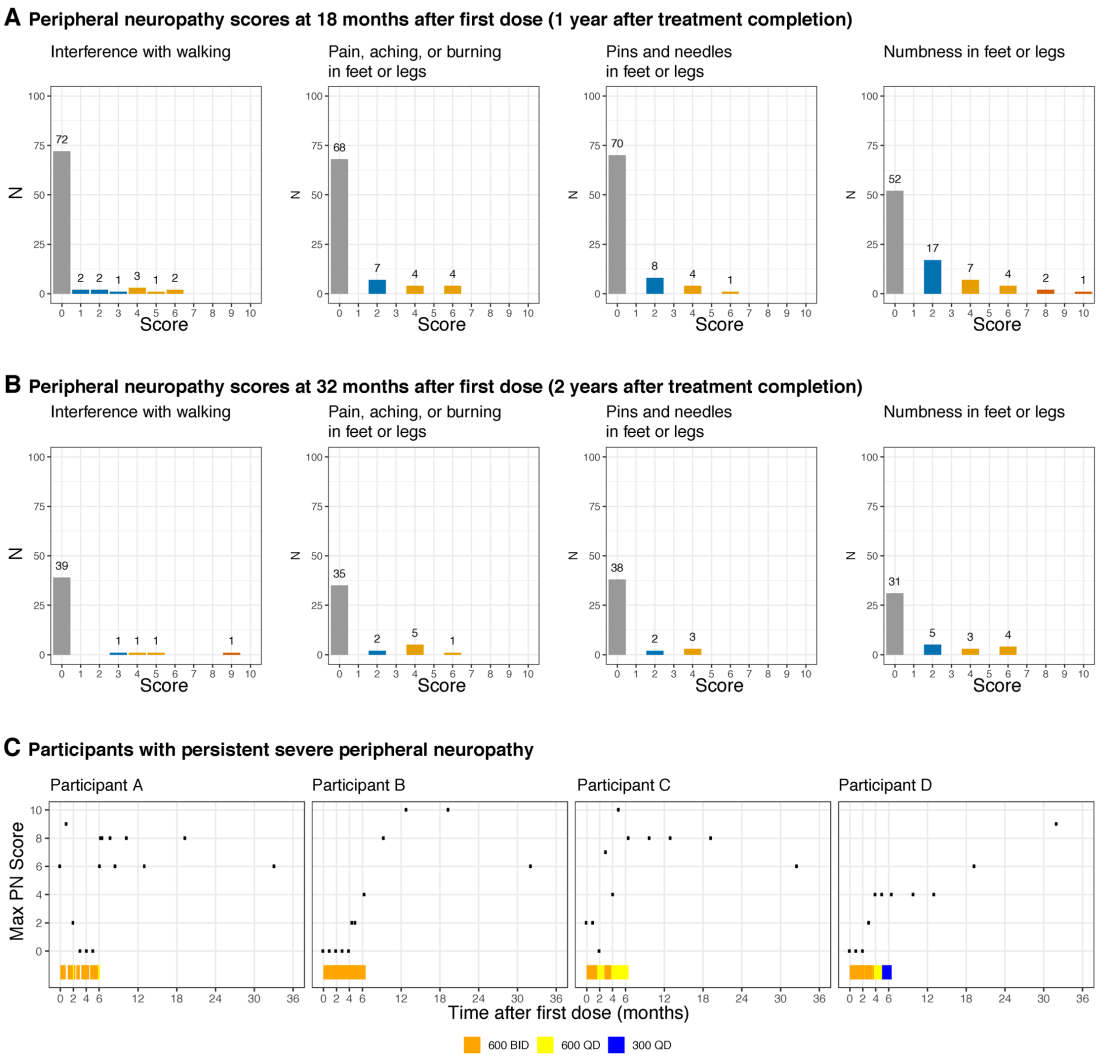



**Figure 5.7** Distribution of subject-elicited peripheral neuropathy symptom scores at 18 and 32 months associated with linezolid treatment (BPaL regimen) for extensively drug resistant tuberculosis. (A) Scores at 18 months after start of treatment (1 year after treatment completion). (B) Scores at 32 months after start of treatment (2 years after treatment completion). (C) Profiles of peripheral neuropathy scores over time for 4 participants with persistent severe peripheral neuropathy at 18 and 32 months after start of treatment. Colored bar at the bottom of each plot represent linezolid dosage pattern over 6 months.

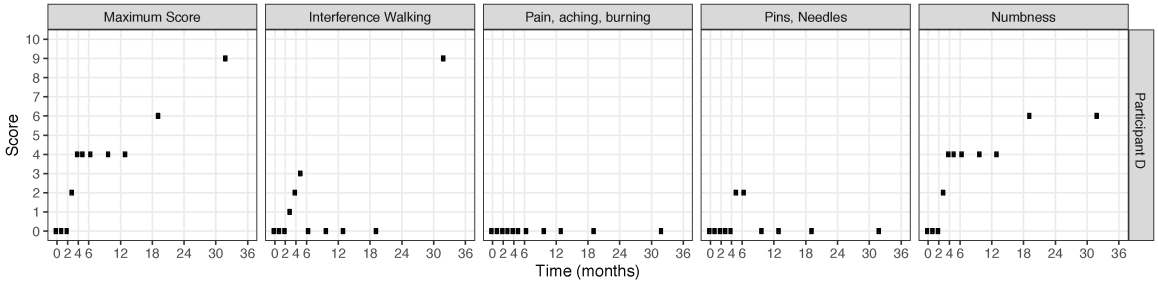



**Figure 5.8** Peripheral neuropathy symptom scores over time for one participant with persistent and increasing peripheral neuropathy. Left panel shows maximum peripheral neuropathy score based on individual scores shown in middle and right panels. Participant D in Figure 5.7.

## Discussion

In this study, we have identified dosing and patient management strategies that may permit safe administration of linezolid without substantial expected reduction in efficacy. We developed the first population PK-toxicodynamic models describing anemia, thrombocytopenia, and peripheral neuropathy associated with high-dose linezolid as part of the FDA-approved, 6-month BPaL regimen in patients with XDR- and TI/NR-MDR TB. Model simulations suggest comparable toxicity rates for QD and BID dosing, but higher rates for higher daily doses. We found that pretreatment and week 4 hemoglobin levels can potentially be used to guide early dose adjustments to prevent severe anemia, and peripheral neuropathy typically improved after linezolid dosage reductions.

The 2-compartment population PK model with nonlinear Michaelis-Menten elimination provided the optimal description of the PK data. The degree of nonlinearity in the model, with 2.7-fold greater AUC at 1200 mg QD than 600 mg QD, was similar to that observed in a noncompartmental PK analysis in TB patients, which showed 2.5-fold greater AUC at 1200 mg daily than 600 mg daily,[18,19] However, others have also shown linear PK behavior and higher elimination at increased linezolid dosages in TB patients.[20-22] Furthermore, it is unknown whether differences in protein binding when linezolid is administered with strong protein binders such as bedaquiline may alter linezolid PK properties. In other non-TB patients, decreased renal function may increase linezolid concentrations, but inclusion of a linear elimination pathway did not improve our model.[23] Overall, based on our simulations, the typical participant AUC at 600 mg and 1200 mg daily doses are consistent with previous analyses of linezolid in TB patients.[12,24]

For anemia, thrombocytopenia, and peripheral neuropathy, linezolid PK was the only predictor of toxicity potency (Table 5.2, final models described in Appendix Table A.7 - Appendix Table A.13). Linezolid troughs have previously been shown to predict linezolid adverse events.

One study with XDR-TB patients showed that mean mitochondrial function, which is associated with linezolid-induced toxicities, decreased with increasing linezolid troughs. In the previous study, a clinically-defined adverse event developed in all patients with linezolid trough level > 2 mg/L but in less than half of the patients with trough level < 2 mg/L.[7] In our analysis, linezolid trough levels at steady state predicted toxicity to platelets but not anemia or peripheral neuropathy, and use of dynamic linezolid profiles accounting for dosage adjustments better predicted all three toxicities (Table 5.2). The low predictive power of linezolid trough levels in our analysis may be explained, in part, by the higher median linezolid trough level (5.7 mg/L) observed at week 2 than the threshold of 2 mg/L cited previously.[7]

Delays between linezolid initiation and onset of toxicity were identified both for anemia, thrombocytopenia and peripheral neuropathy. Hemoglobin was described with an indirect response model that assumes hemoglobin levels are at steady state before treatment onset, decrease during treatment, and return to steady state when linezolid is discontinued. However, the median hemoglobin levels increased from 12.1 g/dL at pretreatment to 13.5 g/dL at end of treatment (Table 5.1); consistent with previous findings that preexisting anemia may resolve with anti-TB treatment, representing a paradoxical contrast between the effects of linezolid on hemoglobin levels from toxicity vs TB improvement.[25,26] A logistic growth function was included as a synthesis component in hemoglobin response in the model to account for increasing hemoglobin levels after starting BPaL treatment (Figure 5.1).

The present models and simulations may be used in the development of treatment protocols and best clinical practices that are needed for linezolid-related toxicities, which are observed at all currently-tested linezolid dosages. Although hematologic and nonhematologic toxicities may be reversible and manageable,[5] predictions for optimal treatment methods may help prevent the development of adverse events, especially at linezolid dosages of 1200 mg daily that may be

required to treat TB. Investigator-reported events and simulations showed that anemia was observed after 8 to 9 weeks of daily linezolid at high doses. Therefore, treatment changes for hemoglobin toxicity should begin before 2 months after starting linezolid therapy. Although linezolid concentration-time profiles affected toxicity, use of linezolid trough levels at 2 weeks had low AUROC for predicting severe anemia (0.56 (90% CI: 0.52-0.60), Figure 5.4) and changes in hemoglobin level at 4 weeks vs pretreatment better predicted the development of severe anemia (0.91 (90% CI: 0.88-0.93), Figure 5.4). This suggest that the effect of linezolid concentrations on hemoglobin levels may be modulated by high variability in individual sensitivity to linezolid potency (coefficient of variation for half-maximal inhibitory concentration, $IC_{50}$, 47%; Appendix Table A.10). Therefore, close monitoring of hemoglobin levels is needed for early identification of linezolid-induced anemia, with weekly monitoring. Changes in hemoglobin level after 1, 2, and 4 weeks of linezolid therapy was evaluated for practical application but we found that changes at 4 weeks optimized the ROC curve for predicting anemia (Figure 5.4Figure 5.4B). The threshold of >10% decrease in hemoglobin level at 4 weeks after starting linezolid may optimize the sensitivity (0.82) and specificity (0.84) of the hemoglobin level in predicting anemia and may prevent 63% of occurrences of severe anemia, with a false-positive rate of only 13% of participants who would undergo unnecessary dosage adjustments.

Severe thrombocytopenia rarely occurred in the Nix-TB trial (1%) so management strategies were not evaluated in the present study. However, weekly monitoring of platelet counts is still recommended because platelets counts did decrease during treatment (median 354 x $10^9$/L at pretreatment vs 262 x $10^9$/L at end of treatment, Table 5.1), although, were still above normal levels.

The most frequent investigator-reported adverse event in the Nix-TB trial was peripheral neuropathy but was frequently reversible with linezolid dosage reductions and discontinuation at

the discretion of the Nix-TB investigators. Although simulations showed that more aggressive dosage reductions were associated with faster reversal of neuropathy (Figure 5.6), most participants had neuropathy reversal within 2 to 6 months after dosage reduction or discontinuation. Therefore, close monitoring of symptoms included in the neuropathy score, with at least monthly documentation, is necessary for early detection of peripheral neuropathy. Dosage adjustments (reduction or discontinuation) may depend on a combination of peripheral neuropathy severity and treatment efficacy. Analysis of optic neuropathy was not included in the present study but is justified because optic neuritis was reported in 2 participants.

Although favorable efficacy of linezolid has been observed at all daily doses tested, it may be difficult to define optimal treatment interventions for linezolid because adverse events may occur regardless of daily dose and treatment duration.[6,24] The present simulations showed that administration of linezolid at lower daily doses reduced the occurrence of adverse events but may compromise treatment efficacy or risk the development of acquired resistance because 19% to 26% of participants did not reach the fAUC/MIC > 100 efficacy target at 600 mg daily dosages (Table 5.3). Further study in the Phase 3 ZeNix trial (NCT03086486), which is a successor trial to Nix-TB that will evaluate 4 parallel treatment groups with varied linezolid daily doses (600 and 1200 mg) and duration (9 and 26 wk), may further clarify the optimal daily dose and duration of linezolid in BPaL combination therapy for XDR-TB, pre-XDR-TB, or TI/NR MDR-TB.[27]

Strengths of the present study include the composition of participants from South Africa, which has among the highest national TB burdens globally,[28] and a high percentage of participants with HIV coinfection. The data used from the Phase 3 Nix-TB trial was rich; including up to 3 PK trough levels per participant, intensive 24-hour PK sampling from 25 participants, a median of 19 hemoglobin levels and platelet counts per participant for 6 months on treatment, and a median of 10 peripheral neuropathy scores per participant for up to 2 years after completion of treatment

(Table 5.1).[4] Therefore, the exposure-toxicity relationships that were defined in our models enable unique evidence-based recommendations about treatment to predict and prevent linezolid-related toxicity.

Limitations of the present study include the evaluation of linezolid as a component of BPaL combination therapy in XDR-TB and TI/NR MDR, which may limit generalizability to other combination therapies or TB populations. As the analyses were limited to data from the Nix-TB trial, we did not consider management for anemia, thrombocytopenia and peripheral neuropathy other than linezolid dosage reduction. In addition, we did not model the effects of dose adjustments on treatment efficacy or benefit-to-risk ratio, and the predicted benefits must be interpreted with caution because of potential loss of efficacy that may exceed the benefit of linezolid dosing reduction or discontinuation. Although this limitation may be mitigated, in part, by the results of our simulations that showed persistent achievement of PK-based efficacy targets (Table 5.3), evaluation of the recommended treatment adjustments is justified in future trials.

In conclusion, the present study provides recommendations for linezolid dosage adjustments based on data-driven models and simulations that may help clinicians predict and minimize toxicity from high-dose long-term linezolid treatment for XDR-TB and TI/NR MDR TB. Model simulations showed that hemoglobin toxicity may be prevented by responding to decreased hemoglobin levels in the early weeks after linezolid initiation. Linezolid-induced peripheral neuropathy frequently was reversible, but no dose adjustments were identified to prevent neuropathy. The quantitative strategies for linezolid toxicities may be applicable toward evaluating toxicity profiles of other antimycobacterial treatment combinations.

## References


1. WHO | Global tuberculosis report 2018. WHO: World Health Organization; 2019.
2. Ahmad N, Ahuja SD, Akkerman OW, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *The Lancet* 2018; **392**: 821-34.
3. World Health Organization. Rapid Communication: Key changes to treatment of multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB). 2018.
4. Conradie F, Diacon AH, Ngubane N, et al. Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *New England Journal of Medicine* 2020; **382**: 893-902.
5. Pretomanid Tablets [package insert]. New York, NY: TB Alliance; 2019.
6. Zhang X, Falagas ME, Vardakas KZ, et al. Systematic review and meta-analysis of the efficacy and safety of therapy with linezolid containing regimens in the treatment of multidrug-resistant and extensively drug-resistant tuberculosis. *Journal of thoracic disease* 2015; **7**: 603-15.
7. Song T, Lee M, Jeon HS, et al. Linezolid Trough Concentrations Correlate with Mitochondrial Toxicity-Related Adverse Events in the Treatment of Chronic Extensively Drug-Resistant Tuberculosis. *EBioMedicine* 2015; **2**: 1627-33.
8. DIVISION OF MICROBIOLOGY AND INFECTIOUS DISEASES (DMID) ADULT TOXICITY TABLE NOVEMBER 2007 DRAFT.
9. Merle CS, Fielding K, Sow OB, et al. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med* 2014; **371**: 1588-98.
10. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med* 2014; **371**: 1577-87.

11. Jindani A, Harrison TS, Nunn AJ, et al. High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N Engl J Med* 2014; **371**: 1599-608.

12. Bolhuis MS, Akkerman OW, Sturkenboom MGG, et al. Linezolid-based Regimens for Multidrug-resistant Tuberculosis (TB): A Systematic Review to Establish or Revise the Current Recommended Dose for TB Treatment. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018; **67**: S327-S35.

13. Diacon AH, De Jager VR, Dawson R, et al. Fourteen-Day Bactericidal Activity, Safety, and Pharmacokinetics of Linezolid in Adults with Drug-Sensitive Pulmonary Tuberculosis. *Antimicrob Agents Chemother* 2020; **64**(4).

14. Tato M, de la Pedrosa EGG, Cantón R, et al. In vitro activity of linezolid against Mycobacterium tuberculosis complex, including multidrug-resistant Mycobacterium bovis isolates. *International Journal of Antimicrobial Agents* 2006; **28**: 75-8.

15. ZYVOXID™ (linezolid), South Africa: Pfizer Pharmaceuticals, 2012.

16. Zong Z, Jing W, Shi J, et al. Comparison of In Vitro Activity and MIC Distributions between the Novel Oxazolidinone Delpazolid and Linezolid against Multidrug-Resistant and Extensively Drug-Resistant Mycobacterium tuberculosis in China. *Antimicrob Agents Chemother* 2018; **62**(8): e00165-18.

17. Yang C LH, Wang D, et al. In vitro activity of linezolid against clinical isolates of Mycobacterium tuberculosis, including multidrug-resistant and extensively drug-resistant strains from Beijing, China. . *Jpn J Infect Dis* 2018; **65**(3): 240-2.

18. Dietze R, Hadad DJ, McGee B, et al. Early and extended early bactericidal activity of linezolid in pulmonary tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 2008; **178**: 1180-5.

19. Alffenaar JWC, Van Altena R, Harmelink IM, et al. Comparison of the pharmacokinetics of two dosage regimens of linezolid in multidrug-resistant and extensively drug-resistant tuberculosis patients. *Clinical Pharmacokinetics* 2010; **49**: 559-65.

20. McGee B, Dietze R, Hadad DJ, et al. Population Pharmacokinetics of Linezolid in Adults with Pulmonary Tuberculosis. *Antimicrobial Agents and Chemotherapy* 2009; **53**: 3981-4.

21. Lee M, Lee J, Carroll MW, et al. Linezolid for Treatment of Chronic Extensively Drug-Resistant Tuberculosis. *New England Journal of Medicine* 2012; **367**: 1508-18.

22. Wasserman S, Denti P, Brust JCM, et al. Linezolid pharmacokinetics in South African patients with drug-resistant tuberculosis and a high prevalence of HIV coinfection. *Antimicrobial Agents and Chemotherapy* 2019; **63**.

23. Matsumoto K, Shigemi A, Takeshita A, et al. Analysis of thrombocytopenic effects and population pharmacokinetics of linezolid: a dosage strategy according to the trough concentration target and renal function in adult patients. *International journal of antimicrobial agents* 2014; **44**: 242-7.

24. Millard J, Pertinez H, Bonnett L, et al. Linezolid pharmacokinetics in MDR-TB: a systematic review, meta-analysis and Monte Carlo simulation. *The Journal of antimicrobial chemotherapy* 2018; **73**: 1755-62.

25. Minchella PA, Donkor S, Owolabi O, Sutherland JS, McDermid JM. Complex Anemia in Tuberculosis: The Need to Consider Causes and Timing When Designing Interventions. *Clinical Infectious Diseases* 2015; **60**: 764-72.

26. Lee SW, Kang YA, Yoon YS, et al. The prevalence and evolution of anemia associated with tuberculosis. *Journal of Korean medical science* 2006; **21**: 1028-32.

27. Development GAfTD. Safety and efficacy of various doses and treatment durations of linezolid plus bedaquiline and pretomanid in participants with pulmonary TB, XDR-TB,

Pre-XDR-TB or non-responsive/intolerant MDR-TB (ZeNix). Web site. January 13, 2020. https://www.clinicaltrials.gov/ct2/show/NCT03086486 (accessed May 28, 2020.

28. GBD Tuberculosis Collaborators. Global, regional, and national burden of tuberculosis from 1990 to 2016: results from the Global Burden of Diseases, Injuries, and Risk Factors (GBD) 2016 Study. . *Lancet Infect Dis* 2018; **18**: 1329-49.

## Chapter 6: Conclusions

The work presented in this dissertation improved the understanding of treatment response following anti-TB treatment. Quantitative, model-based tools were developed to provide evidence-based recommendations on optimal treatment regimens and strategies that i.) maximize durable cure in all patients, ii.) maximize success of late stage regimen development, and iii.) minimize safety concerns associated with a highly potent, but toxic, high-dose linezolid-containing regimen.

In the patient-pooled analysis of recent Phase 3 treatment shortening trials described in Chapter 1 and 2, semi-parametric and parametric survival analysis was performed to identify and quantify major risk factors of treatment outcomes and predict optimal treatment regimens that maximize cure rates for stratified risk groups. An integrated suite of six pragmatic baseline and on-treatment risk factors were identified. A risk stratification algorithm based on these risk factors effectively stratified patients into low, moderate, and high risk groups and a clinical simulation tool provided predictions of optimal treatment regimens. These tools can greatly benefit TB patients who can now be assigned optimal treatment regimens to maximize cure rates based on their phenotypes. In these analysis, minimal non-adherence and missed doses were also significantly associated with high risk of poor outcomes. These findings provide evidence to support 7-days-a-week administration for standard-dose rifampin-based regimens.

To help ensure that the most promising novel regimens are brought through the TB regimen development process, we provide a clinical trial simulation tool that can be used to design optimal clinical trials. The tool integrates the most studied and broadly acceptable intermediate biomarkers assessed in Phase 2B trials, treatment duration, and patient risk factors to predict Phase 3 clinical outcomes. Model simulations suggested that potent TB regimens with culture conversion hazard ratios of 3 may still undertreat high risk groups when using a one-duration-fits-all approach. We provide optimal trial designs for innovative stratified medicine approaches and adjuvant

immunotherapeutic strategies that have potential to increase cure rates for all patients. We have developed a framework that can be extended and adapted to provide more precise predictions of cure rates and further facilitate decisions when more quantitative and sensitive biomarkers are discovered and more data from Phase 3 trials testing novel regimens become available.

A major challenge in TB regimen development process has been the poor sensitivity of current Phase 2 culture-based intermediate biomarkers to predict clinical outcomes. In Chapter 4, proteomic biomarkers, which are quantitative, easy to collect, and available throughout treatment, were evaluated with pharmacokinetic data as predictors of early treatment response by applying advanced nonlinear mixed effect modeling with machine learning approaches. The integration of proteomic biomarkers, pharmacokinetic data, and clinical risk factors improved predictions of treatment response compared to use of these markers alone. As both drug exposure and proteomic biomarkers are easily measurable and available throughout treatment, they can be potentially used to efficiently predict long-term relapse-free outcomes and inform individualized dosing or treatment strategies to maximize cure for all patients.

In Chapter 5, we provide recommendations for linezolid dosage adjustments based on data-driven models and simulations that may help predict and minimize toxicity from high-dose long-term linezolid treatment for XDR-TB and TI/NR MDR-TB. Model simulations showed that hemoglobin toxicity may be prevented by responding to decreased hemoglobin levels in the early weeks after linezolid initiation. Linezolid-induced peripheral neuropathy frequently was reversible and monthly documentation of neuropathic symptoms can prompt linezolid dosage adjustments to markedly decrease the occurrence of adverse events.

In summary, we used quantitative, model-based approaches to characterize treatment response in TB patients and provide evidence-based tools that can be used to improve TB cure.

Overall, the presented work has contributed to identifying shorter, efficacious, and better-tolerated oral regimens for TB that bring us closer to ending the global TB epidemic.

## Appendix

### Appendix Text A.1

A proportional odds model was developed to describe the likelihood of occurrence of the peripheral neuropathy scores over time in the study. The likelihood was estimated using logit transformations to constrain values between 0 and 1 as shown in the equations below:

$$LOGIT = \theta$$

$$L = \frac{e^{LOGIT}}{1 + e^{LOGIT}}$$

where $\theta$ represents the estimated value of the logit parameter for a particular score for peripheral neuropathy and $L$ is the likelihood of a patient experiencing that score. Because there are several possible ordered categorical scores (i.e., 0 = normal, 1 = minimal, 2 = modest, 3 = severe.) the logits were calculated from values which were coded as shown below:

$$B_1 = \theta_1$$

$$B_2 = B_1 + \theta_2$$

where $\theta_1$ is the logit of a score greater than or equal to 1 and $\theta_2$ is the additive logit of a score greater than or equal to 2 for possible ordered categorical scores of 0, 1, or 2. $B_n$ would then be used to calculate the likelihoods as shown below:

$$L_n = \frac{e^{B_n}}{1 + e^{B_n}}$$

where $L_n$ is the likelihood of having greater than or equal to a score $n$.

The probability of a particular score was then calculated as follows:

$$Prob_0 = 1 - L_1$$

$$Prob_1 = L_1 - L_2$$

$$Prob_2 = L_2$$

where $Prob_0$, $Prob_1$, and $Prob_2$ are the probabilities of score 0, 1, and 2, respectively.

Because repeated scores are available for each individual over time an inter-individual error term was added to the logits as follows:

$$LOGIT = \theta + \eta$$

where $\eta = N(0, \omega^2)$ and $\omega^2$ is estimated.

The linezolid drug exposure effect, EFF, was added to the logit terms as follows:

$$LOGIT = \theta + EFF + \eta$$

where EFF function can be described using a linear ($EFF_{linear}$), Emax ($EFF_{EMAX}$), or sigmoidal Emax ($EFF_{SIG-EMAX}$) relationship.

An additional effect compartment was included in the model to account for delays between initiation of linezolid treatment and increased probability of peripheral neuropathy. The transfer of plasma concentration into the effect compartment was described by the following differential:

$$\frac{dC_E}{dt} = k_{e0}(C_p - C_E)$$

where $k_{e0}$ is the first order effect compartment rate and $C_p$ is linezolid concentration in central compartment at time t. The apparent drug concentration in the effect compartment ($C_E$) was then used to describe the effective linezolid concentration in the drug effect functions [e.g., $EFF = \frac{EMAX{C_E}^{HILL}}{{EC_E50}^{HILL}+{C_E}^{HILL}}$].

Participant characteristics (age, sex, weight, BMI, and HIV status) were tested as additional covariates on the logits in an additive manner as shown below:

$$LOGIT = \theta + EFF + COV(x) + \eta$$

where $COV(x)$ is the covariate relationship of interest. In the untransformed domain, this translates to a proportional effect on probability.

**Appendix Table A.1** Definitions of favorable and unfavorable outcomes per clinical trials used in Chapter 1. Outcomes defined as favorable are shown in green and outcomes defined as unfavorable are shown in red. Refer to trial protocols for exact criteria/definition for each outcome category. N refers to number of study participants in MITT analysis population** with corresponding outcome.

| Trial | Outcome | N |
|---|---|---|
| OFLOTUB | Favorable | 1090* |
| | By end of treatment, died | 14 |
| | By end of treatment, adverse event | 2 |
| | By end of treatment, drop-out | 52 |
| | By end of treatment, consent withdrawn | 16 |
| | By end of treatment, treatment failure[‖] | 28 |
| | After end of treatment, TB-recurrence, two positive culture[¶,‖] | 119 |
| | After end of treatment, TB-recurrence, one positive culture[¶,‖] | 21 |
| | After end of treatment, culture negative/unk[‖,‖] | 8 |
| REMoxTB | Culture-negative status at 18 mo. | 1166 |
| | Unable to produce sputum | 2 |
| | Unable to produce sputum at 18 mo. But culture negative status earlier | 115 |
| | Missing data on LJ culture at 18 mo. And MGIT negative | 40 |
| | 6-mo. treatment phase[†], nonviolent death | 18 |
| | 6 mo. treatment phase, adverse reaction | 42 |
| | 6 mo. treatment phase, withdrawal of consent | 34 |
| | 6 mo. treatment phase, relocation | 10 |
| | 6 mo. treatment phase, other investigator decision | 7 |
| | 6 mo. treatment phase, no completion of treatment | 29 |
| | 6-mo. treatment phase, treatment failure, culture confirmed[‖] | 8 |
| | 6 mo. treatment phase, treatment failure, not culture confirmed[‖] | 9 |
| | Follow up, relapse after culture negative status[‖] | 123 |
| | Follow up, retreated for tuberculosis[‖] | 59 |
| | Follow up, no culture negative status ever[‖] | 3 |
| | Follow up, no culture negative status at last visit[‖] | 7 |
| | Follow up, death from tuberculosis or respiratory distress[‖] | 2 |
| RIFAQUIN | Favorable | 302 |
| | During treatment, death | 1 |
| | During treatment, change in treatment due to adverse event | 3 |
| | During treatment, lost to follow up | 11 |
| | During treatment, inadequate treatment | 3 |
| | During treatment, other treatment change | 21 |
| | During treatment, failure (culture confirmed)[‖] | 4 |
| | After treatment, relapse, culture confirmation[‖] | 23 |
| | After treatment, relapse, limited bacteriologic confirmation[‖] | 10 |
| | After treatment, culture positive when last seen[‖] | 2 |
| | After treatment, death due to tuberculosis[‖] | 1 |
| DMID 01-009** | Microbiological cure | 326 |
| | Clinical cure | 28 |

**Appendix Table A.1 continued**

| Trial | Outcome | N |
|---|---|---|
| DMID 01-009** | Bacteriological relapse[ǀ] | 18 |
| | Death | 4 |
| | Lost to follow up | 9 |
| | Other | 1 |

*Six study participants were excluded due inability to verify treatment allocation in source database
[†] In the REMoxTB trial, treatment phase was defined as 32 weeks after randomization.
[ʃ]smear positive/symptoms
[¶]Sixteen TB-recurrences included in the primary analysis for the OFLOTUB trial were MIRU confirmed reinfections
[ǀ]Included in TB related outcome definition.
**For DMID 01-009 trial, data corresponds to time to event analysis population from original publication.

**Appendix Table A.2** Dataset specification and harmonization for Chapters 1 and 2

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Study country | Categorical:<br>1-Benin,<br>2-Botswana<br>3-China,<br>4-Guinea,<br>5-India,<br>6-Kenya,<br>7-Mexico,<br>8-Malaysia,<br>9-Senegal,<br>10- Thailand,<br>11-Tanzania,<br>12-South Africa,<br>13-Zambia,<br>14-Zimbabwe,<br>101- Uganda,<br>102-Philippines,<br>103-Brazil | Yes | Yes | Yes | Yes | 0 (0) | Yes | |
| Age | Number<br>(units, years) | Yes | Yes | Yes | Yes | 5 (<1) | Yes | |
| Sex | Boolean:<br>0-Female,<br>1- Male | Yes | Yes | Yes | Yes | 0 (0) | Yes | |
| Race | Categorical:<br>1-Black,<br>2-Asian,<br>3-Other | No | Yes | Yes | No | 1736 (46) | Yes | Black race was assigned to all study participants in OFLOTUB trial, given all OFLOTUB sites were in Africa. Race was not a significant risk factor in primary analysis, therefore, was not imputed for validation. |
| Weight | Number<br>(units, kg) | Yes | Yes | Yes | Yes | 0 (0) | No | Not included in primary analysis because of correlation with BMI |

**Appendix A.2 continued.**

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Body mass index (BMI) | Number (units, $kg/m^2$) | Calculated- see notes | Calculated- see notes | Calculated- see notes | Calculated- see notes | 291(8) with missing height-see notes | Yes | Body mass index was defined as the weight in kilograms divided by the squared height in meters. Height was not available for 291 study participants in the RIFAQUIN trial, median height from available data was assigned according to sex. |
| HIV status | Boolean: 0-negative, 1-positive | Yes | Yes | Yes | Yes | 9 (<1) | Yes | OFLOTUB trial excluded participants who required concomitant anti-infective treatment or HIV infected participants with WHO stage 3 infection (except those presenting with only the "loss of weight>10% body weight" criterion) and all participants at WHO stage 4, REMoxTB trial excluded HIV infected participants already receiving anti-retroviral therapy or CD4 count less than 250 cells/µL, RIFAQUIN trial excluded HIV infected participants who required anti-retroviral therapy at diagnosis (amended to "already receiving anti-retroviral therapy" during course of trial) or CD4 cell count less than 200 cells/µL (amended to 150 cells/µL during course of trial), and DMID 01-009 trial excluded all HIV infected participants. |

**Appendix A.2 continued.**

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Baseline CD4 cell count | Number (units, cells/μL) | Yes | Yes | Yes | NA | 79 (17 †) | No | See notes for HIV status |
| Smoking | Categorical: 0-No, 1-Before, 2-Ongoing | No | Yes | Yes | No | 1736 (46) | No | |
| Alcohol Use | Boolean: 0-No, 1-Yes | No | No | No | Yes | 3405 (90) | No | |
| Drug Use | Boolean: 0-No, 1-Yes | No | No | No | Yes | 3405 (90) | No | |
| Pharmacogenomics (Isoniazid acetylator) | NA | No | No | No | No | 3791 (100) | No | |
| Baseline cough | Boolean: 0-No, 1-Yes | Yes | Yes | No | Yes | 381 (10) | No | Cough grade data was available in REMoxTB trial. Cough data during treatment and follow-up available in DMID 01-009 trial. |
| Baseline smear grade †† | Categorical: 1-Smear negative or 1+, 2-Smear 2+, 3-Smear 3+ | Yes | Yes | Yes | Yes | 43 (1) | Yes | A conversion chart available in the REMoxTB trial lab manual was used to readjust all smear data to the same grading scale (Additional details in Supplementary Methods). Longitudinal data available in all trials. |
| Baseline cavitation †† | Boolean: 0-Absent, 1-Present | Yes | Yes | Yes | Yes ∫ | 200 (5) | Yes | Follow-up data available in OFLOTUB trial. |

**Appendix A.2 continued.**

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Baseline cavity size †† | Categorical:<br>0- 1-5 cm,<br>1- >5 cm | No | No | Yes | Yes ∫ | 1861 (89 ¶) | No | |
| Baseline lung disease grade †† | Categorical:<br>0-Normal,<br>1-Minimal,<br>2-Moderately Advanced,<br>3-Far Advanced | Yes | No | No | Yes | 2057 (54) | No | Follow-up data available in OFLOTUB and DMID 01-009 trials. |
| Baseline lung disease extent †† | Categorical:<br>0-None Visible,<br>1-Unilateral<br>2-Bilateral | Yes | No | No | No | 2458 (65) | No | Follow-up data available in OFLOTUB and DMID 01-009 trials. |
| Baseline lung zone score †† | Categorical:<br>Factors 0 to 6;<br>0-no area affected,<br>6-all lung regions affected | Yes | No | No | No | 2457 (65) | No | Follow up data available in OFLOTUB trial. |
| On-treatment culture status based on LJ medium †† | Boolean:<br>0-Negative<br>1-Positive | Yes | Yes | Yes ‖ | Yes | Month 2: 524 (14)<br>Month 4: 348 (19**) | No | Longitudinal data available in all trials. |
| On-treatment culture status based on MGIT medium †† | Boolean:<br>0-Negative<br>1-Positive | No | Yes | Yes ‖ | No | Month 2: 2094 (55)<br>Month 4: 1187 (66**) | No | Longitudinal data available in all trials |
| On-treatment culture status based on LJ or MGIT medium (data used in primary analysis) †† | Boolean:<br>0-Negative<br>1-Positive | Yes | Yes | Yes | Yes | Month 2: 308 (8)<br>Month 4: 236 (13**) | Month 2- Yes<br>Month 4- No | Longitudinal data available in all trials |

**Appendix A.2 continued.**

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Drug regimen | Categorical: See description of study arms in original protocols, methods, and supplementary methods | Yes | Yes | Yes | Yes | 6 (<1) | No | Six study participants were excluded from the current analyses due to untraceable regimen assignment in the source database. |
| Drug Dose | Number: See description in original protocols | Yes | Yes | Yes | Yes | 6 (<1) | No | Six study participants were excluded from the current analyses due inability to verify treatment allocation in source database. |
| Treatment Adherence | Number (units, %) | Calculated- see notes | Calculated- see notes | Calculated- see notes | No | 83 (2 ∫∫) | Yes | Treatment adherence was calculated as the number of days that doses were taken divided by the prescribed number of days. |
| Pharmacokinetics | Number: See description in original protocols | Sub-study | No | Sub-study | No | 3207 (86) | No | |

**Appendix A.2 continued.**

| Variable | Data Type and Format | Collected? | | | | No. participants with missing data (%*) | Included in primary analysis? | Notes |
|---|---|---|---|---|---|---|---|---|
| | | OFLOTUB | REMoxTB | RIFAQUIN | DMID01-009 | | | |
| Drug Resistance | Categorical: See description in original protocols | Yes | Yes | Yes | Yes | Variable-based on exclusion/inclusion criteria-see notes | No | According to original trial protocols, OFLOTUB trial withdrew participants resistant to rifampin, REMoxTB trial excluded participants resistant to rifampin or any fluoroquinolone, RIFAQUIN trial excluded individuals resistant to rifampin, isoniazid, or moxifloxacin, and DMID 01-009 trial, excluded individuals resistant to rifampin, isoniazid, ethambutol or pyrazinamide. |

* Percentage based on entire population, which included 3405 participants from REMoxTB, RIFAQUIN, OFLOTUB and 386 from DMID 01-009 studies.
† Percentage based on HIV infected population.
∫ DMID 01-009 study population were non-cavitary disease adults.
¶ Percentage based on study participants with cavitary disease.
‖ Study site specific.
** Percentage based on 6-month control population only.
†† Worst outcomes used for analysis if multiple samples available.
∫∫ Missing population and percentage based on REMoxTB, RIFAQUIN, OFLOTUB datasets only.
NA- not applicable

**Appendix Table A.3** Univariate Cox proportional hazard analysis for merged MGIT and LJ culture data (as used in primary analysis)*, MGIT data only, and LJ data only in the MITT analysis population. Univariate results for experimental group study participants

| Variable | N unfavorable/ N assessable (%) | Hazard Ratio (95% CI) | P value |
|---|---|---|---|
| Month 2 culture negative* | 248/1478 (17) | Reference | Reference |
| Month 2 culture positive* | 115/336 (34) | 2.3 (1.9 - 2.9) | <0.001 |
| Month 2 culture negative-MGIT data only | 77/583 (13) | Reference | Reference |
| Month 2 culture positive-MGIT data only | 141/508 (28) | 2.3 (1.7 - 3.0) | <0.001 |
| Month 2 culture negative- LJ data only | 233/1386 (17) | Reference | Reference |
| Month 2 culture positive-LJ data only | 106/319 (33) | 2.2 (1.8 - 2.8) | <0.001 |

| Variable | N unfavorable/ N assessable (%) | Hazard Ratio (95% CI) | P value |
|---|---|---|---|
| Month 2 culture negative* | 113/998 (11) | Reference | Reference |
| Month 2 culture positive* | 49/285 (17) | 1.6 (1.1 - 2.2) | 0.007 |
| Month 2 culture negative-MGIT data only | 24/310 (8) | Reference | Reference |
| Month 2 culture positive-MGIT data only | 42/296 (14) | 1.9 (1.1 - 3.1) | 0.02 |
| Month 2 culture negative- LJ data only | 108/915 (12) | Reference | Reference |
| Month 2 culture positive-LJ data only | 43/261 (16) | 1.5 (1.0 - 2.1) | 0.04 |

* Culture positivity on either media was used for analyses with preference for solid culture if available.

**Appendix Table A.4** Adjusted hazard ratios for time to stable culture conversion. P value for interaction between response groups and drug/drug exposure not significant at a P = 0.01 level.

| | Train Dataset | Test Dataset | Train + Test Dataset | Independent Dataset |
|---|---|---|---|---|
| **Proteomic signature adjusted for drug (rifampin vs rifapentine)** | | | | |
| **N** | 379[a] | 158[a] | 538[a] | 30 |
| **Stable Culture Conversion Hazard Ratio Median (95% CI)** | | | | |
| Moderate responders relative to slow responders | 1.5 (1.1-2.1) | 2.3 (1.3-3.9) | 1.6 (1.2-2.2) | 1.7 (0.5-5.1) |
| Fast responders relative to slow responders | 2.3 (1.7-3.0) | 2.5 (1.6-3.6) | 2.3 (1.9-2.9) | 2.3 (0.8-6.5) |
| **Proteomic signature adjusted for drug exposure (rifapentine AUC) and clinical risk factors** | | | | |
| **N** | 236[a] | 99[a] | 335[a] | - |
| **Stable Culture Conversion Hazard Ratio Median (95% CI)** | | | | |
| Moderate responders relative to slow responders | 1.2 (0.6-2.5) | 1.1 (0.4-2.7) | 1.1 (0.7-1.9) | - |
| Fast responders relative to slow responders | 2.4 (1.2-4.5) | 1.7 (0.7-4.0) | 2.0 (1.2-3.3) | - |
| [a] Two participants in training dataset and one participants in testing dataset missing time to stable culture conversion | | | | |

**Appendix Table A.5** AUROC values to predict time to stable culture conversion. AUROC (95% confidence interval)

| | | **AUROC at 8 weeks** | | | **AUROC at 12 weeks** | | |
|---|---|---|---|---|---|---|---|
| **Adjusted for:** | **Proteomic Included?** | **Train Dataset** | **Test Dataset** | **Train + Test Dataset** | **Train Dataset** | **Test Dataset** | **Train + Test Dataset** |
| Drug (Rifampin vs Rifapentine) | No | 0.55 (0.51-0.62) | 0.55 (0.45-0.62) | 0.55 (0.51-0.60) | 0.51 (0.46-0.62) | 0.57 (0.42-0.68) | 0.52 (0.45-0.59) |
| Drug (Rifampin vs Rifapentine) | Yes | 0.72 (0.67-0.78) | 0.70 (0.60-0.78) | 0.71 (0.67-0.76) | 0.67 (0.58-0.75) | 0.63 (0.45-0.80) | 0.66 (0.58-0.74) |
| Rifapentine AUC and clinical risk factors | No | 0.70 (0.62-0.77) | 0.74 (0.63-0.84) | 0.69 (0.62-0.75) | 0.72 (0.63-0.82) | 0.87 (0.79-0.96) | 0.77 (0.69-0.84) |
| None | Yes | 0.66 (0.59-0.73) | 0.60 (0.49-0.71) | 0.66 (0.59-0.73) | 0.70 (0.59-0.81) | 0.55 (0.38-0.72) | 0.75 (0.67-0.82) |
| Rifapentine AUC and clinical risk factors | Yes | 0.75 (0.68-0.82) | 0.75 (0.64-0.85) | 0.73 (0.67-0.79) | 0.80 (0.70-0.88) | 0.84 (0.76-0.94) | 0.79 (0.72-0.86) |

**Appendix Table A.6** Relative variable importance for top 5 proteomic predictors in CART analysis. Variable importance is calculated as the sum of the decrease in root mean squared error when variable is used to split a node. Relative importance is scaled to the most important variable and transformed into a percentage.

| **Proteomic signature after adjusting for rifamycin drug only** | |
|---|---|
| **Variable** | **Relative Importance (%)** |
| Week 8 RANTES | 100 |
| Week 8 SAA1 | 52 |
| Week 8 CRP | 51 |
| Baseline RANTES | 45 |
| W8 LBP | 43 |
| **Proteomic signature after adjusting for rifapentine exposure and clinical risk factors** | |
| **Variable** | **Relative Importance (%)** |
| Week 8 SAA1 | 100 |
| Ratio week 8 to baseline SAA1 | 93 |
| Week 8 CRP | 81 |
| Week 8 MMP8 | 78 |
| Week 8 MMP9 | 77 |

**Appendix Table A.7** Differential equations that describe the final linezolid PK model

| | |
|---|---|
| Absorption compartment (A) | $\frac{dA}{dt} = -k_a A$ |
| Central compartment ($A_C$) | $\frac{dA_C}{dt} = k_a A - \left. \frac{\frac{VMAX A_C}{\frac{A_C}{V_c} + K_m}}{V_c} \right/ + \frac{Q}{V_p} A_P - \frac{Q}{V_c} A_C$ <br> $VMAX = CL_{int} K_m$ |
| Peripheral compartment ($A_P$) | $\frac{dA_P}{dt} = -\frac{Q}{V_p} A_P + \frac{Q}{V_c} A_C$ |
| A, $A_C$, $A_P$, amounts in absorption, central, and peripheral compartments; $V_c$, apparent central volume of distribution; $V_p$, apparent peripheral volume of distribution; Q, apparent inter-compartment clearance; $k_a$, absorption rate constant; $K_m$, Michaelis-Menten constant; and $CL_{int}$, Michaelis-Menten intrinsic clearance; $V_{MAX} = CL_{int}K_m$, maximum rate of elimination (saturated elimination). Log-normal inter-individual variability was tested on $k_a$, $K_m$, $CL_{int}$, and $V_c$ The following equation was used: $P_i = TVP exp(\eta_{P_i})$ where $P_i$ is the estimated parameter for individual, $i$; $TVP$ is the typical value (median) for the parameter, and $\eta_{P_i}$ is the inter-individual random effect reflecting the difference between an individual's parameter value and the population's typical value where $\eta_{P_i} = N(0, \omega^2)$ and $\omega^2$ is the variance of $\eta_{P_i}$, which is estimated in model fitting. The coefficient of variation ($CV\% = 100 sqrt(\exp(\omega^2) - 1)$) is reported. A proportional model was used to explain residual variability. The following equation was used: $Y_{i,j} = F_{i,j}(1 + \varepsilon_{i,j})$ where $Y_{i,j}$ $and$ $F_{i,j}$ are the observed and model predicted dependent variable values, respectively, for individual, $i$, and sampling timepoint, $j$; and $\varepsilon_{i,j}$ is the unexplained residual error reflecting the difference between observed and model predicted data where $\varepsilon_{i,j} = N(0, \sigma^2)$ and $\sigma$ is the standard deviation of $\varepsilon_{i,j}$, which is estimated in the model fitting. | |

**Appendix Table A.8** Parameter estimates for final linezolid PK model. CV% is the percent coefficient of variation of the parameter's inter-individual variability, and RSE% is the percent relative standard error of the parameter estimate. See Appendix Table A.7 and Chapter 5: for more details on structural and stochastic models.

| **Parameter (units)** | **Estimate (RSE %)** | **CV % (RSE %)** |
|---|---|---|
| Intrinsic clearance, $CL_{int}$ (L/h) | 8.5 (3) | 5.5 (80) |
| Michaelis-Menten Constant, $K_m$ (mg/L) | 15 (2) | |
| Volume of distribution, $V_C$ (L) | 59 (2) | 66 (27) |
| Rate of absorption, $k_a$, (1/h) | 1.6 (1) | 146 (20) |
| Inter-compartment clearance, Q, (L/h) | 1.3 (1) | |
| Peripheral Volume, $V_p$, (L) | 31.3 (4) | |
| Standard deviation of proportional residual error, $\sigma$ | 0.29 (18) | |
| Inter-occasion variability for $CL_{int}$ (CV%) | 28 (10) | |

**Appendix Table A.9** Differential equations that describe the final linezolid dosing hemoglobin model

| | |
|---|---|
| Hemoglobin compartment (Hb) | $\frac{dHb}{dt} = k_{in}(1 - EFF_{Hb}(C_P)) - k_{out}Hb + TRT_{rate}Hb(\frac{TRT_{cap} - Hb}{TRT_{cap}})$<br><br>$EFF_{Hb}(C_P) = \frac{EMAXC_p{}^{HILL}}{IC50^{HILL} + C_p{}^{HILL}}$ |
| Hb, hemoglobin level; $k_{out}$, first order hemoglobin elimination rate constant; $k_{in}$, zero-order hemoglobin synthesis rate constant, which equals the product of pretreatment hemoglobin, $BASE_{Hb}$ and $k_{out}$; $TRT_{rate}$, rate of increase in hemoglobin, and $TRT_{cap}$, maximum hemoglobin level, for logistic growth function describing improved hemoglobin levels upon initiating anti-TB treatment; $EFF_{Hb}(C_p)$, linezolid concentration-response relationship; $C_p$, concentration of linezolid at time t; EMAX, maximum inhibitory effect; $IC50$, concentration to reach half maximal inhibitory effect; $HILL$, Hill coefficient that describes the shape of the curve. Log-normal inter-individual variability was tested on $k_{out}$, $TRT_{rate}$, $TRT_{cap}$, EMAX, and IC50. The following equation was used: $P_i = TVPexp(\eta_{P_i})$ where $P_i$ is the estimated parameter for individual, $i$; $TVP$ is the typical value (median) for the parameter, and $\eta_{P_i}$ is the inter-individual random effect reflecting the difference between an individual's parameter value and the population's typical value where $\eta_{P_i} = N(0, \omega^2)$ and $\omega^2$ is the variance of $\eta_{P_i}$, which is estimated in model fitting. The coefficient of variation ($CV\% = 100sqrt(\exp(\omega^2) - 1)$) is reported. Correlations between random effects, $a$ and $b$, were predicted using the following formula: $\omega_{a,b}/(\omega_a^2 + \omega_b^2)$, where $\omega_a^2$ is the variance of parameter $a$, $\omega_b^2$ is the variance of parameter $b$, and $\omega_{a,b}$ is the covariance between parameter $a$ and $b$. An additive model was used to explain residual variability. The following equation was used: $Y_{i,j} = F_{i,j} + \varepsilon_{i,j}$ where $Y_{i,j}$ $and$ $F_{i,j}$ are the observed and model predicted dependent variable values, respectively, for individual, $i$, and sampling timepoint, $j$; and $\varepsilon_{i,j}$ is the unexplained residual error reflecting the difference between observed and model predicted data where $\varepsilon_{i,j} = N(0, \sigma^2)$ and $\sigma$ is the standard deviation of $\varepsilon_{i,j}$, which is estimated in the model fitting. | |

**Appendix Table A.10** Parameter estimates for final linezolid dosing hemoglobin model. CV% is the percent coefficient of variation of the parameter's inter-individual variability, and RSE% is the percent relative standard error of the parameter estimate. Sex was implemented into the model to explain the fractional decrease in pretreatment hemoglobin levels using the following linear relationship: $BASE_{Hb,FEMALE} = BASE_{Hb}(1 - \theta_{FEMALE})$. See Appendix Table A.9 and Chapter 5: for more details on structural and stochastic models.

| **Parameter (units)** | **Estimate (RSE %)** | **CV % (RSE %)** |
|---|---|---|
| Pretreatment hemoglobin, $BASE_{Hb}$ (g/dL) | 13 (2) | 12 (9) |
| 1/rate of elimination of hemoglobin [$1/k_{out}$ ] (h) | 3685 (9) | 97 (8) |
| Logistic growth rate for treatment progression*1000, $TRT_{rate}$*1000 (1/h) | 0.44 (18) | 52 (20) |
| Logistic growth capacity for treatment progression, $TRT_{cap}$ (g/dL) | 19 (7) | 28 (21) |
| IC50 (mg/L) | 7.7 (11) | 47 (23) |
| EMAX (%) | 100 FIX | |
| HILL Coefficient | 10 FIX | |
| Fractional decrease in baseline hemoglobin for females, $\theta_{FEMALE}$ | 0.10 (21) | |
| Standard deviation of additive residual error, $\sigma$ (g/dL) | 0.70 (6) | |
| Correlations between random effects: | | |
| $TRT_{rate}$ and $TRT_{cap}$ | -0.65 (21) | |
| $TRT_{rate}$ and $1/k_{out}$ | -0.48 (17) | |
| $1/k_{out}$ and IC50 | -0.57 (20) | |

**Appendix Table A.11** Differential equations that describe the final linezolid dosing platelet model

| | |
|---|---|
| Progenitor cell compartment (Prol) | $\frac{dProl}{dt} = k_{prol}Prol(1 - EFF_{Platelet}(C_P))(Circ_0/Circ)^{\gamma} - k_{tr}Prol$<br><br>$EFF_{Platelet}(C_P) = \frac{EMAXC_p(t)^{HILL}}{IC50^{HILL} + C_p(t)^{HILL}}$ |
| Transit compartment 1 | $\frac{dTransit_1}{dt} = k_{tr}Prol - k_{tr}Transit_1$ |
| Transit compartment 2 | $\frac{dTransit_2}{dt} = k_{tr}Transit_1 - k_{tr}Transit_2$ |
| Transit compartment 3 | $\frac{dTransit_3}{dt} = k_{tr}Transit_2 - k_{tr}Transit_3$ |
| Circulating cells (Platelets) (Circ) | $\frac{dCirc}{dt} = k_{tr}Transit_3 - k_{circ}Circ$ |

Prol, amount of proliferative/progenitor cells; Transit x, transit compartments representing cell maturation process; Circ, circulation compartment; $k_{prol}$, first-order proliferation rate constant, assumed equal to $k_{tr}$; $k_{tr}$, first-order transit rate constant; MTT, mean transit time = (No. transit compartments+1)/$k_{tr;}$ $k_{circ}$, first-order rate of elimination from circulation compartment; $Circ_o$, baseline platelet level; and $\gamma$ homeostatic feedback parameter; $EFF_{Platelet}(C_p)$, linezolid concentration-response relationship; $C_p(t)$, concentration of linezolid at time t; EMAX, maximum inhibitory effect; $IC50$, concentration to reach half maximal inhibitory effect; $HILL$, Hill coefficient that describes the shape of the curve. Log-normal inter-individual variability was tested on $k_{out}$, $TRT_{rate}$, $TRT_{cap}$, EMAX, and IC50. The following equation was used: $P_i = TVPexp(\eta_{P_i})$ where $P_i$ is the estimated parameter for individual, $i$; $TVP$ is the typical value (median) for the parameter, and $\eta_{P_i}$ is the inter-individual random effect reflecting the difference between an individual's parameter value and the population's typical value where $\eta_{P_i} = N(0, \omega^2)$ and $\omega^2$ is the variance of $\eta_{P_i}$, which is estimated in model fitting. The coefficient of variation ( $CV\% = 100sqrt(\exp(\omega^2) - 1)$) is reported. Correlations between random effects, $a$ and $b$, were predicted using the following formula: $\omega_{a,b}/(\omega_a^2 + \omega_b^2)$, where $\omega_a^2$ is the variance of parameter $a$, $\omega_b^2$ is the variance of parameter $b$, and $\omega_{a,b}$ is the covariance between parameter $a$ and $b$. A proportional model was used to explain residual variability. The following equation was used: $Y_{i,j} = F_{i,j}(1 + \varepsilon_{i,j})$ where $Y_{i,j}$ *and* $F_{i,j}$ are the observed and model predicted dependent variable values, respectively, for individual, $i$, and sampling timepoint, $j$; and $\varepsilon_{i,j}$ is the unexplained residual error reflecting the difference between observed and model predicted data where $\varepsilon_{i,j} = N(0, \sigma^2)$ and $\sigma$ is the standard deviation of $\varepsilon_{i,j}$, which is estimated in the model fitting.

**Appendix Table A.12** Parameter estimates for final linezolid dosing platelet model CV% is the percent coefficient of variation of the parameter's inter-individual variability and RSE% is the percent relative standard error of the parameter estimate. See Appendix Table A.11 and Chapter 5: for more details on structural and stochastic models.

| **Parameter (units)** | **Estimate (RSE %)** | **CV % (RSE %)** |
|---|---|---|
| Pretreatment platelet, $Circ_0$ ($10^9$/L) | 340 (6) | 31 (10) |
| Mean transit time, MTT = ($4/k_{tr}$) (h) | 144 (8) | 9 (26) |
| Homeostatic feedback parameter, $\gamma$ | 0.28 (13) | |
| IC50 (mg/L) | 4.1 (18) | 105 (11) |
| EMAX (%) | 8 (14) | 49 (13) |
| HILL coefficient | 3.1 (9) | |
| Standard deviation of proportional residual error, $\sigma$ | 0.17 (4) | |
| Correlations between random effects: | | |
| $Circ_0$ and MTT | 0.36 (50) | |
| $Circ_0$ and IC50 | -0.48 (16) | |
| MTT and IC50 | -0.84 (22) | |

**Appendix Table A.13** Parameter estimates for final linezolid dosing proportional odds peripheral neuropathy model. SD is the standard deviation of the base logit parameter's inter-individual variability and RSE% is the percent relative standard error of the parameter estimate. See Appendix Text A.1 for more details on structural and stochastic models.

| **Parameter (units)** | **Estimate (RSE %)** | **SD (RSE)** |
|---|---|---|
| Base logit value for score ≥ 1 | -1.3 (22) | 1.7 (11) |
| Additive base logit value for score ≥ 4 | -1.1 (11) | |
| Additive base logit value for score ≥ 8 | -1.9 (9) | |
| First order effect compartment rate, $k_{e0}$ (1/h) | 5.9E-5 (3) | |
| Maximum effect of effect compartment concentrations on increasing logit, EMAX | 3.4 (23) | |
| Half maximal effect compartment concentration, $EC_E50$ (mg/L) | 1.3 (33) | |
| HILL coefficient for sigmoidal effect on logit | 4.9 (33) | |

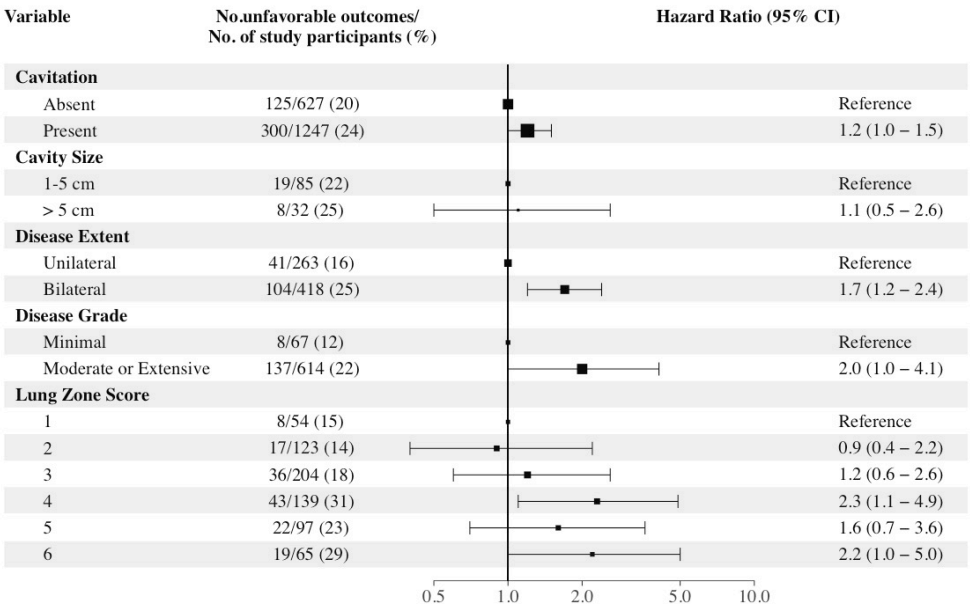
**Appendix Table A.14** Observed and model simulated proportions of maximum subject-elicited symptom scores for peripheral neuropathy over time after treatment with BPaL regimen. Simulations show median and 95% prediction interval.

| **Time (months)** | **Score** | **Observed Proportion** | **Simulated Proportion [Median (95% PI)]** |
|---|---|---|---|
| 0 | Normal, 0 | 0.76 | 0.78 (0.69-0.86) |
| 0 | Minimal, 1-3 | 0.10 | 0.11 (0.06-0.18) |
| 0 | Modest, 4-7 | 0.13 | 0.08 (0.04-0.15) |
| 0 | Severe, 8-10 | 0.01 | 0.03 (0-0.06) |
| 3 | Normal, 0 | 0.58 | 0.56 (0.48-0.66) |
| 3 | Minimal, 1-3 | 0.18 | 0.16 (0.10-0.25) |
| 3 | Modest, 4-7 | 0.18 | 0.18 (0.12-0.24) |
| 3 | Severe, 8-10 | 0.06 | 0.09 (0.04-0.15) |
| 6 | Normal, 0 | 0.31 | 0.41 (0.31-0.50) |
| 6 | Minimal, 1-3 | 0.23 | 0.17 (0.12-0.23) |
| 6 | Modest, 4-7 | 0.28 | 0.23 (0.17-0.33) |
| 6 | Severe, 8-10 | 0.18 | 0.17 (0.12-0.28) |
| 19 | Normal, 0 | 0.52 | 0.55 (0.43-0.63) |
| 19 | Minimal, 1-3 | 0.23 | 0.15 (0.10-0.25) |
| 19 | Modest, 4-7 | 0.20 | 0.19 (0.10-0.30) |
| 19 | Severe, 8-10 | 0.04 | 0.11 (0.04-0.17) |
| 32 | Normal, 0 | 0.60 | 0.67 (0.58-0.81) |
| 32 | Minimal, 1-3 | 0.12 | 0.14 (0.06-0.22) |
| 32 | Modest, 4-7 | 0.26 | 0.12 (0.07-0.22) |
| 32 | Severe, 8-10 | 0.02 | 0.07 (0.02-0.12) |

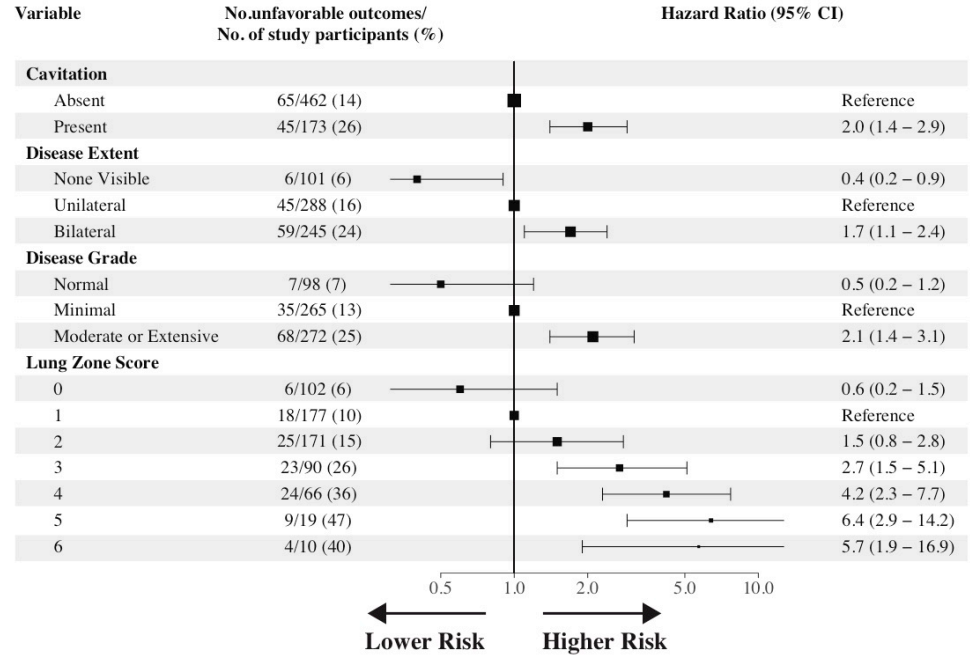
**Appendix Table A.15** Area under the receiver operating characteristic curves for univariate models for prediction of reported anemia adverse events after treatment with BPaL regimen

| **Model** | **AUROC (90% confidence interval)** |
|---|---|
| BMI | 0.50 (0.39-0.61) |
| Weight | 0.51 (0.37-0.60) |
| HIV status | 0.53 (0.43-0.63) |
| Sex | 0.54 (0.55-0.64) |
| Age | 0.58 (0.46-0.69) |
| Pretreatment Hb levels | 0.56 (0.44-0.67) |
| Week 1 Hb levels | 0.53 (0.44-0.67) |
| Percent change in Hb at Week 1 | 0.50 (0.38-0.53) |
| Week 2 Hb levels | 0.63 (0.52-0.74) |
| Percent change in Hb at Week 2 | 0.62 (0.51-0.73) |
| Week 4 levels | 0.73 (0.64-0.82) |
| Percent change in Hb at Week 4 | 0.71 (0.61-0.81) |
| Observed linezolid troughs, week 2 | 0.52 (0.41-0.63) |
| Simulated SS Cmin | 0.61 (0.51-0.72) |
| Simulated SS CMAX | 0.62 (0.51-0.72) |
| Simulated SS AUC24 | 0.60 (0.49-0.70) |

***Baseline***

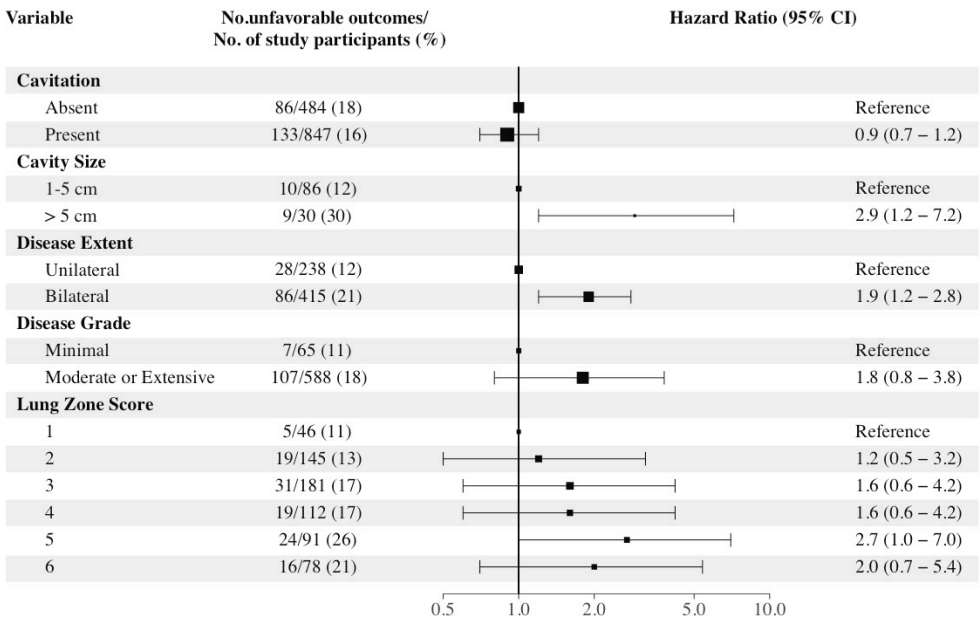



***End of Treatment***

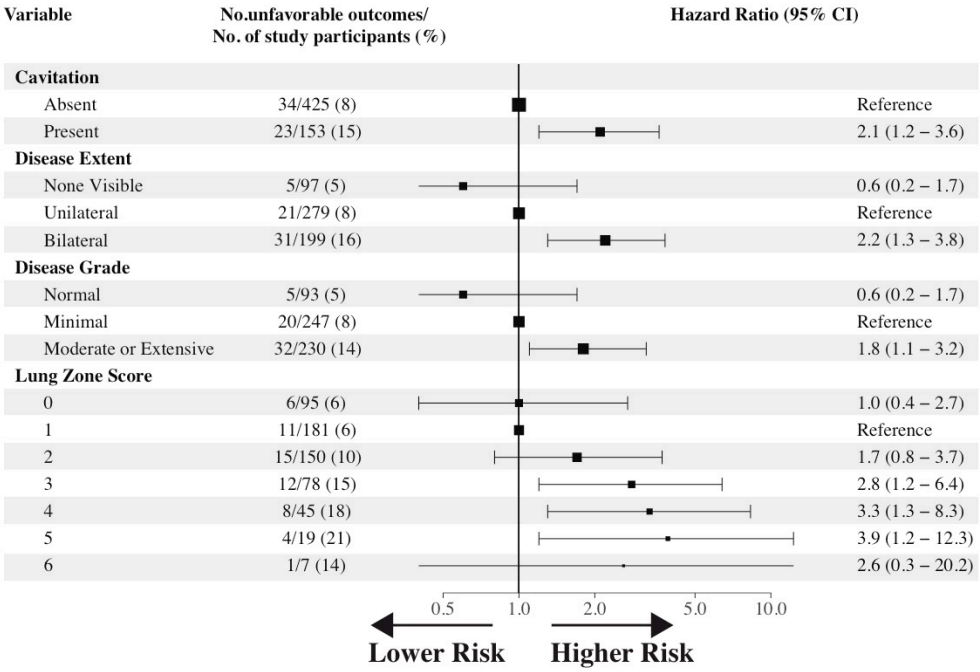



**Appendix Figure A.1** Univariate Cox proportional hazard analysis with X-Ray measurements in the MITT analysis population for the experimental group in pooled patient-level analysis. Measurements collected at baseline are shown in the top panel and measurements collected at the end of treatment are shown in the bottom panel. Cavity size at baseline was collected in the RIFAQUIN trial. Disease Extent, Disease Grade, and Lung Zone Score was collected in the OFLOTUB trial. Hazard ratios with 95% Wald confidence interval are reported.

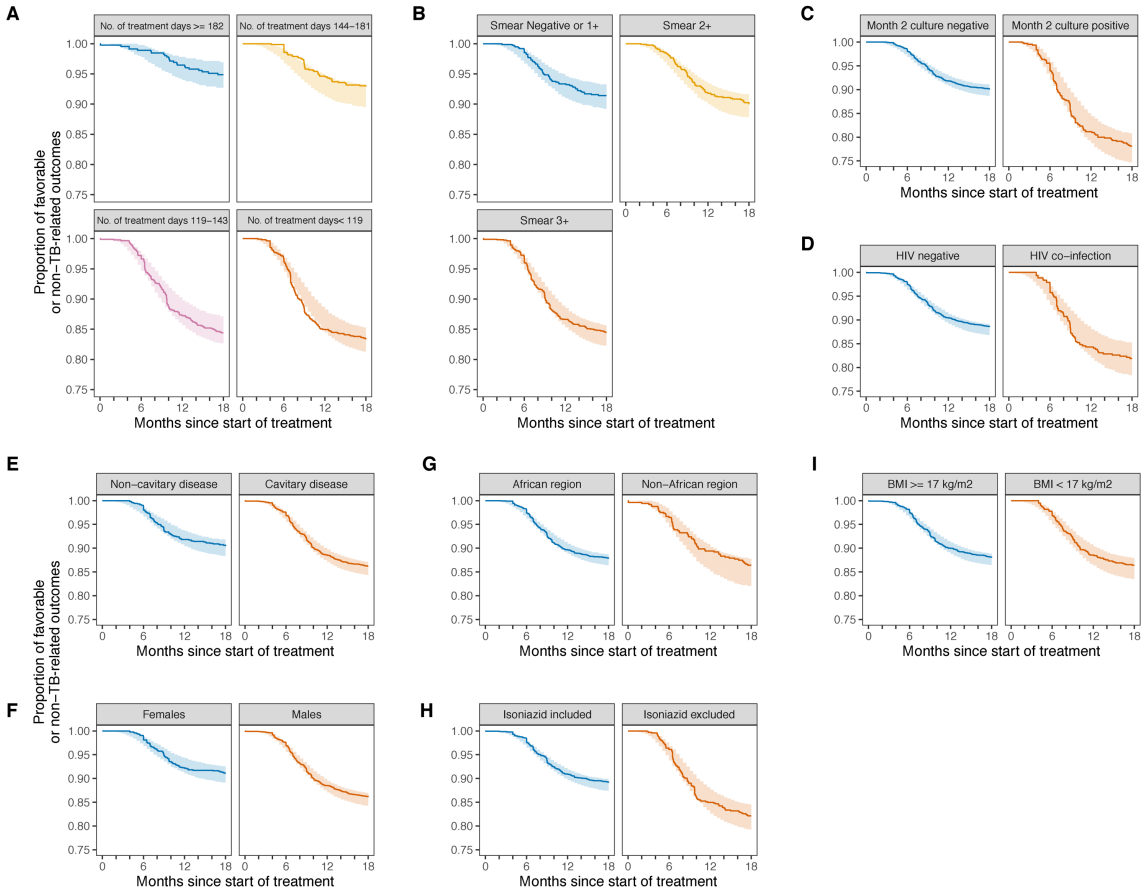
***Baseline***



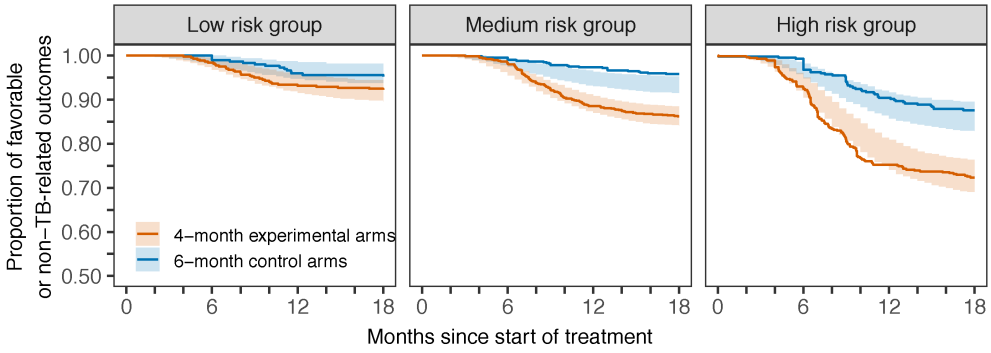

***End of Treatment***



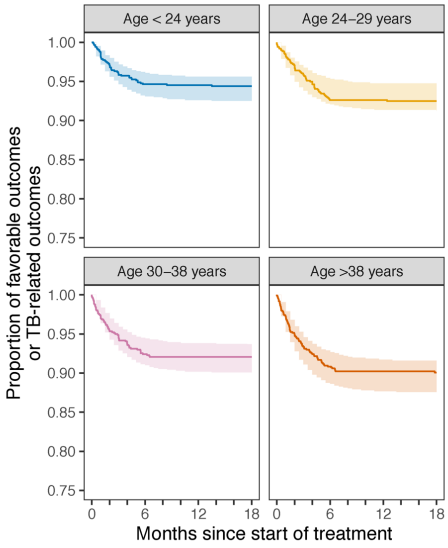

**Appendix Figure A.2** Univariate Cox proportional hazard analysis with X-Ray measurements in the MITT analysis population for the control group in pooled patient-level analysis. Measurements collected at baseline are shown in the top panel and measurements collected at the end of treatment are shown in the bottom panel. Cavity size at baseline was collected in the RIFAQUIN trial. Disease Extent, Disease Grade, and Lung Zone Score was collected in the OFLOTUB trial. Hazard ratios with 95% Wald confidence interval are reported.



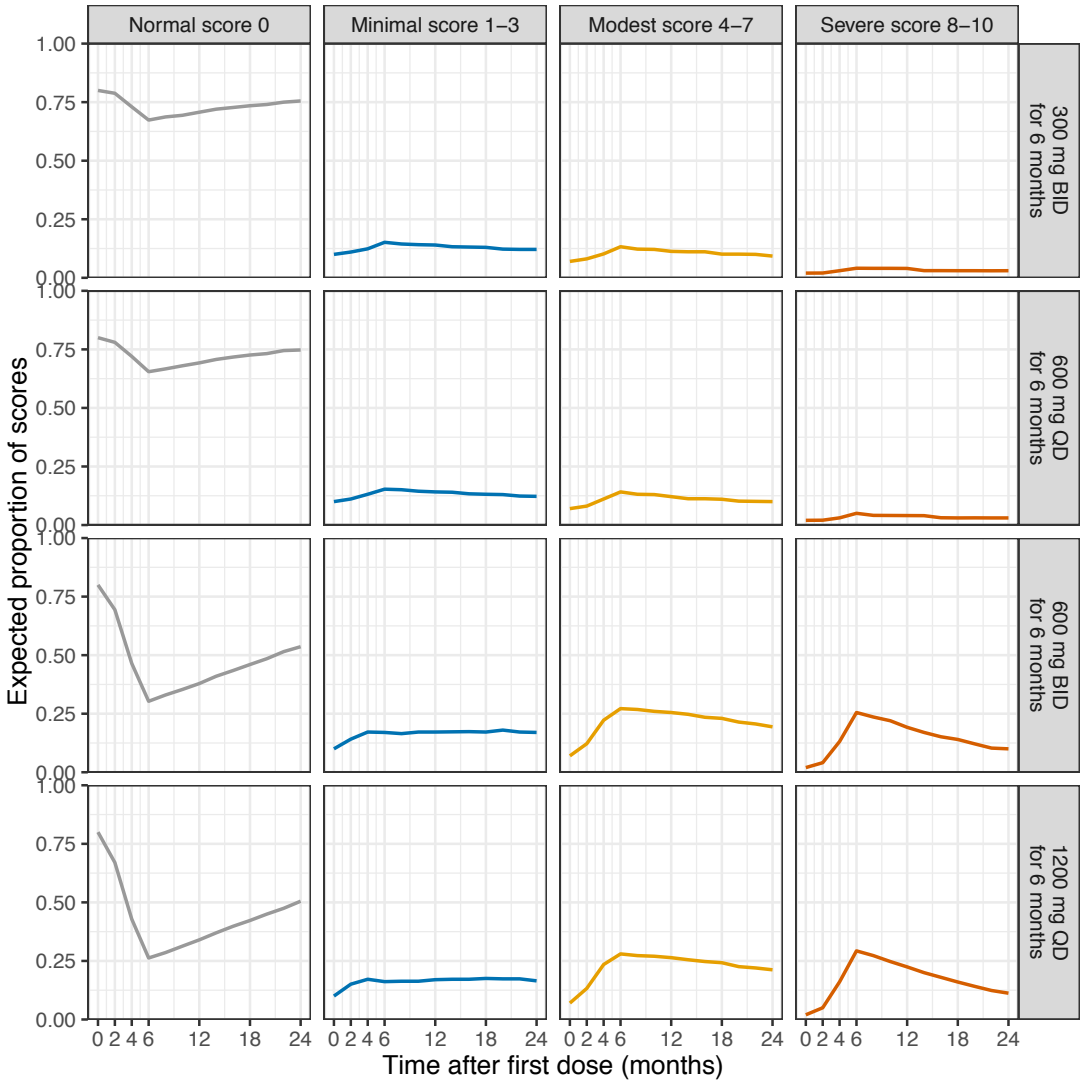

**Appendix Figure A.3** Kaplan-Meier visual predictive checks for final model describing TB-related outcomes stratified by covariates of interest: (A) number of treatment days, (B) baseline smear grade, (C) month 2 culture status, (D) HIV status, (E) baseline cavitation status, (F) sex, (G) clinic site, (H) inclusion of isoniazid, and (I) BMI.




**Appendix Figure A.4** Kaplan-Meier visual predictive checks for final model describing TB-related outcomes stratified by risk groups: low risk (left panel), medium risk (center panel), high risk (right panel) as defined by target cure rate of 93% at 18 months since start of treatment.




**Appendix Figure A.5** Kaplan Meier visual predictive checks for final model describing non-TB-related outcomes stratified by age.




**Appendix Figure A.6** Simulated proportions of peripheral neuropathy scores following various linezolid dosage administrations in BPaL regimen

**Publishing Agreement**

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
5FE4EC7F6C4140E...
Author Signature

6/8/2020
Date