

Precision-Enhancing Risk Stratification Tools for Selecting Optimal Treatment Durations in Tuberculosis Clinical Trials

Marjorie Z. Imperial^{1,2}, Patrick P. J. Phillips^{2,3}, Payam Nahid^{2,3}, and Radojka M. Savic^{1,2,3}

¹Department of Bioengineering and Therapeutic Sciences; ²University of California, San Francisco, Center for Tuberculosis, and ³Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, San Francisco, San Francisco, California

ORCID IDs: 0000-0001-7434-3012 (M.Z.I.); 0000-0002-6336-7024 (P. P.J.P.); 0000-0003-2811-1311 (P.N.); 0000-0003-3143-5579 (R.M.S.).

Abstract

Rationale: No evidence-based tools exist to enhance precision in the selection of patient-specific optimal treatment durations to study in tuberculosis clinical trials.

Objectives: To develop risk stratification tools that assign patients with tuberculosis into risk groups of unfavorable outcome and inform selection of optimal treatment duration for each patient strata to study in clinical trials.

Methods: Publicly available data from four phase 3 trials, each evaluating treatment duration shortening from 6 to 4 months, were used to develop parametric time-to-event models that describe unfavorable outcomes. Regimen, baseline, and on-treatment characteristics were evaluated as predictors of outcomes. Exact regression coefficients of predictors were used to assign risk groups and predict optimal treatment durations.

Measurements and Main Results: The parametric model had an area under the receiver operating characteristic curve of 0.72. A six-item risk score (HIV status, smear grade, sex, cavitary disease

status, body mass index, and Month 2 culture status) successfully grouped participants into low (1,060/3,791; 28%), moderate (1,740/3,791; 46%), and high (991/3,791; 26%) risk, requiring treatment durations of 4, 6, and greater than 6 months, respectively, to reach a target cure rate of 93% when receiving standard-dose rifamycin-containing regimens. With current one-duration-fits-all approaches, high-risk groups have a 3.7-fold (95% confidence interval, 2.7–5.1) and 2.4-fold (1.9–2.9) higher hazard risk of unfavorable outcomes compared with low- and moderate-risk groups, respectively. Four-month regimens were noninferior to the standard 6-month regimen in the low-risk group.

Conclusions: Our model discrimination was modest but consistent with current models of unfavorable outcomes. Our results showed that stratified medicine approaches are feasible and may achieve high cure rates in all patients with tuberculosis. An interactive risk stratification tool is provided to facilitate decision-making in the regimen development pathway.

Keywords: tuberculosis therapeutics; risk stratification; stratified medicine; optimal treatment duration; clinical trial design

Innovation in tuberculosis (TB) therapy is desperately needed. Current TB drug development programs are focused on identifying shorter one-size-fits-all TB treatment regimens that maximize treatment

completion without compromising on overall cure rates (1, 2). However, numerous translational gaps hinder the drug development pathway. Improved and innovative tools and approaches are

necessary to accelerate the identification of optimal treatment regimens and durations for all patients with TB (1, 3).

Whereas current practice guidelines highlight individual risk factors (bacterial

(Received in original form January 15, 2021; accepted in final form July 28, 2021)

Ⓓ This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0. For commercial usage and reprints, please e-mail Diane Gern.

Supported by the Critical Path Institute and Bill and Melinda Gates Foundation (OPP1776259) and the National Institute of Allergy and Infectious Diseases (1R01AI135124-01A1). The study sponsors had no role in study design, data collection, data analysis, data interpretation, or writing of the paper.

Author Contributions: M.Z.I. and R.M.S. designed and performed the analysis. M.Z.I., P.P.J.P., P.N., and R.M.S. interpreted the results, discussed implications, and contributed to the writing and revision of the manuscript.

Correspondence and requests for reprints should be addressed to Radojka M. Savic, Ph.D., University of California, San Francisco, 1700 4th Street, Rm 503C, UCSF Box 2552, San Francisco, CA 94158. E-mail: rada.savic@ucsf.edu.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 204, Iss 9, pp 1086–1096, Nov 1, 2021

Copyright © 2021 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.202101-0117OC on July 29, 2021

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: It has taken over 40 years to reduce the duration of tuberculosis (TB) treatment from 6 to 4 months, underscoring the urgent need for innovation in TB regimen development programs. Current one-size-fits-all approaches impede the identification of new regimens that would be curative if used with greater precision. The TB field should consider the well-documented evidence of diversity of disease burden and severity, which are the main drivers of unfavorable outcomes, to inform the design of the next generation of clinical trials that move the field beyond one-size-fits-all approaches to TB care.

What This Study Adds to the

Field: We pooled individual-level data from four phase 3 trials to develop evidence-based tools capable of stratifying patients into risk groups and informing optimal treatment duration for each strata to test in future clinical trials. Our tools can be used as a clinical trial design resource to inform *a priori* decisions regarding optimal durations for new regimens being considered for phase 3 clinical trials and to design novel phase 3 clinical trials that account for major risk factors.

burden, extent of cavitory disease, culture positivity at 8 weeks, etc.) that, based on *post hoc* analyses, suggest an extension in treatment duration may be warranted, there have been no tools developed that indicate the likelihood of achieving a cure based on an integrated suite of baseline risk factors, with or without on-treatment risk factors (4, 5). Such tools could support stratified medicine principles for TB care, whereby stratifying duration according to major risk factors can maximize cures for all patients and derisk identification of new TB regimens in the drug development pathway. Moreover, there are no tools that could estimate the likelihood of a durable cure when treatment is shortened to durations of less than 6 months.

As short and ultrashort duration regimens, such as those in TB Trials Consortium Study 31/A5349 (NCT02410772) (6) and the Two-month Regimens Using Novel Combinations to Augment Treatment Effectiveness for drug-sensitive TB (TRUNCATE TB) trial (NCT03474198) (7), are evaluated in the treatment of patients with TB, 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 four large contemporary phase 3 trials to develop and validate a data-driven framework aimed to maximize the success of late-stage clinical trials. Specifically, we developed quantitative risk stratification tools that assign patients into various risk groups of unfavorable outcomes and inform the selection of optimal treatment duration for each patient strata to study in clinical trials. Our tools can be used to inform *a priori* decisions regarding optimal durations for new regimens being considered for phase 3 clinical trials and to design novel phase 3 clinical trials that account for major risk factors. Some of the results of this study have been previously reported in the form of abstracts (8, 9).

Methods

Study Design

Individual-level data ($n = 3405$) from three international, randomized phase 3 trials (Ofloxacin-Containing, Short-Course Regimen for the Treatment of Pulmonary Tuberculosis [OFLOTUB] trial, NCT00216385 [10]; Rapid Evaluation of Moxifloxacin in TB [REMoxTB] trial, NCT00864383 [11]; and High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis [RIFAQUIN] trial, ISRCTN44153044 [12]) that compared 4-month fluoroquinolone-containing regimens to the standard 6-month regimen for treatment of drug-susceptible TB was used for model development. A fourth trial conducted by the Division of Microbiology and Infectious Diseases (DMID) of the National Institute of Allergy and Infectious Diseases, DMID 01-009 (NCT00130247, $n = 386$) (13), was used as an independent dataset for external validation, which tested a 4-month standard regimen (no fluoroquinolone). For each of these

regimens, rifampin was administered at the standard dose of 10 mg/kg for the complete duration of the regimen or rifampin was administered at the standard dose during the intensive phase and then replaced by rifapentine at 900 mg twice weekly in the continuation phase of treatment (4-month regimen in RIFAQUIN study). Additional information on study design for these trials is available in the original publications.

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 time point. Because of the composite definitions used to label unfavorable outcomes, we separated outcomes into two groups and developed separate models for 1) time to TB-related outcomes and 2) time to non-TB-related outcomes. TB-related outcomes included treatment failures, deaths owing 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. This approach allowed for the evaluation of the dichotomy between TB-related and non-TB-related outcomes, such that we would not expect common TB predictors (e.g., disease burden and severity) to be associated with non-TB-related outcomes if they are not related to TB. In addition, potential interventions for each outcome can be assessed. For each of the models, time was censored at time of alternative outcome (i.e., when modeling 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 (14).

Model Development and Evaluation

Parametric time-to-event models were used to describe time to TB-related outcomes and time to non-TB-related outcomes. Predictors of hazard risk parameters were tested for each model in a stepwise manner, in which exposure and regimen composition factors were first tested followed by baseline and on-treatment factors (see Supplemental Methods in the online supplement). The model building procedure was guided by Kaplan-Meier visual predictive checks to

assess calibration and area under the receiver operating characteristic curve (ROC AUC) for discrimination using model development and independent validation datasets.

Risk Stratification Algorithm and Optimal Treatment Duration

Exact regression coefficients of baseline and on-treatment predictors of TB-related outcomes were used to derive a risk score, $Risk\ Score_i$, for each individual, i . $Risk\ Score_i$ and the final model for time to TB-related outcomes were used to calculate the optimal treatment duration for each individual, $TRTDURATION_i$, required to reach a specified target cure rate, $CURE_{target}$. 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 $TRTDURATION_i$ are available in the online supplement.

The definitions of low-, moderate-, and high-risk groups were based on the predicted optimal treatment duration of the standard regimen (i.e., isoniazid, rifampin, pyrazinamide, and ethambutol) required to achieve less than or equal to 7% of TB-related outcomes ($CURE_{target} = 0.93$) at 18 months since the start of treatment. This target matches the pooled Kaplan-Meier estimate of TB-related outcomes at 18 months reported in the control groups from the original trials (see Supplemental Methods, Table E2, and Figure E2B in the online supplement). The low-risk group was defined as requiring less than or equal to 18 weeks of treatment, the moderate-risk group as requiring 19–24 weeks of treatment, and the high-risk group as requiring more than 24 weeks of treatment. The risk stratification algorithm was validated with the DMID 01-009 study data by comparing observed Kaplan-Meier estimates to target cure rates after treatment with 4-month and 6-month regimens for each risk group. Hazard ratios (HRs) were used to compare success rates in the 4-month and 6-month regimens for each risk group. As a second validation exercise, we performed an analysis with a random sample of 70% of the population from all four phase 3 trials (OFLOTUB, REMoxTB, RIFAQUIN, and DMID 01-009) for model development and the remaining 30% for validation.

Noninferiority Analysis

Data from all four phase 3 trials were pooled for noninferiority analyses between the 4-month experimental regimen and 6-month

control regimen in each risk group. The absolute difference in percentage of TB-related outcomes was calculated using inverse probability study-weighted Kaplan-Meier estimates at 18 months after start of treatment (15). Noninferiority was assessed using the upper bound of the two-sided 90% confidence interval (CI), determined by bootstrapping 500 samples, and a noninferiority margin of 6.6 percentage points—a margin that had been used in the most recent phase 3 trial (6).

Implementation of the Model and Algorithm into Interactive Tool

An interactive risk stratification tool based on the final model for TB-related outcomes and risk stratification algorithm was developed into a web application using the Shiny package in R (version 1.3.2). The tool has two modules: 1) the Risk Stratification Module uses available information on patient characteristics to assign risk groups and predict optimal treatment durations for the subgroup of interest, and 2) the Clinical Trial Design Module performs model simulations that can inform optimal treatment durations to test in clinical trials based on the study design (e.g., one-size-fits-all, subgroup analysis, enrichment, or risk stratification study designs). Additional information on the development and use of the tool is available in the Supplemental Methods and web application at <http://saviclab.org/tb-risk>.

Results

Data Characteristics

The model development dataset included 3,405 participants with drug-susceptible TB. Baseline characteristics did not differ between experimental and control groups (Table 1) (16). In the 4-month experimental group, 1,257/2,001 (63%) of participants 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 1; Figure E1 in the online supplement). Month 2 culture conversion rates were higher in the 4-month experimental group than 6-month control group (Table 1, $P = 0.01$). Of the 3,405 participants, 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 control 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 E2).

Model Development and Evaluation

The hazard risk for TB-related outcomes was best described with a surge function (see Supplemental Methods). A decreased number of treatment days and exclusion of isoniazid increased the hazard risk of TB-related outcomes (29% [percent relative standard error (%RSE) = 9] increase per 28-day decrease in number of treatment days; 32% [48] increase for exclusion of isoniazid; Table 2). Baseline factors that increased hazard risk included HIV coinfection (86% [%RSE = 29] increase), higher smear grade (68% [36] increase for smear 3+ relative to smear 1+ or negative), male sex (64% [32] increase), presence of cavitory disease (26% [57] increase), and lower body mass index (BMI) (18% [41] increase per 5 kg/m² decrease). Inclusion of Month 2 culture status improved discrimination with an increase in ROC AUC from 0.69 (95% CI, 0.66–0.72) to 0.72 (0.69–0.75) (Table 2). Calibration of the final predictive model was good (Figures E3–E5).

A Gompertz function was used to describe the hazard risk of non-TB-related outcomes (see Supplemental Methods). 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 and Figures E3 and E6). Because the final model for non-TB-related outcomes was 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.

Risk Stratification Algorithm and Optimal Treatment Durations

Optimal treatment duration was predicted based on a six-item hazard risk score: HIV status, baseline smear grade, sex, baseline cavitory disease, baseline BMI, and Month 2 culture status. The derivations and final formulas to calculate individual risk scores ($Risk\ Score_i$) and optimal treatment durations ($TRTDURATION_i$) are presented in the Supplemental Results. Based on the predicted optimal treatment durations to reach a 93% target cure rate, 794/3,405 (23%) participants in the model

Table 1. Baseline, On-Treatment, and Regimen Characteristics of Study Participants Included in the Model Development Population

Characteristic	Model Development Dataset		Independent Dataset	
	4-Mo Experimental Group (n = 2,001)	6-Mo Control Group (n = 1,404)	4-Mo Experimental Group (n = 193)	6-Mo Control Group (n = 193)
Site region				
Sub-Saharan Africa	1,653 (83)	1,228 (88)	80 (41)	79 (41)
India	228 (11)	114 (8)	0 (0)	0 (0)
Asia	120 (6)	62 (4)	46 (24)	46 (24)
South America	0 (0)	0 (0)	67 (35)	68 (35)
Sex, F, n (%)	592 (30)	415 (30)	76 (39)	76 (39)
Age, yr*				
Median	30	29	29	27
Interquartile range	24–39	24–38	23–38	22–36
Range	16–81	17–77	18–59	18–59
Weight, kg				
Median	52	52	54	55
Interquartile range	46–58	47–58	49–62	49–61
Range	35–98	35–137	35–98	32–90
Body mass index [†]				
Median	18.4	18.3	20.3	19.5
Interquartile range	16.9–20.2	16.9–20.1	18.7–22.1	18.5–22.1
Range	12.0–40.7	12.1–50.9	14.0–33.3	12.1–37.7
HIV positivity, n (%) [‡]	248 (12)	220 (16)	0 (0)	0 (0)
Cavitary disease, n (%) [§]	1,247 (62)	847 (60)	0 (0)	0 (0)
Smear, n (%)				
Negative or 1+	483 (24)	317 (23)	111 (58)	115 (60)
2+	503 (25)	404 (29)	32 (17)	36 (18)
3+	988 (49)	667 (48)	50 (26)	42 (22)
Regimen composition				
Isoniazid	1,257 (63)	1,404 (100)	193 (100)	193 (100)
Rifapentine	193 (10)	0 (0)	0 (0)	0 (0)
Moxifloxacin	1,312 (66)	0 (0)	0 (0)	0 (0)
Gatifloxacin	689 (34)	0 (0)	0 (0)	0 (0)
Treatment duration (d) ^{¶**}				
Median	119	175	112	168
Interquartile range	114–119	169–182	111–114	167–170
Range	2–202	4–239	53–142	142–196
Number of treatment days ^{**††}				
Median	114	144	—	—
Interquartile range	96–119	144–182	—	—
Range	1–120	1–189	—	—
Cumulative rifamycin dose, mg ^{**‡‡}				
Median	57,600	86,400	—	—
Interquartile range	51,600–71,400	79,200–108,600	—	—
Range	450–72,000	450–113,400	—	—
Month 2 culture positivity ^{§§}	336 (17)	285 (20)	0 (0)	0 (0)

Model development dataset includes OFLOTUB (Ofloxacin-Containing, Short-Course Regimen for the Treatment of Pulmonary Tuberculosis), REMoxTB (Rapid Evaluation of Moxifloxacin in TB), and RIFAQUIN (High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis) trial data, and independent dataset for external validation includes DMID (Division of Microbiology and Infectious Diseases) 01-009 trial data.

*Age was missing for five study participants.

[†]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 heights for females and males were used to calculate body mass index.

[‡]HIV status was missing for nine study participants.

[§]Cavitary disease status was missing for 200 study participants.

^{||}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.

[¶]Treatment duration, defined as the number of days the participant was on treatment, was missing for 117 study participants.

**For the independent dataset (DMID 01-009 trial), number of treatment days was not available. However, all study participants were required to have completed a minimum of 112 doses of anti-tuberculosis treatment within 18 weeks, and then participants were randomized to stop treatment or to receive an additional 2 months of the continuation phase (isoniazid and rifampin) for a total of 162 doses. Treatment was administered 7 days per week, with at least five doses administered by directly observed therapy.

^{††}Number of treatment days, defined as the total number of treatment days drugs were administered, was missing for 38 study participants.

^{‡‡}Cumulative rifamycin dose, defined as number of treatment days multiplied by individual rifamycin daily dose, was missing for 38 study participants.

^{§§}Month 2 culture was missing for 308 study participants.

Table 2. Estimated Parameters for Models Describing TB-related Outcomes and Non-TB-related Outcomes

	TB-related Outcome Model without Month 2 Culture*	TB-related Outcome Model with Month 2 Culture*	Non-TB-related outcome model
ROC AUC (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)	Estimate (%RSE)	Estimate (%RSE)
Baseline hazard [†]	10 ^{-4.0} (11)	10 ^{-4.1} (11)	0.03 (8)
Shape parameter [†]	0.52 (24)	0.52 (24)	0.38 (6)
Shape parameter 2 [†]	3.9 (26)	3.9 (27)	—
Covariate effects [‡]			
Percent increase in baseline hazard			
Per 28-d decrease in number of treatment days	28 (10)	29 (9)	—
For Month 2 culture positivity	—	145 (19)	—
For HIV coinfection	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 cavitory 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-yr increase in age	—	—	23 (29)

Definition of abbreviations: %RSE = percent relative standard error of the parameter estimate (typical value or median); BMI = body mass index; ROC AUC = area under the receiver operating characteristic curve; RSE = relative standard error; TB = tuberculosis.

*Final model adjusted for region of clinic site (sub-Saharan Africa vs. non-sub-Saharan Africa).

[†]Hazard of TB-related outcomes was described with the surge function, and hazard of non-TB-related outcomes was described with the Gompertz function. Additional details are in the Supplemental Methods.

[‡]Covariate effects added using linear relationships. For continuous covariates, the following relationship was used:

$P(x) = TVP(1 + \theta(COV - COV_{median})/100)$, where TVP is the typical value for parameter P , θ 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: $P(x) = TVP(1 + \theta(COV)/100)$, where TVP is the typical value for parameter P , and θ 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) refers to increased hazard risk of unfavorable outcomes (TB- or non-TB-related, respectively) in this model.

development population were assigned to the low-risk group, with risk scores ranging from 0 to 1.67; 1,624/3,405 (48%) participants were assigned to a moderate-risk group, with risk scores ranging from 1.68 to 3.20; and 987/3,405 (29%) participants were assigned to a high-risk group, 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 Figures 1A and 1B, respectively. Figure 1C illustrates the distribution of different risk factors across the three risk strata. Participants 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 estimates shown in Figure 2 and Figure E5. Participants 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 participants treated with a 6-month regimen resulted in cure rates above 93%, with 4-month regimens leading to significantly higher risk of TB-related outcomes than the 6-month regimen (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). After adjustment for regimen, high-risk groups have a 3.7-fold (95% CI, 2.7–5.1) and 2.4-fold (95% CI, 1.9–2.9) higher hazard risk of unfavorable outcomes compared with low- and moderate-risk groups, respectively. No interaction between regimens and risk groups was identified, suggesting that the risk of TB-related outcomes increases in higher risk groups independent of treatment duration (P value for interaction = 0.4).

Validation of Model and Risk Stratification Algorithm

The final TB-related outcome model and risk stratification algorithm were externally validated using an independent dataset available from the DMID 01-009 trial that included 386 participants with drug-susceptible, noncavitory TB disease at baseline and culture conversion at Month 2. This independent dataset represents a subpopulation of primarily lower risk, with 266/386 (69%) participants in the low-risk group, 116/386 (30%) in the moderate-risk group, and 4/386 (1%) in the high-risk group (Figure E7). The TB-related outcome model had similar discrimination and calibration with the independent dataset as compared with the model development dataset (ROC AUC, 0.78; 95% CI, 0.65–0.90; Figure E8). The observed Kaplan-Meier estimates of TB-related outcomes confirmed that patients in the low-risk group can be treated with a 4-month regimen, and patients in the moderate-risk group require at least 6 months of treatment to reach 93% target cure rates (Figure 3). No TB-related outcomes

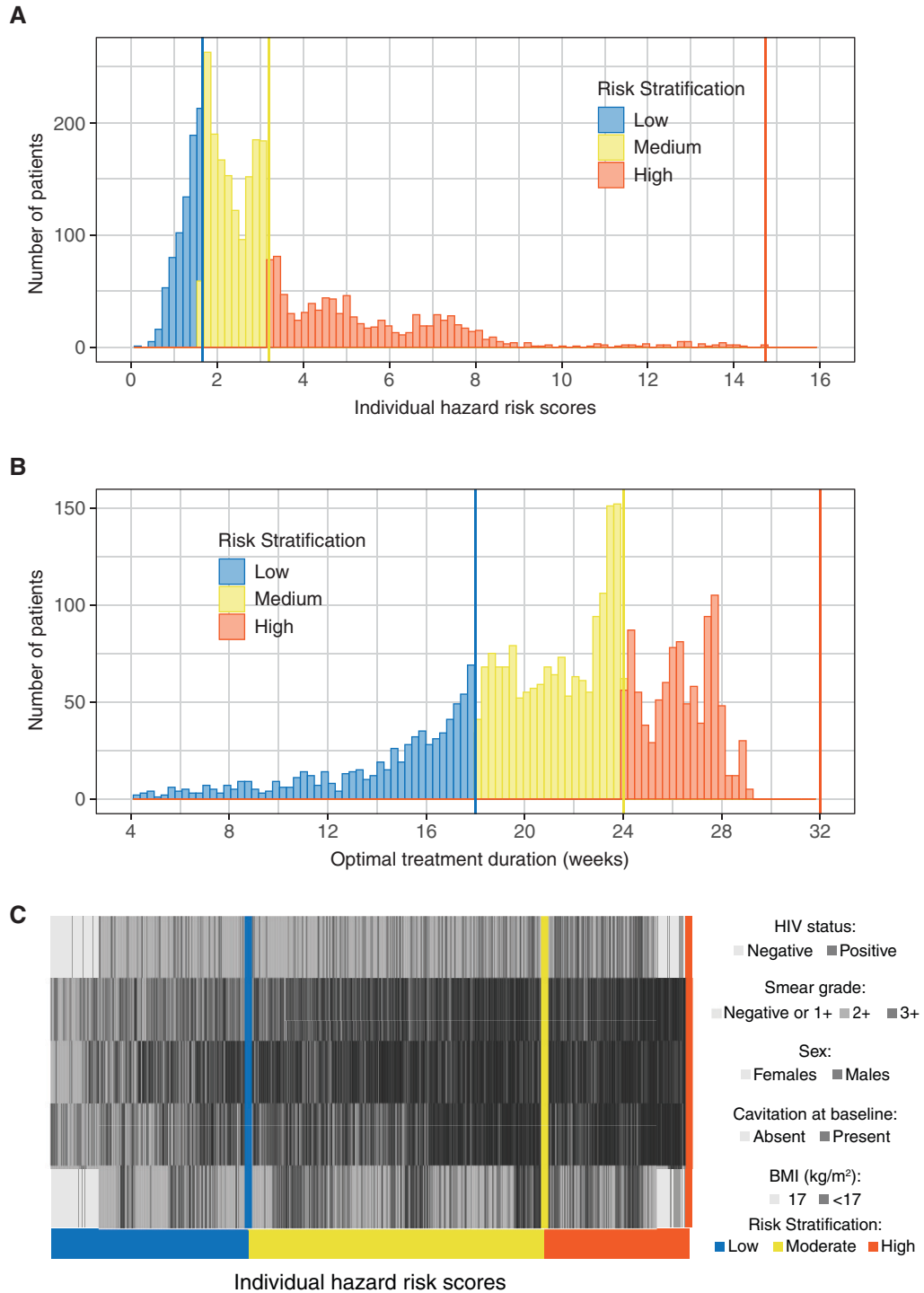


Figure 1. Distribution of individual risk scores, optimal treatment durations, and risk factors for target cure of 93% in the model development population. (A) Distribution of individual risk scores stratified by low-, moderate-, and high-risk groups. (B) Distribution of predicted optimal treatment durations for target cure rate of 93% 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. The low-risk group was defined as patients requiring less than or equal to 18 weeks of treatment, the moderate-risk group as requiring 19–24 weeks of treatment, and the high-risk group as requiring more than 24 weeks of treatment for a target cure rate of 93%. BMI = body mass index.

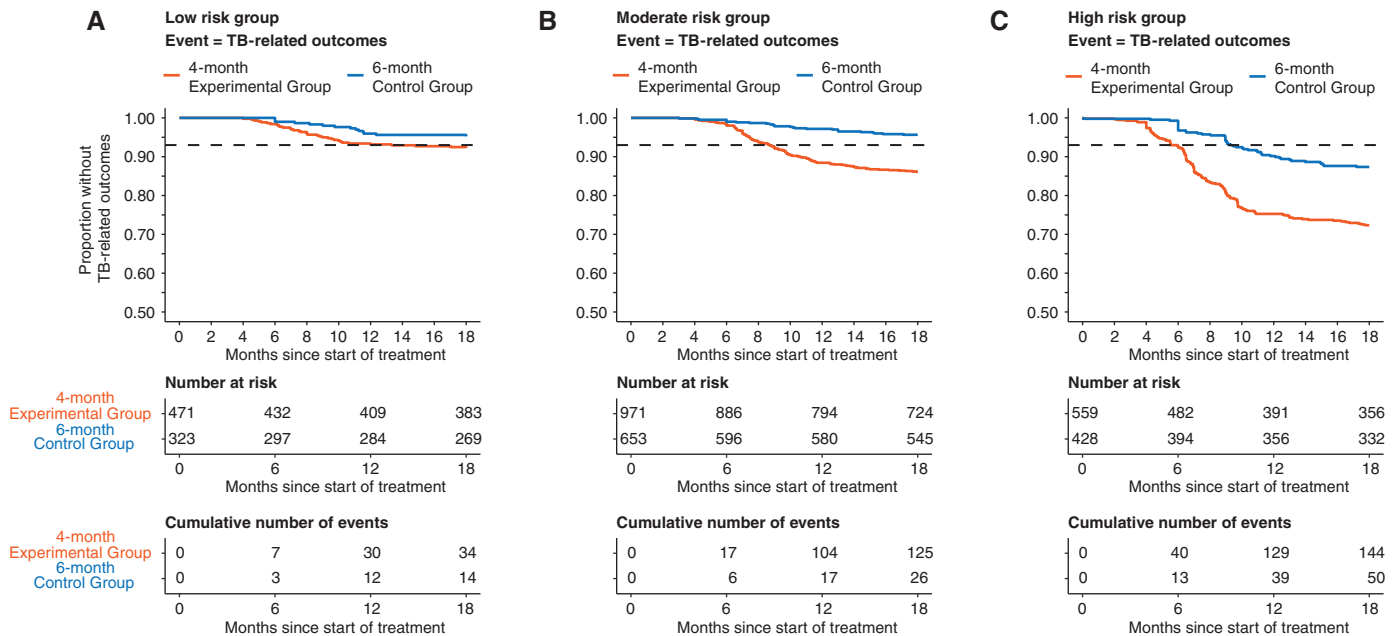


Figure 2. Kaplan-Meier estimates to validate calibration of risk stratification algorithm using model development population. (A) Low-risk group stratified by regimen duration. (B) Moderate-risk group stratified by regimen duration. (C) High-risk group stratified by regimen duration. Dashed line shows target cure rate of 93%. TB = tuberculosis.

were reported in the four patients categorized in the high-risk group. The second validation exercise using a random split of the population for model development and validation had similar exact regression coefficients in the model, proportions of the population assigned to each risk group, and Kaplan-Meier estimates for each risk group compared with the primary analysis (Table E3 and Figures E9 and E10).

Noninferiority Analysis

Combining all four phase 3 trials, in the low-risk group (1,060/3,791; 28%), the 4-month regimens were noninferior to the 6-month control regimen, with a difference in study adjusted Kaplan-Meier estimate of TB-related outcomes of 2.6 (90% CI, 0.2–5.1) (Figure 4). Conversely, in the moderate-risk (1,740/3,791; 46%) and high-risk (991/3,791; 26%) groups, the difference in study adjusted Kaplan-Meier estimate of TB-related outcomes was 9.5 (90% CI, 7.2–11.8) and 16 (90% CI, 11.6–20.3), respectively, both favoring the 6-month control regimen.

Interactive Risk Stratification Tool for Clinical Trial Design

We developed an evidence-based interactive risk stratification tool that can generate critical knowledge essential for regimen optimization in a clinical trial setting by

highlighting those subgroups of patients who are at higher risk of unfavorable outcomes and may require treatment adjustments to reach trial objectives (e.g., identify noninferior or superior regimens). Specifically, it can be used to stratify patients into risk groups of unfavorable outcomes, inform the selection of optimal treatment durations for each risk group to test with new regimens, and inform the design of novel late-stage clinical trials that account for major risk factors (e.g., patient phenotype enrichment or risk stratification studies). Input parameters include arguments about study design, patient characteristics, and patient adherence. The tool can handle missing data by performing simulations with bootstrapped populations from a subset of the model development population with the same available risk factors (Figure E11). Additional details and instructions for its use are available in the Supplemental Methods and web application hosted at <http://saviqlab.org/tb-risk>, with a snapshot shown in Figure 5.

Discussion

The current one-size-fits-all approach to TB regimen development impedes the identification of new regimens that would be

curative if used with greater precision. New clinical trial data has emerged from Study 31/A5349 with a landmark achievement, in which a 4-month high-dose rifapentine regimen with moxifloxacin successfully showed noninferior results to the 6-month standard regimen using a one-size-fits-all approach (6). Still, it has taken over 40 years to reduce the duration of treatment from 6 to 4 months, underscoring the formidable barrier to successfully shortening treatment durations for TB and the urgent need for innovation in TB therapy and regimen development programs. The diversity of disease and large spectrum of patient phenotypes is regularly considered in other diseases when optimizing effective treatment programs, particularly in oncology (17). In that manner, the TB field should consider the well-documented evidence of diversity of disease burden and severity, which are the main drivers of unfavorable outcomes (4, 16, 18), to inform the design of the next generation of clinical trials and regimens that move the field beyond one-size-fits-all approaches to TB care. To support this, we developed a risk stratification algorithm that successfully stratified patients with drug-susceptible TB into low- (1,060/3,791; 28%), moderate- (1,740/3,791; 46%), and high-risk (991/3,791; 26%) groups. In conjunction, through risk stratification, we

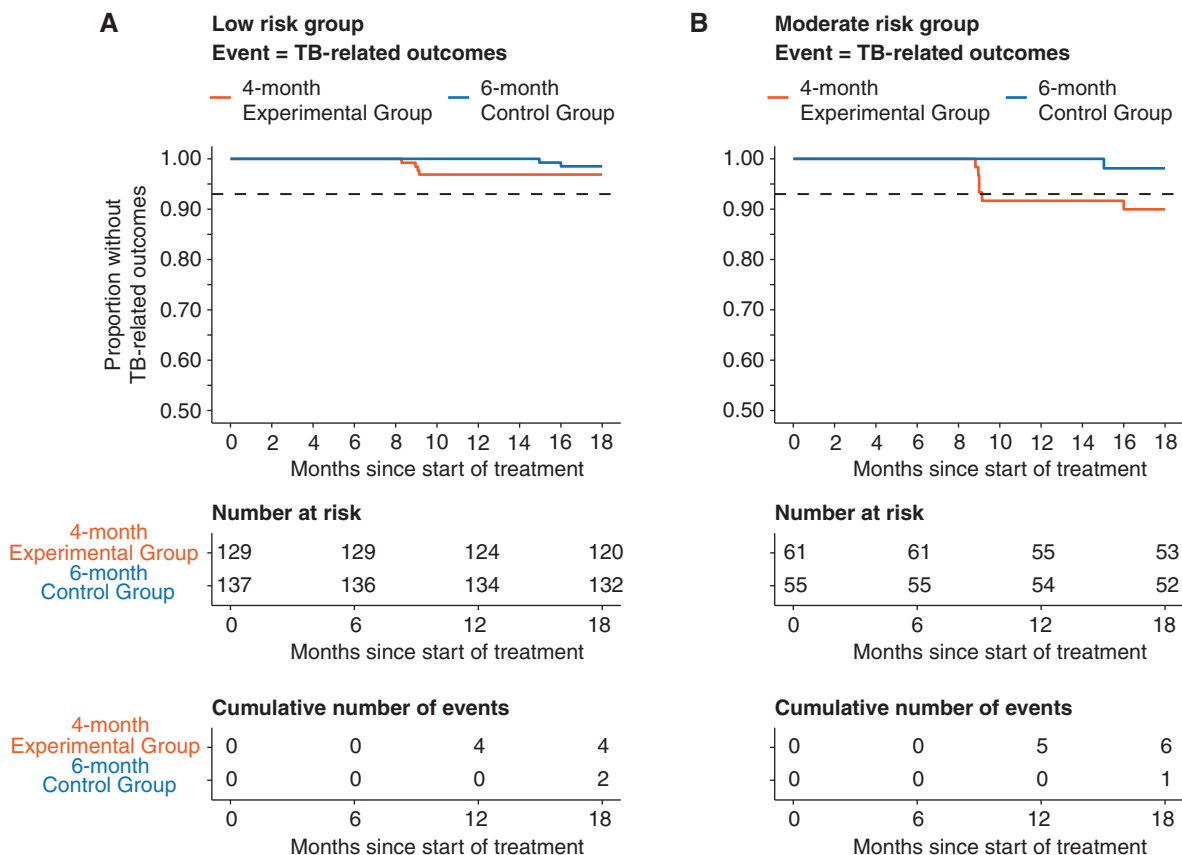


Figure 3. Kaplan-Meier estimates to validate calibration of risk stratification algorithm using an independent dataset. (A) Low-risk group stratified by regimen duration. (B) Moderate-risk group stratified by regimen duration. Only three individuals in the 4-month experimental group and one individual in the 6-month control group, none of which had a TB-related unfavorable outcome, were categorized as high risk in the independent dataset, so a Kaplan-Meier graph is not shown. Dashed line shows target cure rate of 93%. TB = tuberculosis.

are able to predict the optimal treatment durations of standard-dose (10 mg/kg) rifamycin-containing regimens for each stratum, in which low-risk patients can be treated with a 4-month regimen,

moderate-risk patients with a 6-month regimen, and high-risk patients likely with regimens exceeding 6 months without compromising on cure rates. Based on our results, we developed an interactive risk

stratification tool as a clinical trial design resource that can provide evidence-informed recommendations on optimal treatment interventions to be tested in future clinical trials.

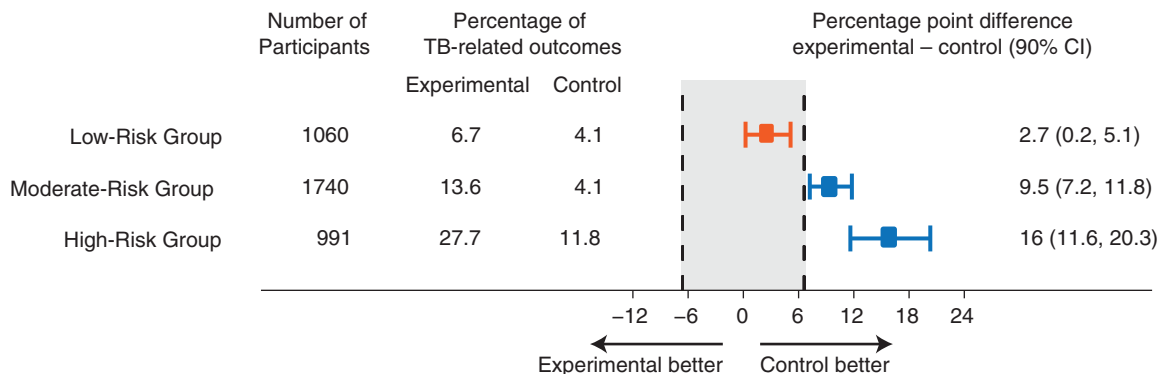


Figure 4. Difference in percentage of TB-related outcomes between the 4-month experimental group and the 6-month control group according to risk groups defined by the risk stratification algorithm. The noninferiority analysis is based on the pooled model development and independent datasets (all four phase 3 trials). The 90% confidence intervals of the differences in percentage of unfavorable outcomes were calculated with inverse probability study weighted Kaplan-Meier estimates at 18 months from 500 bootstrap samples. Red squares denote experimental subgroups that were noninferior to the control subgroups, and blue squares denote subgroups that did not show noninferiority. CI = confidence interval; TB = tuberculosis.

Risk Stratification Tool

[About](#)
[Risk Stratification Module](#)
[Clinical Trial Design Module](#)

Welcome to UCSF's Risk Stratification Tool for Tuberculosis Clinical Trial Design!

This evidence-based interactive tool uses a validated model and risk stratification algorithm to stratify patients with tuberculosis (TB) into risk groups and inform selection of optimal treatment durations for each risk group to test in late-stage TB clinical trials. Our tool can be used by researchers and clinical trialists to inform *a priori* decisions regarding optimal durations for new TB regimens being considered for phase 3 clinical trials and to design novel phase 3 clinical trials that account for major risk factors. We provide a clinical trial design resource that can inform decisions that facilitate early and effective deployment of the best regimens for the right patients.

Risk Stratification Module

The Risk Stratification Module uses available information on patient characteristics to assign risk groups and predict optimal treatment durations for the subgroup of interest.

High risk group

***Optimal treatment duration = 29 weeks**

*Optimal treatment duration for daily (1/7 days per week) regimen with standard dose rifampin, isoniazid, pyrazinamide, and ethambutol or a fluoroquinolone.

[Go to Risk Stratification Module](#)

Clinical Trial Design Module

The Clinical Trial Design Module performs model simulations that can inform optimal treatment durations to test in clinical trials based on the study design (e.g. one-size-fits-all, subgroup analysis, enrichment, or risk stratification study designs).

■ Cavitary disease absent
 ■ Cavitary disease present

Smear Grade	Cavitary disease absent (N)	Cavitary disease present (N)
Smear negative or 1+	281	379
Smear 2+	360	435
Smear 3+	364	1073

[Go to Clinical Trial Design Module](#)

Figure 5. Interactive risk stratification tool. The “About” page in the web application that displays information on the Risk Stratification and Clinical Trial Design Module is shown. UCSF = University of California, San Francisco.

Our risk stratification algorithm uses six markers of risk that are routinely collected in clinical trials: HIV status, baseline smear grade, sex, baseline cavitary disease, baseline BMI, and Month 2 culture status. The risk stratification algorithm successfully grouped and validated low-risk participants eligible for 4-month standard-dose rifamycin-containing regimens and moderate-risk participants requiring at least 6 months. The high-risk participants had suboptimal relapse rates with 6-month standard-dose rifamycin-containing regimens, but it was not validated whether regimens exceeding 6 months would result in better treatment outcomes, as longer treatment durations were not tested in the original studies. Nevertheless, 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 with standard rifamycin doses in clinical trials. For example, the observed proportion of favorable outcomes for this group treated with the 6-month control was 88% (378/428)

compared with the low- and moderate-risk groups at 96% (936/976) (Figure 2). Possible alternative interventions that can be tested in clinical trials 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 (19–21). These potentially more effective regimens may also allow for ultrashort treatments for low- and moderate-risk groups in clinical trials.

Presently, only two separate studies have investigated the relationship between treatment duration and rates of relapse. In one, a meta-regression model developed from published historical data to predict rates of relapse using treatment duration and the proportion of participants with negative culture at Month 2 was capable of predicting the expected rates of relapse in the 4-month experimental regimens from the REMoxTB and RIFAQUIN trials (22). In a second study, a translational pharmacokinetic–pharmacodynamic model derived from preclinical mice data was used to predict the results of a number

of clinical trials with reasonable success (23). However, both models predicted wide confidence and prediction intervals, suggesting that other important factors were 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. Our tools are now capable of quantitatively predicting, with good precision, rates of TB-related outcomes and confidently providing recommendations on optimal treatment durations in stratified groups.

The composite definition of unfavorable outcomes was not standardized across the trials included in our analysis. To alleviate this issue, TB-related and non-TB-related outcomes were modeled separately to determine whether different risk factors affect each outcome. Indeed, treatment- and disease-specific risk factors only affected TB-related outcomes, so proposed treatment interventions would only improve relapse, treatment failures, and TB-related deaths. Still, non-TB-related outcomes are critical to assess because they routinely contribute to

the assessment of overall efficacy and are undesirable because of the risk of disease transmission and emergence of drug-resistant strains (6, 10–12, 24). The only risk factor of non-TB-related outcomes was older age. Hence, to improve non-TB-related outcomes, interventions may instead be focused on informing patients about the potential risk of inadequate treatment, use of maximal efforts by clinicians and researchers to contact participants (e.g., phone calls or home visits) that miss routine visits, and close monitoring of adverse events. In any case, this pooled analysis highlights the need for standard definitions of endpoints and analysis methods in TB clinical trials. A new framework that focuses on TB-related outcomes and provides a standardized language to help articulate the question of interest, analysis, and interpretation in clinical trials has been proposed in a recent addendum (25) to the International Council for Harmonisation of Technical Requirements for Pharmaceutical for Human Use (ICH) E9 Statistical Principles for Clinical Trials (26) and is now being considered and implemented in future TB clinical trials (e.g., Duration Randomized Anti-Multidrug-resistant-TB And Tailored Intervention Clinical Trial [DRAMATIC trial], NCT03828201). In this regard, focusing on TB-related outcomes may be more relevant to future trials.

Our study has limitations. First, our parametric model had modest discrimination (ROC AUC = 0.72) when including all potential risk factors of poor outcomes collected in each trial. However, our goal is not to assign a specific duration for each individual patient but rather to stratify patients into risk groups that can be assigned appropriate durations. In addition, our model performance is consistent with current microbiological markers (e.g., culture conversion and smear grade) as predictors of treatment outcome (27–31). Certainly, more quantitative and sensitive measures of disease burden and severity (e.g., cycle threshold in Gene Xpert and lipoarabinomannan levels in sputum) are now becoming available and may one day replace current markers (32–37). The model described herein provides the framework that can subsequently be revised to account for more robust markers as additional data become available. Second, our independent dataset for external validation represented a subpopulation of primarily lower risk. Thus, we also performed the analysis with a

random sample of the population for model development and validation, which had similar results as our primary analysis (Table E3 and Figures E9 and E10). Third, all tested regimens in the studies included in our analysis were rifamycin-based regimens at standard-suboptimal doses. Predicted optimal treatment durations will likely be underestimated when high-dose rifamycin-containing regimens are considered, especially with the landmark clinical trial data that recently emerged from Study 31/A5349 (6). This model will be continually revised as new clinical trial data become available. Caution is also advised if generalizing our findings to regimens of other compositions, as predictors of relapse may be different. Finally, the tools necessary to measure risk markers are already in use in many settings, but in some (e.g., high-TB-burden settings), the proposed risk markers may not be routinely available with limited access to diagnostics, particularly chest radiographs, routine cultures, and HIV testing, among other factors. The World Health Organization has endorsed wider and quality-assured use of chest radiography for TB detection in combination with laboratory-based diagnostic tests (38). As such, we prefer to present the risk stratification tools inclusive of these data. Nevertheless, we have implemented our interactive tool so it can handle missing values using two approaches. First, a simplified model, excluding 2-month culture as a risk marker, can be used to make predictions with similar discriminatory ability as the full model (Table 2, TB-related Outcome Model without Month 2 Culture). Second, predictions can be made with missing markers by performing simulations with bootstrapped populations from the subset of the model development population with the same available risk factors, such that the bootstrapped populations will be based on a pool of patients with similar risk (Supplemental Methods and Figure E11). Future trials that test stratified medicine approaches to TB care should also evaluate newer measures of risk (e.g., cycle threshold in GeneXpert), which would allow for tools to be refined and expanded, offering additional characteristics and options for determining risk. Overall, our tools are intended as clinical trial design resources that use information routinely collected in contemporary clinical trials. They are not intended for programmatic use at this time, though they could be in the future as we incorporate additional data from newer phase 3 trials.

Strengths of our analyses include the inclusion of four large datasets from phase 3 trials conducted across diverse populations in high-TB-burden settings; our predictive model is evidence based, is fully parametric with minimal assumptions about the shape of hazard risk, and has similar predictive performance in the model development and validation datasets as that of other models of risk of relapse (22); our stratification algorithm is based on routinely collected makers in clinical trials; and our interactive risk stratification tool handles complex calculations and missing data.

In conclusion, we developed a parametric model with performance consistent with current microbiological markers as predictors of treatment outcome and provide a risk stratification algorithm capable of assigning patients into risk groups and informing optimal treatment durations for each risk group. Furthermore, an evidence-based interactive risk stratification web application is provided as a clinical trial design resource that will allow for more informed and accelerated decision-making in the regimen development pathway. Importantly, our results support the idea of stratified medicine approaches for TB care: a paradigm shift in overall objectives that is patient centered and enhances cure rates for the most severe TB cases while reducing duration, toxicity, and cost to programs and patients for the less severe TB cases. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the study participants, the staff at clinical sites, all investigators for each of the four trials, and the data management team at Critical Path Institute. The standardized, individual, deidentified data for the OFLOTUB (ClinicalTrials.gov number NCT00216385), REMoxTB (ClinicalTrials.gov number NCT00864383), and RIFAQUIN (ISRCTN number 44153044) trials that support the development of this tool 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 (ClinicalTrials.gov number NCT00130247) data that support the findings of this study are available from the Tuberculosis Research Unit at Case 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.

References

- Lienhardt C, Nahid P. Advances in clinical trial design for development of new TB treatments: a call for innovation. *PLoS Med* 2019;16:e1002769.
- World Health Organization. Target regimen profiles for TB treatment: candidates: rifampicin-susceptible, rifampicin-resistant and pan-TB treatment regimens. Geneva: World Health Organization; 2016.
- Ginsberg AM, Spigelman M. Challenges in tuberculosis drug research and development. *Nat Med* 2007;13:290–294.
- Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, 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. *Clin Infect Dis* 2016;63:e147–e195.
- World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and patient care (2017 update). Geneva: World Health Organization; 2017.
- Dorman SE, Nahid P, Kurbatova EV, Phillips PPJ, Bryant K, Dooley KE, et al.; AIDS Clinical Trials Group; Tuberculosis Trials Consortium. Four-month rifampentine regimens with or without moxifloxacin for tuberculosis. *N Engl J Med* 2021;384:1705–1718.
- Papineni P, Phillips P, Lu Q, Cheung YB, Nunn A, Paton N. TRUNCATE-TB: an innovative trial design for drug-sensitive tuberculosis. *Int J Infect Dis* 2016;45:404.
- Imperial M, Phillips PPJ, Nahid P, Savic RM. Predicting optimal treatment durations for tuberculosis patients: risk stratification algorithm and simulation tool. Presented at TBScience 2020. October 2020, Seville, Spain [virtual].
- Imperial M, Phillips PPJ, Nahid P, Savic RM. Predicting optimal treatment durations for tuberculosis patients: risk stratification algorithm and simulation tool. Presented at American Society for Clinical Pharmacology and Therapeutics Annual Meeting. March 2021.
- Merle CS, Fielding K, Sow OB, Grinafon M, Lo MB, Mthiyane T, et al.; OFLOTUB/Gatifloxacin for Tuberculosis Project. A four-month gatifloxacin-containing regimen for treating tuberculosis. *N Engl J Med* 2014;371:1588–1598.
- Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, et al.; REMoxTB Consortium. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* 2014;371:1577–1587.
- Jindani A, Harrison TS, Nunn AJ, Phillips PPJ, Churchyard GJ, Charalambous S, et al.; RIFAQUIN Trial Team. High-dose rifampentine with moxifloxacin for pulmonary tuberculosis. *N Engl J Med* 2014;371:1599–1608.
- Johnson JL, Hadad DJ, Dietze R, Maciel EL, Sewali B, Gitta P, et al. Shortening treatment in adults with noncavitary tuberculosis and 2-month culture conversion. *Am J Respir Crit Care Med* 2009;180:558–563.
- Kleinbaum DG, Klein M. Survival analysis: a self-learning text. Berlin: Springer Science and Business Media, Inc.; 2005.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005;24:3089–3110.
- Imperial MZ, Nahid P, Phillips PPJ, Davies GR, Fielding K, Hanna D, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat Med* 2018;24:1708–1715.
- World Health Organization. Cancer control: knowledge into action: WHO guide for effective programmes. Diagnosis and treatment. Geneva: World Health Organization; 2008.
- Fox W. Whither short-course chemotherapy? *Br J Dis Chest* 1981;75:331–357.
- Strydom N, Gupta SV, Fox WS, Via LE, Bang H, Lee M, 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 Med* 2019;16:e1002773.
- Dorman SE, Savic RM, Goldberg S, Stout JE, Schluger N, Muzanyi G, et al.; Tuberculosis Trials Consortium. Daily rifampentine for treatment of pulmonary tuberculosis: a randomized, dose-ranging trial. *Am J Respir Crit Care Med* 2015;191:333–343.
- Velásquez GE, Brooks MB, Coit JM, Pertinez H, Vargas Vásquez D, Sánchez Garavito E, et al. Efficacy and safety of high-dose rifampin in pulmonary tuberculosis: a randomized controlled trial. *Am J Respir Crit Care Med* 2018;198:657–666.
- 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.
- Bartelink IH, Zhang N, Keizer RJ, Strydom N, Converse PJ, Dooley KE, et al. New paradigm for translational modeling to predict long-term tuberculosis treatment response. *Clin Transl Sci* 2017;10:366–379.
- World Health Organization. Global tuberculosis report 2018. Geneva: World Health Organization; 2019.
- European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Amsterdam: European Medicines Agency; 2020 [accessed 2020 Dec 9]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.
- European Medicines Agency. ICH E9 statistical principles for clinical trials. Amsterdam: European Medicines Agency; 1998.
- 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.
- Phillips PPJ, Mendel CM, Burger DA, Crook AM, Nunn AJ, Dawson R, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Med* 2016;14:19.
- Alene KA, Viney K, Yi H, McBryde ES, Yang K, Bai L, et al. Comparison of the validity of smear and culture conversion as a prognostic marker of treatment outcome in patients with multidrug-resistant tuberculosis. *PLoS One* 2018;13:e0197880.
- Horne DJ, Royce SE, Gooze L, Narita M, Hopewell PC, Nahid P, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. *Lancet Infect Dis* 2010;10:387–394.
- Rockwood N, du Bruyn E, Morris T, Wilkinson RJ. Assessment of treatment response in tuberculosis. *Expert Rev Respir Med* 2016;10:643–654.
- Wallis RS, Kim P, Cole S, Hanna D, Andrade BB, Maeurer M, et al. Tuberculosis biomarkers discovery: developments, needs, and challenges. *Lancet Infect Dis* 2013;13:362–372.
- Blakemore R, Nabeta P, Davidow AL, Vadwai V, Tahirli R, Munsamy V, et al. A multisite assessment of the quantitative capabilities of the Xpert MTB/RIF assay. *Am J Respir Crit Care Med* 2011;184:1076–1084.
- Hanrahan CF, Theron G, Bassett J, Dheda K, Scott L, Stevens W, et al. Xpert MTB/RIF as a measure of sputum bacillary burden: variation by HIV status and immunosuppression. *Am J Respir Crit Care Med* 2014;189:1426–1434.
- Shenai S, Ronacher K, Malherbe S, Stanley K, Kriel M, Winter J, 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.
- Kawasaki M, Echiverri C, Raymond L, Cadena E, Reside E, Gler MT, 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 Med* 2019;16:e1002780.
- Friedrich SO, Rachow A, Saathoff E, Singh K, Mangu CD, Dawson R, et al.; Pan African Consortium for the Evaluation of Anti-tuberculosis Antibiotics (PanACEA). Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. *Lancet Respir Med* 2013;1:462–470.
- World Health Organization. WHO consolidated guidelines on tuberculosis. Module 2: Screening. Systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.