**A Patient-Level Pooled Analysis of Treatment Shortening Regimens for Drug-Susceptible Pulmonary Tuberculosis**

Marjorie Z. Imperial1\*, Payam Nahid1\*, Patrick P. J. Phillips1, Geraint R. Davies2, Katherine Fielding3, Debra Hanna4,5, David Hermann5, Robert S. Wallis6, John L. Johnson7, Christian Lienhardt8, Rada M. Savic, Ph.D.1\*\*

1University of California, San Francisco, USA. 2University of Liverpool, Liverpool, UK. 3London School of Hygiene and Tropical Medicine, London, UK. 4Critical Path Institute, Tuscon, AZ, USA. 5Bill and Melinda Gates Foundation, Seattle, WA, USA. 6Aurum Institute, Johannesburg, South Africa. 7Case Western Reserve University and University Hospitals Cleveland Medical Center, OH, USA. 8Global Programme for Tuberculosis, World Health Organization, Geneva, Switzerland.

\*Co-primary authors

**\*\*Corresponding Author:**

Rada M. Savic, Ph.D.

University of California, San Francisco

1700 4th St, Rm 503C, UCSF Box 2552

San Francisco, CA 94158, United States

Phone: +1 415 502-0640

Email: rada.savic@ucsf.edu

**Abstract**

Tuberculosis kills more people than any other infectious disease. Threepivotal 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, phenotypes that are hard-to-treat, and evaluate the impact of adherence and dosing strategy on outcomes. In 3405 participants included in analyses, baseline smear grade of 3+ relative to <2+, HIV seropositivity and adherence of $\leq $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 with greater precision to improve outcomes, providing a stratified medicine approach as an alternative to the “one-size-fits-all” treatment currently used worldwide.

**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 (OFLOTUB1, ClinicalTrials.gov number, NCT00216385; REMoxTB2, NCT00864383; RIFAQUIN3, 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 under-treatment 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.

**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 modefiied intention to treat (MITT) analyses of the OFLOTUB1, REMoxTB2, and RIFAQUIN3 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 01-0099) includes 193 study participants treated with a 4-month experimental regimen (no fluoroquinolone) and 193 study participants treated with the standard 6-month regimen (Figure1). 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); 12% of the participants were HIV-infected.

**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-monthexperimental regimens and 1311/1404 (93%) participants were allocated to the control regimens (Supplementary Tables 1-3). In participants assigned to 4-monthexperimental 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/m2 decrease; 95% CI 1.1-1.7) (Figure 2a, Supplementary Table 4).

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/m2 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/m2 decrease; 95% CI, 1.0-2.0) remained as factors associated with high risk (Figure 2b, Supplementary Table 5). In the per-protocol analysis, results were similar in the experimental and control groups when compared to the primary modified intent-to-treat analysis (Supplementary Table 6).

**Non-inferiority 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 non-inferiority (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 eastimate 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 non-inferior to the 6-month control regimen (Figure 3a). 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 non-inferior 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 3b).

**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 4a).

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 anlaysis 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 4b, Supplementary Table 7). 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 4c, Supplementary Table 7).

**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 response5,10,11, 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, Supplementary Table 8-10. 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 (Supplementary Figure 1 and Supplementary Figure 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 4a). 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.12,13 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 completion14,15, however, such tools can be limited by issues of scale-up, generalizability and cost. 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.16,17 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.18 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 straitification algorithm can be developed (for example, smear grade and BMI, as shown in Figure 3a) 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,19,20 however, data collection was not standardized across the included trials. 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 (Supplementary Figure 1 and Supplementary Figure 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.

**Acknowledgements:**

The analyses conducted for this study were sponsored by the World Health Organization and the Bill & Melinda Gates Foundation (A127340). Partial support for PN and PPJP was also received from NIH NIAID (1R01AI104589). The authors alone are responsible for the views expressed in this publication, and they do not necessarily represent the decisions, policy, or views of the WHO. Data used in the preparation of this article were obtained from Platform for Aggregations of Clinical TB Studies (TB-PACTS; https://c-path.org/programs/tb-pacts/). The Platform for Aggregations of Clinical TB Studies is a public­ partnership launched in May 2015 by Critical Path Institute (C‑Path), the Special Programme for Research and Training in Tropical Diseases (TDR), the Global Alliance for TB Drug Development (TB Alliance), and St. George’s, University of London.

We thank the study participants; the staff at clinical sites; all investigators for each of the four trials; the data management team at Critical Path Institute (K. Romero, B. Stafford, and D. Hartley); C. Mendel, J. Neaton, and A. Vernon for their review of the manuscript; and the TBReFLECT Steering Committee.

**Author contributions:**  M.Z.I and R.M.S designed and performed the analysis. M.Z.I, P.N., P.P.J.P and R.M.S interpreted the results and prepared the manuscript. D.Ha provided OFLOTUB, REMoxTB, and RIFAQUIN data. J.L.J provided DMID 01-009 data. P.P.J.P and K.F provided data management and statistical support. All authors discussed the results and implications and contributed to the final manuscript.

**Competing Interests:** The authors declare no competing risks.

**References:**

1. Merle, C. S. *et al.* A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N. Engl. J. Med* **371,** 1588–1598 (2014).

2. Gillespie, S. H. *et al.* Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N. Engl. J. Med.* **371,** 1577–1587 (2014).

3. Jindani, A. *et al.* High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N. Engl. J. Med.* **371,** 1599–1608 (2014).

4. Geneva: World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. *Licence CC BY-NC\_SA 3.0 IGO* (2017).

5. Nahid, P. *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.* **63,** e147–e195 (2016).

6. Jo, K.-W. *et al.* Risk factors for 1-year relapse of pulmonary tuberculosis treated with a 6-month daily regimen. *Respir. Med.* **108,** 654–659 (2014).

7. Benator, D. *et al.* Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial. *Lancet (London, England)* **360,** 528–34 (2002).

8. TBTC Study 31: Rifapentine-containing Tuberculosis Treatment Shortening Regimens. Available at: https://clinicaltrials.gov/ct2/show/NCT02410772.

9. Johnson, J. L. *et al.* Shortening Treatment in Adults with Noncavitary Tuberculosis and 2-Month Culture Conversion. *Am J Respir Crit Care Med* **180,** 558–563 (2009).

10. Savic, R. M. *et al.* Defining the optimal dose of rifapentine for pulmonary tuberculosis: Exposure-response relations from two phase II clinical trials. *Clin. Pharmacol. Ther.* **102,** 321–331 (2017).

11. Alipanah, N. *et al.* Treatment of non-cavitary pulmonary tuberculosis with shortened fluoroquinolone-based regimens: a meta-analysis. *Int. J. Tuberc. Lung Dis.* **20,** 1522–1528 (2016).

12. Vernon, A. A. & Iademarco, M. F. In the Treatment of Tuberculosis, You Get What You Pay for…. *Am. J. Respir. Crit. Care Med.* **170,** 1040–1042 (2004).

13. Chang, K. C., Leung, C. C., Yew, W. W., Ho, S. C. & Tam, C. M. A Nested Case–Control Study on Treatment-related Risk Factors for Early Relapse of Tuberculosis. *Am. J. Respir. Crit. Care Med.* **170,** 1124–1130 (2004).

14. Ngwatu, B. K. *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* **51,** (2018).

15. DiStefano, M. J. & Schmidt, H. mHealth for Tuberculosis Treatment Adherence: A Framework to Guide Ethical Planning, Implementation, and Evaluation. *Glob. Heal. Sci. Pract.* **4,** 211–221 (2016).

16. Fox, W., Ellard, G. A. & Mitchison, D. A. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *Int. J. Tuberc. Lung Dis.* **3,** S231-79 (1999).

17. Fox, W. Whither short-course chemotherapy? *Br. J. Dis. Chest* **75,** 331–57 (1981).

18. Kendall, E. A. *et al.* Priority-Setting for Novel Drug Regimens to Treat Tuberculosis: An Epidemiologic Model. *PLOS Med.* **14,** e1002202 (2017).

19. CDISC-Tuberculosis Therapeutic Area User Guide v2.0. (2016). Available at: https://www.cdisc.org/standards/therapeutic-areas/tuberculosis.

20. Taichman, D. B. *et al.* Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors. *N. Engl. J. Med.* **374,** 384–386 (2016).

**Figure legends**

**Figure 1 Analysis and Validation Populations**

Individual patient data from three trials were pooled for analysis. The original results were published in Merle et al 2014 (OFLOTUB), Gillespie et al. 2014 (REMoxTB), and Jindani et al 2014 (RIFAQUIN). Data from a fourth trial, DMID 01-009, was used for external validation and previously published in Johnson et al. 2009. The modified intent-to-treat population was used for the analysis. \*For the validation dataset, the time to event analysis population in the original publication was used. †REMoxTB consisted of two 4-month experimental groups.

**Figure 2. Multivariate hazard ratios for unfavorable outcomes**

**a.** Multivariate analysis for experimental group with (top panel) baseline predictors and (bottom panel) baseline and on-treatment predictors. **b.** Multivariate analysis for control group with (top panel) baseline predictors and (bottom panel) baseline and on-treatment predictors. All analyses were adjusted for country and effect sizes are available in Supplementary Table 4 and Supplementary Table 5. Hazard ratios with 95% Wald confidence intervals are reported. The size of the square denotes the relative sample size according to variable.

\* Age <$ $30 years, 179/916 (20%) unfavorable outcomes and Age $\geq $ 30 years, 237/927 (26%) unfavorable outcomes. † Age <$ $30 years, 136/830 (16%) unfavorable outcomes and Age $\geq $ 30 years, 181/838 (22%) unfavorable outcomes; BMI $\geq $ 17 kg/m2, 226/1247 (18%) unfavorable outcomes and BMI < 17 kg/m2, 91/421 (22%) unfavorable outcomes. ∫ Age <$ $30 years, 92/657 (14%) unfavorable outcomes and Age $\geq $ 30 years, 121/654 (19%) unfavorable outcomes; BMI $\geq $ 17 kg/m2, 156/989 (16%) unfavorable outcomes and BMI < 17 kg/m2, 57/322 (18%) unfavorable outcomes. ¶ BMI $\geq $ 17 kg/m2, 102/901 (11%) unfavorable outcomes and BMI < 17 kg/m2, 36/285 (13%) unfavorable outcomes.

**Figure 3. Difference in proprotion 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.

**Figure 4. Analysis of 7/7 and 6/7 dosing strategies and impact of adherence in the control group.**

1. 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 size for country are available in Supplementary Table 7. In panels b and c, the hazard ratios with 95% Wald confidence interval are reported. \*Hazard ratio with 95% Wald confidence interval for 6/7 dosing strategy relative to 7/7 dosing strategy for fully adherent population after adjustment for county and treatment duration. †Treatment Duration < 182 days, 21/110 (19%) unfavorable outcomes and Treatment Duration ≥ 182 days, 40/577 (7%) unfavorable outcomes. ∫ Treatment Duration < 169 days, 21/155 (14%) unfavorable outcomes and Treatment Duration ≥ 169 days, 42/443 (9%) unfavorable outcomes.

**Table 1. Baseline Characteristics of Study Participants in the Modified Intent-to-Treat Analysis**

|  |
| --- |
| **Table 1. Baseline Characteristics of Study Participants in the Modified Intent-to-Treat Analysis** |
|  |  | **Analysis Dataset** **(OFLOTUB, REMoxTB, RIFAQUIN)** | **Validation Dataset****(DMID01-009)** |
|   |   | **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. (%)\*  |   |   |  |  |
|   | Black or African American  | 1326 (66) | 1066 (76) | - | - |
|   | Asian  | 349 (17) | 178 (13) | - | - |
|   | Other  | 326 (16) | 160 (11) | - | - |
| Age- yrs†  |   |   |  |  |
|   | 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∫ |   |   |  |  |
|   | 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. (%)¶ | 248 (12) | 220 (16) | 0 (0) | 0 (0) |
|  | CD4 cell count ‖ |   |   |  |  |
|   | Median | 363 | 317 | - | - |
|   | Interquartile range  |  265-493 | 241-444 | - | - |
|  | $\leq $ 300 - no. (%)  | 74 | 81 | - | - |
|  | > 300 - no (%)  | 135 | 99 | - | - |
| Cavitation- no. (%)\*\* | 1247 (62) | 847 (60) | 0 (0) | 0 (0) |
| Smear- no. (%)†† |   |   |  |  |
|   | 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) |
| \* Race was missing for all OFLOTUB study participants, black race was assigned to all study participants given all OFLOTUB sites were in Africa. †Age was missing for 5 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 height for females and males were used to calculate body mass index. ¶ Human immunodeficiency virus (HIV) status was missing for 9 study participants.‖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.\*\* Cavitation 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. |

**Methods**

**Study Design**

This study utilized individual patient data from four recent, international, randomized phase 3 trials (OFLOTUB, ReMOXTB, RIFAQUIN and DMID 01-009)1–3,9 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 DMID01-009 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 committies 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 Supplementary Table 1 and Supplementary Table 2, and access to original databases is available through TB-PACTS. Data from the DMID 01-009 were obtained directly from the sponsor. Further information on data acquisition and availability is described in the Life Sciences Reporting Summary.

**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 REMox 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 Supplementary Table 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. Sensitivity analyses were performed to evaluate the inclusion of all MIRU-confirmed reinfections, classified as unfavorable (Supplementary Table 11) or favorable, or completely removed from the analysis. 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 (Supplementary Table 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, Supplementary Figure 3). 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 Supplementary Table 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%, Supplementary Table 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 (Supplemental Table 12). 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 (summary on analysis populations available in Supplemental Table 3). There were no major correlations between predefined set of baseline and on-treatment predictors (Supplementary Figure 3)

**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 summarized in Supplementary Table 6). 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.21 The absolute difference in proportion of unfavorable outcomes was calculated using inverse probability study weighted Kaplan-Meier estimates22 at 24 months after start of treatment to include maximal patient-years of follow up and retain maximal data. 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 available through the Platform for Aggregation of Clinical TB Studies (TB-PACTS, https://c-path.org/programs/tb-pacts/). DMID-009 trial data is available from the sponsor (John L. Johnson, Tuberculosis Research Unit).

**References**

21. Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J. & Drazen, J. M. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N. Engl. J. Med* **357,** 2189–2194 (2007).

22. Xie, J. & Liu, C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat. Med.* **24,** 3089–3110 (2005).