

Improving treatment outcome assessment in a mouse tuberculosis model

Bas C. Mourik¹, Robin J. Svensson², Gerjo J. de Knecht¹, Hannelore I. Bax³, Annelies Verbon³, Ulrika S.H. Simonsson², Jurriaan E.M. de Steenwinkel¹

¹ *Dept. of Medical Microbiology & Infectious Diseases, Erasmus MC, Rotterdam, the Netherlands*

² *Dept. of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*

³ *Dept. of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands*

ABSTRACT

Preclinical treatment outcome evaluation of tuberculosis (TB) occurs primarily in mice. Current designs compare relapse rates of different regimens at selected time points, but lack information about the correlation between treatment length and treatment outcome, which is required to efficiently estimate a regimens' treatment-shortening potential. Therefore we developed a new approach. BALB/c mice were infected with a *Mycobacterium tuberculosis* Beijing genotype strain and were treated with rifapentine-pyrazinamide-isoniazid-ethambutol (R_pZHE), rifampicin-pyrazinamide-moxifloxacin-ethambutol (RZME) or rifampicin-pyrazinamide-moxifloxacin-isoniazid (RZMH). Treatment outcome was assessed in n=3 mice after 9 different treatment lengths between 2-6 months. Next, we created a mathematical model that best fitted the observational data and used this for inter-regimen comparison. The observed data were best described by a sigmoidal E_{max} model in favor over linear or conventional E_{max} models. Estimating regimen-specific parameters showed significantly higher curative potentials for RZME and R_pZHE compared to RZMH. In conclusion, we provide a new design for treatment outcome evaluation in a mouse TB model, which (i) provides accurate tools for assessment of the relationship between treatment length and predicted cure, (ii) allows for efficient comparison between regimens and (iii) adheres to the reduction and refinement principles of laboratory animal use.

Published in Nature Scientific Reports

2018, 5714. doi:10.1038/s41598-018-24067-x

INTRODUCTION

Tuberculosis (TB) claimed 1.7 million lives in 2016, which is more than any other infectious disease caused by a single pathogen (1). The global treatment success rate for drug-susceptible TB is 83%, which falls short of the $\geq 90\%$ target rate set by the WHO (1). Moreover, treatment success rates against multi-drug resistant (MDR; 52%) and extensively drug resistant (XDR; 28%) TB are markedly lower (1). These rates emphasize the need for more effective anti-TB drug regimens that can improve treatment success. In addition, new anti-TB regimens should allow for shortening of the current 6-months treatment length to increase compliance and minimize further drug resistance development.

Recently, a large clinical Phase III trial failed to reduce anti-TB treatment length from six to four months by substituting conventional anti-TB drugs with moxifloxacin (2). This trial was conducted based on promising results from clinical Phase IIa/b trials and preclinical experiments in mouse TB models (3-5). Overall, this has led to the conclusion that early surrogates for treatment efficacy assessments as measured in clinical phase IIa/b trials are unreliable predictors for cure in TB (6, 7). This has further inspired efforts to improve preclinical mouse TB models aimed at evaluating treatment outcomes (8).

Preclinical evaluation of TB treatment outcome occurs primarily in mouse models (9). The conventional design involves a two-step approach. During the first step, early treatment efficacy is measured by determining mycobacterial load reductions in the lungs of small groups of mice ($n=3-5$) at preset time points until culture conversion is reached (10-15). In the second step, relapse of infection is evaluated for regimens that resulted in culture negativity. This occurs by determining lung culture status three months after treatment has ended in larger groups of mice ($n=12-30$), after 1-3 selected treatment lengths) (10-15).

This conventional design seems to have several drawbacks. Most importantly, it allows for relapse rate comparison between regimens at selected time points, but does not provide an individual regimen's correlation between treatment length and treatment outcome. This correlation is required to efficiently estimate a regimen's treatment-shortening potential. The conventional design also has limited screening potential for regimens with unknown efficacy, as prior knowledge on when a regimen will reach culture-conversion is required before relapse can be evaluated. Lastly, recent clinical and preclinical observations suggest that early treatment efficacy assessment as measured in step one of the conventional design has limited predictive value for treatment outcome after a full course of anti-TB treatment (6, 7, 16).

In the current study we propose an alternative design for treatment outcome assessment in our mouse TB model. We increase the number of treatment schedules assessing outcome three months after the end of treatment regardless of culture status at the end of treatment, but decrease the number of mice per treatment length ($n=3$ instead of $n=12-30$). This way of data collection allows for mathematical modeling of the observational data optimized for establishing a robust and informative link between treatment length and cure.

The mathematical modeling is based on conventional logistic regression, but is designed to be more informative. This approach differs from survival-, or time-to-event analysis, because the bacterial burden is determined after a fixed period of time after stop of treatment. Therefore the time of culture-conversion relative to stop of treatment is unknown.

In silico simulations of the mathematical model can be used to visualize and accurately quantify the association between treatment length and predicted treatment outcome for each regimen. Advantages include the possibility to compare the curative potential of different anti-TB regimens with each other over time instead of at selected time points only and simultaneously assess the treatment-shortening potential of each individual regimen.

MATERIAL AND METHODS

Mice, infection and mycobacterial strain

Specified pathogen-free female BALB/c mice, aged 13-15 weeks, were infected by intratracheal installation of $1.0-1.8 \times 10^5$ drug-susceptible *Mycobacterium tuberculosis* Beijing VN 2002-1585 (BE-1585) under general anesthesia as described previously (16, 17). The mice were housed and experiments were conducted in the Erasmus MC animal biosafety level III facility.

Ethical approval

All protocols were approved by the Erasmus MC animal ethics committee under DEC number 117-12-13 and EMC number 2887, and were in accordance with the rules laid down in the Dutch Animal Experimentation Act and the EU Animal Directive 201/63/EU.

Treatment

Treatment consisted of either of three regimens: (i) rifapentine, pyrazinamide, isoniazid and ethambutol (R_pZHE), (ii) rifampicin, pyrazinamide, moxifloxacin and ethambutol

(RZME) or (iii) rifampicin, pyrazinamide, moxifloxacin and isoniazid (RZMH). The first two months of each regimen consisted of treatment with all four drugs (intensive phase) followed by four months of treatment with rifapentine and isoniazid for the R_pZHE regimen, rifampicin and moxifloxacin for RZME and rifampicin, moxifloxacin and isoniazid for RZMH. All drugs were administered 5 days a week via oral gavage in their human pharmacokinetic equivalent dose: rifampicin: 10 mg/kg, rifapentine: 10 mg/kg, moxifloxacin: 200 mg/kg, isoniazid: 25 mg/kg, ethambutol: 100 mg/kg, pyrazinamide: 150 mg/kg (18, 19).

Treatment outcome evaluation

Treatment was initiated 2 weeks after infection and was stopped between 2 and 6 months with intervals of 2 weeks (i.e. nine different treatment lengths per drug regimen). The protocol was designed to include three (n=3) mice per treatment length. A sample size of n=3 was found to be sufficient to detect a 50% difference in potency between different treatments and was expected to give reasonably high precision in model parameters, according to a statistical power calculation (described in **supplementary data file S1**).

One 'backup' mouse was added per regimen to reduce the impact of unexpected animal loss. All mice were sacrificed 3 months post-treatment to determine mycobacterial loads in the lungs as described previously (16).

Statistical analysis

The statistical analysis involved the development of a logistic regression model based on the observational data. These data were treated as a binary outcome variable of either cure (defined as a negative solid culture 3 months post-treatment) or failure (defined as a positive solid culture 3 months post-treatment). The independent variable was treatment length. The data were analyzed using the non-linear regression software NONMEM (version 7.3) with simultaneous estimation of all model parameters (20). Only if parameters were estimated close to a parameter boundary (as described below) they were fixed to the value of the respective boundary. NONMEM maximizes the likelihood of a model to fit the observational data. In NONMEM the model fit (defined as the likelihood of the model to describe the observational data) was assessed using the objective function value (OFV), which is equal to -2 times the log value of the likelihood. In order to generate a model that best described (fitted) the data, the OFV between models was compared using the likelihood ratio test (LRT). To this aim, for each model comparison a reduced model and a full model were evaluated where the full model always included more model parameters than the reduced model. The null hypothesis was that the full model did not provide better fit than the reduced model. Testing was performed at the

5% significance level which corresponds to a drop in the OFV of at least 3.84 points with one degree of freedom. Data handling and graphical analysis were conducted in R (version 3.3.0) (21).

The model development was divided into two parts; in the first part, an appropriate relation between probability of cure and treatment length was identified (regardless of drug regimen). In the second part we explored if this relation between probability of cure and treatment length was significantly different between the drug regimens.

The starting point for the first part of model development was a base model which assumed that the probability of cure was identical regardless of treatment lengths according to:

Equation 1: $p_{failure} = 1 - p_{cure} = p_{base}$

In this model, $p_{failure}$ and p_{cure} are the predicted probabilities of failure and cure respectively and p_{base} is the base probability of failure. The p_{base} parameter was constrained to be between 0 and 1. First, this base model was compared to a model assuming linear increase in cure rate with treatment length according to:

Equation 2: $p_{failure} = 1 - p_{cure} = p_{base} \times (1 + Slope \times T)$

In this model 'Slope' is the linear increase in probability of cure with treatment length (T). The 'Slope' parameter was constrained to be between 0 and 1 divided by the maximum treatment duration of 6 months. Secondly, an E_{max} model was tested according to:

Equation 3: $p_{failure} = 1 - p_{cure} = p_{base} \times \left(1 - \frac{E_{max} \times T}{T_{50} + T}\right)$

In this model ' E_{max} ' is the maximal achievable probability of cure and ' T_{50} ' is the treatment length at which half the E_{max} is seen. The E_{max} parameter was constrained to be between 0 and 1. Lastly, a sigmoidal E_{max} model was tested according to:

Equation 4: $p_{failure} = 1 - p_{cure} = p_{base} \times \left(1 - \frac{E_{max} \times T^y}{T_{50} + T^y}\right)$

In this model 'y' is a shape parameter controlling the steepness of the curve produced by the E_{max} equation.

In the second part of model development, we explored if the identified relation between cure and treatment length from the first part of model development was significantly different for the different drug regimens by comparing the model parameters of the different drug regimens (Slope, E_{\max} , T_{50} or γ , depending on the model). This was done in a step-wise approach, here exemplified for a sigmoidal E_{\max} model, which includes the three model parameters E_{\max} , T_{50} and γ . Firstly, one model was fitted to explore if E_{\max} for each regimen was significantly different from the other two regimens. This procedure was repeated for the T_{50} and γ parameters, thus resulting in nine different models. Secondly, the models that did not significantly improve the fit (OFV drop of less than 3.84 points) were not evaluated further. Of the remaining models that did result in an OFV drop of at least 3.84 points, the model with the lowest OFV was accepted. Thirdly, the accepted model with the greatest drop in OFV was combined with the remaining models that also improved the fit significantly (i.e. whose OFV drop was lesser than the accepted model but at least 3.84 points). If this combination improved the fit significantly, it was accepted as the new model. This whole three-step procedure was repeated until no significant improvement was seen anymore, which was defined as the final model.

In addition to assessment of OFV, model selection was guided by parameter uncertainty and visual predictive checks (VPC) generated using PsN (<http://psn.sourceforge.net/> [cited 19-12-2016]) and Xpose (<http://xpose.sourceforge.net/> [cited 19-12-2016]) using 1000 simulated datasets. The VPC is a visual diagnostic which shows how well data simulated from a model agree with the observed data.

Simulations

The observational data using $n=3$ animals can only theoretically generate cure rates of 0%, 33%, 67% or 100%. Therefore, we used the mathematical model to simulate treatment outcome from 1000 mice per time point to increase the resolution in the predicted cure rates (i.e. to allow cure rate to continuously range between 0-100%). Simulations were performed using Monte Carlo sampling from a random uniform distribution ranging from 1 to 0. This was also used to determine the model-predicted treatment length required for each regimen to achieve 85%, 90% or 95% cure, respectively.

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary data files).

RESULTS

Observed treatment outcome

A schematic overview of our method of data collection compared to the conventional design is found in **Figure 1**. The observed proportions of cured animals for the different treatment lengths for the R_pZHE, RZME and RZMH regimens are shown in **Table 1**. R_pZHE started to show cure rates above 0% after 2.5 months of treatment and showed 100% cure after 4 months. RZME displayed similar kinetics and also showed complete cure rates from 4 months of treatment onwards. In contrast, RZMH only started to show cure rates above 0% after 4 months of treatment and did not reach complete cure even after 6 months of treatment.

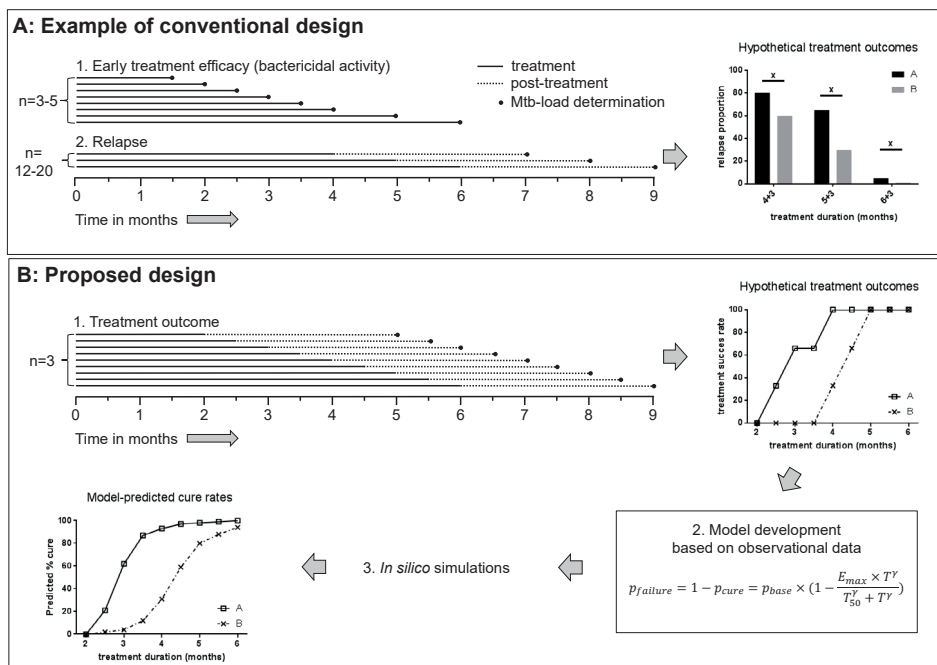


Figure 1. Schematic examples of the conventional design and the proposed design for treatment outcome evaluation in mouse TB models

A) Shows an example of the conventional design in which bactericidal activity is determined by measuring reductions in Mtb-loads in the lungs until culture negativity is reached, followed by cross-sectional evaluation / comparison of relapse rates 3 months post-treatment (x in upper right figure).

B) Shows our proposed design in which treatment outcome is determined regardless of lung culture-status at stop of treatment. This allows for more informative mathematical modeling of the data and subsequent simulation of large numbers of mice to generate a high resolution correlation between treatment duration and treatment success.

Model development

Part I: Relation between treatment length and probability of cure

The observational data from **Table 1** were first converted into a dataset used for modeling (**supplementary data file S2**). Compared to the base model which does not assume any relationship between cure and treatment length, a linear relationship between treatment length and cure gave a significant improvement in model fit compared to the base model ($p < 0.001$, OFV drop of 32.2 points). An E_{\max} relationship between treatment length and cure did not improve model fit compared to a linear relationship (OFV increased with 15.9 points) and was rejected. However, a sigmoidal E_{\max} relationship improved model fit significantly compared to a linear relationship between treatment length and probability of cure ($p = 0.001$, OFV drop of 13.3 points). Thus, the sigmoidal E_{\max} model was identified as appropriate and was brought forward to the second part of model development.

Notably, the baseline probability (p_{base}) in this sigmoidal E_{\max} model was estimated very close to 1, which resulted in an unstable model (not possible to obtain any parameter uncertainty). Fixing p_{base} to 1 could correct for this without affecting the OFV.

Table 1. Observational data on cure

Treatment length	R _p ZHE	RZME	RZMH
2 months	0/3	0/3 ^a	0/3
2.5 months	1/3	0/3	0/3
3 months	2/3	3/3	0/3
3.5 months	2/3	2/3	0/3
4 months	3/3	2/2 ^b	1/3
4.5 months	2/2 ^c	3/3	2/3
5 months	3/3	3/3	3/3
5.5 months	3/3	2/2 ^b	3/3
6 months	3/3	3/3	3/4 ^d

^a 0/3 = 0 of 3 mice was cured (culture-negative lungs 3 months post-treatment) after indicated treatment duration; ^b Animal died of a non-tuberculosis cause prior to time point; ^c The plates for colony counting were contaminated and no counting could be performed; ^d The backup mouse included for the RZMH regimen was still alive at the 6 month time point. R=rifampicin, R_p= rifapentine, Z= pyrazinamide, M= moxifloxacin, H=isoniazid, E= Ethambutol

Part II: inter-regimen differences

The generated sigmoidal E_{\max} relationship in part 1 of model development assumed a similar relationship between treatment length and probability of cure for all three regimens tested. In order to detect if the selected relationship deviated significantly between the different regimens, we determined if implementing drug regimen-specific

model parameters (including γ , E_{\max} and T_{50}) improved model fit and could detect significant differences between the different regimens.

Initially, the following models improved the model fit to the observational data significantly: Model 1: separate T_{50} for RZMH ($p < 0.001$, OFV drop of 18.6 points), Model 2: separate E_{\max} for RZMH ($p = 0.00298$, OFV drop of 8.82 points), Model 3: separate γ for RZMH ($p = 0.0408$, OFV drop of 4.18 points) and Model 4: separate T_{50} for RpZHE ($p = 0.0377$, OFV drop of 4.32 points). Model 1 (separate T_{50} for the RZMH regimen) had the lowest OFV and was therefore accepted. When combined with model 2-4, no significant improvements were observed. Therefore only model 1, which included simultaneous estimation of separate T_{50} parameters for RZMH only and for RZME and RpZHE, respectively was selected.

Notably, The E_{\max} parameter for the sigmoidal E_{\max} model with a separate T_{50} for the RZMH regimen was estimated very close to 1 which also resulted in an unstable model. Fixing E_{\max} to 1 could correct for this and improved the model fit slightly ($p = 0.827$, OFV drop of 0.048 points).

Taken together, the final model included a sigmoidal E_{\max} relationship where the probability of cure increased with treatment length. The E_{\max} parameter had the value of 1 which implies that all included regimens can achieve 100% cure if the treatment length is sufficiently long. The baseline probability (p_{base}) also had a value of 1 which implies that at very short or no treatment duration at all (i.e. $T = 0$) treatment failure will occur in all mice. Apart from p_{base} and E_{\max} all parameters were simultaneously estimated.

Importantly, our finding that estimating a separate T_{50} for the RZMH regimen significantly improved our model fit indicates that RZMH has reduced curative potential compared to the other regimens. The final model parameters are shown in **Table 2**. The final model code is supplied in **supplementary data file S3**.

Model validation

To verify the model, a visual predictive check (VPC) was performed where the observational data and simulated data (presented as 95% confidence interval based on 1000 simulated datasets) were compared in the same plot (**Figure 2**). As can be seen in the VPC, the observed proportions of cure fell within the 95% confidence of the simulated data. The confidence intervals may appear large at some time points which is due to the relatively low number of animals per time point and thus, given the data, the model can describe the observed data well.

Table 2. Final parameter estimates

Parameter	Description	Parameter estimate	Standard error (%CV) ^a
p_{base}	Baseline probability of no cure	1 FIX	-
E_{max}	Maximum achievable probability of cure	1 FIX	-
$T50_{R_pZHE/RZME}$ (months) ^b	The treatment time at which half the E_{max} is reached for R_pZHE and $RZME$	2.87	5.4
$T50_{RZMH}$ (months) ^b	The treatment time at which half the E_{max} is reached for $RZMH$	4.35	6.0
γ	Shape factor	9.82	23.0

R_p = rifapentine, Z = pyrazinamide, M = moxifloxacin, H =isoniazid, E = Ethambutol; CV coefficient of variance; ^a The standard errors were calculated using the covariance step in NONMEM; ^b $T50$ was significantly different between treatment arms (no statistically significant differences were found in other parameters)

Part I: Cure rate predictions based on model simulation

Simulations of high numbers of mice ($n=1000$ per arm) using the developed model enabled us to provide a high-resolution estimate of the predicted cure rates of each regimen for different treatment lengths as shown in **Figure 3**. For R_pZHE and $RZME$ this estimates that mice must be treated at least 3.5 months to reach 85% cure and 4 months to reach 90% or 95% cure. In contrast, mice must be treated with $RZMH$ for at least 5.5 months to reach 85% or 90% cure and a full 6 months to reach 95% cure.

Part II: Model comparison to conventional relapse assessment other mouse TB models

Next, we aimed to evaluate if the predicted cure rates generated in our model were comparable to observational data obtained from other mouse TB models using pulmonary infection. These data are shown in **Table 3** (5, 12, 13, 22, 23). A direct advantage of our model-based approach is the possibility to compare our predicted cure rates for any treatment length evaluated in other mouse TB models (**Fig. 3** and **Table 3**).

Our model predicted a cure rate of 60% for $RZME$ after three months of treatment. This was lower than the observed cure rates from three other mouse TB models, which were 80-100% (**Table 3**: models 2, 4, 5), but higher than the observed cure rate of 40% in model 6 (12). After four months of treatment no data on $RZME$ was available in other mouse TB models and we could only compare our data to the RZM regimen. A 97% predicted cure rate for $RZME$ in our model showed similar cure rates as observed for RZM in models 1 and 4 and a 13% higher cure rate than RZM in model 3 (**Table 3**). After five months of treatment, results were similar compared to one other mouse TB model, which also showed 100% cure.

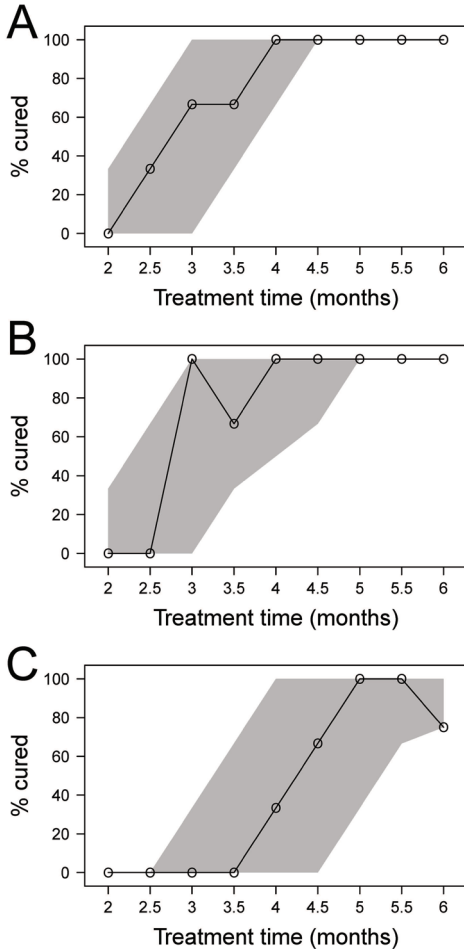


Figure 2. Visual predictive check (VPC) of the final model for each regimen

A) rifapentine, pyrazinamide, isoniazid and ethambutol (R_pZHE), **B)** rifampicin, pyrazinamide, moxifloxacin and ethambutol (RZME) and **C)** rifampicin, pyrazinamide, moxifloxacin and isoniazid (RZMH). The open circles connected by the solid black lines are the observed probabilities of cure following different treatment lengths and the shaded areas are the 95% non-parametric confidence interval of the predicted cure rates following different treatment lengths.

For RZMH, our predicted cure rate of 2% after three months of treatment was lower compared to the observed cure rates of 27%, 93% and 80% in models 2, 4 and 5, respectively, but higher than the 0% cure observed in model 6 (**Table 3**). After four months, RZMH in our model could only be compared to RZM in other mouse TB models. Our predicted cure rate of 29% for RZMH at this point was markedly lower than the cure rates observed for RZM of 100-95%, 84% and 95% in models 1, 3 and 4 respectively. In this regard it is of note to mention that after three months of treatment RZMH also showed inferior results compared to RZM in mouse TB models 1 and 4 and inferior results compared to RZME in mouse TB models 2,4 and 6 (**Table 3**).

Taken together, our finding that RZMH has significantly lower curative potential compared to RZME is reflected in trends observed in other mouse TB models. Moreover,

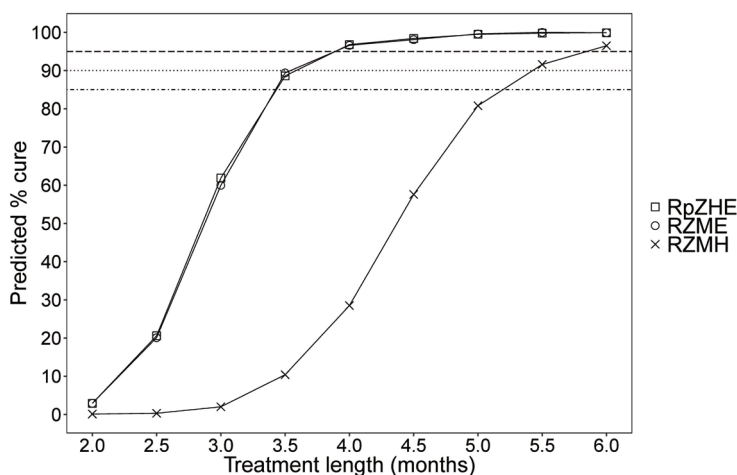


Figure 3. Model-predicted cure at different treatment lengths for each regimen

The black horizontal lines indicate 95% (dashed line), 90% (dotted line) and 85% (dashed-dotted line) cure rates. R=rifampicin, R_p= rifapentine, Z= pyrazinamide, M= moxifloxacin, H=isoniazid, E= Ethambutol

Table 3. Comparison of our model-based predictions of cure rates with observational data

	Regimen	% cured at:			Ref.
		3 months	4 months	5 months	
Our model BALB/c, Beijing, HDIT	2 R _p ZHE/ 1,2,3 R _p H ^a	62%	97%	100%	
	2 RZME/ 1,2,3 RM ^a	60%	97%	100%	
	2 RZMH/ 1,2,3 RMH ^a	2%	29%	81%	
Model 1 BALB/c, H37Rv, HDA	2 R _p ZM/ 1,2 R _p M	100%			22
	3,4,5 RZM	75%	100%	100%	5
	2 RZM/ 1,2,3 RM	83%	100%	100%	5
Model 2 BALB/c, H37Rv, LDA	2 RZM/ 2,3 RM		95%	100%	23
	2 RZME/ 1 RM	80%			12
Model 3 BALB/c, Erdman, HDA	2 RZMH/ 1 RMH	27%			12
	2 RZM/ 2 RM		84%		13
Model 4 BALB/c, Erdman, LDA	3 RZME	100%			12
	3 RZMH	93%			12
	2 RZM/ 1 RM, 2 RM	100%	95%		13
Model 5 C3HeB/FeJ ^b , H37Rv, LDA	2 RZME/ 1 RM	80%			12
	2 RZMH/ 1 RMH	80%			12
Model 6 C3HeB/FeJ ^b , Erdman, LDA	3 RZME	40%			12
	3 RZMH	0%			12

^a Predicted cure for 3,4 and 5 months of treatment is shown as estimated in **Figure 3**, ^b C3HeB/FeJ mice can develop cavitating lesions that more closely resemble human disease, Abbreviations for route of infection: HDIT = high dose intratracheally, HDA = high dose aerosol, LDA = low dose aerosol, R=rifampicin, R_p= rifapentine, Z= pyrazinamide, M= moxifloxacin, H=isoniazid, E= Ethambutol

the discrepancy between our predicted cure rates for RZMH compared to RZM in other mouse TB models, suggests a negative effect of H on the efficacy of RZM in mouse TB models.

DISCUSSION

In this study we demonstrated that a model-based analysis of observational *in vivo* data on TB treatment outcomes can be used to generate a high resolution association between treatment length and probability of cure. The developed model could detect statistically significant differences in the curative potentials of R_pZHE and RZME compared to RZMH, which could not have been identified based on the observational data alone. Validation of our model against other mouse TB models supported a negative effect of isoniazid on the efficacy of RZM in mouse studies.

In our model RZMH showed significantly reduced curative potential compared to RZME. Interestingly, a similar trend was observed in other mouse TB models where RZMH consistently showed a trend towards inferior results compared to RZME and/or RZM (12). One explanation for this phenomenon might be a species-dependent, antagonistic effect of isoniazid on the therapeutic efficacy of rifampicin. Rifampicin is more essential for cure than isoniazid in mice (16). It has been demonstrated that concomitant administration of isoniazid negatively affects the pharmacokinetics of rifampicin by lowering the highest observed plasma concentration (C_{max}) and area under the plasma concentration-time curve (AUC) (24). However, pharmacokinetics is an unlikely cause in our model as isoniazid co-administration previously did not affect rifampicin C_{max} and AUC compared to rifampicin monotherapy (25). Also in patients no clinically significant pharmacokinetic interactions between isoniazid and rifampicin have been reported (26). Nevertheless, addition of isoniazid (H) to the combination of rifampicin and pyrazinamide (RZ) significantly reduced bactericidal activity and cure in other mouse TB models (24, 27, 28). In addition, an earlier study in our mouse TB model showed that RZ-treated mice had higher cure rates than mice treated with RH or RHZ after a six-months treatment course (95% vs. 87% and 80%, respectively) (16). This previous comparison between RZ, RH and RHZ using the conventional design as shown in **Figure 1** did not yield significant differences, but the observed inferiority of RZMH compared to RZME in the current study supports earlier observations of an antagonistic effect of isoniazid on the therapeutic efficacy of rifampicin in mice.

Advantages of the combination of animal research and mathematical modeling are the ability to detect significant differences in the curative potential of different regimens,

and the ability to compare our data with other studies that evaluated treatment outcome after any given treatment length as demonstrated in **Table 3**. In addition, animal research experiments should always strive towards the 3R-principles of replacement, reduction and refinement (29). Our method adheres to the reduction and refinement principles. Firstly, the implementation of mathematical modeling and simulations can be considered a refinement as it enabled us to detect significant differences between regimens and allowed efficient comparison with other mouse TB models, which could not be derived from our observational data alone. Secondly, our approach enables assessment of treatment outcome without requiring early treatment efficacy data. This reduces the total number of mice required (**Fig. 1**).

Early treatment efficacy as measured through bactericidal activity might be of limited predictive value for treatment outcome in TB (16). However, it remains an important screening tool in the setting of early drug discovery. The similar principle of observational data and mathematical modeling can be applied to bactericidal activity experiments as well using 'culture negativity' as outcome parameter in order to improve data interpretation.

One initial concern with the proposed design was that with only $n=3$ mice per time point, the treatment outcome in a single mouse on a crucial time point might have a disproportional impact, e.g. if in the RZMH group 4/4 mice would be cured after 6 months or if only 2/3 mice would be cured in the RZME group after 6 months (**Table 1**). However, sensitivity analysis of such scenarios did not alter the conclusions based on the model (results not shown). This can be explained by the notion that the fit of a model involves all mice evaluated at all time points and thus reduces the impact of potential outliers at a single time point.

A common method to analyze binary data is standard logistic regression but in this work we applied a new alternative to standard logistic regression. The main advantage with our new method is that it is more widely applicable than standard logistic regression. Observational data may not always behave similar to a logistic curve and in such situation our new method will outperform logistic regression. Additionally our new method can detect differences in the maximum probability of cure which standard logistic regression cannot provide. Furthermore, if different mouse models are compared, the treatment failure rate at no treatment may be different (i.e. different p_{base} between mouse models) which is another example of a scenario that can be handled using our approach but not using conventional logistic regression.

A potential improvement of our model in its current form might be evaluation of the (re) growth curve of *M. tuberculosis* during treatment failure. In the current design, the data were analyzed as a binary outcome because cure or failure was based on the absence or presence of mycobacteria in the lungs at a single time point three months after stop of treatment. If mycobacterial loads were measured at multiple time points after stop of treatment, e.g. after one, two and three months, as opposed to only three months, a time-to-event approach could have been used to analyze the data. A time-to-event analysis is considered more informative than analyzing the data as a binary outcome because it can provide information on the time course of cure or relapsing treatment failure. This could allow for better estimation of treatment success rates, but would also require substantially more mice.

In conclusion, we provide a new design for treatment outcome evaluation in our mouse TB model, which (i) provides accurate tools for assessment of the relationship between treatment length and predicted cure, (ii) allows for efficient comparison between regimens, (iii) can be readily compared to other studies and (iv) adheres to the reduction and refinement principles of laboratory animal use.

ACKNOWLEDGEMENTS

Authors thank Carla Roodbol, Marian ten Kate, Aart van der Meijden and Sanne van den Berg for their technical assistance and Sake de Vlas for his scientific comments and critical reading of the manuscript. Research was conducted on behalf of the PreDiCT-TB Consortium (<http://predict-tb.eu>).

FUNDING

This work was supported by the Innovative Medicines Initiative Joint Undertaking (115337), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

REFERENCES

1. WHO Global TB report, available from http://www.who.int/tb/publications/global_report/en/.
2. Gillespie, S. H. *et al.* Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* **371**, 1577-1587, doi:10.1056/NEJMoa1407426 (2014).
3. Conde, M. B. *et al.* Moxifloxacin versus ethambutol in the initial treatment of tuberculosis: a double-blind, randomised, controlled phase II trial. *Lancet* **373**, 1183-1189, doi:10.1016/s0140-6736(09)60333-0 (2009).
4. Dorman, S. E. *et al.* Substitution of moxifloxacin for isoniazid during intensive phase treatment of pulmonary tuberculosis. *Am J Respir Crit Care Med* **180**, 273-280, doi:200901-0078OC [pii] 10.1164/rccm.200901-0078OC (2009).
5. Nuernberger, E. L. *et al.* Moxifloxacin-containing regimens of reduced duration produce a stable cure in murine tuberculosis. *Am J Respir Crit Care Med* **170**, 1131-1134, doi:10.1164/rccm.200407-885OC 200407-885OC [pii] (2004).
6. Phillips, P. P., Fielding, K. & Nunn, A. J. An evaluation of culture results during treatment for tuberculosis as surrogate endpoints for treatment failure and relapse. *PLoS One* **8**, e63840, doi:10.1371/journal.pone.0063840 (2013).
7. Phillips, P. P. *et al.* A new trial design to accelerate tuberculosis drug development: the Phase IIC Selection Trial with Extended Post-treatment follow-up (STEP). *BMC medicine* **14**, 51, doi:10.1186/s12916-016-0597-3 (2016).
8. Nuernberger, E., Sizemore, C., Romero, K. & Hanna, D. Toward an Evidence-Based Nonclinical Road Map for Evaluating the Efficacy of New Tuberculosis (TB) Drug Regimens: Proceedings of a Critical Path to TB Drug Regimens-National Institute of Allergy and Infectious Diseases In Vivo Pharmacology Workshop for TB Drug Development. *Antimicrob Agents Chemother* **60**, 1177-1182, doi:10.1128/aac.02041-15 (2016).
9. Gumbo, T., Lenaerts, A. J., Hanna, D., Romero, K. & Nuernberger, E. Nonclinical models for antituberculosis drug development: a landscape analysis. *The Journal of infectious diseases* **211 Suppl 3**, S83-95, doi:10.1093/infdis/jiv183 (2015).
10. Tasneen, R. *et al.* Contribution of Oxazolidinones to the Efficacy of Novel Regimens Containing Bedaquiline and Pretomanid in a Mouse Model of Tuberculosis. *Antimicrob Agents Chemother* **60**, 270-277, doi:10.1128/aac.01691-15 (2015).
11. Ahmad, Z. *et al.* Contribution of moxifloxacin or levofloxacin in second-line regimens with or without continuation of pyrazinamide in murine tuberculosis. *American journal of respiratory and critical care medicine* **188**, 97-102 (2013).
12. Li, S. Y. *et al.* Evaluation of moxifloxacin-containing regimens in pathologically distinct murine tuberculosis models. *Antimicrob Agents Chemother* **59**, 4026-4030, doi:10.1128/aac.00105-15 (2015).
13. De Groote, M. A. *et al.* Comparative studies evaluating mouse models used for efficacy testing of experimental drugs against *Mycobacterium tuberculosis*. *Antimicrobial agents and chemotherapy* **55**, 1237-1247 (2011).
14. Gupta, S. *et al.* Acceleration of tuberculosis treatment by adjunctive therapy with verapamil as an efflux inhibitor. *American journal of respiratory and critical care medicine* **188**, 600-607, doi: 10.1128/AAC.00595-10 (2013).
15. Dutta, N. K. & Karakousis, P. C. PA-824 is as effective as isoniazid against latent tuberculosis infection in C3HeB/FeJ mice. *Int J Antimicrob Agents* **44**, 564-566, doi:10.1016/j.ijantimicag.2014.07.012 (2014).

16. Mourik, B. C. *et al.* Assessment of Bactericidal Drug Activity and Treatment Outcome in a Mouse Tuberculosis Model Using a Clinical Beijing Strain. *Antimicrob Agents Chemother* **61**, doi:10.1128/aac.00696-17 (2017).
17. de Steenwinkel, J. E. *et al.* Drug susceptibility of Mycobacterium tuberculosis Beijing genotype and association with MDR TB. *Emerg Infect Dis* **18**, 660-663, doi:10.3201/eid1804.110912 (2012).
18. Ahmad, Z. *et al.* Comparison of the 'Denver regimen' against acute tuberculosis in the mouse and guinea pig. *J Antimicrob Chemother* **65**, 729-734, doi:10.1093/jac/dkq007 (2010).
19. Rosenthal, I. M. *et al.* Dose-ranging comparison of rifampin and rifapentine in two pathologically distinct murine models of tuberculosis. *Antimicrob Agents Chemother* **56**, 4331-4340, doi:10.1128/aac.00912-12 (2012).
20. Beal, S., Sheiner, L., Boeckmann, A. & Bauer, R. NONMEM User's Guides.(1989–2009), Icon Development Solutions, Ellicott City, MD, USA, 2009. *Book NONMEM User's Guides.(1989–2009). Ellicott City, MD, USA: Icon Development Solutions (2009).*
21. R Core Team, R. R: A language and environment for statistical computing. (2013).
22. Rosenthal, I. M. *et al.* Daily dosing of rifapentine cures tuberculosis in three months or less in the murine model. *PLoS medicine* **4**, e344, doi:10.1371/journal.pmed.0040344 (2007).
23. Nuermberger, E. *et al.* Powerful bactericidal and sterilizing activity of a regimen containing PA-824, moxifloxacin, and pyrazinamide in a murine model of tuberculosis. *Antimicrob Agents Chemother* **52**, 1522-1524, doi:10.1128/aac.00074-08 (2008).
24. Grosset, J., Truffot-Pernot, C., Lacroix, C. & Ji, B. Antagonism between isoniazid and the combination pyrazinamide-rifampin against tuberculosis infection in mice. *Antimicrob Agents Chemother* **36**, 548-551 (1992).
25. de Steenwinkel, J. E. *et al.* Optimization of the rifampin dosage to improve the therapeutic efficacy in tuberculosis treatment using a murine model. *Am J Respir Crit Care Med* **187**, 1127-1134, doi:10.1164/rccm.201207-1210OC (2013).
26. Yew, W. W. Clinically significant interactions with drugs used in the treatment of tuberculosis. *Drug Saf* **25**, 111-133 (2002).
27. Almeida, D. *et al.* Paradoxical effect of isoniazid on the activity of rifampin-pyrazinamide combination in a mouse model of tuberculosis. *Antimicrob Agents Chemother* **53**, 4178-4184, doi:AAC.00830-09 [pii], 10.1128/AAC.00830-09 (2009).
28. Grosset, J. *et al.* Modeling early bactericidal activity in murine tuberculosis provides insights into the activity of isoniazid and pyrazinamide. *Proceedings of the National Academy of Sciences* **109**, 15001-15005, doi: 10.1073/pnas.1203636109. (2012).
29. Flecknell, P. Replacement, reduction and refinement. *Altex* **19**, 73-78 (2002).