Title: The relative fitness of drug resistant *Mycobacterium tuberculosis*: a modelling study of household transmission in Lima, Peru

Authors: Gwenan M. Knight*[a,b], Mirko Zimic[c], Sebastian Funk[d], Robert H. Gilman[c,e], Jon S. Friedland[b,f], Louis Grandjean[c,f,g]

Affiliations
[a] National Institute of Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Commonwealth Building, Hammersmith Campus, Imperial College London, Du Cane Road, London, W12 0NN, United Kingdom.
[b] Infectious Diseases and Immunity, Commonwealth Building, Hammersmith Campus, Imperial College London, Du Cane Road, London, W12 0NN, United Kingdom.
[c] Laboratorio de Bioinformática y Biología Molecular, Facultad de Ciencias, Universidad Peruana Cayetano Heredia, 31 Av. Honorio Delgado 430, Distrito de Lima, Peru.
[d] Centre for the Mathematical Modelling of Infectious Diseases, Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom.
[e] Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St, Baltimore, Maryland 21205, United States of America.
[f] Wellcome Trust Imperial College Centre for Global Health, St Mary's Hospital Campus, Imperial College London, Praed Street, London, W2 1NY, United Kingdom
[g] TB Centre, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, United Kingdom

***Corresponding author:** Gwenan M Knight,
Address: National Institute of Health Research Health Protection Research Unit in Healthcare Associated Infection and Antimicrobial Resistance, Commonwealth Building, Hammersmith Campus, Imperial College London, Du Cane Road, London, W12 0NN, United Kingdom.
Telephone number: +44 (0)2033131280
Email: g.knight@imperial.ac.uk

Classification: Biological Sciences – Medical Sciences

Short title: The relative fitness of MDR-TB

Keywords: drug-resistance, tuberculosis, fitness

Abstract

The relative fitness of drug resistant versus susceptible bacteria in an environment dictates resistance prevalence. Estimates for the relative fitness of resistant *Mycobacterium tuberculosis* (*Mtb*) strains are highly heterogeneous, many are derived from *in-vitro* experiments and few studies have estimated relative fitness in the field. Measuring fitness in the field allows us to determine how the environment (including factors such as host genetics, TB treatment regimen etc.) influences the ability for bacteria to survive, be transmitted and cause secondary cases of disease.

We designed a household structured, stochastic mathematical model to estimate the fitness costs associated with multi-drug resistance (MDR) carriage in *Mtb* strains in Lima, Peru over a 3-year period. By fitting the model to data from a large prospective cohort study of TB disease in household contacts we estimated the fitness, relative to susceptible strains with a fitness of 1, of MDR-*Mtb* strains to be 0.33 (95% credible interval: 0.17-0.54) or 0.39 (0.26-0.58), if only transmission or progression to disease, respectively, was affected by MDR. The relative fitness of MDR-*Mtb* increased to 0.57 (0.43-0.73) when the fitness cost was modelled to influence both transmission and progression to disease equally.

We found the average relative fitness of MDR-*Mtb* circulating within households in Lima, Peru between 2010-2013 to be significantly lower than susceptible-*Mtb* circulating at the same time and location. If these fitness levels do not change, then existing TB control programmes are likely to keep MDR-TB prevalence at the current levels in Lima, Peru.

Significance Statement

The relative fitness of drug resistant strains is a key determinant of resistance prevalence. We used a new quantitative framework to directly estimate the average relative fitness cost of multi-drug resistant (MDR-) *Mycobacterium tuberculosis* (*Mtb*) strains from a household study in Lima, Peru. This showed that the relative fitness of MDR-*Mtb* was substantially lower than that of drug-susceptible *Mtb* strains circulating at the same time and location, suggesting that with current control methods MDR-TB diseases levels may be held stable in Lima, Peru.

Introduction

Mycobacterium tuberculosis (*Mtb*) is a highly prevalent bacterium, thought to infect just under a quarter of the world's population (1). Treatment of tuberculosis (TB) disease is not simple and drug-susceptible tuberculosis (DS-TB) requires a multiple drug regimen taken for at least 6 months (2). Multidrug-resistant tuberculosis (MDR-TB) treatment regimens are significantly longer, cause serious side effects and are very expensive (3). Whilst currently 5% of all TB cases globally are estimated to be MDR-TB (2), predicting the future burden of DS- and MDR-TB is essential for TB control programmes.

One key parameter that determines the future prevalence of drug resistant TB is the relative fitness of drug resistant *Mtb* strains as compared to drug susceptible *Mtb* strains (4-7). Fitness is a complex, environment-dependent trait that can be defined as the ability of a pathogen to survive, reproduce, be transmitted and cause secondary cases of disease. These abilities are affected by multiple environmental factors such as a host's genetics, the current TB treatment regimen and other risk factors for transmission, which are all time-varying. The importance of this parameter has been highlighted by several mathematical models which show how even small changes in its value can predict widely varying future levels of MDR-TB burden (4-6, 8, 9). Thus, gaining environment dependent, accurate estimates of fitness is of critical importance.

Within *Mtb,* it has been shown that the appearance of drug resistance mutations affects fitness (10-12). These previous studies have shown that resistant *Mtb* is, usually, less fit than susceptible *Mtb* under a range of fitness definitions: either by demonstrating a lower growth rate *in vitro* (e.g. (13)), less progression to disease after inoculation in guinea pigs (e.g. (14)) or a lower chance of causing secondary cases of disease (e.g. (12, 15)). The latter definition is important for epidemiological predictions of burden, whilst the first provides the potential underlying biological cause. The epidemiological fitness of a *Mtb* strain can be split into an ability to (1) cause secondary infections (transmission) and (2) cause subsequent active disease (progression). For example, resistant *Mtb* may be transmitted equally as well, but subsequent disease rates in those infected may be lower or less severe. For *Mtb* this split is especially pertinent due to the importance of the latent, non-infectious, stage of disease.

Also highly important for *Mtb* is the spatial location of transmission (16). Few studies have considered the critical influence of household structure on transmission of *Mtb*. To our knowledge, no studies have considered the spread of drug-resistant tuberculosis in the context of a household-structured stochastic mathematical model.

The difference in definitions of fitness and corresponding experimental data makes translation from data analysis to predictive mathematical modelling difficult. Here, we tackle this problem by fitting a mathematical model to a detailed data set on the transmission of *Mtb* strains collected in a large cohort study of households undertaken in Lima, Peru between 2010 and 2013 (17). We derive estimates of fitness in this specific setting with different fitness definitions (either effects on transmission and/or progression to disease) and test the robustness of these estimates under a range of assumptions. These parameters will allow for better predictions of future MDR-TB levels and an improved understanding of MDR-TB spread.

Results

Fit to the data

Model structures 1-3 could all replicate the data from the household study (Figure 2). The MCMC trace and density plots of the posterior distributions are shown in SI Text.

Parameter estimates

The estimates of the external force of infection for DS- and MDR-TB were similar across the three models (Table 3, Figure 3). The per capita transmission rate of DS-TB within households was also similar across the three models. The relative fitness of MDR-*Mtb* was similar for Model 1 and 2, but increased in Model 3, as might be expected as in this third model the reduction in fitness is applied to two rates. For Model 1, that is assuming a resistance phenotype affects transmission, the relative fitness of MDR-*Mtb* was estimated to be 0.33 (95% CI: 0.17-0.54) vs. DS-*Mtb* with a fitness of 1. In Model 2, where a resistance phenotype affected disease progression, a similar relative fitness was estimated: 0.39 (0.26-0.58). If both rates were affected, then the relative fitness of MDR-*Mtb* was estimated to be 0.57 (0.43-0.73) (Table 3, Figure 3).

Comparing the external force of infection for DS- vs. MDR-TB we found that the ratio of the two was around 0.5 (median estimate 0.42 / 0.54 / 0.55 from the three models). This single value for the external force of infection (*foi*) represents a complex set of processes (contact patterns, length of infectiousness etc.) and so cannot be used to determine relative fitness. However, the ratio is in the range that supports our estimates of the relative fitness from the internal household model.

Probability of remaining free from tuberculosis

We explored the probability of remaining free from tuberculosis as was presented from the original study (Figure 2 in (17)). By comparison we had highly similar dynamics to the study (SI text Figure S5)

Scenario analysis

Our five scenarios gave very similar estimates for the relative fitness of MDR-*Mtb* (a range of medians from 0.22 – 0.41, SI text). This suggests that the estimates of relative fitness are robust to: increasing the initial proportion of households that were initially infected with latent MDR-*Mtb* from 2% to 10% (in the pre-study), setting TB incidence to high or low levels (see SI text for parameter details), extending the initial run-in period from 10 to 30 years or removing the saturation of transmission within households.

Discussion

Our results suggest that the average relative fitness of MDR-*Mtb* strains in circulating in households in Lima, Peru in 2010-2013 was substantially lower than that of drug susceptible strains (~40-70% reduction). When the effect of resistance was measured as an effect on transmission only, then the relative fitness of MDR-*Mtb* strains was lower than if the effect of resistance lowered only the progression rate to disease. When a resistance phenotype was assumed to affect both transmission and progression to disease rates, then the relative fitness of MDR-TB strains was higher at ~60%. These costs to resistance carriage were observed despite the longer infectious period of drug-resistant TB arising from delayed diagnosis and treatment.

The strengths of this study are that we were able to fit a stochastic household-level model to detailed location-specific data, accounting for accurate distributions of both household size and study follow-up time. We were also able to differentiate between internal and external transmission, matching the resistance typing data from the household study (17). This model and its MCMC fitting algorithm can be applied to other settings and then used as the basis for predictions of future levels of DS- and MDR-TB. In particular, this novel way of estimating fitness costs, by fitting dynamic transmission models to resistance-specific incidence data could be used for other TB prevalent settings or for other bacteria. Furthermore, the estimates given can be directly translated into dynamic transmission models for prediction whilst previous estimates, for example of differences in growth rates have less clear epidemiological translations.

Our modelling analysis is limited by homogeneity - of both hosts and strains. The characteristics of the DS- and MDR-TB contacts under consideration in the underlying household study were highly similar (17). Thus, as our estimate is of a relative fitness we believe that including host differences in our model may have had little effect on our relative results. Strain heterogeneities however, mean that our result is (potentially) an average across many different drug resistant strains. It is known that differences in resistance and compensatory mutation combinations result in a diversity in fitness across strains (13). This diversity is highly important for predictions of MDR-TB levels in the future (18). Our estimate must therefore be taken as a population average in Lima, at a certain time and indicative of the mean fitness rather than an indicator of the range of potential fitness in the population. If one highly fit MDR-TB strain were to emerge (or were already present), then future prevalence predictions based on our (mean) estimate could be an underestimate. We also assumed that transmission of the strains was internal if the resistant phenotype was the same as the index, but external if different. We made this same assumption for both susceptible and resistant strains. Hence, our estimates would only be affected if we thought that a different percentage of infection for susceptible strains was occurring outside the household than for resistant strains and at the moment we cannot determine this.

Our Model 1, where a transmission effect is assumed, is the most similar to previous models of MDR-TB transmission (6, 9, 19). However, our MDR-TB fitness predictions are at the lower end of the range seen previously (10). This may reflect the situation in Peru where there is a strong tuberculosis control infrastructure with a well-developed MDR-TB treatment program and a growing economy. These two factors may have combined to limit the spread of MDR-TB and hence prevent the adaptation of MDR-TB to a higher fitness. At the bacterial level, compensatory fitness mutations that could influence the ability of drug resistant *Mtb* strains to spread may not have emerged or not been allowed to spread. Calibrating the model to other settings would help clarify this issue. Alternatively, it may be that our estimates are providing, for the first time, a better direct translation of fitness from epidemiological data to a transmission model parameterisation.

There is a paucity of evidence for whether differences in TB disease prevalence in general are due to infection or progression to disease (20). In particular, for resistant strains it is unclear where the effect of becoming resistant should be applied in the natural history of tuberculosis infection. Both Snider and Teixeira (21, 22) demonstrated similar levels of tuberculin skin test (TST) conversion among MDR- and DS-TB household contacts but lower levels of disease in contacts of those with MDR-TB. This was also seen in a recent study in children (23), whilst a higher prevalence of TST positivity was found in household contacts of MDR-TB patients than contacts of newly diagnosed TB patients in Viet Nam (24). This evidence combines to suggest that the fitness cost to resistance, if any, was to be observed on the progression to disease. We make this assumption in our Model 2, where the hypothesis is that those with active TB disease, whether due to resistant or susceptible bacteria, have a similar

bacterial load and hence ability to transmit successfully. However, once successfully established in a new host, resistant bacteria may be less able to combat the immune system and establish a disease state. This has been assumed in a previous model of HIV and MDR-TB interaction (25).

The relative fitness estimates from Model 1 and 2 are similar, whilst the relative fitness estimates from Model 3 are higher, potentially due to the effect being on two processes and hence multiplicative. To determine which model structure is more appropriate, data on infection as well as disease status in contacts of index cases would be required. As mentioned above, TST surveys have found similar rates of infection in contacts of DS-TB and MDR-TB cases (21, 23) but lower rates of disease in contacts of MDR-TB cases, suggesting that Model 2 may be the most biologically appropriate choice (where progression to disease but not transmission rates are affected by becoming resistant).

Previous models have assumed that resistant strains could become more fit (i.e. have a relative fitness greater than 1), whilst we capped the relative fitness of the resistant strains at 1, due to the data from the household cohort (17). Our posterior parameter distributions for the estimated relative fitness parameter (reflected in the 95% CI for *f*, see SI text) suggest that this is a valid assumption for the resistant strains circulating at this time in Lima. Importantly, all our estimates are of "relative" fitness, and therefore should be robust to changes in natural history assumptions as these would affect both drug susceptible and resistant strain transmission.

Future work will include adding in detail on host and strain heterogeneity to the model. However, as described above the former characteristics distribution was similar across DS- and MDR-TB households whilst the latter requires data that is currently unavailable. Data collection of strain heterogeneity along with active contact tracing and an understanding of where and from whom transmission occurs would drastically improve our understanding of fitness and hence improve estimates of future MDR-TB levels. As we were fitting to a specific household-based study we modelled the community effect on transmission as external force of infection rather than explicitly including a community compartment as has been done for previous models of TB spread (26). Exploring the external infection methods and potential changes in this force of infection over time (i.e. making it dynamic) would allow for models that can predict levels of MDR-TB in Lima. Importantly, this model provides a key estimate of a parameter that is needed for many existing mathematical model structures, where only a single "resistant" strain is included. Future predictive transmission modelling using our relative fitness estimates are likely to suggest that if treatment objectives are maintained and this fitness measure remains constant, that MDR-TB prevalence will remain under control in Lima in the short term.

In conclusion, if the fitness cost of drug resistance in *Mtb* is exerted on the rate of progression to active disease rather than transmission, we estimate that the relative fitness is between 30-40% relative to drug susceptible strains (at 100%). Importantly this paper provides direct transmission model estimates, using a novel method, of the relative fitness levels of drug resistant *Mtb* strains. If these fitness levels do not change, then the existing TB control programmes are likely to keep MDR-TB prevalence at their current levels in Lima, Peru. These estimates now need to be gained for *Mtb* in other settings and the values used in models to explore future global burden.

Materials and Methods

Data

The details of the study and participants can be found in (17). Briefly, 213 and 487 households were recruited with an index case of diagnosed MDR- or DS-TB respectively during 2010 to 2013. Households were followed up for variable periods of time up to a maximum of 3 years (SI text Figure S1). During the study households were visited every 6 months, and household contacts were monitored for TB disease. It was found that 35/1055 (3.32%, 95% CI [2.32, .4.58]), of the MDR-TB contacts, and 114/2356 (4.84%, 95% CI [4.01, 5.78]), of the DS-TB contacts developed TB disease, suggesting that DS-TB has higher fitness. There were no significant differences between cohorts by HIV status, age, gender or household size (17).

The specific data used to calibrate the model was 1) the incidence of MDR-TB and 2) DS-TB in households with an index DS-TB case and 3) the incidence of MDR-TB and 4) DS-TB in households with an index MDR-TB case (Table 1). We assumed that transmission of the strains was internal if the resistant phenotype was the same as the index, but external if different. The percentages of incident cases with resistance profiles matching the index was used to multiply the incidence levels accordingly.

Model structure

The mathematical model was a standard two-strain dynamic TB model (Figure 1), with transmission modelled at the level of the household. A Gillespie stochastic simulation algorithm in *R* (27) was developed using the R package "GillespieSSA" (28). Using a stochastic transmission model was important as the model was implemented independently in households where the small populations mean stochastic effects are highly important. We assumed that saturation of transmission could occur and hence scaled our transmission rate by the size of the household (number of people), assuming households have the same ventilation level (or at least that this did not vary by index case *Mtb* resistance status) and within-household homogeneous mixing (29). This assumption means that in households with more people, household members are assumed to have lower individual chance of infection from an active disease case than in smaller households, due to decreased exposure. This has been observed for another airborne pathogen, influenza (30) and was explored in sensitivity analysis. All natural history parameters were taken from the literature, are listed in Table 2 and the dynamics explained in the legend to Figure 1.

Four parameters were estimated from the data (Table 1): (1) the per capita transmission rate of DS-TB within households (β_S), (2) the relative fitness of MDR-*Mtb* strains vs. DS-*Mtb* strains (f) expressed as an effect on transmission or progression or both, and the external (to households) force of infection (foi) of (3) DS-TB foi_s and (4) MDR-TB foi_r.

Three model formulations

Resistant strains were allowed to have an equal or lower fitness relative to susceptible strains. The mechanisms behind this reduction were estimated to affect two different rates: the transmission rate, the rate of progression to disease, or both (Figure 1). We assumed that the fitness of the resistant strains could not rise above that of susceptible strains due to the data from the household cohort (17). Model 1 (transmission fitness cost model) assumed that fitness costs directly affected the number of secondary infections by reducing the transmission parameter for MDR-*Mtb* ($0 < f_1 < 1$, $f_2 = 1$, Figure 1). This is the standard assumption for the effect of resistance on fitness for transmission dynamic models of *Mtb* (6, 9, 19) and other pathogens (31). Model 2 (progression fitness cost model) assumed that although MDR-TB transmission occurred at the same rate as DS-TB, there is a fitness cost to progression to disease ($f_1 = 1$, $0 < f_2 < 1$, Figure 1). Model 3 assumed that there was a fitness cost to both transmission and progression, and that the cost was the same for both processes ($0 < f_1 = f_2 < 1$, Figure 1). We could not explore a Model with fitness affecting both processes at differing levels as we did not have data on levels of infection. Without this data, a model with high transmission fitness cost but low progression cost would be equally as likely as a model with a low transmission fitness cost but a high progression cost and hence would be uninformative. Note that fixing either f_1 or f_2 equal to one is the same as ignoring this parameter altogether and leaving the multiplied rate at its background level as they are both scalar constant parameters with no units.

Model simulation

The model initially sampled 700 household sizes from the distribution of household sizes in the trial (32), and initial conditions of latent infection taken from data (2, 33) (SI text). The model was then simulated for 10 years with a sampled set of the four unknown parameters (pre-study period). A random time point from over this 10-year period in which there was at least one active case with the same sensitivity as the initial case in the household (i.e. DS-TB or MDR-TB) was taken to be the time the household entered the study and the active index case was detected. This allows for simulation of changes in latency in the household and provides initial conditions dependent upon each parameter sample.

The above randomly sampled time point of entry to the study was taken to be the initial conditions for the simulation of the model that was fitted to the household study (17) (study period). The same values of the four unknown parameters was used as in the pre-study period and the simulation time for each household was randomly sampled from the distribution of follow-up times in the study (SI text Figure S1). The only parameter that changed, to match the altered patient care in the study, was the case detection rate which increased for the study period from the WHO estimates to a screen occurring every 6 months (Table 2).

The TB incidence from the model was calculated by determining the total number of new cases of active TB in all 700 households over the follow-up time, and dividing this by the total number of follow-up years in these households. The total number of follow-up years was a product of the number of household members and the follow-up time for the household taking into account any deaths over this time. We assumed that no-one left the households other than by death (natural or due to TB). For a detailed overview of the process see SI text Figure S2.

Model fitting

Approximate Bayesian Computation (ABC) was paired with Markov chain Monte Carlo (MCMC) methods to estimate the four unknown parameters (34). All other parameters were kept fixed at their baseline value (Table 2). The summary statistic used was the TB incidence from the model falling within the 95% CI for all four TB incidence measures from the data. Uniform priors were assumed for all four parameters (Table 1).

To estimate the standard deviation required for the MCMC for the four unknown parameters, Latin Hypercube Sampling (LHS) from the prior ranges was initially used (Stage A). The empirical standard deviation from the accepted fits was then used as proposal distribution of a Metropolis-Hastings MCMC sampler (Stage B), used to estimate posterior probabilities of the parameters.

We used the sampled trajectories to consider the probability of remaining free of tuberculosis from the model output and compare the general trends to the data (Figure 2 from (17)).

Scenario analysis

A scenario analysis was used to explore the sensitivity of Model 1 results to key natural history parameters. Firstly, we changed the initial proportion of the population latently infected with MDR-*Mtb* from 2% to 10%.

A full sensitivity analysis of the parameters kept fixed in the model fits was not possible due to limitations imposed by computation time. Instead, to determine which further scenarios to explore, we determined the parameters most correlated with TB incidence in our model, and hence likely to have the biggest impact on our model fit and parameter estimates. To determine these parameters, we used LHS to choose 10,000 parameter sets from (uniform) prior distributions for all parameters (Table 2). We then ran Model 1 with these 10,000 parameter sets and determined the parameters that were statistically significantly correlated with any of the four TB incidence outputs (Kendall correlation, $p < 0.01$). These parameters were then used to design two scenarios - one with a combination of these parameters at their prior values which gave highest TB incidence and the combination which gave the lowest TB incidence.

We also increased our 10-year initial run-in period for the population to 30 years and explore the impact on the estimates. Furthermore, we explored removing the assumption of saturating household transmission (per capita transmission rate was then not dependent on household size).

Acknowledgements

This work was funded by the TB Modelling and Analysis Consortium (TBMAC, Bill and Melinda Gates Foundation, OPP1084276). We would like to thank David Moore, Christophe Fraser and Eduardo Gushiken for their input into the initial grant. We are also grateful to anonymous reviewers from TBMAC and to the TB Modelling Group at the London School of Hygiene & Tropical Medicine for their comments. GK is affiliated with the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at Imperial College London in partnership with Public Health England (PHE). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

References

1. Houben RM & Dodd PJ (2016) The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* 13(10):e1002152.
2. WHO (2016) Drug-resistant TB. Available at http://www.who.int/tb/areas-of-work/drug-resistant-tb/en/. Accessed August 17, 2017.
3. Marks SM, *et al.* (2014) Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005-2007. *Emerg Infect Dis* 20(5):812-821.
4. Dye C & Espinal MA (2001) Will tuberculosis become resistant to all antibiotics? *Proc Biol Sci* 268(1462):45-52.
5. Dye C, Williams BG, Espinal MA, & Raviglione MC (2002) Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science* 295(5562):2042-2046.
6. Cohen T & Murray M (2004) Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. *Nat Med* 10(10):1117-1121.
7. Blower SM & Chou T (2004) Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance. *Nat Med* 10(10):1111-1116.
8. Trauer JM, Denholm JT, & McBryde ES (2014) Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. *J Theor Biol* 358:74-84.
9. Dye C & Williams BG (2000) Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci U S A* 97(14):8180-8185.
10. Cohen T, Sommers B, & Murray M (2003) The effect of drug resistance on the fitness of Mycobacterium tuberculosis. *Lancet Infect Dis* 3(1):13-21.
11. Barnett M, Busby SR, & Mitchison DA (1953) Tubercle bacilli resistant to isoniazid: virulence and response to treatment with isoniazid in guinea-pigs and mice. *Br J Exp Pathol* 34(5):568-581.
12. Burgos M, DeRiemer K, Small PM, Hopewell PC, & Daley CL (2003) Effect of drug resistance on the generation of secondary cases of tuberculosis. *J Infect Dis* 188(12):1878-1884.
13. Gagneux S (2009) Fitness cost of drug resistance in Mycobacterium tuberculosis. *Clin Microbiol Infect* 15 Suppl 1:66-68.
14. Mitchison DA (1954) Tubercle bacilli resistant to isoniazid; virulence and response to treatment with isoniazid in guinea-pigs. *Br Med J* 1(4854):128-130.
15. Van Rie A, *et al.* (2000) Classification of drug-resistant tuberculosis in an epidemic area. *Lancet* 356(9223):22-25.
16. Andrews JR, Morrow C, Walensky RP, & Wood R (2014) Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. *J Infect Dis* 210(4):597-603.
17. Grandjean L, *et al.* (2015) Transmission of Multidrug-Resistant and Drug-Susceptible Tuberculosis within Households: A Prospective Cohort Study. *PLoS Med* 12(6):e1001843; discussion e1001843.
18. Knight GM, *et al.* (2015) The Distribution of Fitness Costs of Resistance-Conferring Mutations Is a Key Determinant for the Future Burden of Drug-Resistant Tuberculosis: A Model-Based Analysis. *Clin Infect Dis* 61Suppl 3:S147-154.
19. Kendall EA, Fofana MO, & Dowdy DW (2015) Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. *Lancet Respir Med* 3(12):963-972.
20. Yates TA, *et al.* (2016) The transmission of Mycobacterium tuberculosis in high burden settings. *Lancet Infect Dis* 16(2):227-238.
21. Teixeira L, *et al.* (2001) Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* 5(4):321-328.
22. Snider DE, Jr., Kelly GD, Cauthen GM, Thompson NJ, & Kilburn JO (1985) Infection and disease among contacts of tuberculosis cases with drug-resistant and drug-susceptible bacilli. *Am Rev Respir Dis* 132(1):125-132.

23. Golla V, *et al.* (2016) Effect of drug resistance on risk of *M. tuberculosis* transmission to young children. in *7th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease (The Union)*.
24. Fox GJ, *et al.* (2017) Latent tuberculous infection in household contacts of multidrug-resistant and newly diagnosed tuberculosis. *Int J Tuberc Lung Dis* 21(3):297-302.
25. Sergeev R, Colijn C, Murray M, & Cohen T (2012) Modeling the dynamic relationship between HIV and the risk of drug-resistant tuberculosis. *Sci Transl Med* 4(135):135ra167.
26. Kasaie P, Andrews JR, Kelton WD, & Dowdy DW (2014) Timing of tuberculosis transmission and the impact of household contact tracing. An agent-based simulation model. *Am J Respir Crit Care Med* 189(7):845-852.
27. R (2005) R: A Language and Environment (R Foundation for Statistical Computing, Vienna, Austria).
28. Pineda-Krch M (2010) GillespieSSA: a stochastic simulation package for R.
29. Grijalva CG, *et al.* (2015) A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. *PLoS One* 10(3):e0118457.
30. Cauchemez S, Carrat F, Viboud C, Valleron AJ, & Boelle PY (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med* 23(22):3469-3487.
31. Blower SM, Aschenbach AN, Gershengorn HB, & Kahn JO (2001) Predicting the unpredictable: transmission of drug-resistant HIV. *Nat Med* 7(9):1016-1020.
32. Grandjean L, *et al.* (2011) Tuberculosis in household contacts of multidrug-resistant tuberculosis patients. *Int J Tuberc Lung Dis* 15(9):1164-1169, i.
33. Martinez L, *et al.* (2013) Changes in tuberculin skin test positivity over 20 years in periurban shantytowns in Lima, Peru. *Am J Trop Med Hyg* 89(3):507-515.
34. Marjoram P, Molitor J, Plagnol V, & Tavare S (2003) Markov chain Mote Carlo without likelihoods. *Proc Natl Acad Sci U S A* 100(26):15324-15328.
35. Dye C, Garnett GP, Sleeman K, & Williams BG (1998) Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet* 352(9144):1886-1891.
36. Abu-Raddad LJ, *et al.* (2009) Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proc Natl Acad Sci U S A* 106(33):13980-13985.
37. Sutherland I (1976) Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli. *Adv Tuberc Res* 19:1-63.
38. Vynnycky E (1996) An Investigation of the Transmission Dynamics of M. tuberculosis. (University of London).
39. Vynnycky E & Fine PE (1997) The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 119(2):183-201.
40. Sutherland I (1968) The ten-year incidence of clinical tuberculosis following 'conversion' in 2,550 individuals aged 14 to 19 years. in *Tuberculosis Surveillance and Research Unit Progress Report* ed The Hague.
41. Sutherland I, Svandova E, & Radhakrishna S (1982) The development of clinical tuberculosis following infection with tubercle bacilli. 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. *Tubercle* 63(4):255-268.
42. Horwitz O (1969) Public health aspects of relapsing tuberculosis. *Am Rev Respir Dis* 99(2):183-193.
43. Barnett GD, Grzybowski S, & Styblo K (1971) [The current risk of contracting evolutive tuberculosis, in Saskatchewan, according to the state of previous tuberculin tests and x-ray image]. *Bull Int Union Tuberc* 45:55-79.
44. Lew W, Pai M, Oxlade O, Martin D, & Menzies D (2008) Initial drug resistance and tuberculosis treatment outcomes: systematic review and meta-analysis. *Ann Intern Med* 149(2):123-134.
45. Comstock GW (1982) Epidemiology of tuberculosis. *Am Rev Respir Dis* 125(3 Pt 2):8-15.
46. Murphy BM, Singer BH, Anderson S, & Kirschner D (2002) Comparing epidemic tuberculosis in demographically distinct heterogeneous populations. *Math Biosci* 180:161-185.

47. World Health Organisation (2012) Global Health Observatory data. Available at http://www.who.int/gho/en/. Accessed: August 17, 2017.
48. WHO (2013) *Global Tuberculosis Control*. Available at http://apps.who.int/iris/handle/10665/91355. Accessed: August 17, 2017.
49. WHO (2012) *Global Tuberculosis Control*. Available at *apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf*. Accessed: August 17, 2017.

Figure Legends

Figure 1: A standard natural history, transmission model for two strains (susceptible and resistant) of *Mtb* was used. Uninfected people become infected at a rate dependent on the number of active cases (dynamic transmission). Once infected, the majority of people (85%) are assumed to enter a Latent slow (LS / LR) state. The remainder enter a rapid progression (Latent fast, LFS / LFR) state which has a higher rate of progression to active disease (AS / AR). Resistance mutations are acquired during active disease. Those with active disease recover to the Latent slow state via treatment or natural cure. The fitness cost to resistance is assumed to affect the rate of transmission (f_1) or the rate at which those latently infected with MDR-TB progress to active disease (f_2). Only the effect on primary transmission of f_1 is highlighted here, but reinfection is also decreased. f_1 and f_2 are set at 1 or allowed to vary between 0 and 1 in the three separate models: f_1 in Model 1, f_2 in Model 2 and both f_1 and f_2 in Model 3.

Figure 2: Model fits. Black dots represent Model 1 output that matches to data shown in coloured ranges for each type of household (HH). See SI text Figures S3&S4 for equivalent plots for Model 2&3.

Figure 3: Fitted parameters from each Model. The units for the y-axis of the corresponding plots are: for the external forces of infection ('foi_s' and 'foi_r') proportion infected per year, for the relative fitness (f) there are no units and for the per capita transmission rate ('beta') the units are effective contact rate per year. Model 1 assumes a transmission cost to resistance, Model 2 a disease progression cost and Model 3 assumes an effect on both.

Table 1: Fitted parameters with description, data used for fitting, prior distributions and any differences by model structure. All parameters are fitted to the TB incidence date from the household (HH) study (17). The three models have different assumptions around the effect of decreased fitness, with f varying to be f_1 (affects transmission rate) or f_2 (affects progression to disease rate) (see Figure 1).

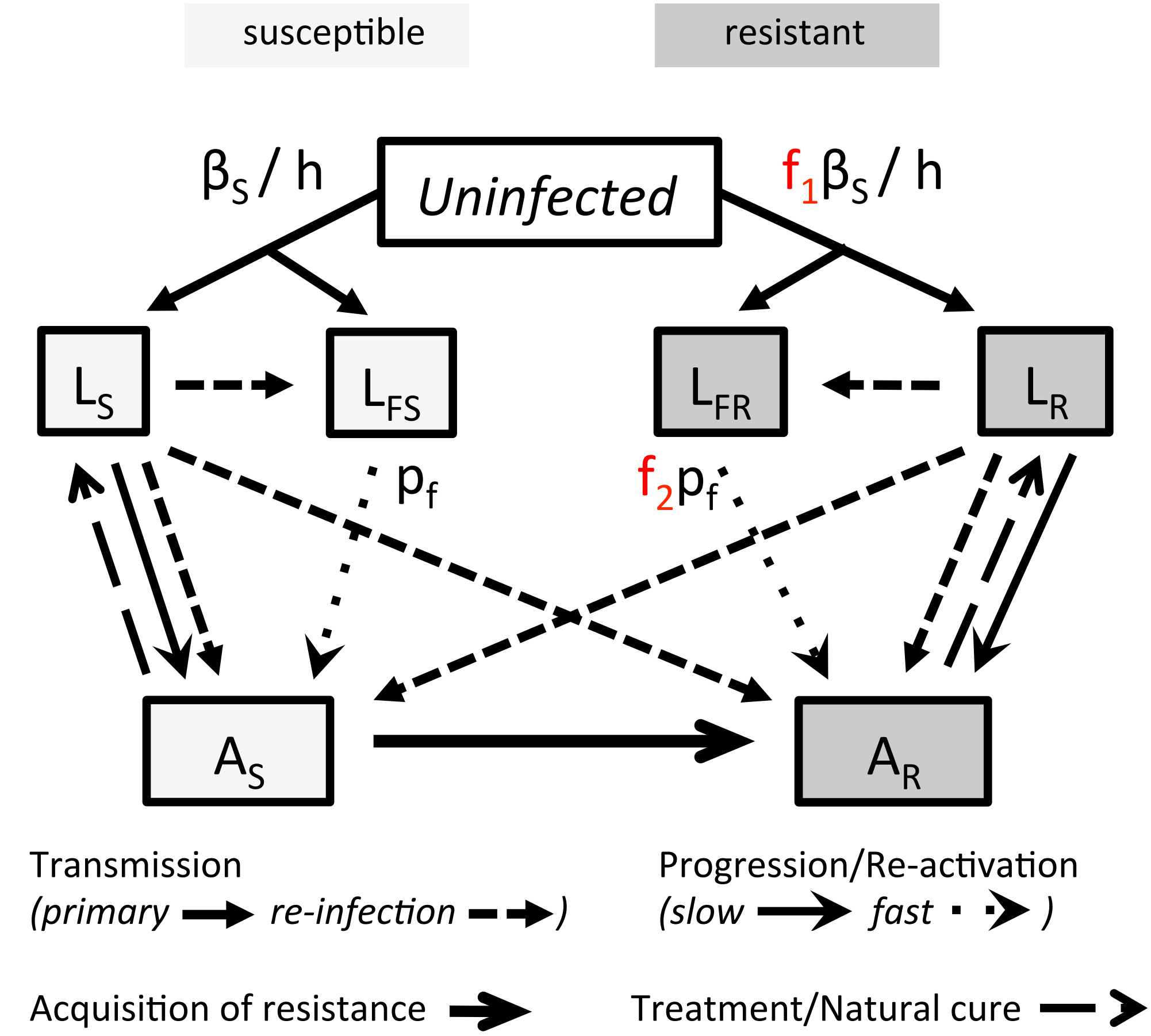
Symbol	Parameter description	Data	Prior Distribution	Model 1	Model 2	Model 3
foi_s	External force of infection of DS-TB	DS-TB incidence in MDR-TB index HH: 4264 [3916, 4338]	Uniform [0; 0:5]	/		
foi_r	External force of infection of MDR-TB	MDR-TB incidence in DS-TB index HH: 87 [13, 435]	Uniform [0, 0:3]	/		
f (f_1, f_2)	Relative fitness of MDR-TB strains compared to DS-TB strains which have a fitness of 1	MDR-TB incidence in MDR-TB index HH: 2112 [1646, 2358]	Uniform [0, 1]	$0 < f_1 < 1$ $f_2 = 1$	$f_1 = 1$ $0 < f_2 < 1$	$f_1 = f_2$ $0 < f_1 < 1$
β_s	Per capita transmission rate of DS-TB within households	DS-TB incidence in DS-TB index HH: 4264 [3916, 4338]	Uniform [90, 140]	/		
β_r	Per capita transmission rate of MDR-TB within households	Calculated from other fitted parameters: $\beta_r = f_1\beta_s$				

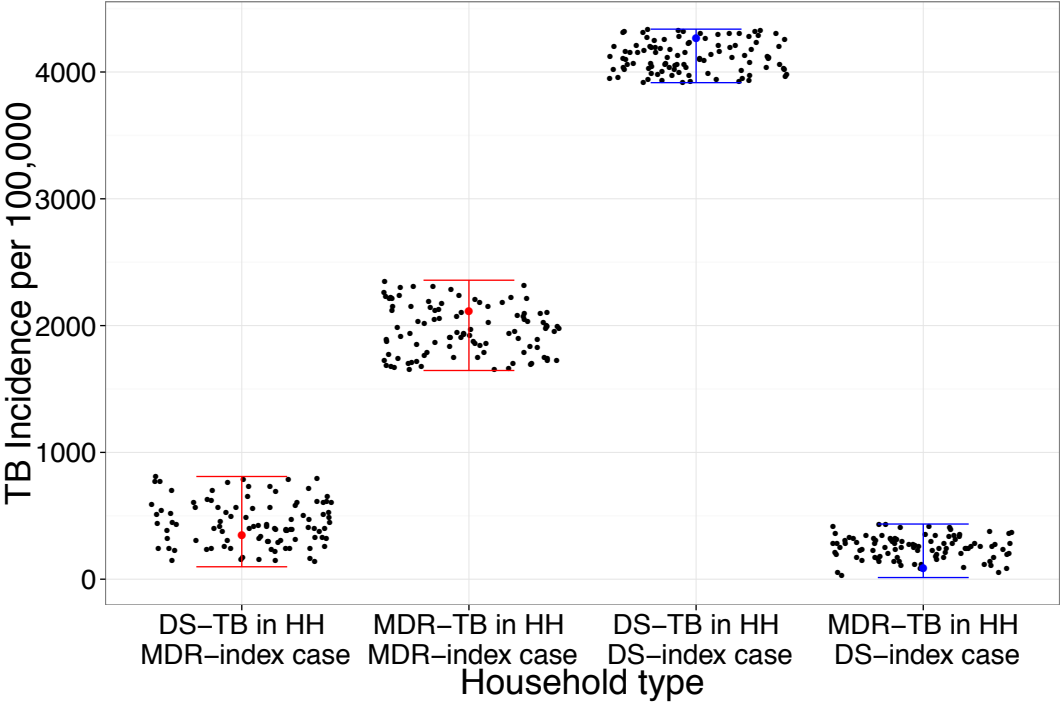
Table 2: Parameter values with description and baseline values. All prior distributions were uniform.

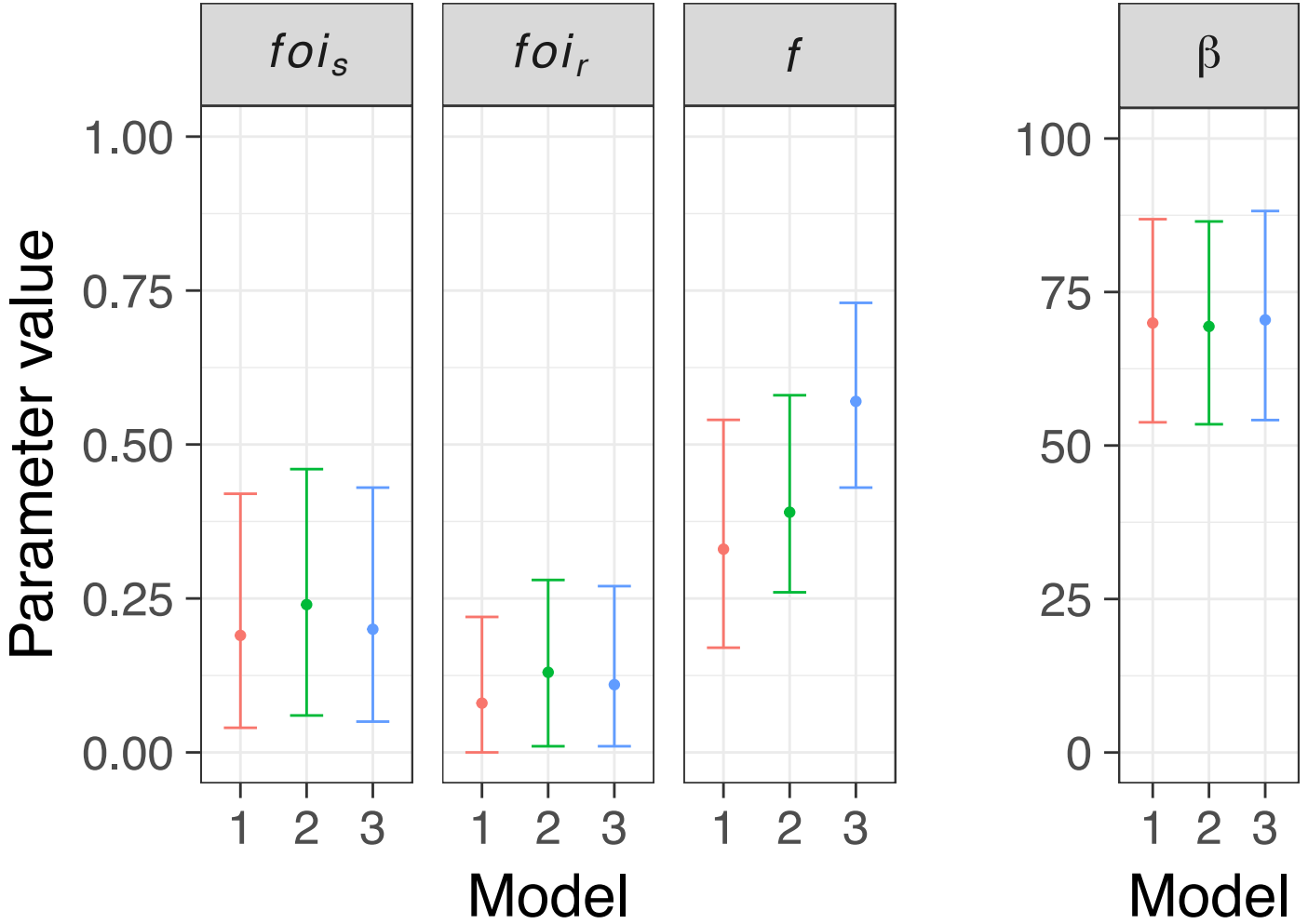
Symbol	Parameter description	Baseline value	Prior distribution	Notes and references
Nr	Number of households with MDR-TB index case	213	/	(32)
Ns	Number of households with DS-TB index case	487	/	(32)
h	Household size	2 - 15	/	(17)
p	Proportion of (re-)infected individuals which progress to the "latent fast" state	0.15	0.08-0.25	(35-37)
χ	Protection from developing active TB upon re-infection	0.35	0.25-0.45	(35, 38-41)
ϕ	Risk of reactivation among those latently infected per year	1.13×10^{-4}	$1 - 3 \times 10^{-4}$	(35, 38, 39, 41-43)
ε	Probability of acquiring new drug resistance during treatment	0:008	0.005-0.01	(44)
d	Proportion of new active cases which directly become infectious	0.5	0.25-0.75	(35, 41, 45, 46)
μ	Background death rate	1/77 = 0.013	0.012-0.014	Inverse of average life expectancy in Peru (47)
μ_A	Additional death rate of those actively infected and infectious per year	0.26	0.2-0.4	(35)
n	Annual risk of natural cure for TB cases (returns to latent state)	0.2	0.15-0.25	(35)
ω_s	Proportion of DS-TB active cases detected and treated per year	0.8; 2	0.5-0.95	For 2012 (2) for prestudy; In study: screen every 6 months
ω_r	Proportion of MDR-TB active cases detected and treated per year	0.64; 2	0.2-0.9	79% of the above 80% (ω_s) found that received DST in 2012 (48); In study: screen every 6 months
$(1 - k_s)$	Proportion of DS-TB active cases started on treatment that are successfully cured	0.74	0.5-0.9	(48, 49) (for midpoint of study)
$(1 - k_r)$	Proportion of MDR-TB active cases started on treatment that are successfully cured	0.6	0.2-0.9	For 2012 (2)
p_f	Progression rate of latent fast individuals to active disease	0.2	0.1-0.9	Duration of fast latency period of 5 years (39)

Table 3: Parameter estimates for the median and 95% credible intervals of the four unknown parameters from at least 100 5,000 MCMC runs. The fitness cost to resistance is assumed to affect transmission in Model 1, progression to active disease in Model 2 and both transmission and progression in Model 3.

Model	foi_s	foi_r	β	f
1	0.19 (0.04 – 0.42)	0.08 (0.00 – 0.22)	69.95 (53.78 – 86.86)	0.33 (0.17 – 0.54)
2	0.24 (0.06 – 0.46)	0.13 (0.01 – 0.28)	69.39 (53.46 – 86.49)	0.39 (0.26 – 0.58)
3	0.20 (0.05 – 0.43)	0.11 (0.01 – 0.27)	70.46 (54.14 – 88.19)	0.57 (0.43 – 0.73)


susceptible
resistant
β_S / h
Uninfected
$f_1\beta_S$ / h
L_S
L_{FS}
L_{FR}
L_R
p_f
f_2p_f
A_S
A_R
Transmission
(primary → re-infection ⇢)
Progression/Re-activation
(slow → fast ⋯➔)
Acquisition of resistance ➔
Treatment/Natural cure ➔


TB Incidence per 100,000
4000
3000
2000
1000
0
DS−TB in HH
MDR−index case
MDR−TB in HH
MDR−index case
DS−TB in HH
DS−index case
MDR−TB in HH
DS−index case
Household type


foi_s
foi_r
f
β
Parameter value
1.00
0.75
0.50
0.25
0.00
100
75
50
25
0
1
2
3
Model

Supporting Information for: The relative fitness of drug resistant *Mycobacterium tuberculosis*: a modelling study of household transmission in Lima, Peru

Contents

1 Time of follow-up in study **2**

2 Detailed overview of simulation **3**

3 Models 2 & 3: Fit to data **4**

4 Probability of remaining free from tuberculosis **6**

5 Trace and density plots for each unknown parameter for main models **7**

6 Result: scenario analysis: Fit to data **10**

7 Trace and density plots for each unknown parameter for scenario analysis **15**

8 Scenario analysis results **20**

1 Time of follow-up in study

Households were followed-up for variable lengths of time in the original household study (Supplementary Figure 1).

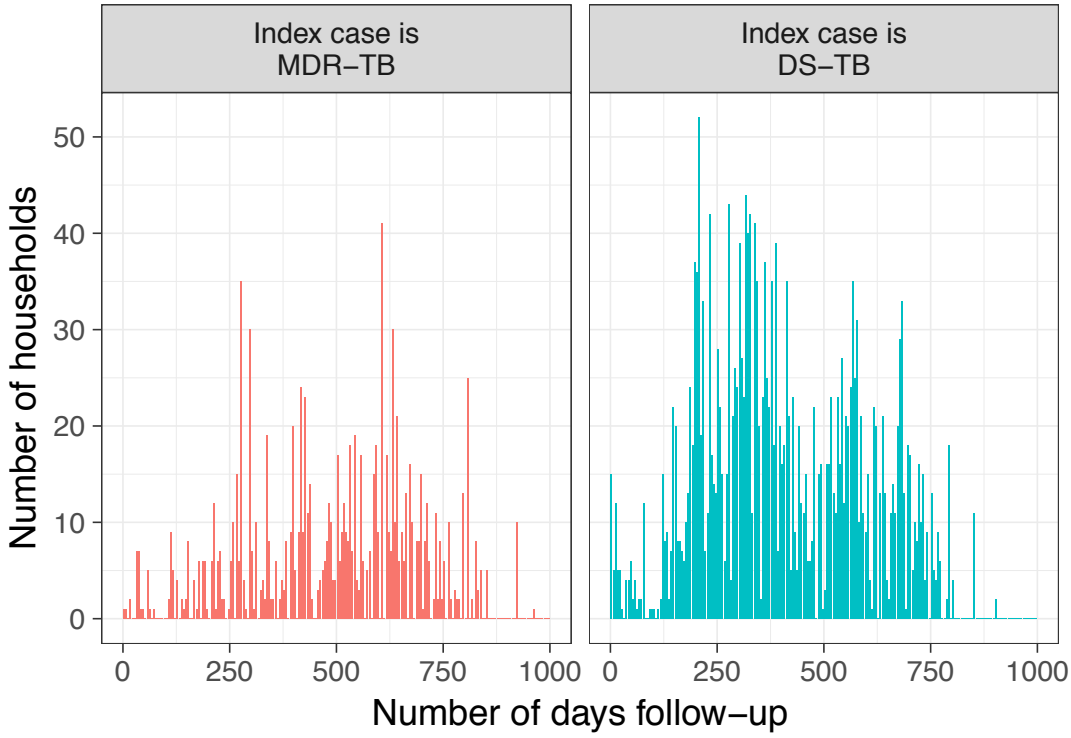


Figure 1: Distribution of follow-up times for households with an index case that was MDR-TB (left) or DS-TB (right).

2 Detailed overview of simulation

A detailed overview of all the stages used in the simulation are provided in Figure 2.

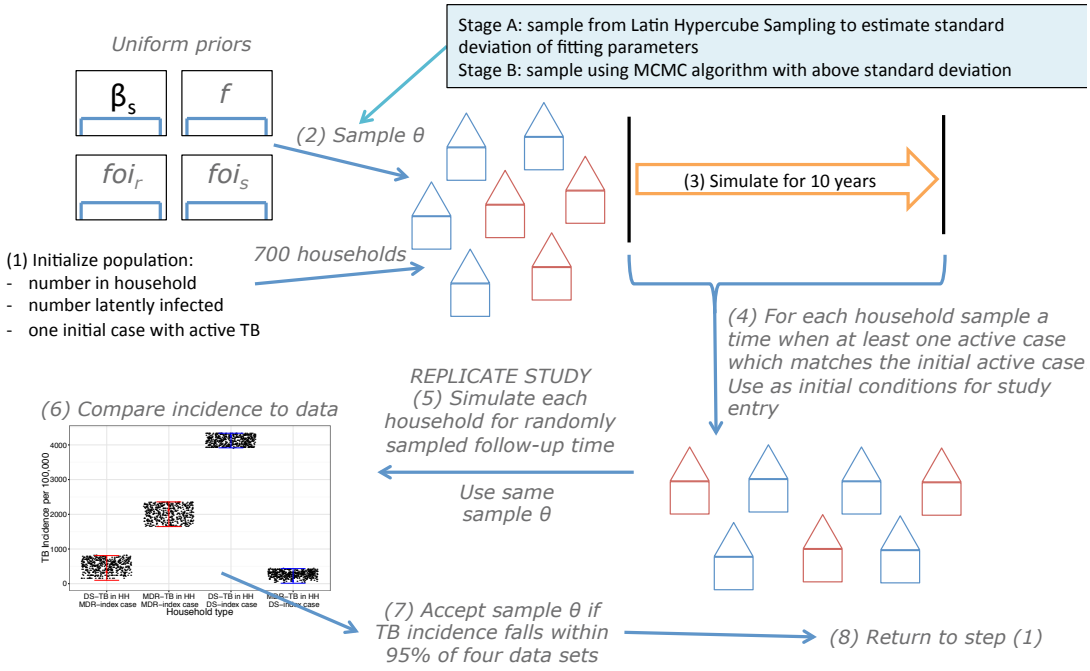


Figure 2: A pictorial representation of the simulation stages.

The model initially sampled 700 household sizes from the distribution of household sizes in the trial [1]. 213 of these had an initial MDR-TB case, 487 an initial DS-TB case. Tuberculin skin test (TST) prevalence surveys across Lima have found 52% (95% CI: 48-57%) to be infected with Mtb [2]. Hence, the number of cases initially latently infected was sampled from a normal distribution with mean 0.5 and standard deviation of 0.1. Informed by the TB prevalence in Lima, it was assumed that initially, 98% of these latent infections were with DS-TB strains, 2% with MDR-TB strains in all households [2]. This proportion was varied in scenario analysis. Random sampling from a binomial distribution, with this 98% DS-TB, determined the distribution of latent DS-TB and MDR-TB cases across the 700 households. The proportion of latent cases that were "latently fast" cases (Figure 1) was taken to be 3% to reflect that although the proportion of new infections that are fast latent is 15%, over time these will change state more rapidly than latent slow.

3 Models 2 & 3: Fit to data

Model 1-3 structures could all replicate the data from the household study as shown in Figure 2 in the main paper and Supplementary Figures 3 & 4.

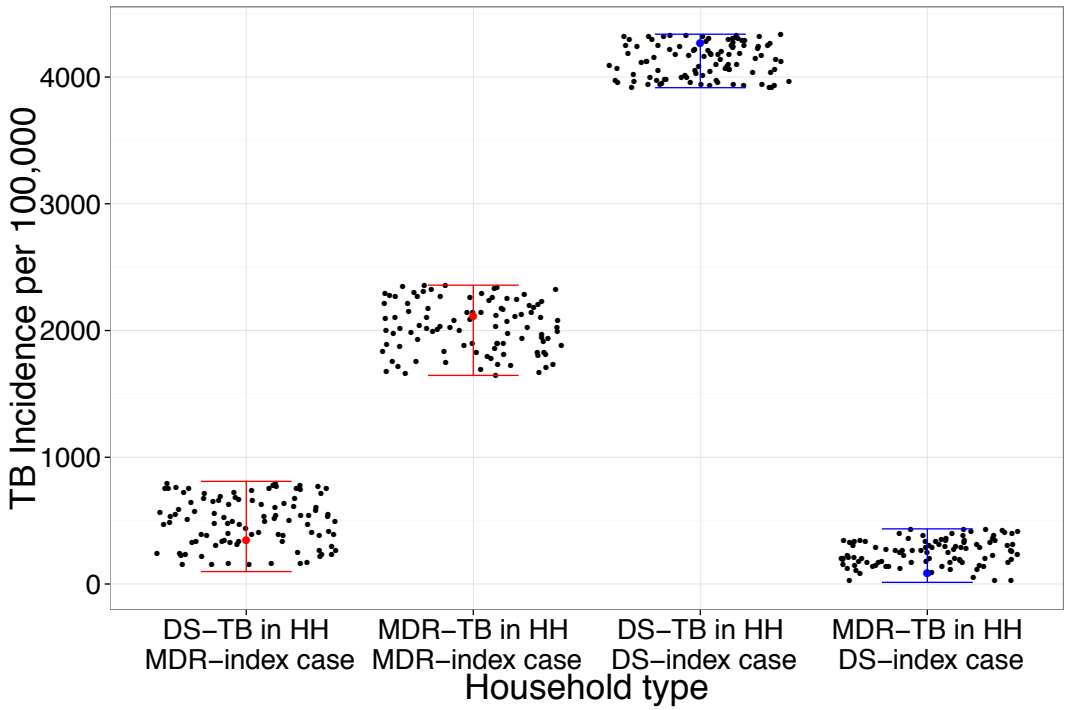


Figure 3: 100 example model fits. Black dots represent Model 2 output that matches to data shown in coloured ranges for each type of household (HH).

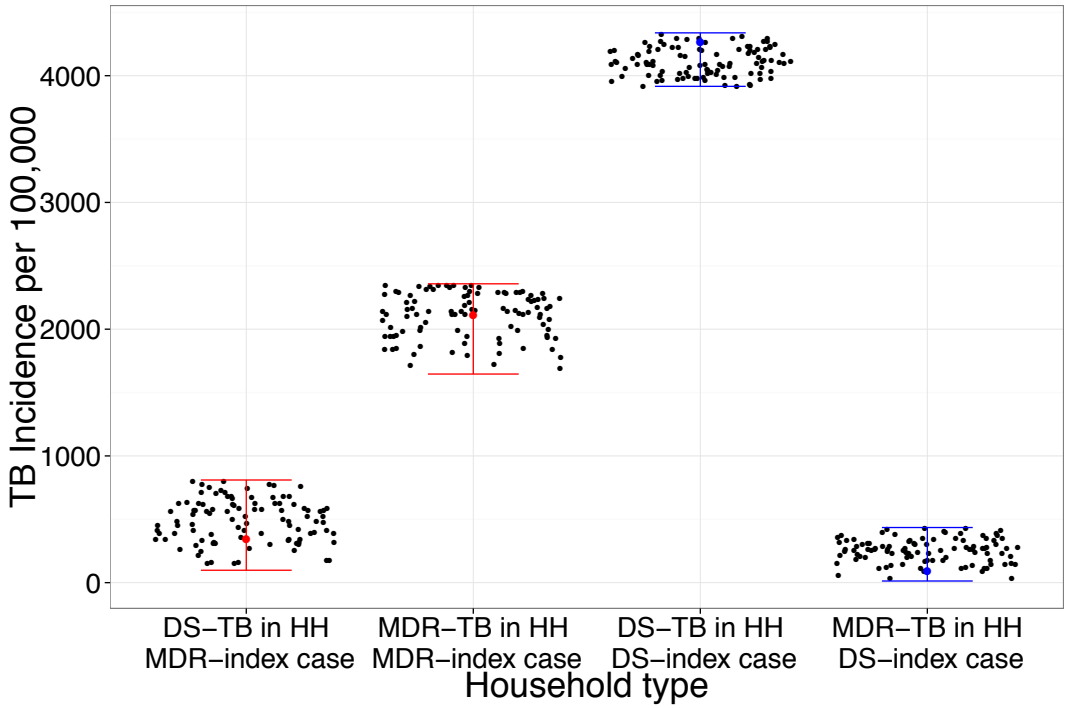
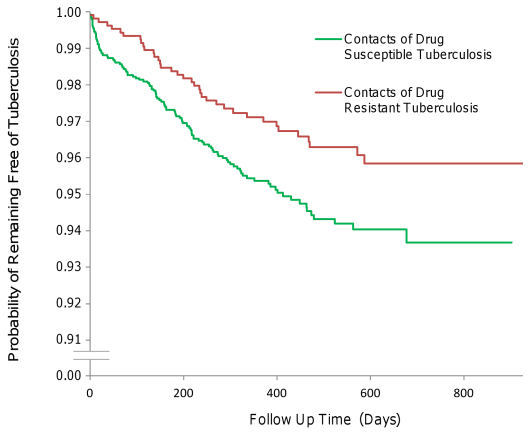


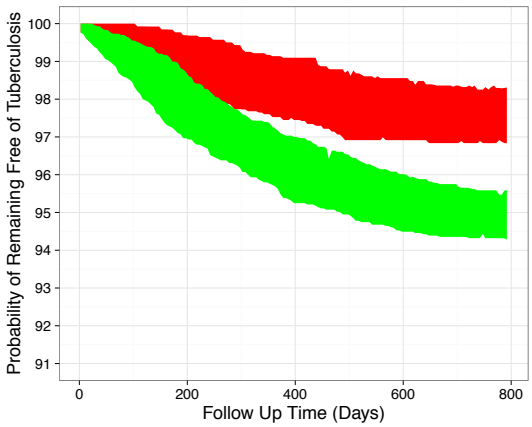
Figure 4: 100 example model fits. Black dots represent Model 3 output that matches to data shown in coloured ranges for each type of household (HH).

4 Probability of remaining free from tuberculosis

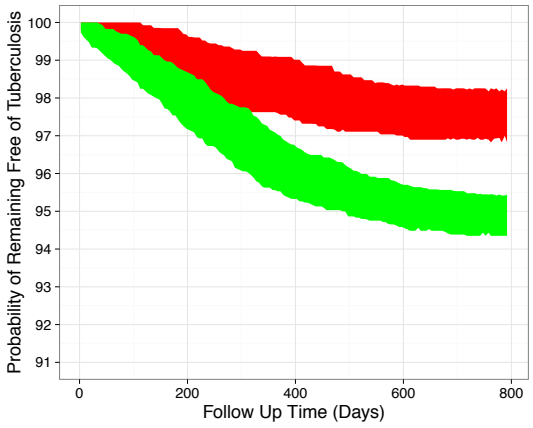
We compared the probability of remaining free from TB in our model to that presented in the original study (Figure 2 in [3]). We had highly similar dynamics to those in the main study (Figure 5).



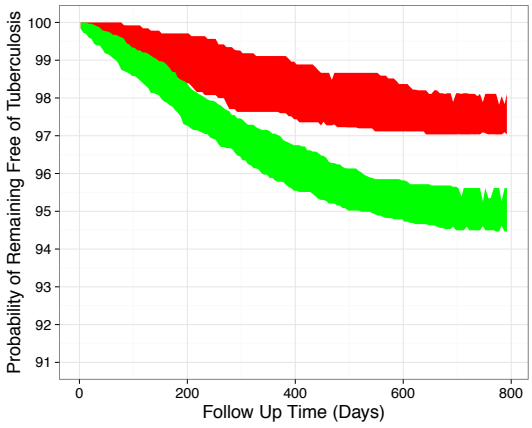
(a) Data from study (17)



(b) Model 1



(c) Model 2



(d) Model 3

Figure 5: Probability of remaining free from tuberculosis for study (a) and three model structures (b-d).

5 Trace and density plots for each unknown parameter for main models

The trace and density for each unknown parameter, from the three models are shown in Supplementary Figures 6-8.

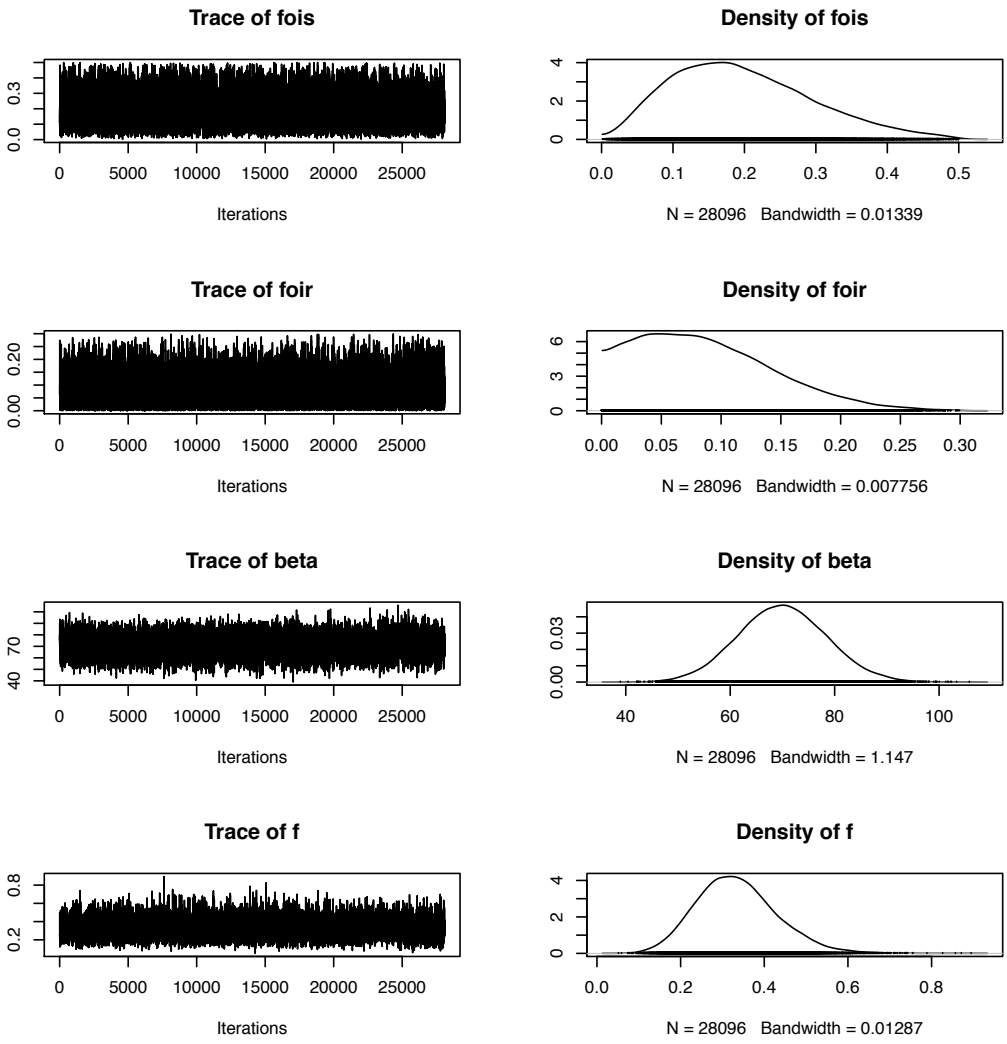


Figure 6: Trace and density plots for Model 1

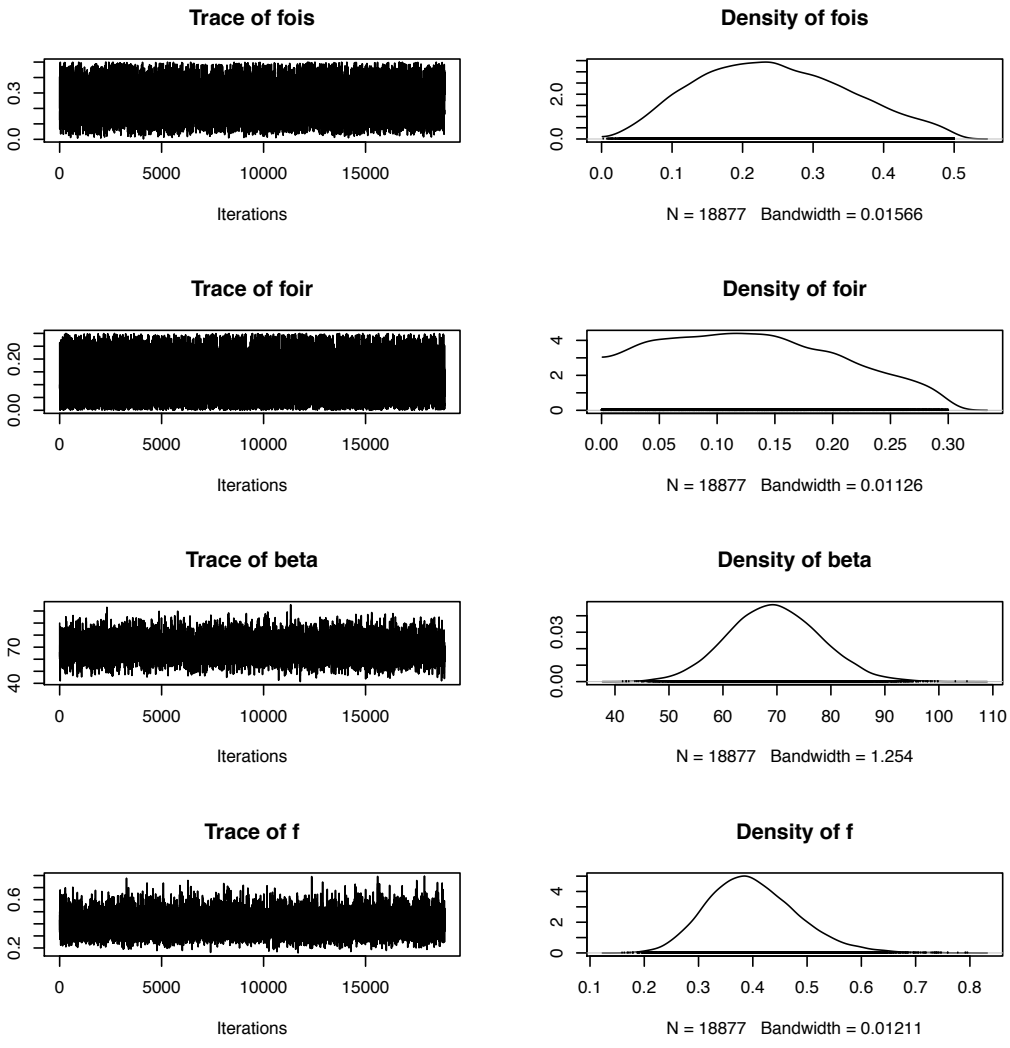


Figure 7: Trace and density plots for Model 2

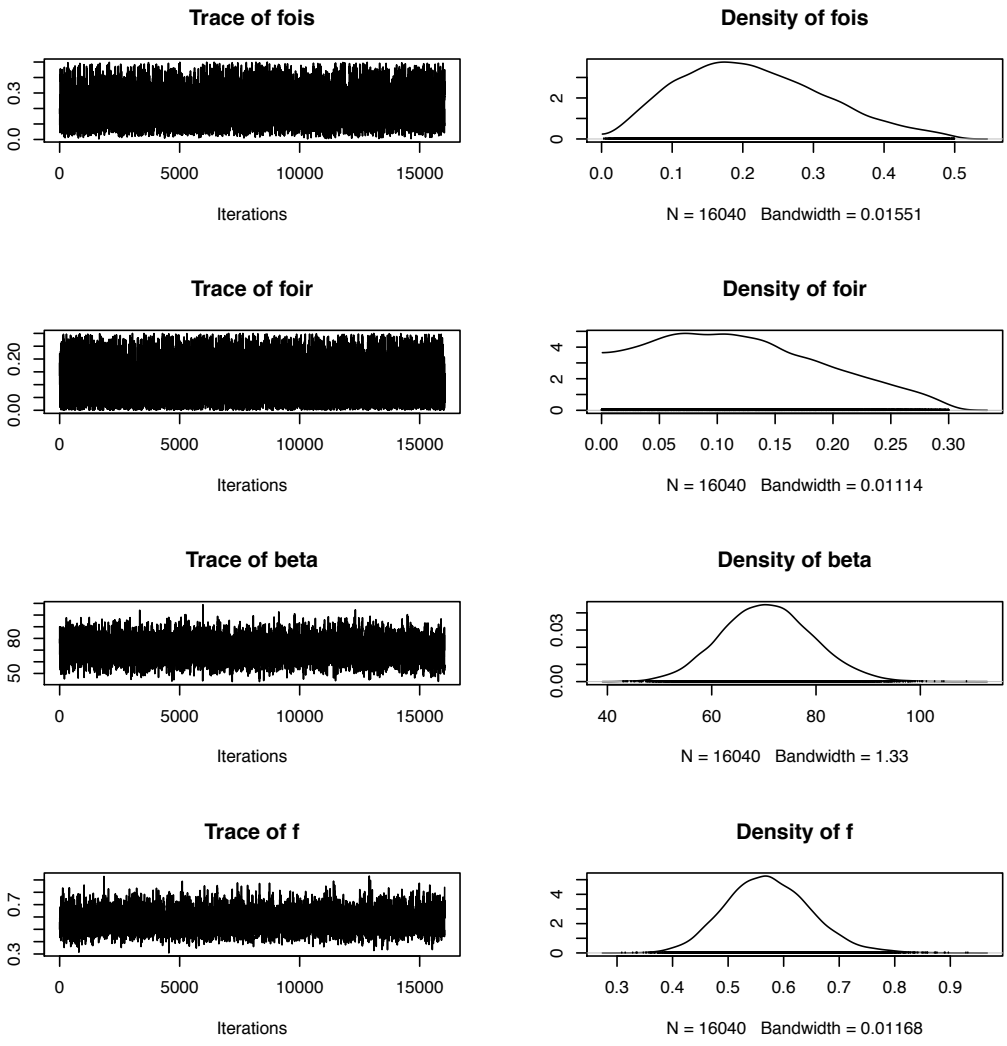


Figure 8: Trace and density plots for Model 3

6 Result: scenario analysis: Fit to data

Scenario analysis used the structure from Model 1 with altered parameters. All four could replicate the data from the household study as shown in Supplementary Figures 9 - 11.

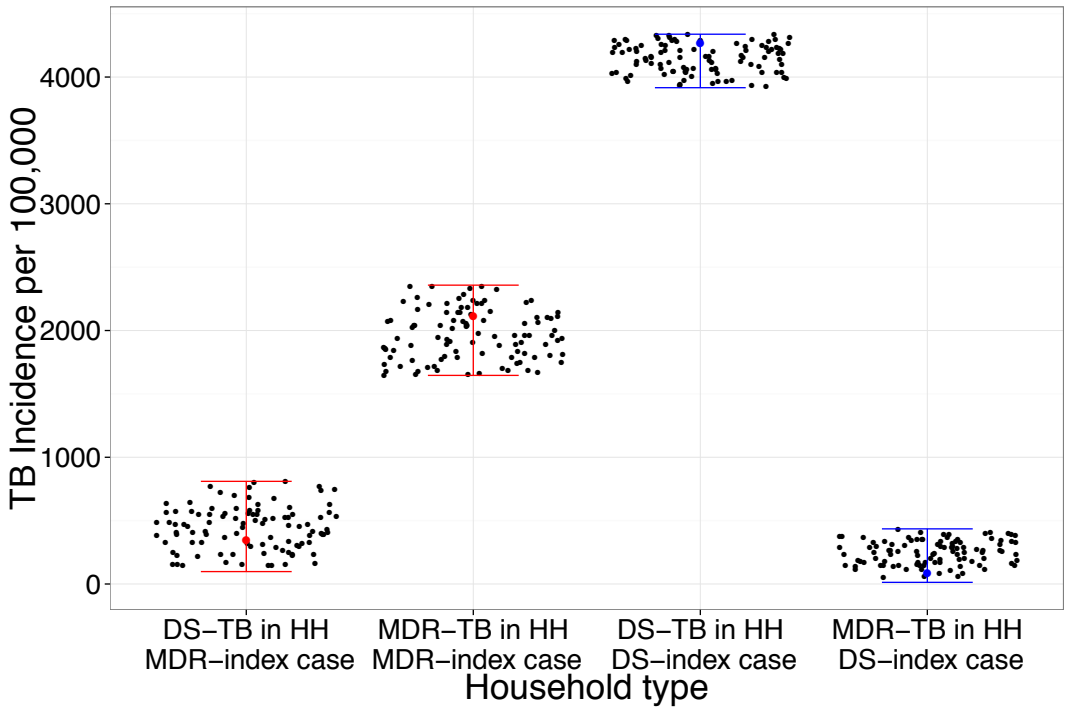


Figure 9: 100 example model fits. Black dots represent Model 1 output with scenario 1 parameters that matches to data shown in coloured ranges for each type of household (HH).

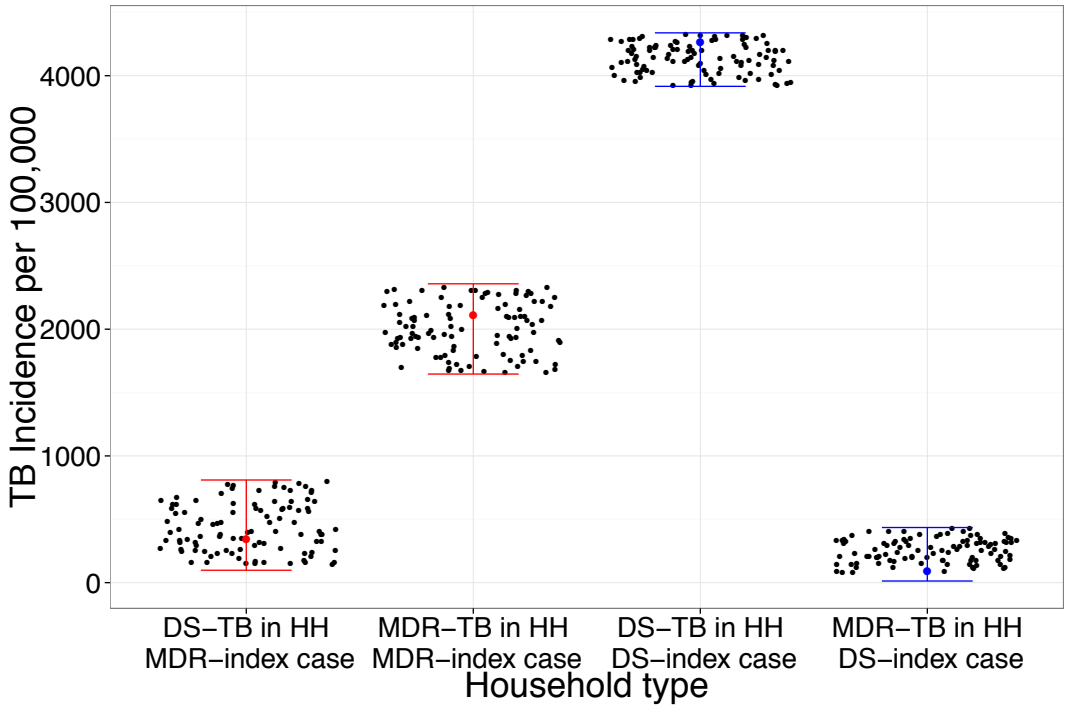


Figure 10: 100 example model fits. Black dots represent Model 1 output with scenario 2 parameters that matches to data shown in coloured ranges for each type of household (HH).

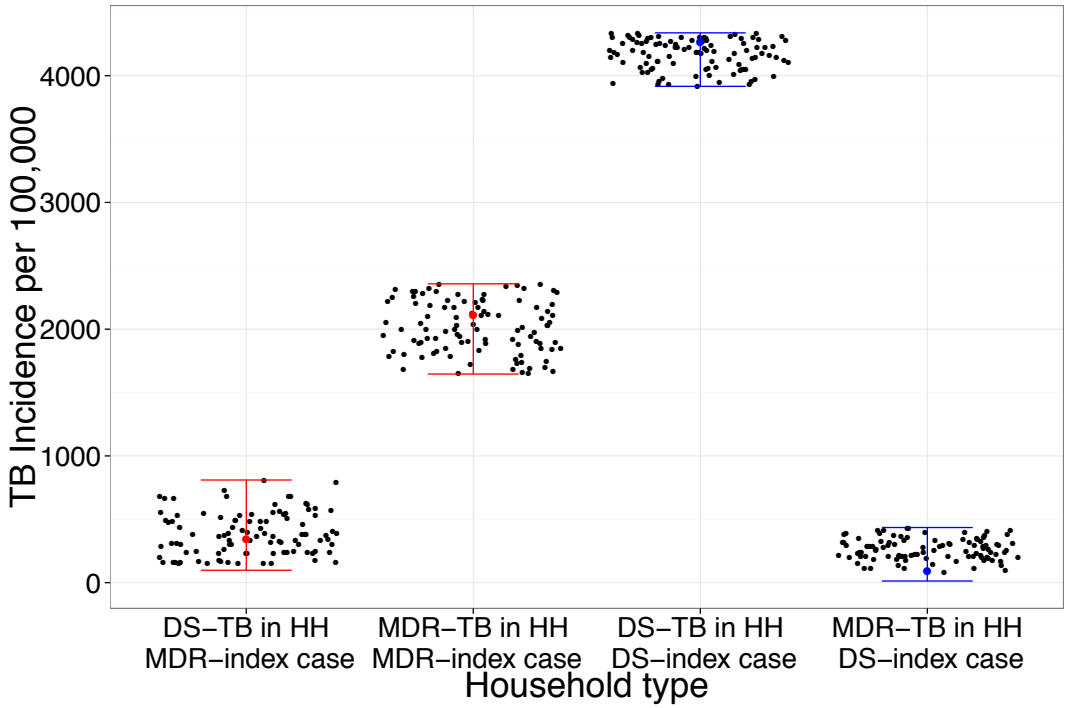


Figure 11: 100 example model fits. Black dots represent Model 1 output with scenario 3 parameters that matches to data shown in coloured ranges for each type of household (HH).

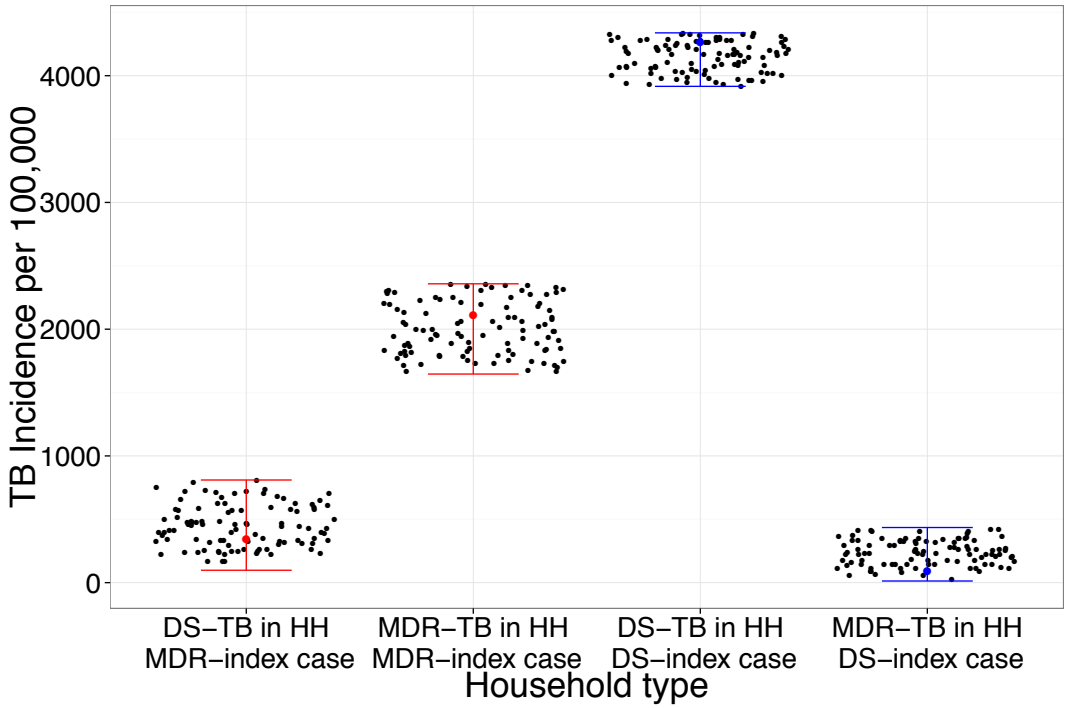


Figure 12: 100 example model fits. Black dots represent Model 1 output with scenario 4 parameters that matches to data shown in coloured ranges for each type of household (HH).

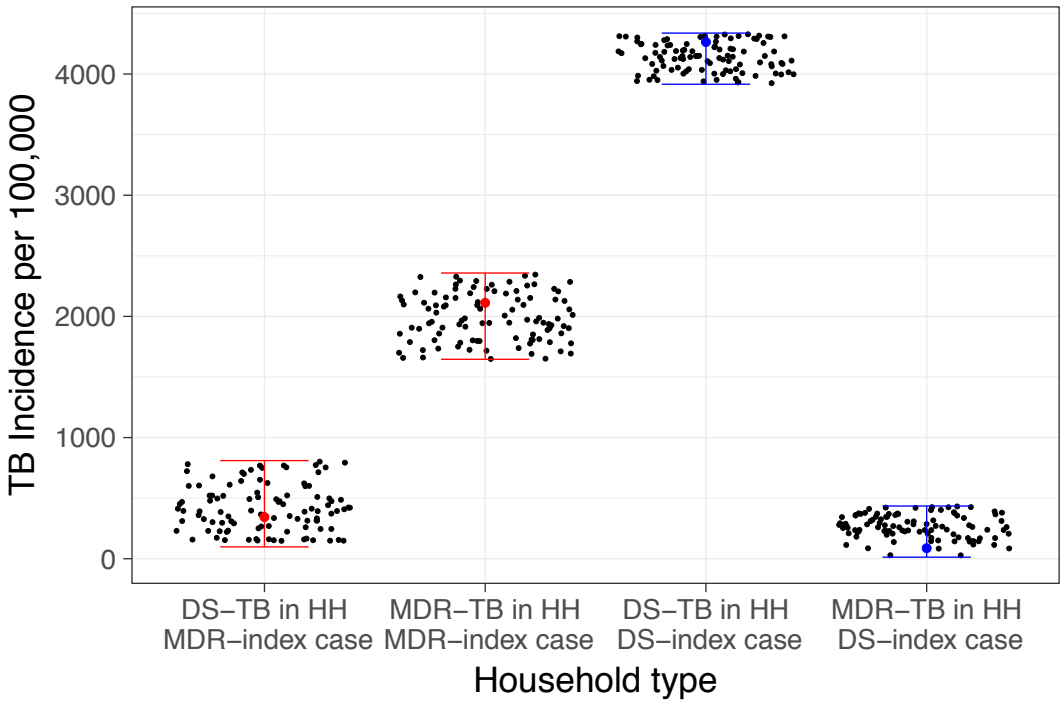


Figure 13: 100 example model fits. Black dots represent Model 1 output with scenario 5 parameters that matches to data shown in coloured ranges for each type of household (HH).

7 Trace and density plots for each unknown parameter for scenario analysis

The trace and density for each unknown parameter, from the three models are shown in Supplementary Figures 14-18.

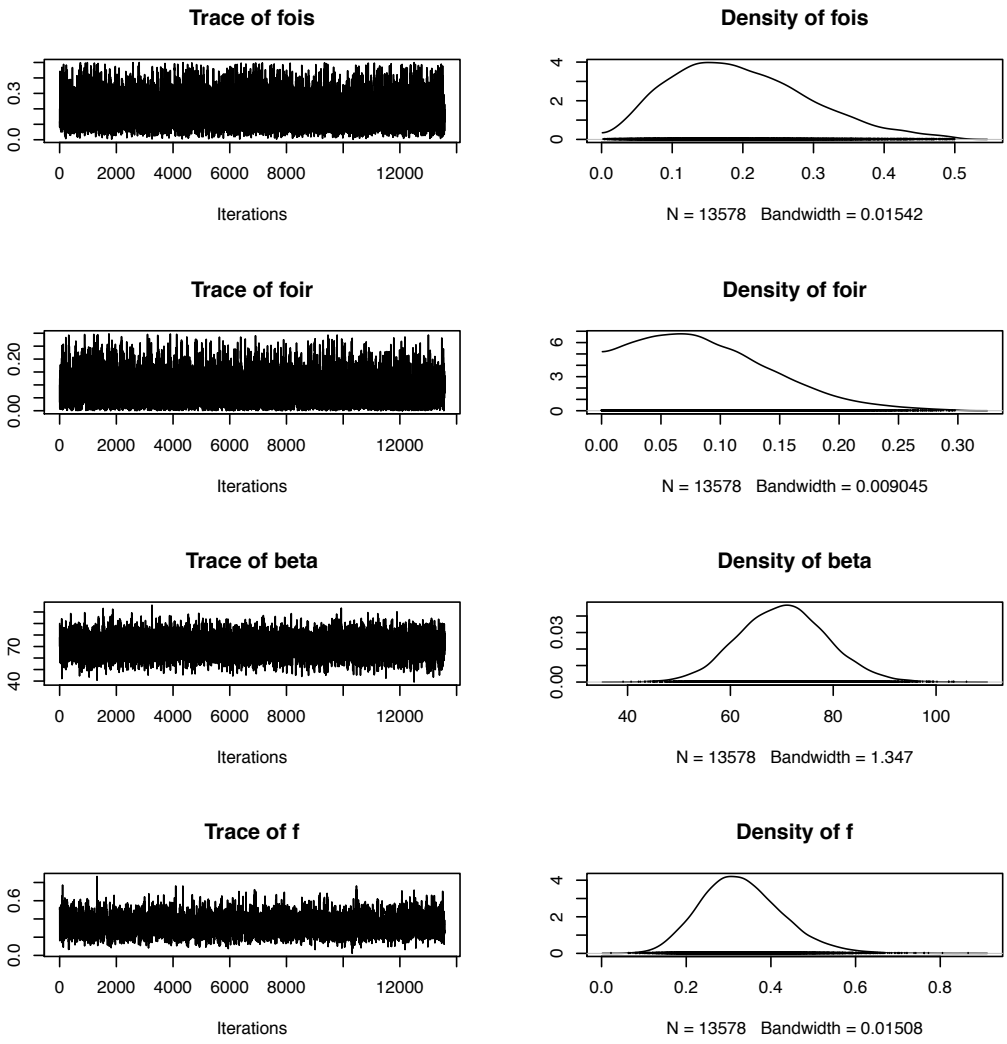


Figure 14: Trace and density plots for Model 1, scenario 1 (latent proportion)

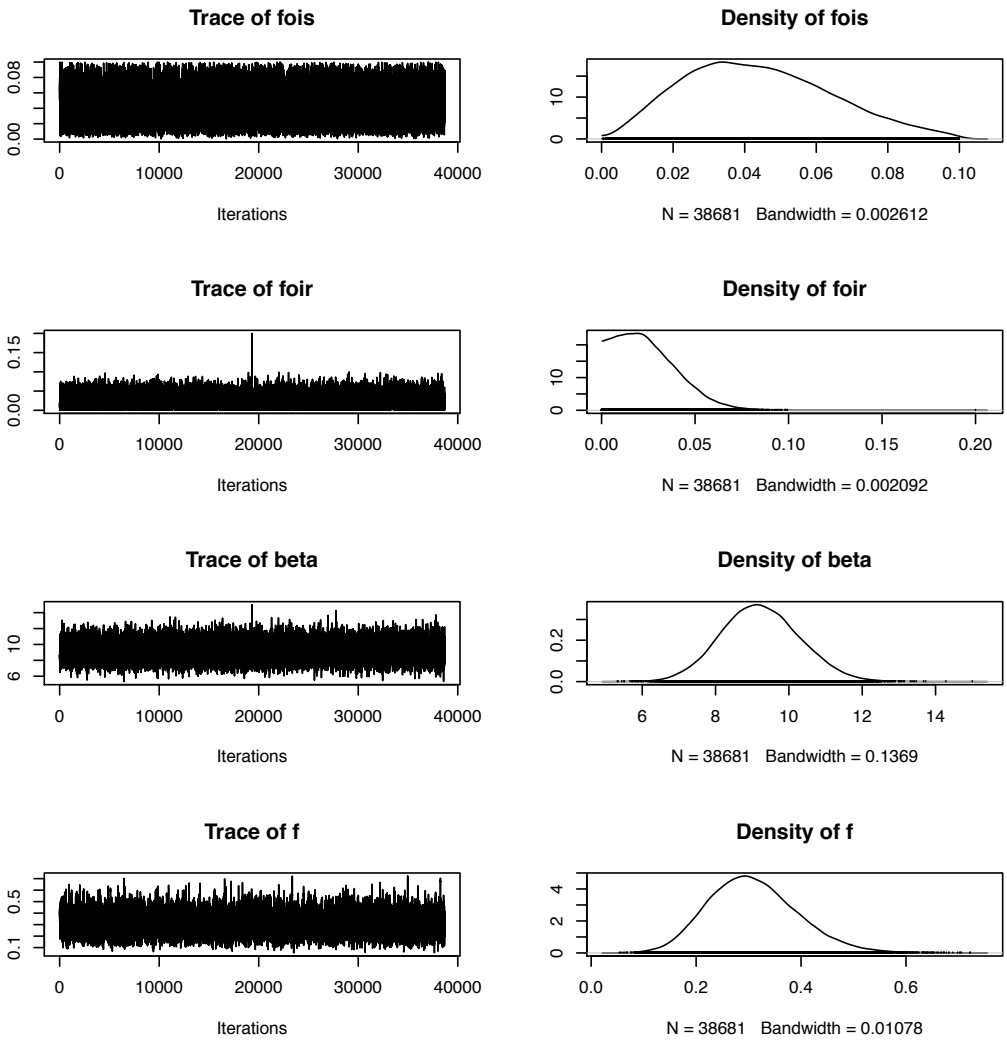


Figure 15: Trace and density plots for Model 1, scenario 2 (high TB incidence)

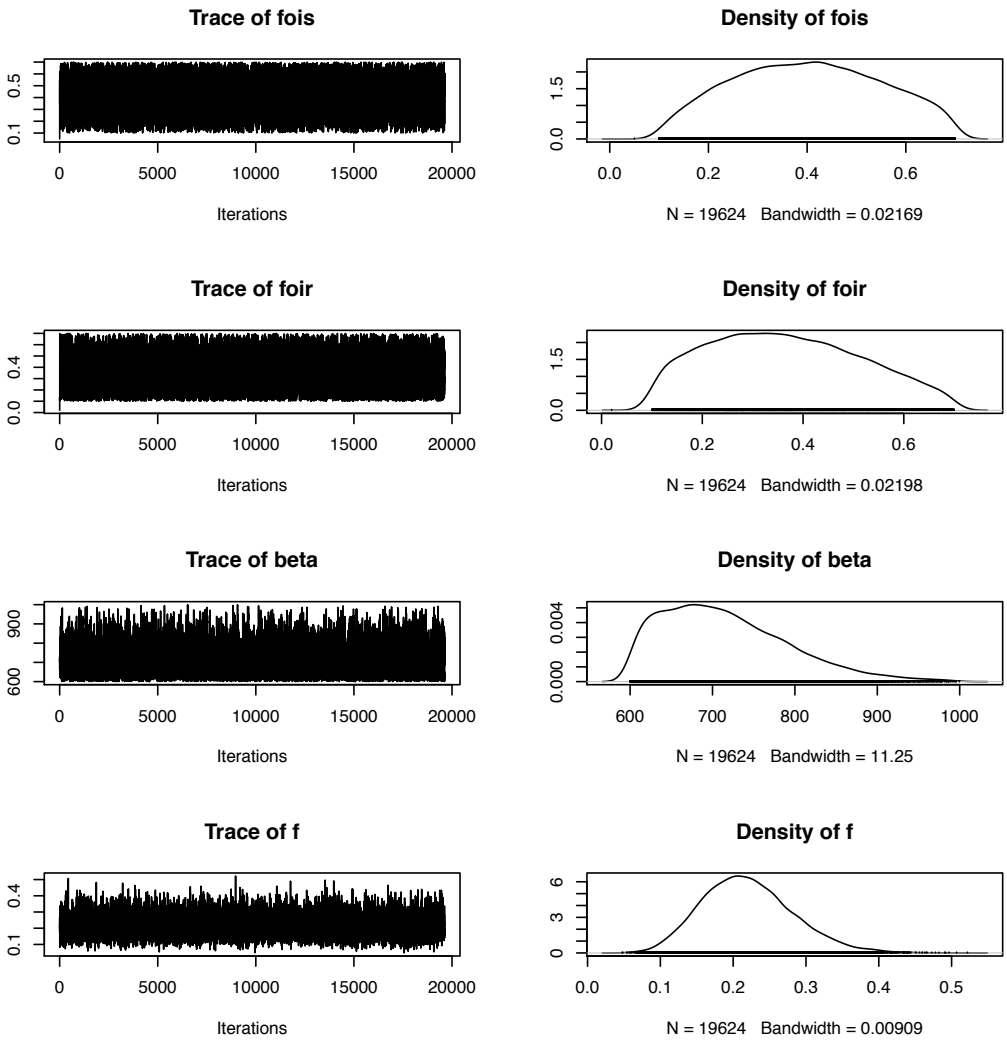


Figure 16: Trace and density plots for Model 1, scenario 3 (low TB incidence)

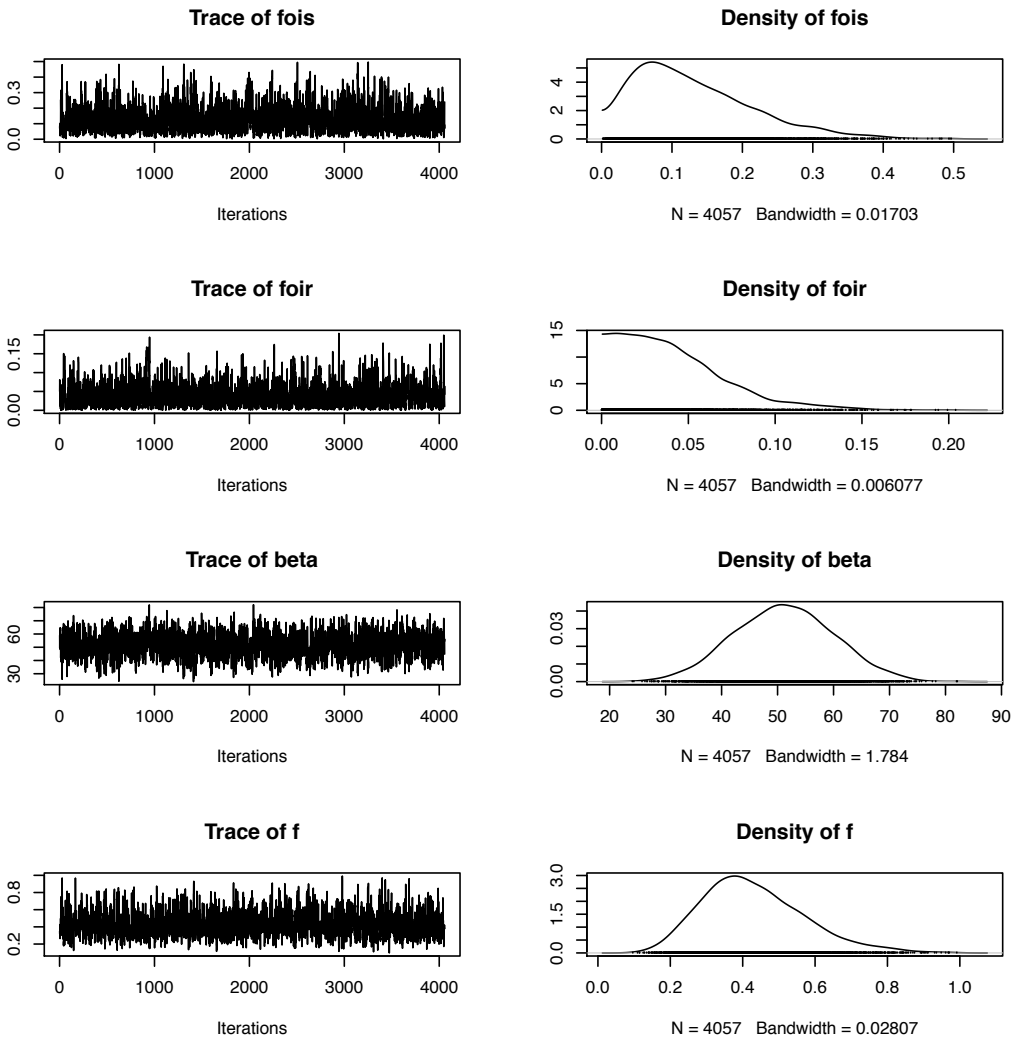


Figure 17: Trace and density plots for Model 1, scenario 4 (30 year burn in)

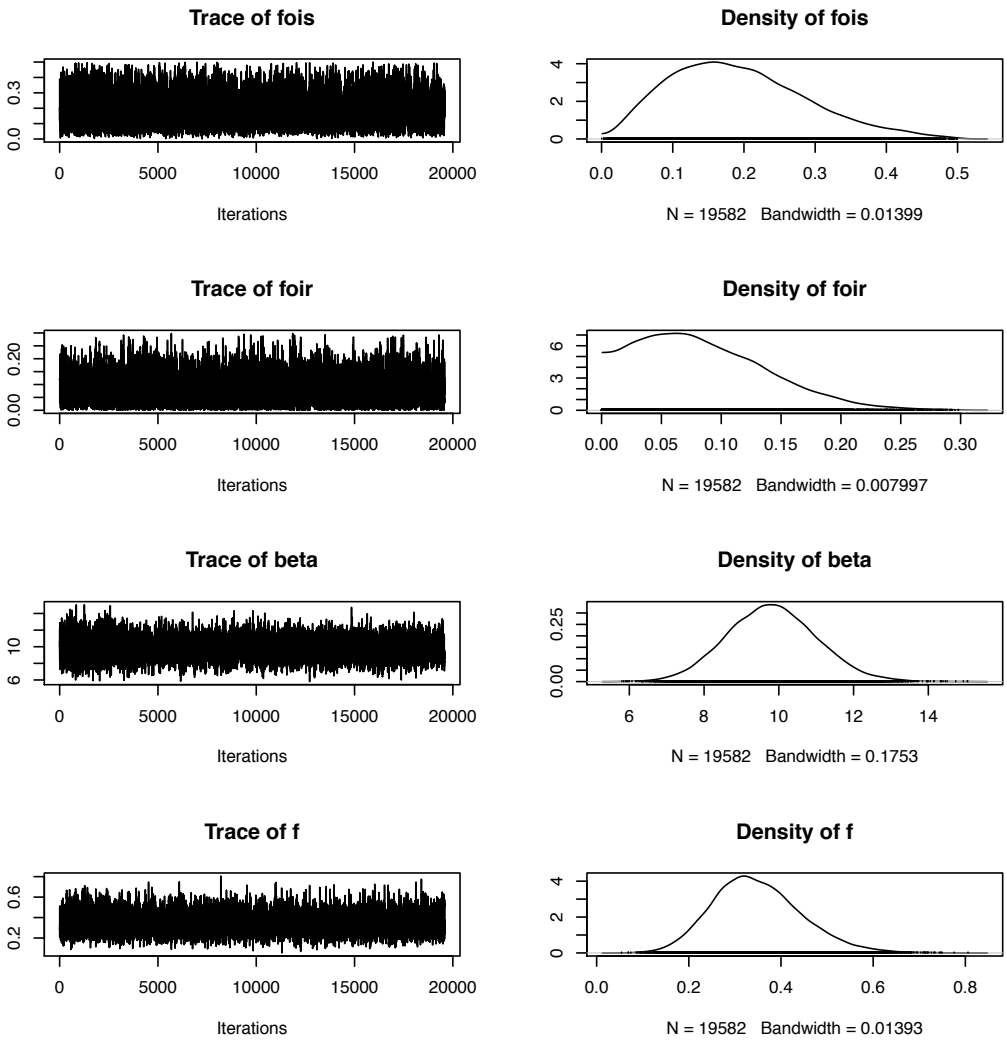


Figure 18: Trace and density plots for Model 1, scenario 5 (ho household saturation)

8 Scenario analysis results

The parameters estimates for the five scenarios are given in Table 1 and Figure 19.

Our first scenario analysis explored increasing the initial proportion of households that were initially infected with latent MDR-Mtb from 2% to 10% (in the pre-study). Fitting the four unknown parameters revealed that this increased MDR-Mtb latency proportion had very little impact on the estimates.

Our correlation analysis revealed four parameters (other than the four unknown parameters) to be correlated with TB incidence: the proportion of (re)infected individuals which progress to "latent fast" (p), the protection from developing active TB upon re-infection (), the proportion of new active cases which directly become infectious (d) and the progression rate of latent fast individuals to active disease (pf). The second scenario set these four parameters to be (p, , d, pf) = (0.25, 0.25, 0.75, 0.9) (high TB incidence) and the third (low TB incidence) to be (0.08,0.45,0.25,0.1). These second and third scenarios affected the estimates for the external force of infection and per capita transmission rate as would be expected due to the nature of the change in the natural history parameters. However, the estimates for the relative fitness (f) remain relatively consistent with our initial parameter set in Model 1 at approximately 0.30. Scenario 3 has a lower mean fitness at 0.22.

The fourth scenario, extended the initial run-in period from 10 to 30 years. All parameter estimates are similar to those of Model 1, including the relative fitness. The 95% credible intervals are larger as would be expected from the larger initial variation that will come from taking initial conditions from a 3x bigger run-in.

The fifth scenario removed the saturating household effect. The parameter estimates from this were also highly similar to the main analysis, except for the per capita transmission parameter, which was lower, reflecting the change to the model structure (no longer divided by household size).

Scenario	foi_s	foi_r	β	f
Model 1	$0.19(0.04-0.42)$	$0.08(0-0.22)$	$69.95(53.78-86.86)$	$0.33(0.17-0.54)$
1 (Greater proportion initially latently infected with MDR-TB)	$0.19(0.04-0.42)$	$0.08(0-0.22)$	$70.25(53.9-87.44)$	$0.32(0.16-0.54)$
2 (High TB incidence natural history parameters)	$0.04(0.01-0.09)$	$0.02(0-0.06)$	$9.2(7.21-11.38)$	$0.3(0.16-0.49)$
3 (Low TB incidence natural history parameters)	$0.4(0.14-0.67)$	$0.36(0.12-0.66)$	$704.47(606.88-896.37)$	$0.22(0.11-0.35)$
4 (30 years burn-in)	$0.11(0.02-0.32)$	$0.03(0-0.12)$	$51.55(34.91-67.04)$	$0.41(0.2-0.73)$
5 (No transmission saturation)	$0.18(0.04-0.41)$	$0.08(0-0.21)$	$9.8(7.54-12.24)$	$0.34(0.18-0.56)$

Table 1: Parameter estimates for the median and 95% credible intervals of the four unknown parameters from at least 100 5,000 MCMC runs for the five scenarios explored within Model 1.

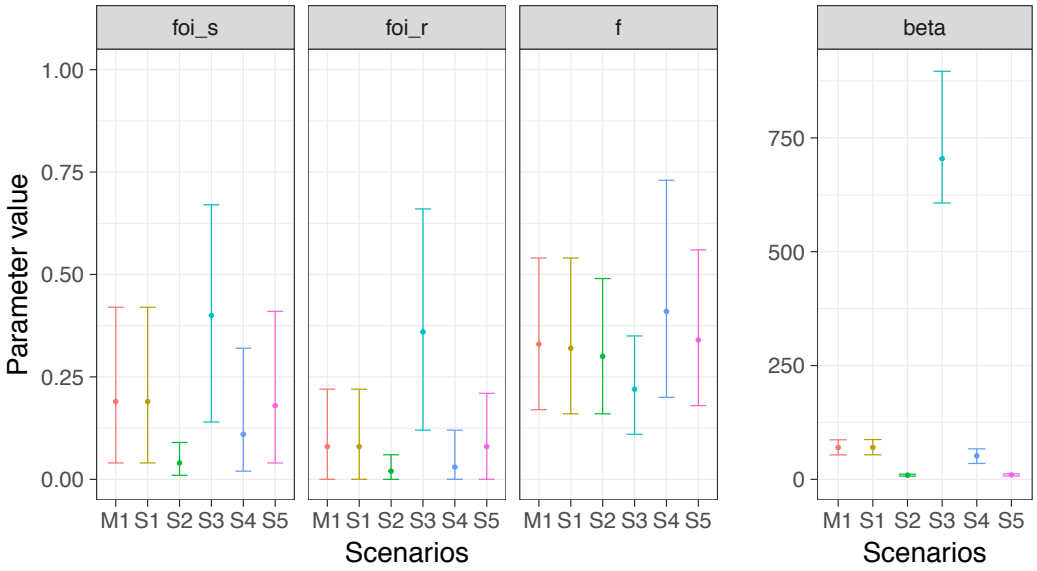


Figure 19: Fitted parameters for Model 1 and the five scenarios (S1-5). The units for the y-axis of the corresponding plots are: for the external forces of infection ('foi_s' and 'foi_r') proportion infected per year, for the relative fitness ('f') there are no units and for the per capita transmission rate ('beta') the units are effective contact rate per year.

References

[1] Grandjean L, et al. (2011) Tuberculosis in household contacts of multidrug-resistant tuberculosis patients. *The International Journal of Tuberculosis and Lung Disease* 15(9):1164–1169.

[2] Martinez L, et al. (2013) Changes in tuberculin skin test positivity over 20 years in periurban shantytowns in Lima, Peru. *The American journal of tropical medicine and hygiene* 89(3):507–515.

[3] Grandjean L, et al. (2015) Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. *PLoS Med* 12(6):e1001843.