

SOFTWARE

Open Access



# Modular programming for tuberculosis control, the “AuTuMN” platform

James McCracken Trauer<sup>1\*</sup>, Romain Ragonnet<sup>2</sup>, Tan Nhut Doan<sup>3</sup> and Emma Sue McBryde<sup>4</sup>

## Abstract

**Background:** Tuberculosis (TB) is now the world’s leading infectious killer and major programmatic advances will be needed if we are to meet the ambitious new End TB Targets. Although mathematical models are powerful tools for TB control, such models must be flexible enough to capture the complexity and heterogeneity of the global TB epidemic. This includes simulating a disease that affects age groups and other risk groups differently, has varying levels of infectiousness depending upon the organ involved and varying outcomes from treatment depending on the drug resistance pattern of the infecting strain.

**Results:** We adopted sound basic principles of software engineering to develop a modular software platform for simulation of TB control interventions (“AuTuMN”). These included object-oriented programming, logical linkage between modules and consistency of code syntax and variable naming. The underlying transmission dynamic model incorporates optional stratification by age, risk group, strain and organ involvement, while our approach to simulating time-variant programmatic parameters better captures the historical progression of the epidemic. An economic model is overlaid upon this epidemiological model which facilitates comparison between new and existing technologies. A “Model runner” module allows for predictions of future disease burden trajectories under alternative scenario situations, as well as uncertainty, automatic calibration, cost-effectiveness and optimisation. The model has now been used to guide TB control strategies across a range of settings and countries, with our modular approach enabling repeated application of the tool without the need for extensive modification for each application.

**Conclusions:** The modular construction of the platform minimises errors, enhances readability and collaboration between multiple programmers and enables rapid adaptation to answer questions in a broad range of contexts without the need for extensive re-programming. Such features are particularly important in simulating an epidemic as complex and diverse as TB.

**Keywords:** Disease transmission, infectious, Tuberculosis, Models, biological, Global health, Software, Tuberculosis, multidrug-resistant

## Background

The latest estimates for tuberculosis (TB) burden unequivocally identify the pathogen as the world’s leading infectious killer [1]. Despite the consistent claim that rates of disease are declining, the last three Global Tuberculosis Reports from the World Health Organization (WHO) estimate a greater burden of disease than that in the preceding year [1–3]. Although this may be attributable to improvements in case detection and better recognition of the scale of the problem, insufficient funding

for both control programs and for research threatens to derail even the modest current progress being made [1, 4]. Despite these obstacles, several technologies are now emerging, including novel diagnostics, medications and treatment regimens [5], while the post-2015 End TB Targets are an ambitious call to action [6]. Therefore, the field of TB control is now in a state of flux, with well-established methods of control competing against new technologies for a share of the limited TB control budgets of highly-endemic countries. Such countries need to understand better the likely epidemiological and economic consequences of programmatic decisions for TB control to guide strategic investment to maximise impact, particularly given that the highest burden countries are almost

\* Correspondence: james.trauer@monash.edu

<sup>1</sup>School of Public Health and Preventive Medicine, Monash University, 99 Commercial Road, Melbourne 3004, Australia

Full list of author information is available at the end of the article



universally those with the lowest budgets. The recent addition of a financial outcome (catastrophic costs) to the established disease burden targets further emphasises the importance of the economics of such decisions.

Mathematical models are powerful tools to evaluate programs for which evaluation through intervention is impractical, unethical or impossible [7]. Despite significant uncertainties regarding the best approach to structuring and parameterising dynamic models of TB transmission [8], such models are frequently and increasingly used to answer key policy questions in TB control at the national or sub-national level [9, 10]. Moreover, if such models are to be applied repeatedly across multiple regions of the world, they must be able to address the dramatic differences in TB drivers by context – such as HIV in Africa, migration in Western nations and drug resistance in Eastern Europe.

Our recent research has focused on simulating programmatic responses to TB to answer questions of relevance to TB control policy in high-burden countries. Such modelling is important to countries in the cyclical process of evaluating current programs, setting future priorities and planning the TB response. Both countries and funders increasingly view mathematical modelling as a key component of this cyclical process, which is described in detail by a recent publication on the TIME Impact model for TB. The TIME paper describes important principles in using modelling as a programmatic tool, including the need to partner with such countries to support the development of local epidemiological expertise and improve understanding of the epidemic [11]. Here we describe our development of a software platform to achieve these goals and our approach to the particular challenges in simulating the complex and heterogeneous global TB pandemic.

## Implementation

### Past approach

Our earlier TB modelling studies involved the construction of code and compartmental structure appropriate to the question at hand, but with limited capacity for model elaborations to be re-used in subsequent applications – an approach we believe to be common in infectious disease modelling. For example, our work in Western Province, Papua New Guinea incorporated a matrix of population mixing to capture differences between districts of the Province [12], while our modelling in Karakalpakstan, Uzbekistan allowed for mis-assignment of patients according to the extent of drug resistance of the infecting organism [9]. While some elements of our code could be re-used in subsequent applications, the need for extensive re-working of such models for each application was a significant inefficiency.

### Modular development

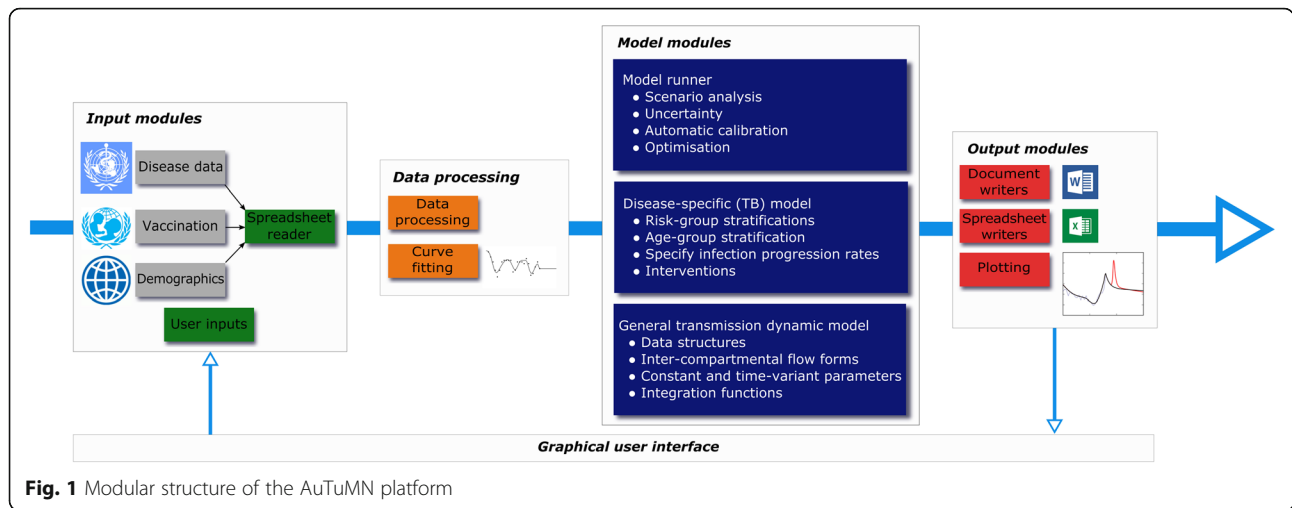
We fundamentally changed our approach to model development, incorporating principles from computer science and coding in open-source object-oriented Python (version 2.7). This is intended to make our software platform more reliable, testable, reusable and interpretable, as well as to expand the possible functions for which the model can be used and facilitate multi-person programming (via GitHub). The modular structure of our software platform (branded “AuTuMN”) is presented in Fig. 1. After user inputs have been received through the Graphical user interface module, the processes involved in model running proceed in a logical sequence from data processing through model execution to creation and display of model outputs. We used an object-oriented programming approach, with each module (represented by a rectangle in Fig. 1) being a single class. Each such module accepts and produces a consistent set of data structures for the linked modules, to avoid a situation in which adjustments to a single module requires adaptations of other modules. Default settings avoid errors when the user selects a certain option (e.g. age stratification) without specifying necessary associated inputs (e.g. the corresponding age breakpoints).

### Coding syntax

In addition to the modular development, we aimed for our code to be stylistically consistent and adhere to sound basic principles of computer programming. In addition to object-oriented and modular programming, these principles include extensive commenting and explanation of code, using functions wherever possible, using loops to avoid code repetition, employing consistent naming of variables and consistent structuring of data and object attributes. Advantages of employing such principles include ensuring that the code is more readable and interpretable by the multiple contributing programmers, reducing the possibility of errors and allowing functions developed for use in a certain part of the code base to be re-used elsewhere.

### Characteristics of the epidemiological and economic models

The general compartmental structure of the TB transmission dynamic model is presented in Fig. 2 and described in Section 1 of Additional file 1. As model stratification is intended to be variable by context and is determined by user inputs, it is not possible to present a universal flow diagram or list of differential equations. However, the structure presented depicts most of the main transition flows between compartments that are universal to all applications of the model. The epidemiological principles of model construction have been described in several of our previous publications and are



retained through all AuTuMN country applications. These include: partial vaccine efficacy [13], waning risk of progression to active disease with time from infection [14, 15], a three year average duration of active disease [16], distinguishing the process of detection from that of determining the drug resistance profile [9], amplification of drug resistance with default from treatment [14, 17] and a partial reduction in transmissibility for organisms with greater levels of drug resistance [9, 12].

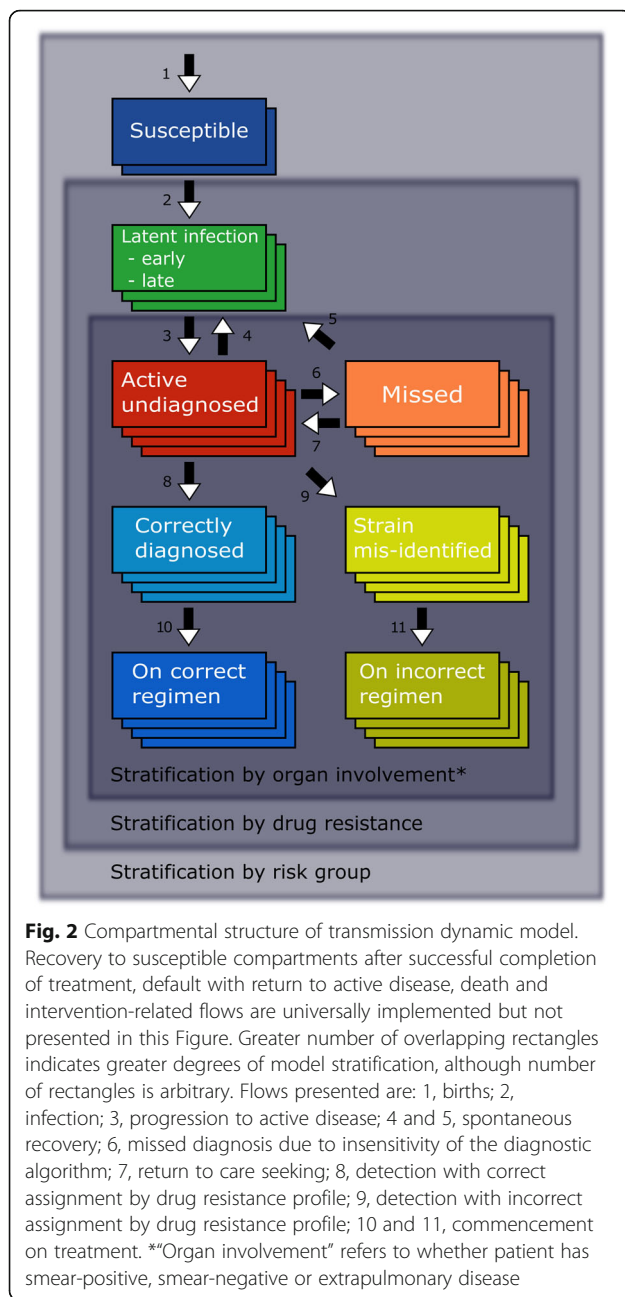
A logistic cost-coverage function links spending on programmatic interventions to coverage and subsequent epidemiological impact. We consider start-up costs as additional fixed costs for programs absent from the baseline scenario, because we are frequently comparing well-established interventions (e.g. BCG vaccination) against the implementation of newly proposed technologies (e.g. molecular diagnostics). Therefore, we assume a maximal cost-coverage gradient at zero spending, after start-up costs have been applied.

**Results**

Earlier iterations of our model (coded in Matlab™ without adherence to all the programming principles described above) were applied to inform TB control programs in five contexts: Western Province of Papua New Guinea [12], Karakalpakstan in Uzbekistan [9], and India, China and South Africa through the TB-MAC-coordinated “Targets” exercise [10, 18]. With support from the Global Fund to Fight AIDS, TB and Malaria, we were then commissioned to undertake policy-relevant modelling for the National Tuberculosis Programs (NTPs) of three further countries: Fiji, the Philippines and Bulgaria. Reports on the epidemiological and economic results for each of these three applications have been submitted to the NTPs of each country and we also aim to publish these in peer-reviewed journals in the coming

months in collaboration with our country partners, including detailed exposition of the epidemiological assumptions and parameter values made in each. The software allows for immediate adaptation of many features of the transmission model to each context through the Graphical user interface, such as addition of risk groups or strains and changes to the purpose of the simulation. The “General transmission dynamic model” module has been released as open-source software (see <https://github.com/popdynamics/popdynamics>), while a repository that presents the framework of each platform module is also available (see [https://github.com/jtrauer/AuTuMN\\_framework](https://github.com/jtrauer/AuTuMN_framework)). These repositories include code, dependencies and detailed user documentation. The characteristics of each module are presented in Table 1, with examples provided in the AuTuMN\_framework repository.

Examples of our coding approach are presented in Table 2 and Section 2 of Additional file 1: (Table S1). Note the consistent naming conventions. For example, parameter name strings are separated by underscores that describe first the source of the parameter (e.g. disease specific = ‘tb’, programmatic = ‘program’), then the nature of the parameter (e.g. rate = ‘rate’, proportion = ‘prop’), then the meaning of the parameter (e.g. spontaneous recovery = ‘recover’, untreated death = ‘death’) and lastly the stratum to which the parameter applies if applicable (e.g. smear-positive/negative/extrapulmonary status = ‘organ’). As the source, meaning, units and applicability of each parameter are implied by its name string, there is less need for detailed comments to explain the meaning of each parameter at each code line. Also, the division of the code into functions responsible for a very small component of the entire platform’s function limits the potential for errors and enables changes to be made that only affect one component of the platform. For example, changes to progression through stages of active



disease could be adjusted without the need to check that all inter-compartmental flows have been appropriately coded and that flows are conserved.

Figure 3 illustrates our approach of automatically fitting time-variant parameter values to data loaded through the Spreadsheet reader module and an example of variation in a programmatic parameter (vaccination coverage) for scenario simulation. After setting time-variant parameters, a Metropolis-Hastings algorithm is used to vary uncertain (time-constant) parameters in order to calibrate the model to epidemiological data on disease burden, with the

progression of parameter values displayed through the Graphical user interface in real time. Progression of model outputs over consecutive runs can also be displayed overlaid on calibration data from the Global TB Report displayed as shaded areas (Fig. 4). A similar approach to parameter variation is also used to quantify uncertainty in future predictions.

Figure 5 presents examples of our approach to implementing logistic cost-coverage curves for interventions and hence estimating costs over time for each program. The logistic function links an intervention's coverage to its associated cost and is presented in detail in Section 3 of Additional file 1. This function incorporates variables such as population sizes that are updated at each iteration of the model integration so that the economic and epidemiological modules interact continuously. The cost-coverage curves can be used in two directions: costing of programs when coverage values are specified, and adjustment of coverage based on input costs.

Outputs from our model include assessment of the baseline epidemiology, such as reconciling data for incidence, mortality, prevalence and notifications or explaining why such indicators appear mutually inconsistent. We also present estimates of the likely impact of planned or considered changes to the programmatic response based on our scenario simulations. Next, the epidemiological and economic differences between the baseline scenario and these counter-factual scenarios are then used for cost-effectiveness analysis. Last, an optimisation algorithm allows for minimisation of disease burden indicators over a specified time horizon, with such indicators including incidence, prevalence, mortality and proportion of TB multidrug-resistant in the overall population or sub-groups, as well as any weighted combination of these indicators. The current approaches to calibration and optimisation are described briefly in Table 1. However, given that each procedure is called from only one function of the Model runner module, alternative techniques could easily be substituted.

### Discussion

We present the development of our novel software platform and mathematical model of TB transmission. The significant change in computer programming philosophy that we adopted to tailor it to TB control simulations helps improve the flexibility, reliability and reusability of our software. Our modular approach allows for compartmentalisation of the software's functions, in order to improve these characteristics of the tool and allow for a broader range of functionality, including stratification by age, risk group, strain and organ involvement, historical understanding of the epidemic, uncertainty, automatic calibration, cost-effectiveness and optimisation. Dynamic models of infectious disease transmission have the major

**Table 1** Characteristics of modules

Module	Characteristics
Graphical user interface	<p>Accepts user inputs to determine:</p> <ul style="list-style-type: none"> <li>• Country to be simulated</li> <li>• Purpose of simulation               <ul style="list-style-type: none"> <li>◦ Uncertainty, optimisation, automatic calibration, scenario analysis                   <ul style="list-style-type: none"> <li>▪ Scenarios required (uncertainty or scenario analysis only)</li> </ul> </li> <li>◦ Saving and loading of previous simulations</li> </ul> </li> <li>• Compartmental model structure, including optional elaborations               <ul style="list-style-type: none"> <li>◦ Multiple health systems differing by quality of care</li> <li>◦ Amplification of drug resistance with treatment default</li> <li>◦ Mis-assignment of drug resistance status of infecting organism</li> <li>◦ Sub-populations and risk groups</li> <li>◦ Heterogeneous mixing between population groups</li> </ul> </li> <li>• Other methodological aspects of run               <ul style="list-style-type: none"> <li>◦ Integration method</li> <li>◦ Integration time step</li> <li>◦ Fitting method of parameters to loaded data</li> <li>◦ Epidemiological and economic parameters</li> </ul> </li> <li>• Outputs required</li> </ul>
Spreadsheet reader	<p>Reads data according to country selected in Graphical user interface          Accepts original format of spreadsheets as input (currently all Microsoft Excel™)          Converts data to consistent format (Python dictionary with years as keys and data entries as values)          Spreadsheets read:</p> <ul style="list-style-type: none"> <li>• Data on aggregate TB burden and programmatic response, WHO [45]</li> <li>• BCG vaccination coverage, UNICEF/WHO [46]</li> <li>• Demographic variables, the World Bank [47]</li> </ul>
Data processing	<p>Creates data structures with a format interpretable by the Model modules</p> <ul style="list-style-type: none"> <li>• Combines and reconciles external inputs from Spreadsheet reader with user inputs from GUI</li> <li>• Calls Curve fitting module to fit functions to reconciled data structures</li> <li>• Derives parameters determined by multiple input parameters           <ul style="list-style-type: none"> <li>◦ e.g. for comorbidities leading to an increased progression rate (such as diabetes) the risk group-specific progression rate is calculated by multiplying the age group-specific progression rate by the relative progression rate attributable to the risk factor</li> </ul> </li> </ul>
Curve fitting	<p>Derives polynomial spline functions to represent parameters (often interventions) scaling over time (See Fig. 3)</p> <ul style="list-style-type: none"> <li>• Fits to input data of dictionaries with keys time (in years) and values parameter values (often intervention coverage as a proportion)</li> <li>• Intervention values remain constant into the future from most recent parameter value           <ul style="list-style-type: none"> <li>◦ That is, the default behaviour (baseline scenario) is all interventions frozen at this value</li> </ul> </li> </ul>
Model runner	<p>Creates and runs model objects according to the purpose selected in the Graphical user interface</p> <ul style="list-style-type: none"> <li>• For scenarios, runs baseline model followed by each requested scenario with interventions as requested</li> <li>• For uncertainty, iteratively runs model, updating uncertainty parameters between each run           <ul style="list-style-type: none"> <li>◦ Currently uses a Metropolis-Hastings algorithm</li> <li>◦ Priors are estimated from the distributions of included epidemiological parameters</li> <li>◦ Posteriors are estimated from a comparison of outputs to WHO data</li> </ul> </li> <li>• For automatic calibration (an extension of uncertainty), iteratively runs model, updates model parameters, starting populations and other epidemiological parameters</li> <li>• For optimisation, estimates epidemiological outputs from proportionate allocation of funding across programs given a certain funding envelope           <ul style="list-style-type: none"> <li>◦ Currently uses SLSQP from the “minimize” function of the scipy.optimize package</li> <li>◦ Considers proportional funding to interventions to be the bounds</li> <li>◦ Considers the function to be minimised to be the epidemiological output of interest (usually incidence or mortality) when scenarios are run with varying funding allocation from a fixed/calibrated baseline</li> </ul> </li> </ul>

**Table 1** Characteristics of modules (*Continued*)

Disease-specific (TB) model <sup>b</sup>	Defines stratifications and their interaction by: <ul style="list-style-type: none"> <li>• Age</li> <li>• Comorbidity and/or population risk group</li> <li>• Organ involvement (smear-positive pulmonary, smear-negative pulmonary and extrapulmonary)<sup>a</sup></li> <li>• Drug resistance of infecting strain<sup>a</sup></li> <li>• Health system quality<sup>a</sup></li> </ul> Sets inter-compartmental flows, for: <ul style="list-style-type: none"> <li>• Ageing</li> <li>• Natural progression through stages of infection and disease</li> <li>• Detection (by each stratum of health system quality, if applicable)</li> <li>• Drug resistance status assignment by health system</li> </ul> Implements interventions selected from scenarios requested in Graphical user interface <ul style="list-style-type: none"> <li>• Estimates economics of interventions using logistic cost-coverage curves and population sizes</li> </ul>
General transmission dynamic model	Defines fundamental structures of transmission dynamic model which are not pathogen-specific (i.e. components common to any deterministic, compartmental, ordinary differential equation-based model of population-level infectious disease transmission), including: <ul style="list-style-type: none"> <li>• Compartments</li> <li>• Inter-compartmental flows</li> <li>• Fixed parameters</li> <li>• Variables to be updated at each integration time step</li> </ul>
Output modules	Creates Word™ and Excel™ tables of epidemiological and economic outputs and graphical (PNG) figures to illustrate: <ul style="list-style-type: none"> <li>• Compartmental model structure (using “graphviz” repository)</li> <li>• Time-variant parameters, including fit to input data (see Fig. 3)</li> <li>• Other illustrations of epidemiological implementation of interventions             <ul style="list-style-type: none"> <li>◦ e.g. visualisation of matrix of population mixing</li> </ul> </li> <li>• Aggregate model outputs compared to Global TB Report estimates, for:             <ul style="list-style-type: none"> <li>◦ Scenarios</li> <li>◦ Uncertainty (see Fig. 4)</li> </ul> </li> <li>• Model outputs by risk groups</li> <li>• Optimised funding distribution across interventions</li> </ul>

<sup>a</sup>Note that not all types of stratification apply to all compartments (e.g. susceptible population not stratified by drug resistance of infecting organism), see Fig. 2.

<sup>b</sup>This module is not a stand-alone class, but instead inherits general methods from the General transmission dynamic model module, adding TB-specific methods to the class. Abbreviations: TB, tuberculosis; PNG, portable network graphics; UNICEF, The United Nations International Children's Emergency Fund; WHO, World Health Organization.

advantage that they are able to capture the non-linear positive feedback loop of greater disease burden leading to more transmission and, in turn, to greater disease burden. However, such models must be fit for purpose and the challenges in developing a realistic tool for simulating TB transmission are considerable. Although mathematical modellers have long used computer programming to produce numerical simulations of infectious disease transmission, this is most often undertaken by small groups of researchers or individual academics and so is often limited to analyses whose end result is expected to be a single journal article.

Model flexibility is critically important in simulating the TB epidemic across multiple settings, because of the heterogeneity in drivers of the TB epidemic – arguably more than for any other disease. For example, HIV coinfection is a critical driver of the high rates of disease in sub-Saharan Africa [19, 20], drug resistance reaches alarming levels in many countries of Eastern Europe [21, 22] and the Asia-Pacific Region is characterised by a huge absolute caseload and a significant proportion of

care delivered outside of the public sector [1, 23]. Therefore, the characteristics of the epidemic are dramatically different in these settings, as are the programmatic interventions that policy-makers from these regions wish to understand better. Moreover, the epidemiology of TB immunity, transmission and progression differs dramatically according to several population strata, adding further to the degree of model complexity necessary to capture the epidemic realistically. Such strata include age groups, which affect immunity following vaccination [24], infectiousness [25] and rates of progression from infection to active disease [15]. Similarly, comorbid conditions that affect rates of progression from infection to active disease can have major epidemiological effects, such as HIV [26] and diabetes [27], while the impacts of factors such as poverty and overcrowding are more difficult to quantify but likely to be equally important [28]. Further, patients with smear-negative pulmonary or extrapulmonary disease are known to have significantly lower infectiousness than smear-positive pulmonary disease [29] and treatment

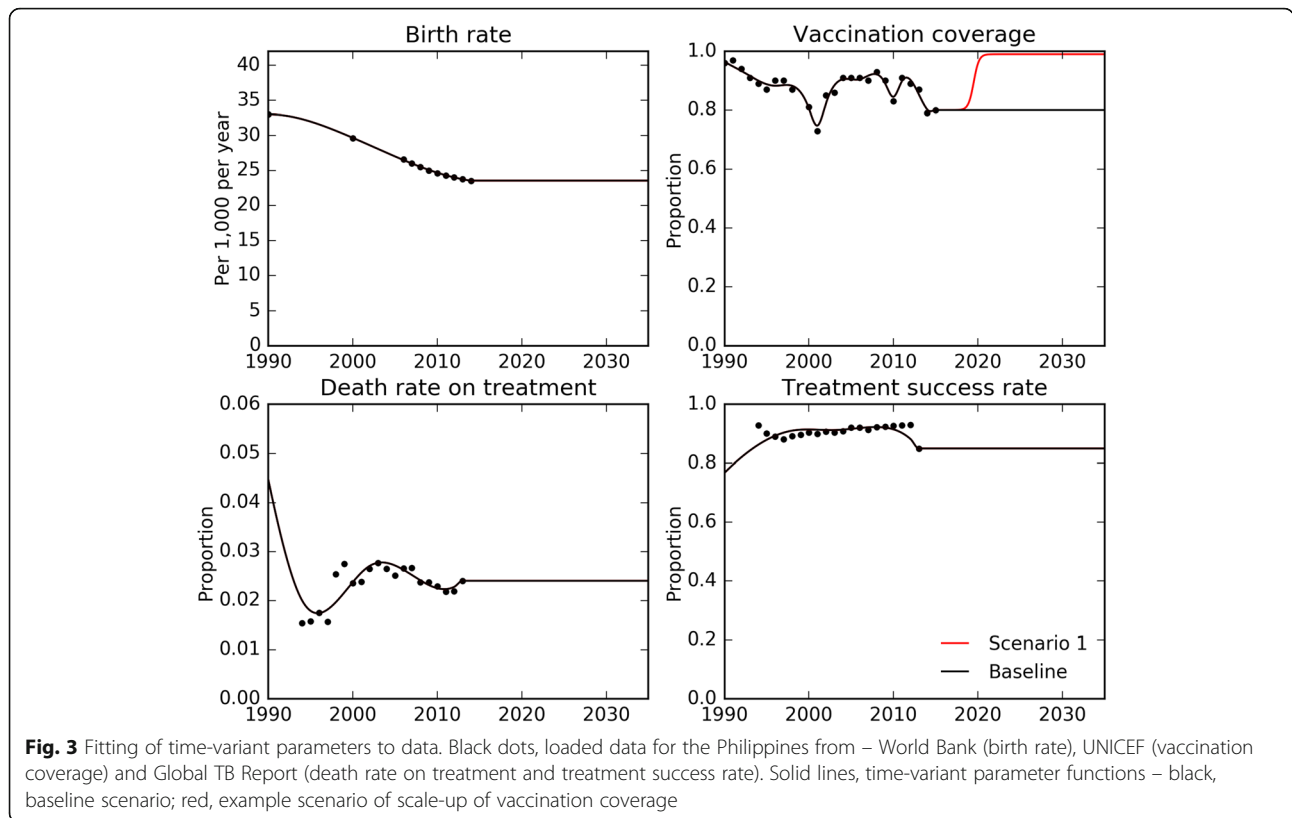
**Table 2** Illustration of approach coding

<p><b>Current Python code</b></p> <pre> def set_natural_history_flows(self):     """     Set flows for progression through active disease to either recovery or death.     """      # Determine the compartments to which natural history flows apply     active_compartments = ['active', 'missed']      # Apply flows     for agegroup in self.agegroups:         for riskgroup in self.riskgroups:             for strain in self.strains:                 for organ in self.organ_status:                     for compartment in active_compartments:                          # Recovery                         self.set_fixed_transfer_rate_flow(compartment + organ + strain + riskgroup + agegroup,   'latent_late' + strain + riskgroup + agegroup,   'tb_rate_recover' + organ)                          # Death                         self.set_fixed_infection_death_rate_flow(compartment + organ + strain + riskgroup + agegroup,   'tb_rate_death' + organ)                     </pre>
<p><b>Code explanation</b></p> <p>This code defines a method to the Disease-specific (TB) Model class that sets inter-compartmental flows relating to spontaneous recovery from active TB and TB-related death before treatment has been commenced by the health system. The method applies equally to patients with active, undetected TB disease, as well as those who had recently presented to the health system for assessment, but for whom the diagnosis had been missed as a result of insufficient sensitivity of the diagnostic algorithm employed.</p> <p>If this instance of the model object is stratified by age group, the flows apply equally to each age group, if the model object is stratified by comorbidity or population sub-group, the flows apply equally to each risk group, if the model object is stratified by drug resistance of the infecting organism, the flows apply equally to each “strain” and if the model object is stratified by smear status or pulmonary/extrapulmonary status, the flows apply differently to each “organ” (such that smear-positive pulmonary TB has a higher case fatality).</p> <p>The method “set_fixed_transfer_rate_flow” is declared in the General transmission dynamic model module and so available to the TB model through nested inheritance. It specifies that this constant inter-compartmental flow must be set to a value contained in the model parameters attribute (obtained from the third argument to this method) and that its value should be deducted from the origin compartment (first argument) and incremented to the destination compartment (second argument). The method “set_fixed_infection_death_rate_flow” has similar characteristics, except for having no destination compartment and so only two input arguments (compartment to which flow applies and value for flow rate).</p> <p>Note that model attributes specifying age, risk, strain and organ stratification are consistently specified lists of strings for each stratum (e.g. [‘age0to5’, ‘age5to15’, ‘age15up’]). Note also that the code will run if any of these stratifications are not requested, in which case the model attribute relating to that stratification is an iterable list containing only an empty string (i.e. [“]).</p>
<p><b>Previous closest equivalent code</b></p> <pre> eps.*LA+nu.*LB+(1-eta).*omega.*T-(gamma+delta+mui).*I;... %5. I compartment     </pre> <p>This line of Matlab code is taken from our group’s first paper describing a simpler compartmental model and its basic characteristics (see Trauer et al.[14]). It specifies the differential equation for the rate of change in the active infectious compartment, and so is not directly equivalent to the Python code presented above. Note that it is difficult to infer the meaning of the code due to the approach to variable naming.</p>

outcomes differ markedly according to the drug resistance of the infecting organism [30].

These differences are important to TB models; both to accurately capture the underlying epidemiology and because interventions often act differentially across such disease categories. However, simultaneous model implementation of several such factors can be challenging and lead to high levels of complexity, as the number of

compartments to track may scale multiplicatively with the number of stratifications included. Therefore, function-based code that consistently and automatically implements such strata into models regardless of their compartmental structure and other stratifications, add much to the flexibility, reliability and reusability of the software platform. For example, ageing rates between sequential age groups are set automatically once the



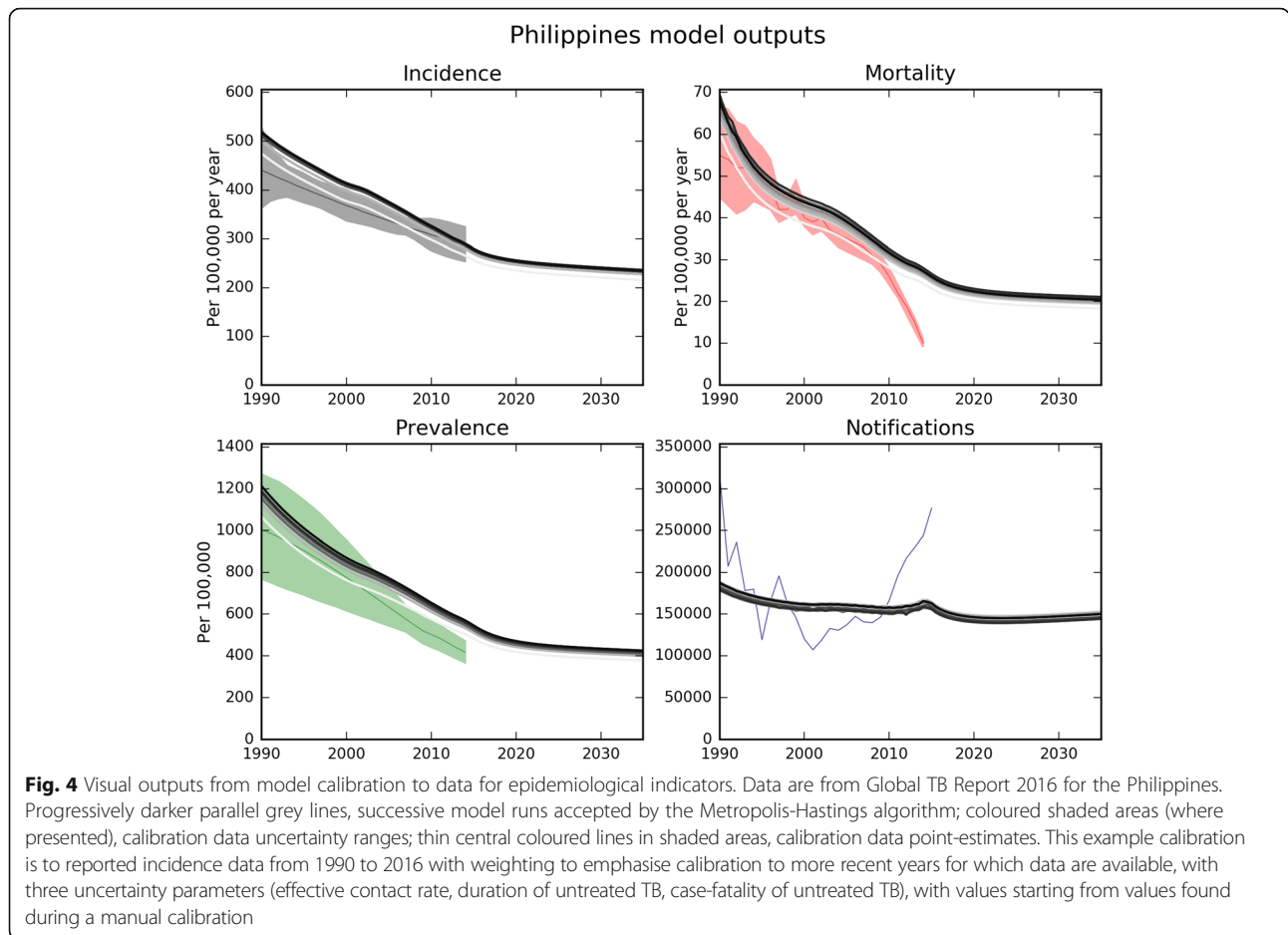
user has requested a set of breakpoints between age groups.

Tuberculosis is an ancient disease, as well as being a slow-moving epidemic due to the propensity of the infecting organism to re-activate after many years to decades of dormancy [31]. Therefore, accurately capturing the historical dynamics of the disease is of greater importance than for many acute infections, as interventions and dynamics are dependent on disease burden many decades earlier and can be markedly affected by demographic trends. In our analyses, we have consistently found the size of the latent pool to be a key driver of the future disease burden, and that this pool of infection limits the effectiveness seen with interventions directed at active cases alone. To address this issue, all our analyses commence from many decades to centuries into the past and aim to capture the dynamics of the TB response from the point at which they commenced. Time-variant parameters representing this response are fit to freely available officially reported data, and a zero value is added automatically at the time the technology first became available. Similarly, loading of disease burden outcomes from the WHO is important to ensure that model calibrations capture the progression of the epidemic over the course of the time period for which data are available. The General transmission dynamic model is intended to apply equally to any infectious

disease and examples are provided for measles, influenza and TB. The issues specific to ancient, slow-moving endemic infections and TB in particular have partially informed the broader modular development of our tool, although most of the modules could be adapted to any infection. We aim to release further modules as possible, such as the Curve fitting and Spreadsheet reader modules, and to extend the capability of the General transmission dynamic model. However, given the major differences between interventions requested, population-level risk factors and epidemiological context in each country, we make significant adaptations to the Disease-specific model module at each application, and so have no short-term plan to release this module.

A number of modular software platforms exist to estimate future burden of disease. Perhaps best known is the Spectrum suite of models, which simulate population demographics and disease progression, although these are not necessarily underpinned by transmission dynamic models [32, 33]. For influenza, complex microsimulation-based models underpinned by realistic synthetic population structures and explicit transmission dynamics have been made available to policy-makers [34, 35]. Models designed to simulate strategies to combat the huge global burden of HIV have also proliferated, and include the Asian Epidemic Model [36, 37], Optima [38] and the Estimation Projection Package [39–41]. For TB, transmission

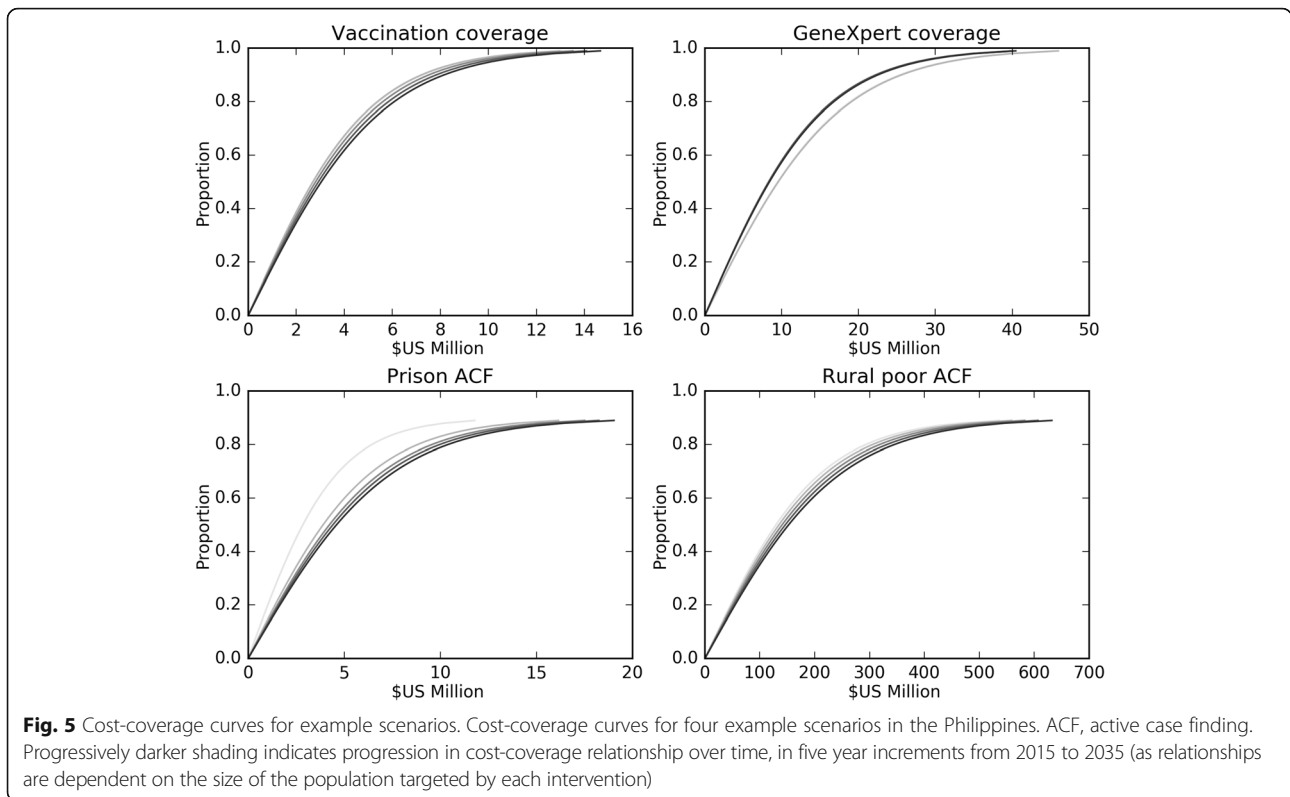




dynamic model-based tools have been tailored to consider specific interventions for control, such as active case finding [42] and diagnostic strategies [43, 44]. However, such tools are not readily adaptable to consider a broader range of interventions or to compare interventions that act at very different points in an individual’s journey through infection, disease, detection and treatment. A more general TB modelling tool that has been used across several countries worldwide is TIME Impact, which incorporates a transmission dynamic model and is integrated into the Spectrum suite of modelling tools [11]. The most important epidemiological advances of AuTuMN by comparison to TIME relate to flexibility. TIME describes a fixed approach to latency (single latent compartment with bypass), drug-resistant strains (MDR-TB and DS-TB), age groups (5-year age bins), organ status (smear-positive and smear-negative) and comorbidities (HIV). By contrast, AuTuMN permits greater versatility in relation to each of these epidemiological stratifications, while additionally incorporating uncertainty, calibration and optimisation capacity through the Model runner module.

Complicated modelling tools have the disadvantages of increased computational expense (i.e. processing time),

potential for errors and increased difficulty for the user to understand and interpret all the dynamics operating within a particular simulation. We minimise the potential for coding errors through consistent coding, minimisation of code repetition and built-in checks throughout our modules. Although we frequently use a highly stratified version of AuTuMN, involving hundreds to thousands of compartments, some or all stratifications can immediately be removed to facilitate the user’s understanding of the underlying epidemiological processes at play. Also, we build in intermediate outputs to model running, in order to facilitate understanding of such processes, such as mixing matrix diagrams and population sizes by sub-group. Despite the rapidly increasing speed of modern computers and methods for decreasing computation time, computation time remains an issue in our model, particularly for uncertainty calculations. Whereas optimisation and scenario analyses typically require the full model to be run forward over the simulation window from a single baseline run (typically twenty to thirty years), uncertainty requires repeated runs from the start of the simulation (typically several decades to centuries). For example, explicit Euler model integration from the start of simulation time with



three organ status strata, two strains (DS-TB and MDR-TB), two health care strata differing by quality and five population risk groups takes around 40 s for a full run and ten seconds for each additional scenario run on a standard personal computer (Intel i7 2.6GHz processor, 8 GB RAM). Run-time scales approximately multiplicatively with the number of stratifications for each type of stratification, while the time taken for initialisation remains approximately unchanged at around 7 s.

### Conclusions

We employed several principles of software engineering to develop a robust and flexible tool to support TB control decisions. The tool is underpinned by a transmission dynamic model and is rapidly adaptable to capture many of the complexities of the TB epidemic across a range of settings and the interventions available in the fight against this critically important infectious disease. Such tools are of increasing importance as TB emerges as the world’s leading infectious threat, ambitious new targets for its control have been set and old technologies vie with new tools for limited budgets.

### Availability of data and materials

The code for the “General transmission dynamic model” module has been released as open-source software, including the full code, dependencies and detailed user documentation.

**Project name** popdynamics.

**Project home pages** <https://github.com/popdynamics/popdynamics>, [https://github.com/jtrauer/AuTuMN\\_framework](https://github.com/jtrauer/AuTuMN_framework)

**Operating system** Platform independent.

**Programming language** Python version 2.7.

**Other requirements** numpy, scipy, graphviz.py, graphviz binary on path: <http://www.graphviz.org/>, matplotlib.

**License** Open source.

### Additional file

**Additional file 1:** Notes on structure of epidemiological model, further code examples and notes on economic model. (DOCX 1113 kb)

### Abbreviations

ACF: Active case finding; AuTuMN: The Australian Tuberculosis Modelling Network (see [www.tb-modelling.com](http://www.tb-modelling.com)); BCG: Bacille Calmette Guérin; HIV: Human immunodeficiency virus; PNG: Portable network graphics; TB: Tuberculosis; TIME Impact: A new user-friendly TB model to inform TB policy decisions, trade mark (see <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0608-4>); TM: Trade mark; UNICEF: United Nations International Children’s Emergency Fund; WHO: World Health Organization

### Acknowledgements

We gratefully acknowledge the staff of the National Tuberculosis Programs, Departments of Health and other affiliated institutions of the countries with which we have worked. In particular, Fiji, Bulgaria and the Philippines.

### Funding

This work was funded through the following grant by the Global Fund to Fight AIDS, TB and Malaria: "AuTuMN Technical Support to Enhance Allocative Efficiency of Country Tuberculosis Program".

### Authors' contributions

JMT, RR and TND coded the software platform. JMT produced the first draft of the manuscript. ESM supervised and coordinated the project. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>School of Public Health and Preventive Medicine, Monash University, 99 Commercial Road, Melbourne 3004, Australia. <sup>2</sup>The Burnet Institute, 85 Commercial Road, Melbourne 3004, Australia. <sup>3</sup>Department of Medicine, Clinical Sciences Building, the Royal Melbourne Hospital, Parkville 3050, Australia. <sup>4</sup>Australian Institute of Tropical Health and Medicine, James Cook University, Townsville 4811, Australia.

Received: 22 March 2017 Accepted: 28 July 2017

Published online: 07 August 2017

### References

- World Health Organization. Global Tuberculosis Report 2016. Geneva: WHO; 2016. Available from: <http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf?ua=1>.
- World Health Organization. Global Tuberculosis Report 2014. Geneva, Switzerland: WHO; 2014. Available from: [http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf).
- World Health Organization. Global tuberculosis report 2015. Geneva, Switzerland: WHO; 2015. Available from: [http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1).
- Treatment Action Group. 2015 Report on Tuberculosis Research Funding Trends, 2005–2014: A Decade of Data. Frick M, editor. New York, USA: Treatment Action Group; 2015. Available from: [http://www.treatmentactiongroup.org/sites/default/files/201511/TB\\_FUNDING\\_2015\\_WEB.pdf](http://www.treatmentactiongroup.org/sites/default/files/201511/TB_FUNDING_2015_WEB.pdf).
- Horsburgh CR, Barry CE, Lange C. Treatment of Tuberculosis. Longo DL, editor. N. Engl. J. Med. Massachusetts Medical Society; 2015 [cited 2017 Feb 25];373:2149–60 Available from: <http://www.nejm.org/doi/10.1056/NEJMra1413919>.
- World Health Organization. WHO End TB Strategy. Global strategy and targets for tuberculosis prevention, care and control after 2015. Available from: [http://www.who.int/tb/post2015\\_strategy/en/](http://www.who.int/tb/post2015_strategy/en/).
- Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N. Mathematical models in the evaluation of health programmes. *Lancet*. 2011;378:515–25. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/21481448>.
- Dowdy DW, Dye C, Cohen T. Data Needs for Evidence-Based Decisions: a Tuberculosis Modeler's "Wish List." *Int J Tuberc Lung Dis*. 2013/06/08. 2013; 17:866–77. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23743307>.
- Trauer JM, Achar J, Parpieva N, Khamraev A, Denholm JT, Falzon D, et al. Modelling the effect of short-course multidrug-resistant tuberculosis treatment in Karakalpakstan, Uzbekistan. *BMC Med*. 2016 [cited 2017 Jan 25]; 14:187 Available from: <http://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-016-0723-2>.
- Houben RM, Menzies NA, Sumner T, Huynh GH, Arinaminpathy N, Goldhaber-Fiebert JD, et al. Feasibility of achieving the 2025 WHO global tuberculosis targets in South Africa, China, and India: a combined analysis of 11 mathematical models. *Lancet glob heal*. 2016;4:e806–15. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27720688>.
- Houben RM, Lalli M, Sumner T, Hamilton M, Pedrazzoli D, Bonsu F, et al. TIME impact - a new user-friendly tuberculosis (TB) model to inform TB policy decisions. *BMC med*. 2016;14:56. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27012808>.
- Trauer JM, Denholm JT, Waseem S, Ragonnet R, McBryde ES. Scenario analysis for programmatically tuberculosis control in Western Province, Papua New Guinea. *Am J Epidemiol*. 2016;183:1138–48. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27199387>.
- Ragonnet R, Trauer JM, Denholm JT, Geard NL, Hellard M, McBryde ES. Vaccination programs for endemic infections: modelling real versus apparent impacts of vaccine and infection characteristics. *Sci rep*. 2015;5: 15468. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26482413>.
- Trauer JM, Denholm JT, McBryde ES. Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. *J Theor Biol*. 2014;358:74–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24878110>.
- Trauer JM, Moyo N, Tay E-L, Dale K, Ragonnet R, McBryde ES, et al. Risk of Active Tuberculosis in the Five Years Following Infection. ... 15%? *Chest*. 2016;149:516–25.
- Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS one*. 2011/04/13. 2011;6:e17601. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21483732>.
- McBryde ES, Meehan MT, Doan TN, Ragonnet R, Marais BJ, Guernier V, et al. The risk of global epidemic replacement with drug resistant M. tuberculosis strains. *Int. J. Infect. Dis. Elsevier*; 2017 [cited 2017 Feb 8];0. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28163165>.
- Menzies NA, Gomez GB, Bozzani F, Chatterjee S, Foster N, Baena IG, et al. Cost-effectiveness and resource implications of aggressive action on tuberculosis in China, India, and South Africa: a combined analysis of nine models. *Lancet glob heal*. 2016;4:e816–26. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27720689>.
- Corbett EL, Marston B, Churchyard GJ, De Cock KM, Salaniponi F, Elzinga G. Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment. *Lancet*. World Health Organization, Geneva; 2006 [cited 2017 mar 6];367:926–37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16546541>.
- Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch intern med*. 2003;163:1009–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12742798>.
- Skrachina A, Hurevich H, Zalutskaya A, Sahalchik E, Astrauko A, van Gemert W, et al. Alarming levels of drug-resistant tuberculosis in Belarus: results of a survey in Minsk. *Eur Respir J*. 2012;39:1425–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22005924>.
- Acosta CD, Dadu A, Ramsay A, Dara M. Drug-resistant tuberculosis in Eastern Europe: challenges and ways forward. *Public Heal. action*. The International Union Against Tuberculosis and Lung Disease; 2014 [cited 2017 mar 6];4:53–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26393095>.
- Uplekar M, Pathania V, Raviglione M. Private practitioners and public health: weak links in tuberculosis control. *Lancet*. 2001;358:912–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11567729>.
- Abubakar I, Pimpin L, Ariti C, Beynon R, Mangtani P, Sterne J, et al. Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette–Guérin vaccination against tuberculosis. *Heal. Technol Assess*. 2013 [cited 2017 mar 6];17:1–372, v–vi. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24021245>.
- Matlow A, Robb M, Goldman C. Infection control and paediatric tuberculosis: A practical guide for the practicing paediatrician. *Paediatr. Child Health*. 2003 [cited 2017 mar 6];8:624–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20019856>.
- Kwan CK, Ernst JD. HIV and tuberculosis: a deadly human syndemic. *Clin. Microbiol. Rev.*. American Society for Microbiology; 2011 [cited 2017 mar 7]; 24:351–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21482729>.
- Lönnroth K, Roglic G, Harries AD. Improving tuberculosis prevention and care through addressing the global diabetes epidemic: from evidence to policy and practice. *Lancet Diabetes Endocrinol*. 2014 [cited 2017 Mar 7];2:730–9 Available from: <http://linkinghub.elsevier.com/retrieve/pii/S2213858714701093>
- Dhedea K, Barry CE, Maartens G. Tuberculosis. *Lancet*. 2016 [cited 2017 Mar 7];387:1211–26 Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0140673615001518>.
- Tostmann A, Kik S V, Kalisvaart NA, Sebek MM, Verwer S, Boeree MJ, et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clin infect dis*. 2008;47: 1135–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18823268>.
- Espinal MA, Kim SJ, Suarez PG, Kam KM, Khomenko AG, Migliori GB, et al. Standard short-course chemotherapy for drug-resistant tuberculosis:

- treatment outcomes in 6 countries. *JAMA*. 2000;283:2537–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10815117>.
31. Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VO, Andersen AB. Molecular evidence of endogenous reactivation of mycobacterium tuberculosis after 33 years of latent infection. *J infect dis*. 2002;185:401–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11807725>.
  32. Stover J, McKinnon R, Winfrey B. *spectrum*: a model platform for linking maternal and child survival interventions with AIDS, family planning and demographic projections. *Int J Epidemiol*. 2010;39(Suppl 1):i7–10. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20348129>.
  33. Stover J. Projecting the demographic consequences of adult HIV prevalence trends: the Spectrum Projection Package. *Sex Transm Infect*. 2004 [cited 2017 mar 7];80:i14–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15249694>.
  34. Grefenstette JJ, Brown ST, Rosenfeld R, DePasse J, Stone NT, Cooley PC, et al. FRED (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC public health*. 2013;13:940. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24103508>.
  35. Chao DL, Halloran ME, Obenchain VJ, Longini Jr. IM. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol*. 2010;6:e1000656. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20126529>.
  36. Brown T, Peerapatanapokin W. The Asian epidemic model: a process model for exploring HIV policy and programme alternatives in Asia. *Sex Transm infect*. 2004;80(Suppl 1):i19–24. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/15249695>.
  37. Gouws E, White PJ, Stover J, Brown T. Short term estimates of adult HIV incidence by mode of transmission: Kenya and Thailand as examples. *Sex Transm infect*. 2006;82(Suppl 3):iii51–55. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/16735294>.
  38. Kerr CC, Stuart RM, Gray RT, Shattock AJ, Fraser-Hurt N, Benedikt C, et al. Optima: a model for HIV epidemic analysis, program prioritization, and resource optimization. *J Acquir Immune Defic Syndr*. 2015;69:365–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25803164>.
  39. Brown T, Grassly NC, Garnett G, Stanecki K. Improving projections at the country level: the UNAIDS estimation and projection package 2005. *Sex Transm infect*. 2006;82(Suppl 3):iii34–40. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/16735291>.
  40. Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, Stover J, et al. Modelling HIV epidemics in the antiretroviral era: the UNAIDS estimation and projection package 2009. *Sex Transm infect*. 2010;86(Suppl 2):ii3–10. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20929855>.
  41. Stover J, Brown T, Marston M. Updates to the Spectrum/estimation and projection package (EPP) model to estimate HIV trends for adults and children. *Sex Transm infect*. 2012;88(Suppl 2):ii11–6. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23172341>.
  42. Nishikiori N, Van Weezenbeek C. Target prioritization and strategy selection for active case-finding of pulmonary tuberculosis: a tool to support country-level project planning. *BMC public health*. 2013;13:97. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23374118>.
  43. Lin HH, Langley I, Mwenda R, Doulla B, Egwaga S, Millington KA, et al. A modelling framework to support the selection and implementation of new tuberculosis diagnostic tools. *Int J Tuberc Lung Dis*. 2011;15:996–1004. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21740663>.
  44. Dowdy DW, Andrews JR, Dodd PJ, Gilman RH. A user-friendly, open-source tool to project impact and cost of diagnostic tests for tuberculosis. *Elife*. 2014;3. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24898755>.
  45. World Health Organization. WHO | Download data as CSV files. WHO. World Health Organization; 2016 [cited 2017 Feb 27]. Available from: <http://www.who.int/tb/country/data/download/en/>.
  46. UNICEF. Immunization coverage by antigen (including trends). 2016 [cited 2017 Feb 27]. Available from: <https://data.unicef.org/wp-content/uploads/2015/12/Immunization-coverage-by-antigen-including-trends.xls>.
  47. The World Bank. World Development Indicators | DataBank. [cited 2017 Feb 27] Available from: <http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

