Bayesian Compartmental Modeling of Tuberculosis

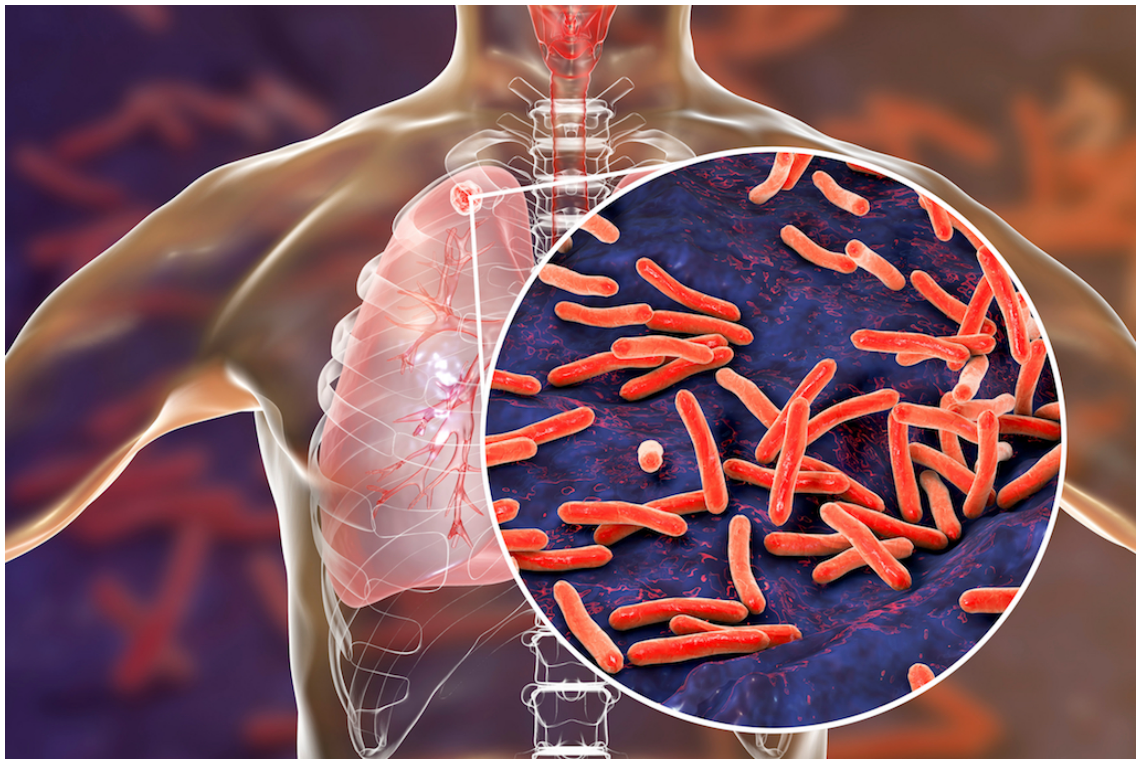
by Joshua-Michael Tomiyama

April 2022

Outline

1. Background on Tuberculosis
2. Introduce Compartmental SEIRS Model Framework
3. Discuss Approximate Bayesian Computation for Bayesian Analysis
4. Apply to US data

Tuberculosis Characteristics



- The bacterium, *Mycobacterium tuberculosis* usually affects the lungs, but can also infect other parts of the body
- Tends to spread through the air not surfaces
- Globally, $\approx$ 1.5 million people die from TB each year
- second leading infectious disease after COVID-19 (above HIV/AIDS)

Tuberculosis Characteristics

- Symptoms include prolonged cough, blood is sputum, chest pain or difficulty breathing, Unintentional weight loss, Fatigue, Fever, etc
- Treatment takes 6 months of antibiotics
- You are able to spread it as long as you have bacteria in your body
- typically not infectious 1-2 weeks after starting treatment
- a study in other countries found it to be a seasonal disease [10]

Goals of Analysis

- Confirm Model of exposure probabilities
- Predict new cases of Tuberculosis
- Predict the peak of Tuberculosis
- Loosely Estimate transition probabilities

SEIR Compartmental Models

- Categorize people into 4 disease states at given time point:
- S - people **susceptible** to disease (no immunity)
- E - people who have been **exposed** to the disease, but not infectious
- I - people who are **infectious** for a disease
- R - people who are **removed** due to developing immunity or dying
- Other compartmental models exist, but are highly dependent on disease characteristics (Vaccination compartment, disease vectors)

Temporal Process Model

SEIR Compartmental Structure

Notation

S	Number of Susceptible Individuals
E	Number of Exposed Individuals
I	Number of Infectious Individuals
R	Number of Removed Individuals
N	total population, fixed
t	index for time
*	Newly exposed, infected, or removed individuals for a given time point

Equations

$$S_{(t+1)} = S_t - E_t^* + S_t^*$$
$$E_{(t+1)} = E_t + E_t^* - I_t^*$$
$$I_{(t+1)} = I_t + I_t^* - R_t^*$$
$$R_{(t+1)} = R_t + R_t^* - S_t^*$$
$$N = S_t + E_t + I_t + R_t$$

Temporal Process Model

$$E_t^* | (S_t, I_t, \pi_t^{(SE)}) \sim \text{Binomial}(S_t, \pi_t^{(SE)})$$
$$I_t^* | (E_t, \pi_t^{(EI)}) \sim \text{Binomial}(E_t, \pi_t^{(EI)})$$
$$R_t^* | (I_t, \pi_t^{(IR)}) \sim \text{Binomial}(I_t, \pi_t^{(IR)})$$

- Examples of risk factors that affect exposure ($\pi_t^{(SE)}$) could be:
  - The total population size of a county
  - The timing of an intervention (vaccine, quarantine)
  - The climate of a region
- In principle, one could also include risk factors that affect the transition from exposed to infectious ($\pi_t^{(EI)}$) or infectious to removed compartments ($\pi_t^{(IR)}$)
- However, the length of time spent in these compartments are strongly characteristic of the disease in question, so additional data provides diminishing returns in estimation.

Transition Probabilities

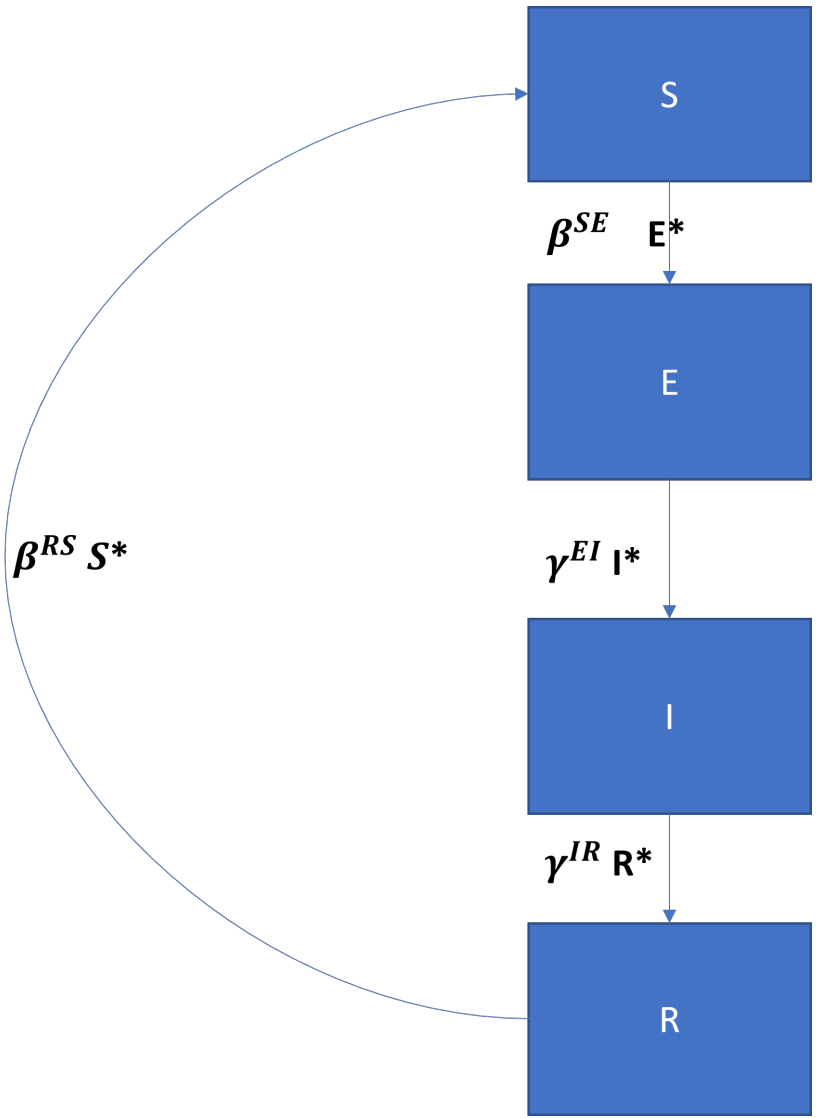
Notation

X_t	A **vector** of exposure covariates
β	A **vector** of exposure/reinfection effects
γ_{EI}	The reciprocal of the average latent period
γ_{IR}	The reciprocal of the average infectious period

Distributional Assumptions

$$\pi_t^{(SE)} = 1 - \exp\{-\frac{I_t}{N}\exp\{\beta^{(SE)\prime} X_t^{(SE)}\}\}$$
$$\pi_t^{(EI)} = 1 - \exp\{-\gamma_{EI}\}$$
$$\pi_t^{(IR)} = 1 - \exp\{-\gamma_{IR}\}$$
$$\pi_t^{(RS)} = 1 - \exp\{-exp\{\beta^{(RS)\prime} X_t^{(RS)}\}\}$$

Diagram



Approximate Bayesian Computation

- A fast approach for fitting Bayesian models is the Approximate Bayesian Computation (ABC) algorithm
- Only requires that the data is able to be simulated, 'likelihood-free'.
- Can be used to conduct Bayesian Inference on models whose likelihoods are computationally expensive or intractable to evaluate

Notation

y	The observed data
$d(\cdot, \cdot)$	A distance measure between two numbers/vectors
$S(\cdot)$	A summary statistic of the data
ε	The max tolerable discrepancy between the two statistics

Do Until N points accepted:

1. Draw $\theta_i \sim p(\theta)$
2. Simulate $x_i \sim p(x|\theta_i)$
3. Reject if θ_i if $d(S(x_i), S(y)) > \varepsilon$

Approximate Bayesian Computation

- You have an sample of $P(\theta|d(S(x_i), S(y)) \leq \varepsilon)$
- If $\varepsilon = 0$, then one is sampling exactly from the posterior distribution
- This is not feasible as the probability of generating the exact same data set is ≈ 0 in most applications.
- Thus, ABC accepts data that is 'close enough' $(\leq \varepsilon)$, where we want ε as small as practically possible
- The rejection algorithm is intuitive, but is very inefficient especially when the prior distribution is diffuse with the true parameter distribution.

Approximate Bayesian Computation

Sequential Monte Carlo Algorithm

1. Pick a decreasing sequence of $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T$
2. If t $= 1$
   - Do until N acceptances
     - Draw $\theta_i \sim \pi(\theta)$
     - Simulate $x_i \sim p(x|\theta_i)$
     - Reject if θ_i if $d(S(x_i), S(y)) > \varepsilon_0$
   - Set $w_i^{(1)} = \frac{1}{N}$; $\tau_1^2 = 2var(\theta^{(1)})$; t $= 2$;
3. If t >1
   - Do until N acceptances
     - Draw θ_i^* from $\theta^{(t-1)}$ weighted by $w_i^{(t-1)}$
     - Draw $\theta_i^{(t)} \sim N(\theta_i^*, \tau^2)$
     - Simulate $x_i \sim p(x|\theta_i)$
     - Reject if θ_i if $d(S(x), S(y)) > \varepsilon_t$
   - Set $w_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N} w_j^{(t-1)} \phi_j(\theta_i^{(t)})}$; $\tau_t^2 = 2var(\theta^{(t)})$; t $=$ t $+ 1$;

$$\phi_j \equiv \text{pdf of } N(\theta_j^{(t-1)}, \tau_{t-1}^2)$$

Notation

M_j	Proposed model j
$A_i^{(M_j)}$	0-1 indicator

Do until N total Draws from each model

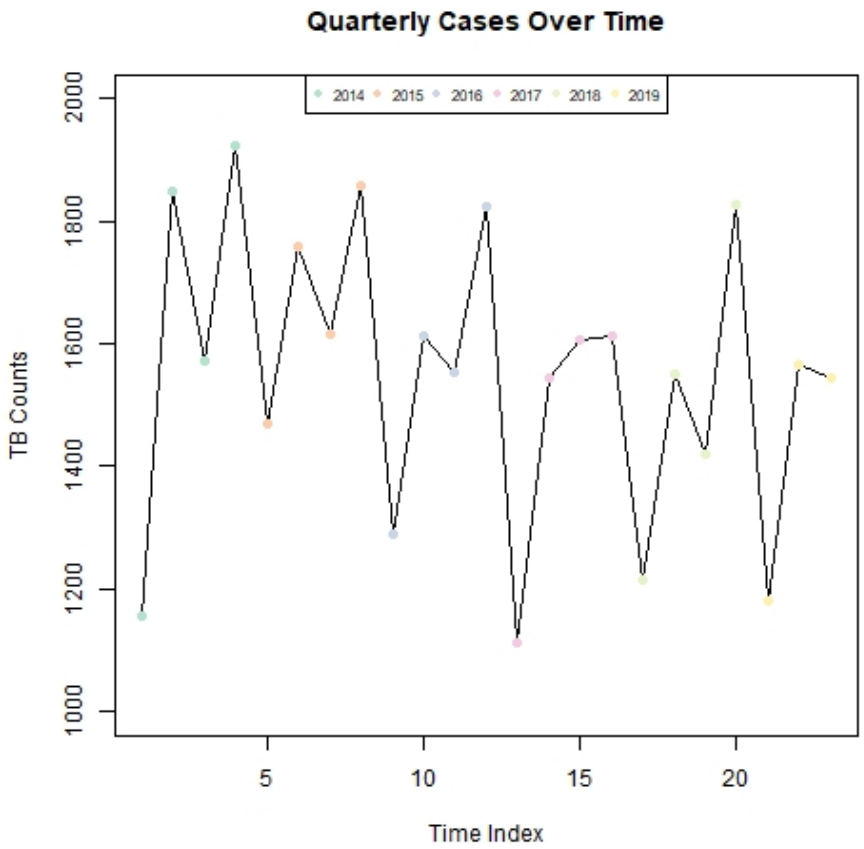
1. Draw a model $m_j \sim P(M_j)$
2. Draw $\theta_i \sim p(\theta|M_j = m_j)$
3. Simulate $x_i \sim p(x|\theta_i, M_j = m_j)$
4. $A_i^{(M_j)} = \mathbb{I}_{d(x_i,y)<\varepsilon}$

$$\frac{1}{N}\sum_i A_i^{(M_j)} \to \mathbb{E}(A_i^{(M_j)}) = P(d(x_i, y)) < \varepsilon | M_j = m_j)$$

- Recall the $BF_{1,2} = \frac{P(y|M_1)}{P(y|M_2)}$
- Sometimes people skip the model draw part and just reweight the sum at the end

Data

Quarterly United States TB Data



- Data is from DATA.CDC.GOV
- Reported quarterly from 2014-2019
- Last year used for prediction

Model Parameters

- Exposure Covariates Used
  - Model 1: Basis Spline expansion of the time index, 3 df
  - Model 2: Trigonometric Temporal Basis
  - $\beta_0 + \beta_1 sin(\frac{2\pi t}{4}) + \beta_2 cos(\frac{2\pi t}{4}) + \beta_3 sin(\frac{2\pi t}{4}) cos(\frac{2\pi t}{4})$
- Initializing Compartments
  - Loosely Fixed on Rough estimates
  - (S0=300e6,E0=1000,I0=9000/4,R0=3*9000/4)

Prior Distribution on Parameters

- $\beta_k^{(SE)} \sim N(0, \text{precision} = 0.5)\ \forall k$
- $\beta^{(RS)} \sim N(\approx 0.0118, \text{precision} \approx 313000)$
  - estimated from a study[12]
- $\gamma_{EI} \sim Gamma(100, -log(1 - 0.536)/100) =$
  - chosen such that 90% chance of becoming infectious by 9 months on average
- $\gamma_{IR} \sim Gamma(10, -log(1 - 0.437)/10)$
  - chosen such that 90% chance of becoming removed by 1 year on average

Results: Posterior Predictive Distribution

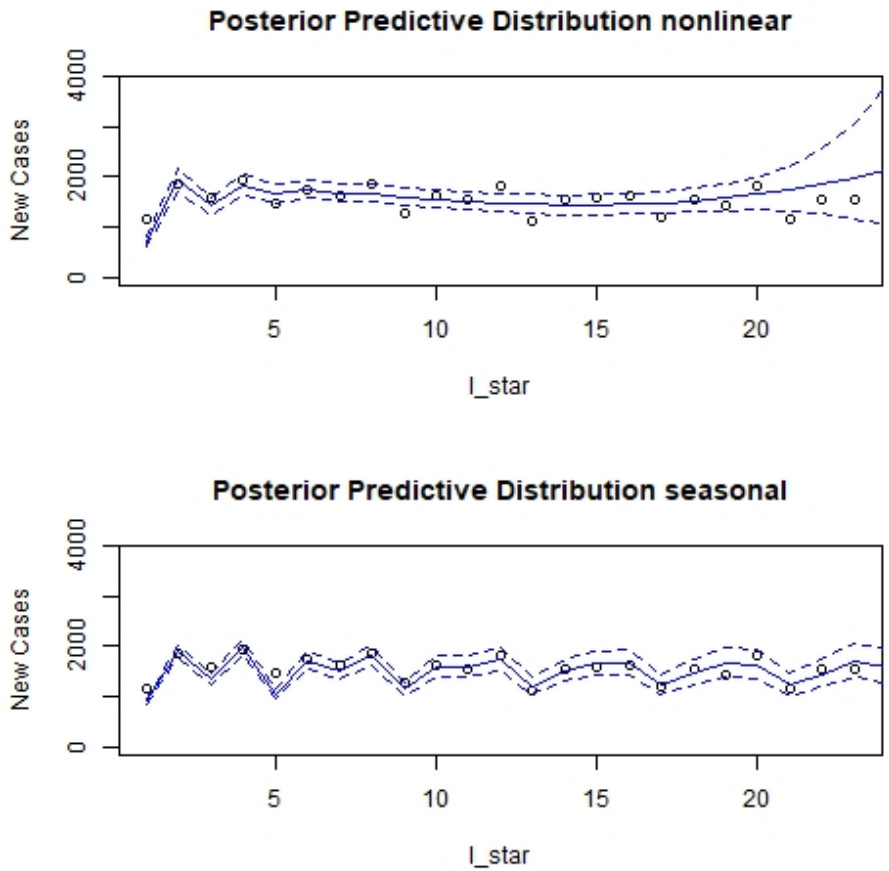


Figure: Posterior Predictive Distribution and 95% CI

$P(max\{I_{21}^*, I_{22}^*, I_{23}^*\} = I_2^*2) = 0.01524$; Num Post Pred sample $= 50000$

Results: Bayes Factor

Table: Terminating Epsilon of models

	Terminating.Epsilon
Nonlinear	950.413
Seasonal	597.479

- Extremely different terminating epsilon
- In general, we use the mean of these two epsilons to assess Bayes factor
- $BF_{seasonal,nonlinear} = Inf$. Does it make sense why?

Results: Exposure Parameters

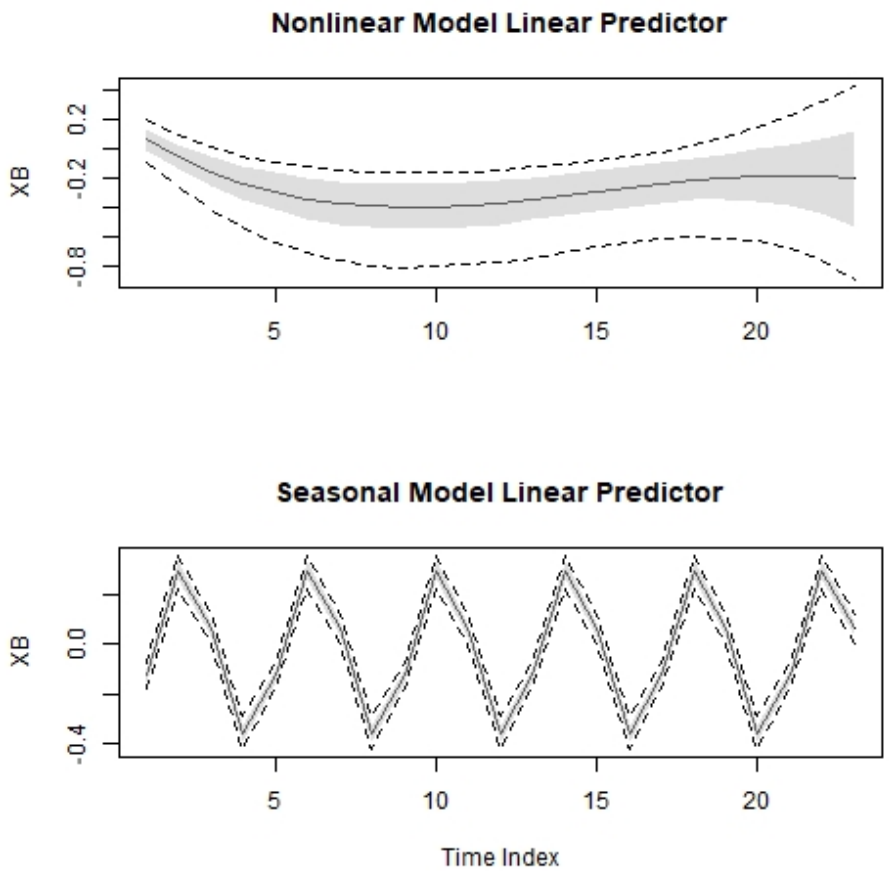


Figure: Plot of exposure over time Index

Results: Transition Parameters

Table: Transition Parameter Estimates

	Mean	SD	95% LB	95% UB
Beta_RS_1	1.200e-02	0.000*	1.2000	1.2000
γ_{EI}	2.457	0.208	2.0780	2.9140
γ_{IR}	3.862	0.402	3.1530	4.7160
$\pi^{(EI)}$	0.912	1.780e-02	0.8747875	0.9457200
$\pi^{(IR)}$	0.977	8.837e-03	0.9572694	0.9910503
$\pi^{(RS)}$	0.636	1.787e-05	0.636	0.636

Table: Mean transition parameters and 95% credible intervals for the seasonal model. SD not truly zero. N = 1000

Discussion

- The posterior update is quite different the the prior beliefs for E-I, I-R compartments
- R-S is pretty much the same
- decent at prediction, not great at finding peaks
- Ignores spatial heterogeneity (Alaska, Hawaii)
- Ignores underreporting

Work Cited I

Mark A. Beaumont.
Approximate bayesian computation in evolution and ecology.
Annual Review of Ecology, Evolution, and Systematics, 41(1):379–406, 2010.

MARK A. BEAUMONT, JEAN-MARIE CORNUET, JEAN-MICHEL MARIN, and CHRISTIAN P. ROBERT.
Adaptive approximate bayesian computation.
Biometrika, 96(4):983–990, 2009.

Grant Brown.
ABSEIR: Posterior Approximation to Fully Bayesian Spatial SEIR Models Using Multithreaded C++, 2019.
R package version 1.3.

Work Cited II

Grant Brown, Aaron Porter, Jacob Oleson, and Jessica Hinman.
Approximate bayesian computation for spatial seir(s) epidemic models.

Spatial and Spatio-temporal Epidemiology, 24, 11 2017.

Grant D. Brown, Jacob J. Oleson, and Aaron T. Porter.
An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two ebola outbreaks.
Biometrics, 72(2):335–343, 2016.

Viral Hepatitis STD Division of STD Prevention, National Center for HIV/AIDS and TB Prevention.
Nndss - table iv. tuberculosis, 2015.
[Online; accessed April 28, 2022].

Work Cited III


Viral Hepatitis STD Division of STD Prevention, National Center for HIV/AIDS and TB Prevention.
Nndss - table iv. tuberculosis, 2016.
[Online; accessed April 28, 2022].


Viral Hepatitis STD Division of STD Prevention, National Center for HIV/AIDS and TB Prevention.
Nndss - table iv. tuberculosis, 2017.
[Online; accessed April 28, 2022].


Viral Hepatitis STD Division of STD Prevention, National Center for HIV/AIDS and TB Prevention.
Nndss - table iii. tuberculosis, 2019.
[Online; accessed April 28, 2022].

Work Cited IV


Auda Fares.
Seasonality of tuberculosis.
J. Glob. Infect. Dis., 3(1):46–55, January 2011.


Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz.
Approximate bayesian computation.
PLOS Computational Biology, 9(1):1–10, 01 2013.


Suzanne Verver, Robin M. Warren, Nulda Beyers, Madalene Richardson, Gian D. van der Spuy, Martien W. Borgdorff, Donald A. Enarson, Marcel A. Behr, and Paul D. van Helden.
Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis.
American Journal of Respiratory and Critical Care Medicine, 171(12):1430–1435, 2005.

Work Cited V

PMID: 15831840.