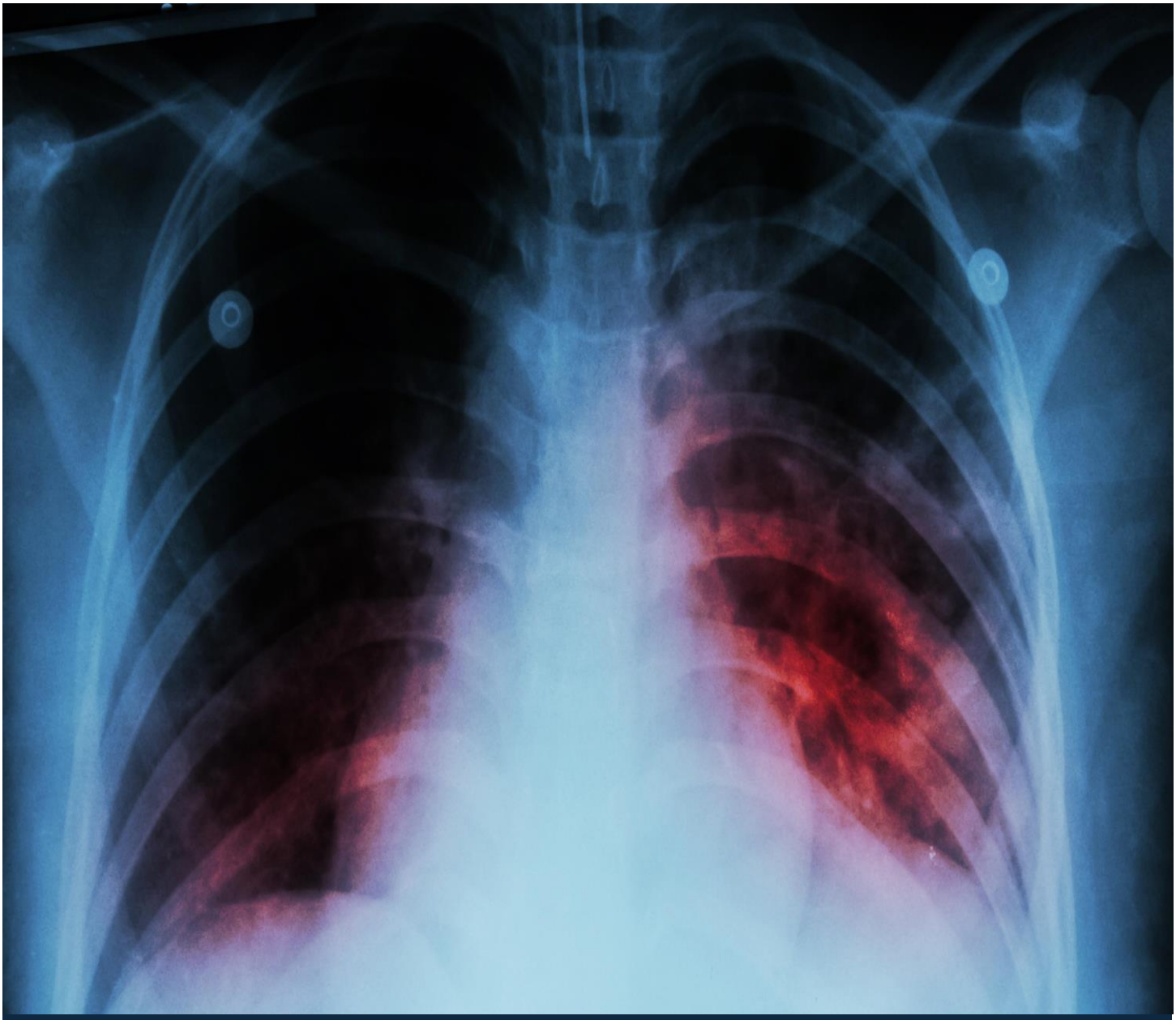


Modelling tuberculosis in Cambodia using a SIRD Model approach

By: Rithy Techavoan Yean

Submitted to Dr. Olga Korosteleva 5/3/24 STAT 495 Statistical Modelling in Epidemiology

Table of Contents

Table of Contents 1

Introduction 2

Methodology 2

Data Collection 3

SIRD Epidemic Model 4

Parameter Estimation 7

Simulating Cambodia's TB Future 8

Dynamics Trajectory 9

Conclusion 10

Works Cited ..... **Error! Bookmark not defined.**

Appendix I: SAS Code 11

Appendix II: R Code 12

Introduction

Tuberculosis, otherwise known as TB, is caused by a bacteria named *Mycobacterium tuberculosis* which targets the lungs (WHO, 2024). It is a disease that is airborne which means it can spread through the air whenever a person who is diagnosed with TB sneezes, spits, or coughs (WHO, 2024). Due to the nature of this disease, a person may just need to inhale a few of its germs and will start developing TB symptoms (WHO, 2024). Some symptoms of TB include chest pain, coughing out blood or mucus, weight loss, and difficulty breathing (WHO, 2024). The World Health Organization states that annually, 10 million people become ill with TB and around 1.5 million people die from TB which makes the disease the most infectious killer (WHO, 2024). The disease is very prevalent in low-income and middle-income countries.

Cambodia, a country located in Southeast Asia, is considered to be one of 30 high TB incident countries prior to 2019. In 2019, 6500 Cambodians died as a result of TB complications and 47,000 Cambodians started to develop TB symptoms (StopTBPartnership, 2024). Therefore, TB can be considered prevalent within Cambodia and serves as a topic of interest to explore using the *Susceptible-Infected-Recovered-Dead (SIRD)* approach.

Methodology

The data used within this analysis are taken from various sources such as STOPTB partnership and a published journal article titled “Progress towards the 2020 milestones of the end TB strategy in Cambodia: estimates of age and sex specific TB incidence and mortality from the Global Burden of Disease Study 2019” written by Jianning Ma, Avina Vongparith, and many other contributing authors. The STOPTB partnership is an organization affiliated with the United Nations Office for Project Services (UNOPs) which was established in 2000. STOPTB partnership partners themselves with many government programs, research agencies, foundations, and NGOs to compile TB data for countries around the world. Information

pertaining to the number of TB cases, number of people on treatment for TB, number of TB recoveries, and number of deaths were taken directly from STOPTB partnership.

The number of cases and number of deaths were then compared with the journal article mentioned above which was an article that analyzed the prevalence and incidence rate of TB within Cambodia to determine the progress that Cambodia has made to reduce TB and whether it can fully eliminate the disease in the future. Finally, information pertaining to the population of Cambodia was taken from *Worldometer,* a reference website that provides real-time population statistics for countries around the world. The combination of data extracted from these websites provided the raw data shown below:

Data Collection

Table 1: Raw data of Cambodian population, number of Cambodians developing TB, number of Cambodians on treatment for TB, number of Cambodians recovered from TB, and number of Cambodians death by TB from 2012 to 2019

Year	Population	Developing TB	On treatment	Recovered	Deaths
2012	14786640	61000	37743	36057	8190
2013	14999683	60000	43059	33048	7890
2014	15210817	58000	43059	40119	7630
2015	15417523	57000	35619	33038	7370
2016	15624584	55000	33453	30413	7120
2017	15830689	53000	34238	32338	6860
2018	16025238	50000	28620	27035	6670
2019	16207746	47000	29906	28561	6500

Table 2: The number of Cambodians living with TB from 2012 to 2019

Year	Developing TB	On treatment	Total
2012	61000	37743	98743
2013	60000	43059	103059
2014	58000	43059	101059
2015	57000	35619	92619
2016	55000	33453	88453
2017	53000	34238	87238
2018	50000	28620	78620
2019	47000	29906	76906

SIRD Epidemic Model

The *Susceptible-Infectious-Recovered-Deceased* model is slightly different from the regular Compartmental SIR Model *Susceptible-Infectious-Removed* because it organizes 'Recovered' and 'Deceased' as their own separate compartments. For the SIRD model, recovered implies individuals that have survived the disease and are now immune.

In the SIRD model, we will assume that those who have recovered from TB will remain immune to it. Interestingly this assumption of immunity has been discussed in one particular scientific journal titled "Genetic Resistance to Mycobacterium Tuberculosis Infection and Disease" authored by Moller, Kinnear, Orlova, and et al. They claim that complete resistance to infection has been observed in some individuals after prolonged Mycobacterium Tuberculosis exposure. However, it should be acknowledged that TB is a disease that is possible to be re-infected as suggested by Diana Rodriguez in her medically reviewed article titled "When tuberculosis comes back" that was published on EveryDayHealth in 2009 (Upham, 2009). EveryDayHealth is a digital media company that primarily publishes content pertaining to consumer health. For this particular analysis, it is assumed that once an infected TB person recovers, they will be immune to the disease.

The model uses the following differential equations:

$$\begin{cases} \dfrac{dS(t)}{dt} = -\dfrac{\beta I(t)S(t)}{N} \\ \dfrac{dI(t)}{dt} = \dfrac{\beta I(t)S(t)}{N} - (\gamma + \mu)I(t) \\ \dfrac{dR(t)}{dt} = \gamma I(t) \\ \dfrac{dD(t)}{dt} = \mu I(\mathrm{t}) \end{cases}$$

where β, γ, μ are the rates of infection, recovery, and mortality.

At time t, $S(t)$ represents the *Susceptible* population which includes those who are able to be infected with TB. $I(t)$ represents the *Infected* population, the number of people who has TB which is the sum of Cambodians developing TB and Cambodians who are on treatment for TB. $R(t)$ represents the *Recovered* population which includes Cambodians that that have fully recovered from TB and are now immune to it. Finally, $D(t)$ represents the number of Cambodians that have died from TB.

The total population at any given time t is denoted by $N(t)$ and is given by:

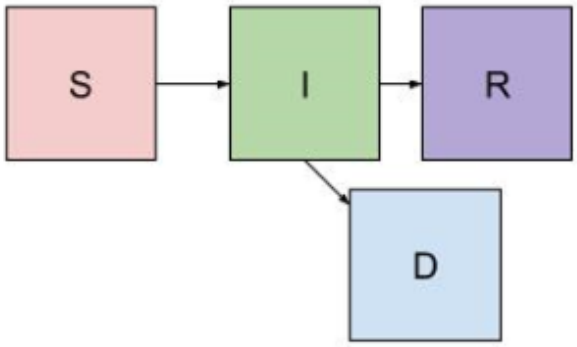
$$N(t) = S(t) + I(t) + R(t) + D(t)$$

This equation suggests that the total population is the sum of all the compartments. This population is assumed to be closed meaning that no migration takes place. In reality, of course, the population dynamics of a country cannot be maintained at a constant rate. This is due to both births, deaths, and people moving out of the country and into the country. However, since the population of Cambodia rises per year from 2012 to 2019, we will still use the slight changes in population to estimate $\beta, \gamma, and\ \mu$ but we will average the results afterwards using methods of moments estimation. It should be noted that the increase in population of Cambodians per year

appears to be less than 1% of the total population per year, and so these slight changes in population can be argued to be negligible.

The *Susceptible* compartment is reduced by the infection rate of β. The *Infected* compartment will consist of people who are developing TB symptoms and those who are on treatment. Seeing as those who develop TB symptoms and those on treatment are sick with TB and so they can spread it. Additionally, the *Infected* compartment is reduced at a rate of γ and μ as individuals within this compartment will either be successfully treated and move to the *Recovered* compartment, or they will be unsuccessfully treated and move to the *Deceased* compartment.

Figure I: Visualization of SIRD compartments



To calculate the *Susceptible* population, we can rearrange the equation above as follows:

$$S(t) = N(t) - I(t) - R(t) + D(t)$$

Using 2012 as the baseline year at time $t = 0$, we can calculate $S(0)$ as follows using the data above, where $N(0) = 14786640,\ I(0) = 98743, R(0) = 36057, D(0) = 8190$:

$$S(0) = N(0) - I(0) - R(0) - D(0)$$

$$= 14786640 - 98743 - 36057 - 8190 \ = 14671517$$

With 2019 being the endline at time $t = 7$, we calculated the *Susceptible* population for all the years within the dataset. Under the assumption of a closed population, we calculate the cumulative values for *Infected, Recovered,* and *Deaths* because the number of people that stays infected, recovered, or died are still within the total population.

Table 3: Cumulative Values of Tuberculosis Cases, Recovered, and Deaths

Time	Population	Susceptible	Infected	Recovered	Deaths
0	14786640	14671517	98743	36057	8190
1	14999683	14765721	201802	69105	16080
2	15210817	14860536	302861	109224	23710
3	15417523	14959883	395480	142262	31080
4	15624584	15064251	483933	172675	38200
5	15830689	15169398	571171	205013	45060
6	16025238	15271987	649791	232048	51730
7	16207746	15364589	726697	260609	58230

Parameter Estimation

It is not very practical to solve the differential equations and so we use a system of difference equations instead. We suppose that that there are discrete time points j that are separated by one unit interval (1,2,3,4,…,etc) then we transform these differential equations into the following difference equations for $j = 0, 1, 2, 3, etc$

$$\begin{cases} S(j+1) = S(j) - \dfrac{\beta S(j) I(j)}{N}, \\ I(j+1) = I(j) + \dfrac{\beta S(j) I(j)}{N} - \gamma I(j) - \mu I(j), \\ R(j+1) = R(j) + \gamma I(j), \\ D(j+1) = D(j) + \mu I(j) \end{cases}$$

If we were to denote the increments by $\Delta S(j) = S(j+1) - S(j),\ \Delta I(j) = I(j+1) - I(j),\ \Delta R(j) = R(j+1) - R(j)$, and $\Delta D(j) = D(j+1) - D(j)$. We can rewrite the following equations as follows:

$$\begin{cases} \dfrac{\Delta S(j)}{S(j)I(j)} = -\dfrac{\beta}{N}, \\ \dfrac{\Delta I(j)}{I(j)} = \dfrac{\beta}{N}S(j) - \gamma - \mu, \\ \dfrac{\Delta R(j)}{I(j)} = \gamma, \\ \dfrac{\Delta D(j)}{I(j)} = \mu \end{cases}$$

Alternatively, using methods of moments estimation, we arrive at the following estimations:

$$\hat{\beta}_{MM} = -N \cdot (sample\ mean\ of\ \frac{\Delta S(j)}{S(j)I(j)})$$

$$\hat{\gamma}_{MM} = sample\ mean\ of \frac{\Delta R(j)}{I(j)}$$

$$\hat{\mu}_{MM} = sample\ mean\ of \frac{\Delta D(j)}{I(j)}$$

Simulating Cambodia's TB Future

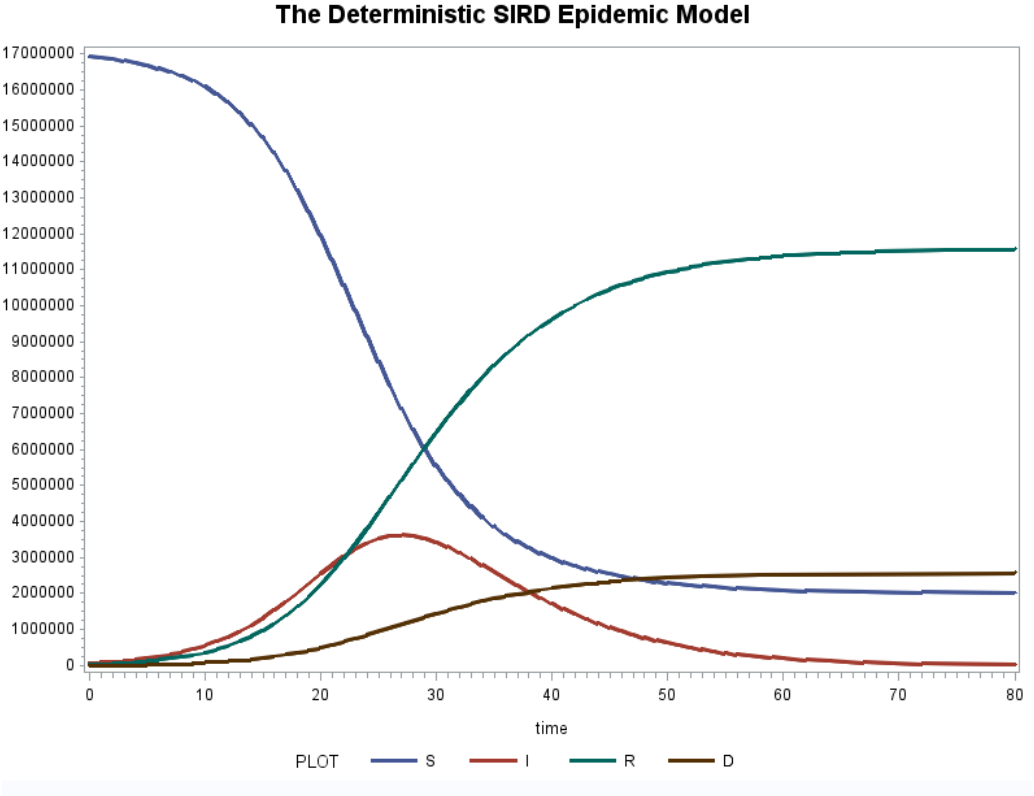
As shown in Appendix I and Appendix II, SAS and R code output estimates for beta, gamma, and mu as follows:

Table 4: Estimates for rate of infection, recovery, and mortality for Tuberculosis

Estimates:	SAS	R
β	0.380282 person^{-1}year^{-1}	0.3636574 person^{-1}year^{-1}
γ	0.12537 person^{-1}year^{-1}	0.1253698 person^{-1}year^{-1}
μ	0.027987 year^{-1}	0.0279869 year^{-1}

Dynamics Trajectory

Figure 2: TB Dynamics Trajectory of Cambodia in 80 years



To conduct a numerical simulation of the SIRD model, a population size of 17,000,000 was used as this is the expected population present within Cambodia at this time. Though we should note that because the model does not consider vital dynamics, the population size used is

constant and will not be subjected to change in the simulation even if time passes. At a constant population, it is clear that over time, the number of infected is expected to rise from 10 years to 40 years but will ultimately remain stagnant from 50 years onwards. The number of deaths will increase from 10 years to 40 years but will also stay stagnant from 50 years onwards. The susceptible population will decrease over time which means that more and more people will get TB which decreases the number of susceptible individuals. Similarly, the recovered curve appears to increase at around 27 years but falls back down and remains stagnant. The behavior of these curves suggests that overall TB cases is expected to decrease over time in Cambodia which could suggest that the country is likely to combat the onset of the disease.

Conclusion

Tuberculosis was considered endemic for Cambodia prior to 2019. In this analysis, we used existing tuberculosis data to estimate the parameters for the infection, recovery, and mortality rate. These parameters were estimated using the method of moments estimation and a simulation was run to project what TB would look like in the future. While it appears that TB is expected to decrease over time, the SIRD model only functioned as a result of a constant population. In reality, such an assumption is difficult because the population changes and so at any given time, there could be more susceptible individuals. This impacts the calculations of the parameters. Moreover, the analysis assumed that those that recover from TB are immune to the disease which is a large assumption because there are cases of latent TB infection. Indicating that the SIRD model may not be the best model to conduct the analysis given. We also assume that healthcare infrastructure does not change and that is not reflected in the calculations. It is clear that medical advancements will continue in the next decades and so people may recover from TB at a much faster rate than provided. Overall, the analysis provides a glimpse on how a epidemic model approach can help map out the spread of TB within a country like Cambodia which has seen many cases of the disease.

Appendix I: SAS Code

```
/*reading data into SAS*/
data SIRD;
input N I R D S;
cards;
14786640    98743       36057       8190  14671517
14999683    201802      69105       16080 14765721
15210817    302861      109224      23710 14860536
15417523    395480      142262      31080 14959883
15624584    483933      172675      38200 15064251
15830689    571171      205013      45060 15169398
16025238    649791      232048      51730 15271987
16207746    726697      260609      58230 15364589
;
run;

/*calculating total population N*/
data SIRTD;
set SIRD;
N = sum(S, I, R, D);
;

/*checking if N was calculated correctly*/
proc print data=SIRTD;
run;

/*method of moments estimation*/
data SIRTD;
set SIRTD;
lag_S = lag(S);
lag_I = lag(I);
beta_est = N*(S - lag(S)) / (lag(S)*lag(I));
gamma_est = (R - lag(R)) / lag(I);
mu_est = (D - lag(D)) / lag(I);
if beta_est ne .;
run;

/*outputting estimates for beta, gamma, and mu*/
proc sql;
select mean(beta_est) as beta_hatMM,
         mean(gamma_est) as gamma_hatMM,
         mean(mu_est) as mu_hatMM from SIRTD;
quit;
```

beta_hatMM	gamma_hatMM	mu_hatMM
0.380282	0.12537	0.027987

```
/*creating a macro function that takes in the estimates as our parameters*/
%macro SIRmodel(N,beta,gamma,mu,t);
data SIRD;
S=&N-47000-29906;
I=47000+29906;
```

```
R=28561;
D=6500;
time=0;
output;
do time=1 to &t;
S+(-&beta*S*I/&N);
I+&beta*(S*I/&N)-&gamma*I-&mu*I;
R+&gamma*I;
D+&mu*I;
output;
end;
run;
symbol interpol=join value=none width=2;
axis label=none;
title "The Deterministic SIRD Epidemic Model";
proc gplot data=SIRD;
plot (S I R D)*time/overlay legend vaxis=axis;
run;
%mend;
%SIRmodel(17000000,0.380282, 0.12537, 0.027987, 80);
```

Appendix II: R Code

```
#reading the data into R

tuberculosis <- read.csv("C:/Users/User/Desktop/Tuberculosis_Project.csv", header = TRUE, sep = ",")


#calculating the change in Susceptible, Infected, Recovered, and Dead

tuberculosis$S.delta<- c(tuberculosis$S[-1],0)-tuberculosis$S

tuberculosis$I.delta<- c(tuberculosis$I[-1],0)-tuberculosis$I

tuberculosis$R.delta<- c(tuberculosis$R[-1],0)-tuberculosis$R

tuberculosis$D.delta<- c(tuberculosis$D[-1],0)-tuberculosis$D

tuberculosis$N.delta<- c(tuberculosis$N[-1],0)-tuberculosis$N

tuberculosis<- tuberculosis[-nrow(tuberculosis),]


#creating a for loop that recursively calculate the changes in Susceptible, Infected, Recovered, and Dead
to calculate estimates for beta, gamma, and mu using method of moments estimation

for (j in 1:nrow(tuberculosis)) {
```

```
tuberculosis$beta.est[j] <- tuberculosis$N*(tuberculosis$S.delta[j] / (tuberculosis$S[j])) *
(1/tuberculosis$I[j])

tuberculosis$response.var[j]<- tuberculosis$I.delta[j]/ tuberculosis$I[j]

tuberculosis$gamma.est[j]<- tuberculosis$R.delta[j]/ tuberculosis$I[j]

tuberculosis$mu.est[j]<- tuberculosis$D.delta[j]/ tuberculosis$I[j]

}


#printing estimates for beta, gamma, and mu

print(beta.hatMM <- mean(tuberculosis$beta.est))
```

```
0.3636574
```

```
print(gamma.hatMM<- mean(tuberculosis$gamma.est))
```

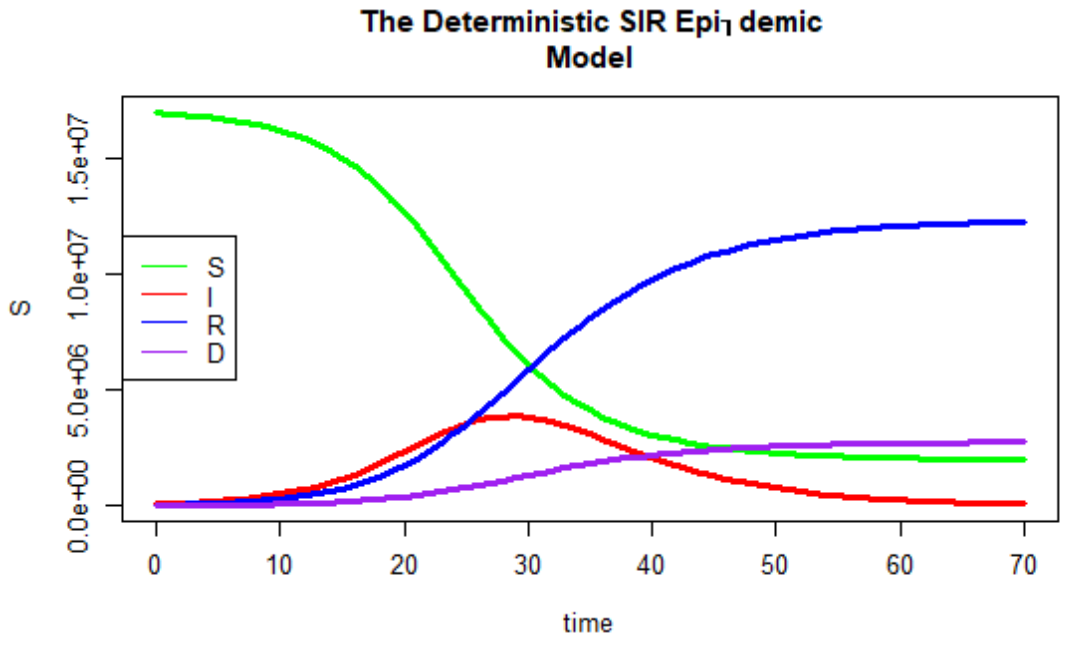
```
0.1253698
```

```
print(mu.hatMM<- mean(tuberculosis$mu.est))
```

```
0.0279869
```

```
#simulating the SIRD curves

N<- 17000000

beta <- 0.3636574

mu<- 0.0279869

gamma<- 0.1253698

S<- c()

I<- c()

R<- c()

D<- c()


S[1]<- N-47000-29906

I[1]<- 47000+29906

R[1]<- 28561

D[1]<- 6500
```

```
for (j in 1:70) {
S[j+1]<- S[j]-beta*S[j]*I[j]/N
I[j+1]<- I[j]+beta*S[j]*I[j]/N- gamma*I[j] - mu*I[j]
R[j+1]<- R[j]+gamma*I[j]
D[j+1] <- D[j]+mu*I[j]
}

time<- 0:70
plot(time,S, type="l", lty=1, lwd=3, main="The Deterministic SIR Epidemic
Model", ylim=c(0,N), col="green")
lines(time,I,lwd=3, col="red")
lines(time,R,lwd=3, col="blue")
lines(time,D, lwd=3, col = "purple")
legend("left", c("S","I","R", "D"), lty=1, col=c("green","red","blue","purple"))
```



Bibliography

- Idowu, Ilekura. “Mathematical Modelling: Modelling the Spread of Diseases with Sird Model.” *Analytics Vidhya*, Analytics Vidhya, 10 Aug. 2021, www.analyticsvidhya.com/blog/2021/08/mathematical-modelling-modelling-the-spread-of-diseases-with-sird-model/. Accessed 2 May 2024.
- Korosteleva, Olga “STAT 495 Lecture Notes”. Accessed 2 May 2024.
- Ma, Jianing, et al. “Progress towards the 2020 milestones of the end TB strategy in Cambodia: Estimates of age and sex specific TB incidence and mortality from the global burden of disease study 2019.” BMC Infectious Diseases, vol. 22, no. 1, 3 Dec. 2022, https://doi.org/10.1186/s12879-022-07891-5.
- Möller, Marlo, et al. “Genetic resistance to mycobacterium tuberculosis infection and disease.” Frontiers in Immunology, vol. 9, 27 Sept. 2018, https://doi.org/10.3389/fimmu.2018.02219.
- “Tuberculosis Situation in 2020.” *CAMBODIA TB Dashboard*, STOP TB PARTNERSHIP, www.stoptb.org/static_pages/KHM_Dashboard.html. Accessed 2 May 2024
- “Tuberculosis.” *World Health Organization*, World Health Organization, www.who.int/health-topics/tuberculosis. Accessed 2 May 2024.
- Upham, Becky, et al. “When Tuberculosis Infection Comes Back - Tuberculosis Center.” *EverydayHealth.Com*, EverydayHealth.com, 16 Dec. 2009, www.everydayhealth.com/tuberculosis/when-tuberculosis-infection-comes-back.aspx.