

Epidemiological modelling of spread of COVID-19 in Germany and South Korea

Devishi Kesar
Georgia Institute of Technology
Atlanta, USA
dkesar3@gatech.edu

Aparna Maddala
Georgia Institute of Technology
Atlanta, USA
amaddala3@gatech.edu

Sai Preetham Peddireddy
Georgia Institute of Technology
Atlanta, USA
speddireddy3@gatech.edu

ABSTRACT

Estimation of the prevalence and infection rates of COVID-19 infections is critical for understanding the overall number of infection rates and pandemic potential of this disease. In our study we use observations of reported infection rates in Germany and South Korea in conjunction with publicly available data, an SEIRD model, to infer critical epidemiological characteristics associated with SARS-CoV-2 and the effectiveness of healthcare countermeasures to combat the spread of the virus. We compared the spread of the virus in Germany and South Korea in terms of similarity and modelled estimations of effects of lockdown and social distancing imposed by Germany and South Korean. This model also showcases the modelled observations on the time taken for the public health interventions to have a significant effect on viral transmission. A compartmental model of phase adjusted estimation of the spread of COVID-19 in Germany and South Korea was used. Comparison of Germany and South Korea was done due to the effective measures taken by these countries and their relative success in containing COVID-19. Our method estimated two dates at which R_t changed significantly within a national epidemic of COVID-19 and compared these dates with the dates that lockdown and social distancing in order to estimate the date at which public health interventions take effect. Then the parameters generated by our analysis between countries were compared. Furthermore, a GUI was developed that allows users to investigate the effect of various public health interventions on case numbers drawn from user-supplied data by inputting the time series data for a region they want to model with infections, recoveries, and fatalities.

KEYWORDS

COVID-19, SEIRD model, health countermeasures, infection rates

ACM Reference Format:

Devishi Kesar, Aparna Maddala, and Sai Preetham Peddireddy. 2020. Epidemiological modelling of spread of COVID-19 in Germany and South Korea. In *ACM*, New York, NY, USA, 9 pages. <https://doi.org/10.1145/1122445.1122456>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BMED 8813-BHI, Group Project

© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00
<https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

A cluster of pneumonia cases of unknown origin was reported in Wuhan, China around late December 2019. On January 7th, 2020, the causative agent was discovered to be a novel coronavirus (SARS-CoV-2) with high genetic similarity to SARS-CoV and MERS-CoV [Team 2020]. The disease associated with the novel coronavirus was named COVID-19. As of February 11th, COVID-19 was estimated to have an international case-fatality rate (CFR) of 2.1%, which is lower than both MERS (34.4%) and SARS (9.6%) [Wu and Mcgoogan 2020]. Despite its low CFR, SARS-CoV-2 is responsible for more deaths than either MERS and SARS because of its high transmissibility, which has led to a much higher number of cases [Team 2020]. As a result of high effective viral transmission rate (R_t), SARS-CoV-2 spread to the entire country of China within 30 days [Wu and Mcgoogan 2020]. COVID-19 was declared a pandemic by WHO on March 11th, and as of April 20th, cases have been reported in 210 countries [Newsroom and Organization [n.d.]], [situation report [n.d.]].

In the absence of a clinically proven treatment for COVID-19 or an effective vaccine, public health interventions like social distancing and border lockdowns remain the only way to combat the spread of the virus. Therefore, there is an urgent public health need for data-driven decision making support for public health and social policies. Germany and South Korea are frequently cited by the media as countries that successfully contained COVID-19 epidemics through public health interventions [Rising 2020], [Submission [n.d.]]. Both Germany and South Korea are characterized by early border lockdowns and a high degree of testing, and South Korea further implemented very efficient contact tracing.

According to Edward Shortliffe, a major barrier to the success of public health informatics is the lack of systems that support adaptive learning and real-time decision making, integrating new information to evaluate clinical or public health interventions [Shortliffe and Cimino 2014]. To address this issue within the context of the current COVID-19 epidemic, we would like to compare the spread of the virus in Germany and South Korea to see if there are any similarities. In doing so, we would like to answer the following questions: a) did the lockdowns and social distancing imposed by Germany and South Korea have similar effects? b) how long did it take for the public health interventions to have a significant effect on viral transmission? To accomplish these goals, we will use a compartmental model to model the spread of COVID-19 in Germany and South Korea.

2 LITERATURE REVIEW

To examine the growth rate of the outbreak, many studies carried out initial modelling to report the reproduction number of

COVID-19 in South Korea using publicly available studies and data. A study by Eunha Shim et al. (2020) using the empirical reporting delay distribution and simulating the generalized growth model, estimated the effective reproduction number based on the discrete probability distribution of the generation interval using an SEIR model [Shim et al. 2020]. Sustained disease transmission of COVID-19 in different regions of Korea was studied and estimation of the effectiveness of the implementation of social distancing measures in South Korea was showcased. They identified four major clusters where the outbreak was severe and estimated the reproduction number at 1.5 as of March 6th 2020. However data according to the dates considered for their model was limited and the recent trends and cases were not taken into account. The effectiveness of countermeasures against the spread of the disease was also taken into account but was unable to be compared with the actual measures taken after February. There were also new policies implemented and a difference in real world data that was different from the data considered here.

There is a stark difference in the parameters to be considered when other factors like population density and mode of spread are taken into account for SARS-CoV-2. There was a study by J Rocklöv et al (2020) where the transmission of the virus on the Diamond Princess was considered and modelled to figure out the spread in dense and confined areas and the effectiveness of countermeasures such scenarios [Rocklöv et al. 2020]. On 3 February, 2020, an outbreak of COVID-19 on cruise ship Diamond Princess was reported with 10 initial cases, following an index case on board around 21-25th January. By 4th February, public health measures such as removal and isolation of ill passengers and quarantine of non-ill passengers were implemented. The study estimated the basic reproduction number from the initial period of the outbreak using SEIR modeling. The study also additionally estimated a scenario in absence of countermeasures, and established a model to study the impact of differential contact rates among the groups and the difference in spread with and without any precautions and countermeasures taken. It was seen that Basic reproduction rate was initially 4 times higher on-board compared to the R_0 in the epicentre in Wuhan, but the countermeasures lowered it substantially. Infection rates with and without control parameters were compared and plotted and showed a positive correlation in the decrease of cases when precautions and lockdown policies were put into place. We also used a similar model to this to calculate our R_0 number and took into account the effectiveness of different measures as they have modelled. However in this paper, incubation rates, potential delay in testing availability of fast test results for better tracking the spread were not taken into account.

Isolation of cases and contact tracing is used to control outbreaks of infectious diseases, and has been used for coronavirus disease 2019 (COVID-19) too in most countries and regions around the world. Joel Hellewell et al (2020) studied whether this strategy will achieve control and takes into account the characteristics of both the pathogen and the response [Hellewell et al. 2020]. They used a stochastic mathematical model to assess if isolation and contact tracing are able to control onward transmission from imported cases of COVID-19. Various scenarios that varied in the number of initial

cases, the basic reproduction number (R_0), the delay from symptom onset to isolation, the probability that contacts were traced, the proportion of transmission that occurred before symptom onset were considered in the study. The paper quantified the weekly maximum number of cases traced to measure feasibility of public health effort. The delay between symptom onset and isolation had the largest role in determining whether an outbreak was controllable when R_0 was 1.5. The paper also quantified probability of control decreases with long delays from symptom onset to isolation, fewer cases ascertained by contact tracing, and increasing transmission before symptoms. This model can be modified to reflect updated transmission characteristics and more specific definitions of outbreak control to assess the potential success of healthcare countermeasures to combat the virus which we considered. However it was a relatively simplified model and most additional parameters like quarantine rate and region lockdown were not measured objectively by modelling.

Similarly another study examined the early transmission dynamics of the infection and evaluated the effectiveness of control measures for assessing the potential for sustained transmission to occur in new areas [Kucharski et al. 2020]. Combining a mathematical model of severe SARS-CoV-2 transmission with four datasets from within and outside Wuhan, they estimated how transmission in Wuhan varied between December, 2019, and February, 2020 and then used these estimates to assess the potential for sustained human-to-human transmission to occur in locations outside Wuhan if cases were introduced. This study took into account the possibility that newly introduced cases might generate outbreaks in other areas which has been proven true recently given the data and also these assumptions were taken as a part of our modelling. For this study, an SEIR model with extended parameters such as incubation period was used similar to the models that we would be using. However their study assumed homogeneity and generalization which may not be true for real world data and also it needed expansion to different datasets and different regions as it was limited to Wuhan in China.

Another recent study modelled the COVID-19 coronavirus epidemic in South Korea, Italy, France, and Germany; where they used early reported case data to predict the cumulative number of reported cases to a final size [Magal and Webb 2020]. The key features of their model are the timing of implementation of major public policies restricting social movement, the identification and isolation of unreported cases, and the impact of asymptomatic infectious cases, all of which are the factors that we are taking into account in our modelling. They also Divided the epidemic into three different phases similar to our model. Taking into account their model's results for the countries of South Korea and Germany, They calculated the actual number of reported cases, the calculated number of cases estimated and the calculated number of unreported cases estimated. They found that for South Korea with its early and effective countermeasures against COVID-19, Their calculated number of cases was very similar to the reported actual number of cases and they also gave a rough estimation of the total number of unreported cases of the recent date according to data collection date in March. There was also a prediction of number of

cases for Germany where they showcased the calculated number of reported cases at the end of the epidemic with the actual number of cases estimated. This paper used a similar model to our and gave some initial parameters that were being considered by them as well estimated the population infection rates for us to compare to using our model.

3 METHODS AND SYSTEM DESIGN

3.1 Data Processing

Worldometers.info was used to retrieve reported numbers of active cases, recoveries, and fatalities per day for Germany and South Korea [Worldometer [n.d.]]. The Worldometers data source was chosen because it provided information on recoveries since mid-February, while other leading data sources did not add recovery data until much later. Additionally, for countries such as Italy, Worldometers provides information on hospitalization and severe case numbers. Like other papers have done, we only analyze cases that occur after the date when 100 cases were confirmed in that geographic area [Shet et al. 2020], [Roser et al. 2020]. Media reports from the Robert Koch Institute and Korean Ministry of Health (MOHW) were used to cross-reference null and conflicted values. The final values were tabulated and compared against our model predictions (Figure 1). The resulting data had a temporal resolution of daily measurements.

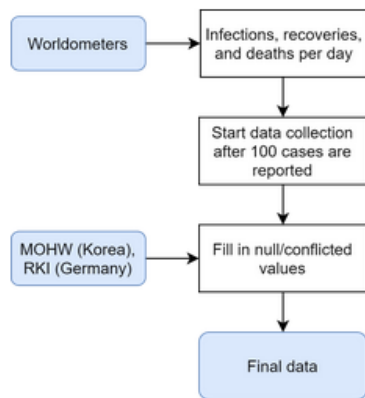


Figure 1: Workflow for creation of time series data for Germany and South Korea.

3.2 Informatics System

We constructed a basic SEIRD model to predict the cases of COVID-19 in Germany and South Korea (Figure 3). In order to parameterize this model, we first took a survey of parameters from literature (Table 1). We iterated through the range of literature values we encountered for each parameter to find initial values for our model that minimized the root sum of squares (RSS) between the actual and predicted cases, and these values were used to construct heatmaps. Finally, we used a truncated Newton optimization method to minimize the RSS between actual and predicted cases, with initial values

selected by the heatmaps [Dembo and Steihaug 1983]. These optimized parameters were compared with the literature values (Figure 2).

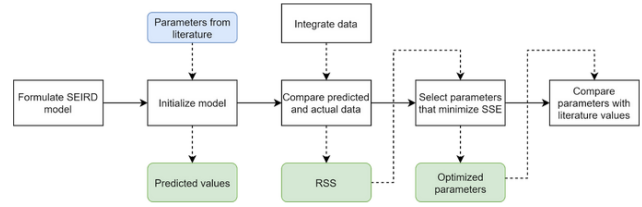


Figure 2: Initialization of an SEIRD compartmental model with parameters drawn from literature, comparison of predictions with the actual data using the root sum of squared error (RSS), and selection of the parameters that minimized RSS using heatmaps and a truncated Newton optimization method was done.

3.3 Prediction Methodology

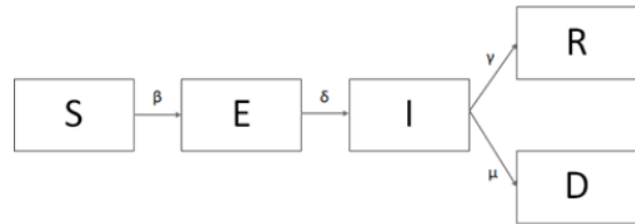


Figure 3: A 5-compartment SEIRD model was used for our analysis. β represents the contact rate between susceptible and infected individuals. δ represents the lag time between exposed and infected individuals. Infected individuals can either follow two outcomes: recovery or fatality. γ represents the recovery rate, while μ represents the mortality rate.

We use a simple SEIRD compartmental model to reflect the spread of COVID-19 in our two countries of interest. The differential equations for the SEIRD model are adapted from Rocklöv et al. and are displayed below [Rocklöv et al. 2020].

We added complexity to our model by utilizing a phase-adjusted analysis of viral transmission in each country; i.e. the effective transmission rate of the virus (R_t) was said to decrease as public health interventions were imposed by each country [Wang et al. 2020]. We model a country's epidemic in three distinct phases: a first phase corresponding to initial spread of the virus before any public health interventions, a second phase corresponding to decreased spread after the first public health intervention (usually border lockdown, but occasionally increased social distancing), and a third phase corresponding to extremely low viral transmission after both border lockdown and social distancing. Dates of public health interventions in Germany and South Korea were taken from media sources [tim 2020], [Welle [n.d.]]. While we are most concerned with public policy interventions, our analysis may also

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \delta E$$

$$\frac{dI}{dt} = \delta E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$\frac{dD}{dt} = \mu I$$

Parameter	Units	Formula
Mortality rate (μ)	1/days	$\frac{\text{reported fatalities}(\text{individuals})}{\text{reported cases}(\text{individuals})} * \frac{1}{\text{infectious period}(\text{days})}$
Recovery rate (γ)	1/days	$(1 - \frac{\text{reported fatalities}(\text{individuals})}{\text{reported cases}(\text{individuals})}) * \frac{1}{\text{infectious period}(\text{days})}$
Contact rate (β)	1/days	$R_t * \gamma$
Exposure rate (δ)	1/days	$\frac{1}{\text{incubation period}(\text{days})}$

Table 1: Definition of parameters in our compartmental model.

reveal changes in viral transmission caused by behavior changes in the populace.

Our model contained the following assumptions:

- We assume a fixed population size, so we do not account for birth, immigration, or mortality from any cause other than COVID-19.
- While we account for changes in effective transmission rate between phases, we assume that R_t stayed constant in each phase.
- We do not allow for the reinfection of individuals that have previously contracted COVID-19 and recovered.
- We assume that there is homogenous mixing of the population, even when social distancing is in place.

3.4 Performance Metric

Similar to Mulder et al. 2020, we used the root sum of squared errors (RSS) to compare our model's predictions to the actual case numbers data [Mulder 2020]. We calculated the root sum of squared errors as follows: the root sum of squared errors over n days is equal to the square root of the sum of the squared difference between the actual infected and predicted infected, actual recovered and predicted recovered, and actual dead and predicted dead values from day 0 to day n . As explained above, we only calculated RSS for

Parameter	Minimum Value	Maximum Value	Step
R_t for phase 1	1	20	4
R_t for phase 2	1	10	3
R_t for phase 3	1	NA	NA
Exposure rate (δ)	1/k; minimum value for k is 1	maximum value for k is 20	Step of 4 for k
Recovery rate (γ)	1/k; minimum value for k is 1	maximum value for k is 20	Step of 4 for k
Mortality rate (μ)	$1 * 10^{-k}$; minimum value of k is 0	maximum value for k is 7	Step of 1 for k
Date of phase 2 (in days after 100th case)	10	40	10
Date of phase 3 (in days after 100th case)	Start of phase 2 + 10 days	NA	NA

Table 2: Ranges tested for each parameter during the iterative optimization. The table gives the minimum value, maximum value, and steps used when iterating through possible values of each parameter. The parameters without maximum values and steps were assumed to take a single value and later optimized through the truncated Newton method.

the period of time after 100 cases were confirmed for each country.

$$RSS = \sum_{t=0}^n \sqrt{(I_t - \hat{I}_t)^2 + (R_t - \hat{R}_t)^2 + (D_t - \hat{D}_t)^2}$$

3.5 Parameter Selection

In order to parameterize our SEIRD model, we first took a survey of parameters from literature (Table 3). We iterated through the range of literature values we encountered for each parameter to find the initial values for our model that minimized the root sum of squared error (RSS) between the actual and predicted data (Table 2). This allowed us to find an approximate value for each parameter that minimized the RSS. Furthermore, the RSS values generated during this iterative process were used to construct heatmaps. Two specific heat maps were generated for each country: A comparison of R_t for phases 1 and 2, and a comparison of mortality rate and recovery rate (Figure 5, 7).

Parameter	Literature value for Germany	Literature value for South Korea
Initial population size	83,736,779	51,256,387
Lockdown date	March 25th, 24 days after 100th case [Welle [n.d.]]	February 23rd, 4 days after 100th case [tim 2020]
R_0	4.43 [Yuan et al. 2020]	1.5 [Shim et al. 2020]
Exposure rate (δ)	0.153 [Bulletin [n.d.]]	0.120 [Do et al. 2017]
Recovery rate (γ)	0.048 [Nguyen 2020]	0.079 [oso 2020]
Mortality rate (μ)	0.0013 [Lipsitch 2020]	0.016 [Dudel et al. 2020]

Table 3: Initial parameters from literature for Germany and South Korea.

We used the truncated Newton optimization method to further refine our parameters, with the objective function for the optimization set as the minimization of the RSS [Dembo and Steihaug 1983]. The truncated Newton method allowed us to set boundaries for each parameter being optimized, which we used to ensure that all parameters were non-negative and fell between certain reasonable ranges.

4 RESULTS

When we utilized parameters taken from literature in our compartmental model, the predicted case numbers were much lower than the actual case numbers (Figure 4). The iterative process described in the methods was used to find an approximate range for each of the parameters in our model. For Germany, we discovered that R_t of 13 for phase 1, an R_t of 1 for phase 2, a recovery rate of 0.0625, and a mortality rate of 10^{-6} resulted in the lowest RSS values (Figure 5). For South Korea, we discovered that an R_t of 17 for phase 1, an R_t of 1 for phase 2, a recovery rate of 0.0588, and a mortality rate of 0.001 result in the lowest RSS values (Figure 7). When the values returned from the iterative process were fed into the truncated Newton optimization method, the parameters in Table 4 were generated.

Since our model cannot predict the effective transmission rate (R_t) of future phases, it can only be used to make short-term predictions, and these predictions only hold as long as we assume that the last phase of public health interventions (border lockdown, social distancing, mask usage) continue. Comparisons of our model predictions to the actual data are shown in Figures 6 and 8.

While our model has reasonable accuracy for fatalities throughout all 3 phases of the epidemic in South Korea, we are not able to predict infections and recoveries in phase 1 and phase 2 of the lockdown very well in South Korea (Figure 8). However, the model predictions ultimately converge to the actual data towards the end of phase 3 of the epidemic in South Korea. In contrast, our model fares much better in Germany, as we are able to predict infections, recoveries, and fatalities fairly accurately in all three phases of our model (Figure 6).

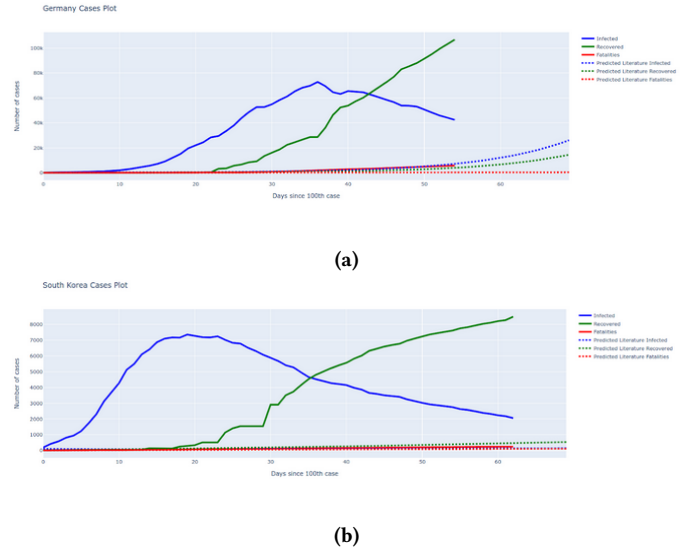


Figure 4: Predicted case numbers for Germany and South Korea utilizing consolidated literature parameters. Our SEIRD compartmental model was initialized using parameters taken from several literature sources. Since most of the papers we examined did not calculate R_t for each phase, only a single value for R_0 , we assumed that the reported R_0 value was constant over 60 days. a) Predicted values for Germany: $R_0 = 4.43$, $\gamma = 0.048$, $\delta = 0.153$, $\mu = 0.0013$. Even without accounting for a decrease in R_t , the resulting model vastly underrepresents the number of cases of COVID-19 in Germany (RSS = 4.493e5). b) Predicted values for South Korea: $R_0 = 1.5$, $\gamma = 0.079$, $\delta = 0.12$, $\mu = 0.016$. The resulting model also vastly underrepresents the number of cases of COVID-19 in South Korea (RSS = 5.1070e4).

Parameter	Germany	South Korea
R_t for phase 1	13.0	17.0
R_t for phase 2	4.0	1.0
R_t for phase 3	0.5767	0.8747
Start of phase 2, days since 100th case	20.0	10.0
Start of phase 3, days since 100th case	30.0	20.0
Exposure rate (δ)	0.200	0.200
Recovery rate (γ)	0.0603	0.0588
Mortality rate (μ)	0.0013	0.0011

Table 4: Finalized parameters selected by truncated Newton optimization. The table above gives the optimized parameters for both Germany and South Korea as selected by the truncated Newton optimization method.

In comparing our model's parameters between Germany and South Korea, we see that South Korea appears to start out with a higher R_t for phase 1 (17) than Germany (13) (Table 4). However, by phase 2, South Korea's R_t decreased sharply to 1, while Germany's R_t only

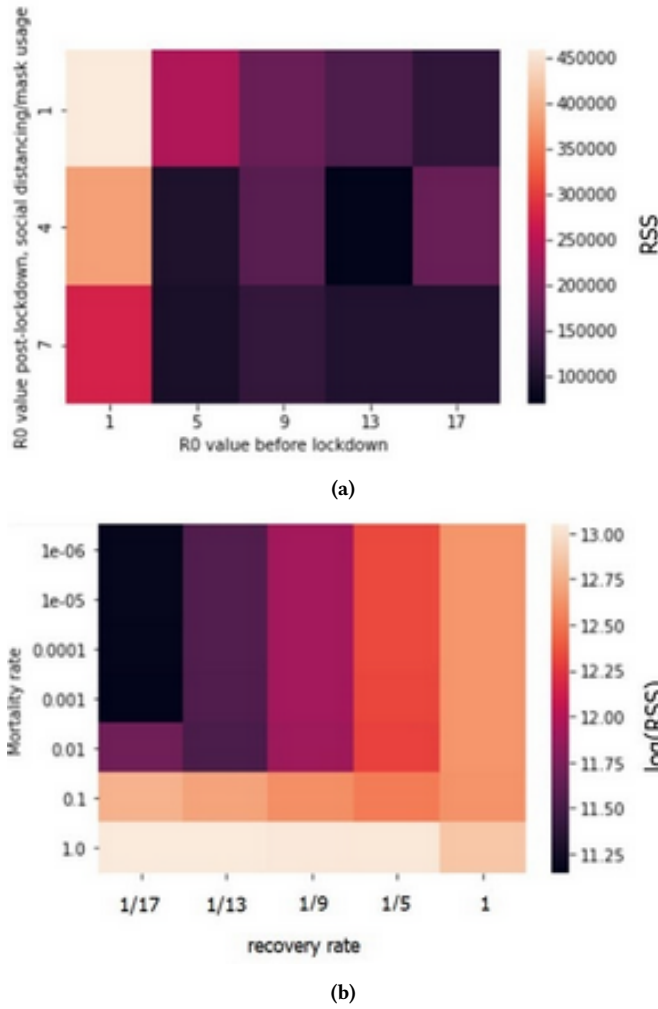


Figure 5: Heatmaps generated when iterating through parameters for Germany. a) The lowest RSS for the Germany data results from an R_t of 13 for phase 1 and an R_t of 1 for phase 2. b) For the recovery and mortality rates, a recovery rate of 0.0625 and a mortality rate of 10^{-6} result in the lowest RSS values.

decreased to 4. In contrast, the third epidemic phase in Germany seems to be characterized by a sharper decrease in case numbers than in South Korea, with an R_t for phase 3 in Germany of 0.5767 compared to an R_t for phase 3 in South Korea of 0.8747. The remaining parameters appear to be relatively similar in both countries (Table 4).

5 DISCUSSION

Our analysis provides an estimation of the start dates of the second and third phases of a COVID-19 epidemic in a geographical area. By comparing the phase start dates with the dates of public health interventions, we can estimate how long it takes for measures like border lockdown and social distancing to take effect. For example,

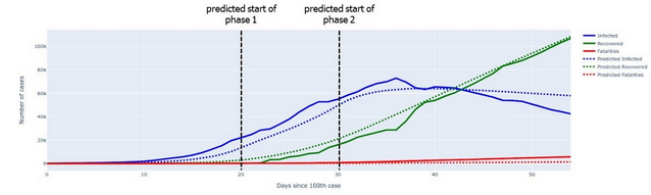


Figure 6: A comparison of actual Germany data and predicted case numbers derived from optimized model parameters. Phase 2 in Germany was shown to begin 10 days after the 100th case was reported, while phase 3 began 10 days after phase 2. R_t 1 = 13.0, R_t 2 = 4.00, R_t 3 = 0.5767, γ = 0.0603, δ = 0.200, μ = 0.0013. These parameters resulted in a relatively good fit to the actual data, especially in phase 3 of lockdown (RSS = 4.7918e4).

South Korea announced its border lockdown 4 days after its 100th case was confirmed [tim 2020]. We calculate that the second phase of the South Korea epidemic began 10 days after the 100th case was confirmed (Figure 7, 8). Our start date suggests that the effects of South Korea's border lockdown on R_t were apparent 6 days after the lockdown was announced. Surprisingly, a similar analysis implies that Germany saw a decrease in effective transmission rate before its first national public health policy, the announcement of social distancing on March 22nd [Welle [n.d.]]. While Germany announced social distancing 24 days after its 100th case was confirmed, we estimate that the second phase of the epidemic in Germany began 20 days after the 100th case was confirmed (Figure 5, 6) [Welle [n.d.]]. Our findings suggest that German citizens were utilizing individual public health measures like self-imposed social distancing and increased hygiene to effectively reduce viral transmission before border lockdown. South Korea saw a greater decrease in R_t from the first to the second phase than Germany, suggesting that South Korea's first public health intervention was more effective than Germany's. German public policies may have been less effective because the German government imposed its first national public health policy relatively late. Consequently, the second phase of the epidemic started later in Germany than in South Korea. Furthermore, South Korea locked down its borders before imposing social distancing, while Germany mandated social distancing before imposing a border lockdown. South Korea's proactive approach may also explain why South Korea's case numbers are so much smaller than Germany's total case numbers; at South Korea's epidemic peak on March 11th, it reported approximately 7500 cases, while at Germany's peak on April 16th, it reported approximately 73000 cases. Germany's epidemic peak was therefore an entire order of magnitude higher than South Korea's. Ultimately, we suggest that South Korean policies were more effective than German policies in containing the spread of COVID-19. Our model fit better to the first and second phases of the lockdown in Germany than in South Korea (Figures 7, 8). However, our model resulted in higher RSS values in Germany than South Korea. This difference is simply an artifact of the much higher number of actual infections, recoveries and deaths in Germany. Therefore, it is difficult to compare RSS values between Germany and South Korea. The initial R_t values produced by our

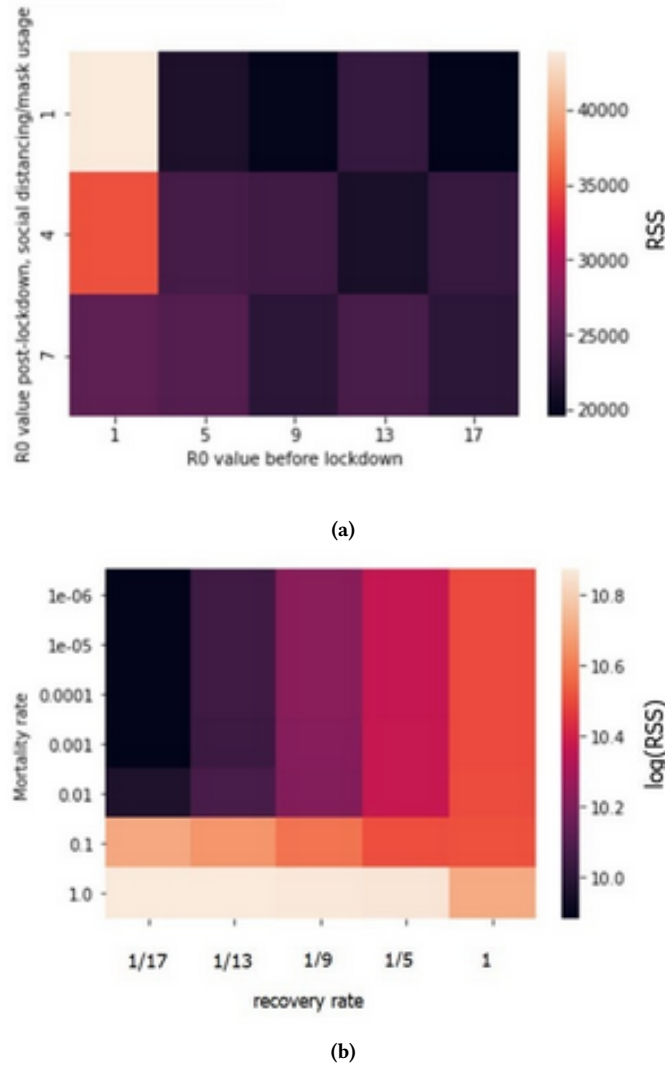


Figure 7: Heatmaps generated when iterating through parameters for South Korea. a) The lowest RSS for the South Korea data is produced by an R_t of 17 for phase 1 and an R_t of 1 for phase 2. b) For the recovery and mortality rates, a recovery rate of 0.0588 and a mortality rate of 0.001 result in the lowest RSS values.

model were 17 for South Korea and 14 for Germany. These initial R_t values are within the same range as the R_0 value calculated for the Diamond Princess cruise ship (14.8) [Rocklöv et al. 2020]. It is unrealistic for the effective transmission rate of COVID-19 to be higher in a country than a small, confined space like a cruise ship, which suggests that our analysis is seriously flawed. When we utilized parameters taken from literature in our compartmental model, the predicted case numbers were much lower than the actual case numbers (Figure 4). This suggests two things: that the parameters drawn from literature, especially the estimated transmission rates, lead to underestimations of the actual case numbers, and/or that

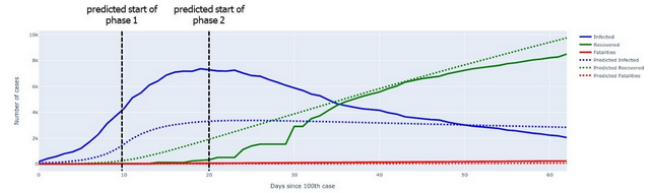


Figure 8: A comparison of actual South Korea data and predicted case numbers derived from optimized model parameters. Phase 2 in South Korea was shown to begin 10 days after the 100th case was reported, while phase 3 began 10 days after phase 2. R_t 1 = 17.0, R_t 2 = 1.00, R_t 3 = 0.875, γ = 0.0588, δ = 0.200, μ = 0.00113. These parameters resulted in a relatively good fit to the actual data, especially in phase 3 of lockdown (RSS = 1.9197e4).

our model assumptions do not hold true in the real world. There is reason to believe that the R_0 values for Germany and South Korea from literature that we used to initialize our model, which ranged from 1.5-4, are underestimates. For example, Sanche et al. estimate the initial R_0 of COVID-19 in Wuhan as 5.7, which is far above this commonly accepted range [Sanche et al. 2020]. On the other hand, Mulder et al. also utilized Worldometer data, and a similar method for calculating RSS, which resulted in a similarly high estimate of an R_0 of 20 in South Korea [Mulder 2020]. The high commonality between our results and theirs suggest a shared flaw in our data source or models rather than an implementation issue. Our work contains several weaknesses, many of which may explain the unexpectedly high R_t values generated by our model. First of all, since we directly integrated data from Worldometers, we did not consider the effects of underreporting in our model. Magal and Webb 2020 suggest that, while the actual reporting rate of COVID-19 cases are unknown, unreported cases could greatly influence the trajectory of epidemics in Germany and South Korea [Magal and Webb 2020]. Asymptomatic and mild cases are especially prone to underreporting. Since we did not account for reporting or testing rates in our model, our modeling approach was very sensitive to biases in our data source. The effects of this can be seen most likely in South Korea, where cases seemed to explode at an unprecedented rate as detection rates increased. Consequently, we calculated an R_t of 17 in phase 1 (Table 4). Some of our assumptions are valid: for example, we can assume that immigration is negligible or that the birth rates and death rates for both countries are approximately equal. The following assumptions of our model were violated by our data: while we accounted for changes in effective transmission rate between phases, we assumed that R_t stayed constant among each phase of lockdown. We did not account for the reinfection of individuals that have entered the recovered compartment, although this has been reported in several sources. Although several papers report asymptomatic infection of COVID-19, suggesting that individuals may be infectious even when in the infectious period, we did not account for this possibility [Team 2020]. South Korea's initial cases were greatly influenced by several super-spreaders, but our model did not have a way to account for this behavior, causing us to overestimate R_t values in the first phase. Finally, we assume

that populations undergo homogenous mixing even after social distancing is imposed. According to Mercer et al., basic reproduction number is rarely measured directly, and R_0 values generated by models are highly dependent on model structures and assumptions [Mercer et al. 2011]. R_0 will fluctuate if the rate of human–human or human–vector interactions varies over time or space. Although measuring the true R_0 value is possible during an outbreak of a newly emerging infectious pathogen that is spreading through a wholly susceptible population, rarely are there sufficient data collection systems in place to capture the early stages of an outbreak when R_0 might be measured most accurately. As a result, R_0 values are often overestimated due to many factors, including the nature of outbreaks that are used for estimation, incorrectly accounting for imported cases, underestimation of active cases due to low testing rates, and outbreaks arising in subpopulations with higher transmission rates. As we did not account for these factors within our model, our R_t values are likely gross overestimates. Finally, the use of a five-compartment model may have also limited the utility of our analysis. A more complex model, like an SEIRHQD model, may account for variable transmission in subsets of the population and therefore be more suitable for COVID-19. For example, self-isolated individuals may exhibit lower viral transmission than the overall population.

Novelty Introduced

We adapted the common methodology of phase adjusted estimation in order to estimate effective transmission rates of COVID-19 in Germany and South Korea. We specifically aimed to compare these two countries because of their relative success in containing COVID-19. Furthermore, we propose a novel method to estimate two dates at which R_t changed significantly within a national epidemic of COVID-19. By comparing these dates with the dates that lockdown and social distancing were announced, we can estimate the date at which public health interventions take effect. While our preliminary results are not comparable with outside sources, we can compare the parameters generated by our analysis between countries.

We have further developed a GUI that allows users to investigate the effect of various public health interventions on case numbers drawn from user-supplied data. We require the user to input time series data for a region they want to model, with columns for infections, recoveries, and fatalities. Users must also enter the initial population size of their region of interest. The GUI provides optimized parameters for the region's spread of COVID-19: start date for phase 2, start date for phase 3, R_t for phase 1, R_t for phase 2, R_t for phase 3, exposure rate, recovery rate, and mortality rate. Additionally, it produces heatmaps that depict the optimal values for R_t for phase 1, R_t for phase 2, mortality rate, and recovery rate, and it calculates the RSS that results from these optimal parameters. Ultimately, our analysis can help users understand how R_t varies across phases of an epidemic, reveal whether public health interventions imposed upon the region had a major effect on the effective reproductive rate of the virus, and pinpoint the dates when R_t changed the most.

6 CONCLUSION

Ultimately, we modeled the spread of COVID-19 in a phase-adjusted manner for both South Korea and Germany. Our model used a basic optimization method to estimate parameters (R_t for three phases, recovery rate, infection rate, and mortality rate) for both countries. Furthermore, we compared the optimal parameters generated by our model to the parameters of SEIRD models reported in literature. Our approach allows us to estimate the dates of significant changes in effective transmission rate of COVID-19, which we can compare to the date of public health interventions in each country. The parameters generated by our model may not be comparable with outside sources, but we can easily compare the parameters generated by our analysis between countries. Consequently, we were able to compare the effect of public health interventions in Germany and South Korea using our model, concluding that South Korea's first lockdown resulted in a greater decrease in the R_t due to the earlier lockdown and use of border lockdown before social isolation. Our project allowed us to investigate epidemiological modeling in the context of a major global health crisis, and we observed the influence of mathematical models on policy decisions. Additionally, we identified variables that increase the complexity of a particular disease beyond a simple SEIRD model. In the case of COVID-19, underreporting, asymptomatic transmission, and a lack of serological testing prevent scientists from reliably modeling the spread of COVID-19. The ongoing public health crisis has prevented literature from reaching a consensus on best practices or important parameters relevant to the disease, which meant that we relied partially on media sources to develop our understanding. Similarly, a major challenge in our project involved the design of our compartmental model. We needed to clearly state our parameters and their units, list the assumptions of our model, find reliable literature parameters to initialize our model, and diagnose problems within our model.

ACKNOWLEDGMENTS

We would like to thank Dr. Wang, Ryan Hoffman and Li Tong for their constant support and encouragement

REFERENCES

2020. Early Epidemiological and Clinical Characteristics of 28 Cases of Coronavirus Disease in South Korea. *Osong Public Health and Research Perspectives* 11, 1 (2020), 8–14. <https://doi.org/10.24171/j.phrp.2020.11.1.03>
2020. The Latest: Bars, clubs in Seoul closed to halt virus spread. <https://www.washingtontimes.com/news/2020/apr/7/the-latest-new-zealand-official-demoted-for-lockdo/>
- WHO Bulletin. [n.d.]. https://www.who.int/bulletin/online_first/20-255695.pdf
- Ron S. Dembo and Trond Steihaug. 1983. Truncated-newton algorithms for large-scale unconstrained optimization. *Mathematical Programming* 26, 2 (1983), 190–212. <https://doi.org/10.1007/bf02592055>
- Mijin Do, Jongtae Kim, and Boseung Choi. 2017. A study of epidemic model using SEIR model. *Journal of the Korean Data and Information Science Society* 28, 2 (2017), 297–307. <https://doi.org/10.7465/jkdi.2017.28.2.297>
- Christian Ducl, Tim Riffe, Enrique Acosta, Alyson A Van Raalte, Cosmo Strozza, and Mikko Myrskylä. 2020. Monitoring trends and differences in COVID-19 case fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. (Jan 2020). <https://doi.org/10.31235/osf.io/j4a3d>
- Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Cmmid Ncov Working Group, and et al. 2020. Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts. (Nov 2020). <https://doi.org/10.1101/2020.02.08.20021162>

- Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, and Rosalind M Eggo. 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. (Feb 2020). <https://doi.org/10.1101/2020.01.31.20019901>
- Marc Lipsitch. 2020. Estimating case fatality rates of COVID-19. *The Lancet Infectious Diseases* (2020). [https://doi.org/10.1016/s1473-3099\(20\)30245-0](https://doi.org/10.1016/s1473-3099(20)30245-0)
- Pierre Magal and Glenn Webb. 2020. Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany. (2020). <https://doi.org/10.1101/2020.03.21.20040154>
- G. N. Mercer, K. Glass, and N. G. Becker. 2011. Effective reproduction numbers are commonly overestimated early in a disease outbreak. *Statistics in Medicine* (2011). <https://doi.org/10.1002/sim.4174>
- Leslie Mulder. 2020. Use of simulated annealing to determine the operational parameters of the SEIR model for the coronavirus for various jurisdictions. (2020). <https://doi.org/10.2471/blt.20.260513>
- APO Group Africa Newsroom and World Health Organization. [n.d.]. Coronavirus - Africa: WHO Director-General's opening remarks at the media briefing on COVID-19. <https://who.africa-newsroom.com/press/coronavirus-africa-who-directorgenerals-opening-remarks-at-the-media-briefing-on-covid19>
- Trung Nguyen. 2020. Germany COVID-19 social distancing simulation. <https://towardsdatascience.com/covid-19-social-distancing-simulation-f091a58732f9>
- David Rising. 2020. Germany and South Korea have done well fighting coronavirus. Here's why. <https://www.chicagotribune.com/coronavirus/ct-nw-south-korea-germany-coronavirus-covid-19-20200422-ca2wbvagu5dl7kjqaq5bg7jnpe-story.html>
- J Rocklöv, H Sjödin, and A Wilder-Smith. 2020. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *Journal of Travel Medicine* (2020). <https://doi.org/10.1093/jtm/taaa030>
- Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina, and Joe Hasell. 2020. Coronavirus Pandemic (COVID-19) - Statistics and Research. <https://ourworldindata.org/coronavirus>
- Steven Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, and Ruian Ke. 2020. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases* 26, 7 (2020). <https://doi.org/10.3201/eid2607.200282>
- Anita Shet, Debashree Ray, Neelika Malavige, Mathuram Santosham, and Naor Bar-Zeev. 2020. Differential COVID-19-attributable mortality and BCG vaccine use in countries. (Jun 2020). <https://doi.org/10.1101/2020.04.01.20049478>
- Eunha Shim, Amna Tariq, Wongyeong Choi, Yiseul Lee, and Gerardo Chowell. 2020. Transmission potential of COVID-19 in South Korea. (2020). <https://doi.org/10.1101/2020.02.27.20028829>
- Edward Hance, Shortliffe and James J. Cimino. 2014. *Biomedical informatics: computer applications in health care and biomedicine*. Springer.
- WHO situation report. [n.d.]. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200420-sitrep-91-covid-19.pdf?sfvrsn=fcf0670b_4
- Internal Submission. [n.d.]. In Germany and South Korea, preparation and quick action aid coronavirus fight. <https://www.japantimes.co.jp/news/2020/04/29/world/germany-south-korea-preparation-coronavirus/>
- Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. 2020. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) - China, 2020. <http://weekly.chinacdc.cn/en/article/doi/10.46234/cdcw2020.032>
- Huwen Wang, Zezhou Wang, Yinqiao Dong, Ruijie Chang, Chen Xu, Xiaoyue Yu, Shuxian Zhang, Lhakpa Tsamag, Meili Shang, Jinyan Huang, and et al. 2020. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. *Cell Discovery* 6, 1 (2020). <https://doi.org/10.1038/s41421-020-0148-0>
- Deutsche Welle. [n.d.]. Coronavirus: What are the lockdown measures across Europe?: DW: 14.04.2020. <https://www.dw.com/en/coronavirus-what-are-the-lockdown-measures-across-europe/a-52905137>
- Worldometer. [n.d.]. Coronavirus Cases:. <https://www.worldometers.info/coronavirus/>
- Zunyou Wu and Jennifer M. McGoogan. 2020. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. *Jama* 323, 13 (Jul 2020), 1239. <https://doi.org/10.1001/jama.2020.2648>
- Jing Yuan, Minghui Li, Gang Lv, and Z. Kevin Lu. 2020. Monitoring Transmissibility and Mortality of COVID-19 in Europe. *International Journal of Infectious Diseases* (2020). <https://doi.org/10.1016/j.ijid.2020.03.050>