

https://dspace.org/

Modelling the spread of COVID-19 in South Africa using stratified compartmental models in the period March 2020 - August 2020

Mbayise, Elona

2022
https://hdl.handle.net/10539/34497

Downloaded from a DSpace repository. https://dspace.org/



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Modelling the spread of COVID-19 in South Africa using stratified compartmental models in the period March 2020 - August 2020

by

Elona Mbayise

Supervisor

Raeesa Docrat

.

.

A Dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

30 April 2022

Declaration

I declare that this dissertation is my own unaided work. It is being submitted for the degree of Mathematical Statistics Masters at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.



Signed on the 30th day of November 2021 in Johannesburg.

Acknowledgements

I would first like to thank my supervisor, Raeesa Docrat, whose expertise was invaluable in my thesis. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would also like to thank Charl Janse Van Rensburg and Nada Abdelatif for their guidance and advice throughout the process.

I am very thankful to Canada's International Development Research Centre (IDRC) (Grant No. 109559-001) for funding my studies and my research.

Finally, I would like to thank my mother and sister for their unending support and encouragement.

Abstract

The novel coronavirus strand (SARS-CoV-2) first appeared in Wuhan, China in December 2019 and caused the respiratory syndrome COVID-19. A unique feature of COVID-19 is its non-uniform effect on populations. The effects of COVID-19 are more severe amongst older age groups and people with co-morbidities as seen by the higher mortality, infection and hospitalisation rates observed amongst these groups. This study models the spread of COVID-19 in South Africa between March-August 2020 using stratified compartmental models to capture this population heterogeneity.

To understand the unique dynamics of COVID-19, an age stratified and co-morbidity stratified compartmental model was built with additional compartments and parameters. A sensitivity analysis was then performed to determine the models' sensitivity to initialisation date and lockdown level to determine the optimal pandemic start date and to identify the effects of harsh lockdown restrictions on infections, hospitalisations and deaths. A parameter sensitivity analysis was also conducted to determine the parameters that needed to be re-estimated to improve model accuracy and to identify the age groups which were driving infections, hospitalisations, and deaths. These analyses showed that a prolonged harsh lockdown would have reduced infections by approximately 46% and delayed the infection peak by 4-6 months. The analyses also showed that hospitalisations were driven by the 61-75 age group while infections and deaths were driven by the 76-90 age group. In addition, the model was most sensitive to infection duration, death rate and proportion of asymptomatic infection. These parameters were then re-estimated using Gauteng hospitalisation data and literature to account for the effects of age and co-morbidity and a χ^2 Goodness of Fit test proved that making parameters age/co-morbidity dependent improved the accuracy of all models.

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Aims and Objectives . . . 3
- 1.4 Summary . . . 3

2 Literature Review **5**
- 2.1 Compartmental Modelling in Epidemiology . . . 5
- 2.2 Deterministic versus Stochastic Models . . . 6
- 2.3 Multi-Compartmental Models . . . 7
- 2.4 Parameter and Assumption Adjustments . . . 8
- 2.5 Stratified Models . . . 10

3 The Dataset **12**

4 Methodology **18**
- 4.1 Basic SIR Model . . . 18
- 4.2 Basic Reproduction Number . . . 19
- 4.3 Stratified Compartmental Model . . . 20
- 4.4 Age Stratified Model . . . 21
- 4.5 Co-morbidity Stratified Model . . . 27
- 4.6 Sensitivity Analysis . . . 30
  - 4.6.1 One-by-one sensitivity analysis . . . 30
  - 4.6.2 Multivariate sensitivity analysis . . . 31
  - 4.6.3 Sensitivity to initialisation date . . . 31
  - 4.6.4 Sensitivity to lockdown level . . . 32
  - 4.6.5 Sensitivity to transmission matrix β . . . 32
- 4.7 Re-estimation of Parameters . . . 37
- 4.8 Comparison of Models and Prediction . . . 38

5 Analysis **40**
- 5.1 Sensitivity Analysis . . . 40
  - 5.1.1 Sensitivity to initialisation date . . . 40
  - 5.1.2 Sensitivity to lockdown restrictions . . . 42
  - 5.1.3 One-by one sensitivity analysis . . . 45
  - 5.1.4 Multivariate sensitivity analysis . . . 48
  - 5.1.5 Sensitivity of transmission matrix β . . . 50
- 5.2 Parameter re-estimation . . . 52

5.2.1 Re-estimating ρ_2, ρ_4 and μ_i 52
5.2.2 Re-estimating p 54
5.3 Fitting model to the second wave 55

6 Conclusion **57**
6.1 Discussion 57
6.2 Limitations 58
6.3 Further work 58

References **60**

A Understanding the parameters **64**

B R Code **66**

List of Figures

1 Observed COVID-19 cases during the first infection wave . . 13
2 Hospitalisations by age 14
3 Deaths by age 15
4 Co-morbidities of hospitalised COVID-19 patients 16
5 Co-morbidity specific death distributions 17
6 Basic SIR Model 18
7 Age stratified COVID-19 Model 22
8 The (a) Population pyramid and (b) Contact matrix for South Africa 26
9 Co-morbidity stratified model for healthy stratum 28
10 Co-morbidity stratified model for co-morbidity stratum . . . 28
11 Co-morbidity models with and without an extended Level 5 lockdown 43
12 Age models with and without an extended Level 5 lockdown . 44
13 Sensitivity of parameters to hospitalisations 47
14 Sensitivity of parameters to total infections 48
15 Sensitivity of parameters to total deaths 49
16 Transmission matrices from Methods 1 and 2 of estimating β 52

List of Tables

1 Age stratified model Parameters 23
2 Expected age stratified model parameter changes 25
3 Expected co-morbidity stratified model parameter changes . . 29
4 Model Parameters 41
5 Determining the optimal start date 41
6 Multivariate sensitivity analysis results 49
7 Determining the age groups driving hospitalisations, infections and deaths 50
8 Age specific ρ_2, ρ_4 and μ_i 53
9 Co-morbidity specific deaths 54
10 Age-specific p 55
11 Co-morbidity specific deaths 55
12 Effects of changes in parameter estimates 65

Acronyms

AIC - Akaike Information Criterion
BIC - Bayesian Information Criterion
GDP - Gross Domestic Product
GoF - Goodness of Fit
HIV - Human Immunodeficiency Virus
ICU - Intensive Care Unit
MSIR - Maternal antibody Susceptible Infected Recovered
NGM - Next Generation Matrix
NICD - National Institute of Communicable Diseases
PPE - Personal Protective Equipment
SEIR - Susceptible Exposed Infected Recovered
SEIS - Susceptible Exposed Infected Susceptible
SIR - Susceptible Infected Recovered
SIRS - Susceptible Infected Recovered Susceptible
SIS - Susceptible Infected Susceptible
WHO - World Health Organisation

1 Introduction

1.1 Background

The novel coronavirus strand (SARS-CoV-2) first appeared in Wuhan, Hubei province, China in December 2019 and caused the respiratory syndrome COVID-19 (World Health Organisation, 2020). Due to the highly infectious nature of the virus, it spread globally over a short period of time and COVID-19 was declared a global pandemic by the World Health Organization (WHO) in March 2020 (Wu, Leung, & Leung, 2020). Since then, the COVID-19 pandemic has had devastating health, economic, social and psychological effects on countries across the world (McKibbin & Fernando, 2020). Exponentially increasing numbers of infections and fatalities were recorded globally putting severe pressure on health systems. Secondary effects of the pandemic were seen as a result of restrictions such as nationwide lockdowns. These restrictions resulted in an increase in unemployment (Blustein et al., 2020), an increase in the number of people below the poverty line, declines in mental health (Xiong et al., 2020), closing of businesses and crumbling economies (McKibbin & Fernando, 2020).

One of the greatest challenges facing leaders during this pandemic is striking a balance between saving lives and saving livelihoods. This requires an understanding of the dynamics of COVID-19 for timely impositions and relaxations of restrictions. Understanding the dynamics of the corona virus requires mathematical and statistical modelling. Possible methods for understanding the spread and dynamics of COVID-19 include the use of compartmental models. Compartmental models are used frequently in Epidemiology to understand the spread of infectious diseases (Brauer, 2008). The purpose of these models is to predict how a disease spreads, the duration of an epidemic, the number of people infected and the severity of an epidemic. Under these models, people in a closed population are classified as Susceptible, Infected or Recovered/Immune to a certain infectious disease over a given time frame (Cooper, Mondal, & Antonopoulos, 2020). Transitions can be made between these compartments at rates dependent on the dynamics of the disease in question. Differential equations are used to determine the rate of change from one compartment to another (Gibson & Renshaw, 1998). Compartmental models have been used to simulate the spread of diseases such as mumps, measles, rubella, HIV/AIDS (Silva & Torres, 2017) and more recently COVID-19.

A unique feature of COVID-19 is it's non-uniform effect on populations. Older age groups and people with co-morbidities are affected more severely and hence have notably higher mortality rates (Balabdaoui & Mohr, 2020). The most common compartmental model in Epidemiology is the SIR model where individuals in a population are assigned to the Susceptible (S), Infectious (I) or Recovered/Immune (R) (SIR) compartment. Rates of transition between compartments are calculated from historical data. The SIR model is less than optimal in modelling the spread of COVID-19 as it has insufficient compartments to fully capture the dynamics of the disease. The compartments under the COVID-19 model are therefore: Susceptible, Exposed, Asymptomatic, Infected, Infectious and hospitalised, Infectious in ICU, Recovered and Dead. Another downfall of the SIR model is its inability to capture COVID-19's non-uniform effect on people with respect to age (Cooper et al., 2020) and co-morbidity. Stratified models are a solution for capturing this heterogeneity in populations. In the context of modelling COVID-19, stratified models are used to understand the dynamics of the virus for each stratum. This involves stratifying the population by age and co-morbidity then determining the mortality rate, reproduction number, transition rates and expected pandemic duration for each age and co-morbidity stratum. These models are then compared to the original compartmental model formed prior to stratification.

1.2 Problem Statement

The effects of COVID-19 have been felt globally. A notable attribute of COVID-19 is its non-uniform attack on people of different ages and people with different co-morbidities. Compartmental models have been used frequently in Epidemiology to understand the spread and dynamics of COVID-19. An assumption of these models thus far has been that populations are homogenous and hence the effects of COVID-19 are uniform across age demographics and underlying illnesses. This research uses stratified compartmental models to understand the dynamics of COVID-19 for various age groups and co-morbidities to account for the non-uniform attack of COVID-19, specifically in South Africa.

1.3 Aims and Objectives

The aim of this study is to model the spread of COVID-19 in South Africa using stratified compartmental models. The objectives are:

1. To build compartmental models and estimate parameters based on historical data and literature.
2. To build stratified compartmental models to identify the effects of COVID-19 on people from different age groups and people with various underlying illnesses during the first infection wave.
3. To conduct a sensitivity analysis to identify how sensitive the models are to changes in parameter estimates, initialisation dates and lock-down levels.
4. To fit the models and re-estimate parameters using hospitalisation data and literature.
5. To compare standard compartmental models with age/co-morbidity stratified compartmental models using goodness of fit tests. This is done to determine the effect of stratifying populations and the viability of doing so. These effects include the mortality rate, hospitalisation rate, reproduction number and transition rates.
6. To make final predictions about the COVID-19 dynamics across age demographics and underlying illnesses for future infection waves/-surges. Vaccinations were ignored for the purposes of this study.

1.4 Summary

The Literature Review in Chapter 2 gives an overview of compartmental modelling in Epidemiology, specifically in the context of COVID-19. This involves the history, applications and extensions of the basic SIR (Susceptible - Infected - Recovered) model. The review also explores using compartmental modelling as a means of capturing population heterogeneity and unique disease dynamics through parameter adjustments, additional compartments and assumption adjustments.

The dataset comprising of observed active COVID-19 cases is described in Chapter 3. The chapter also includes an Exploratory Data Analysis (EDA) used to gain an understanding of the distribution, volatility and rate of

increase of observed COVID-19 infections. The EDA includes a written analysis and graphical representations of the observed data.

Chapter 4 describes the methodology used for this study. Sections 4.1 - 4.5 outline the compartments, parameters, differential equations and assumptions used to build the basic, age stratified and co-morbidity stratified models. These sections also explore the influence of contact network heterogeneity on transmissibility especially in developing countries with young populations. Section 4.6 outlines the methods used to determine the models' sensitivity to parameter changes (One-by-one and Multivariate), transmission probabilities, initialisation dates and lockdown levels. In addition, two methods are proposed to estimate transmission matrix β. The first method assumes that susceptibility and transmissibility are highest for younger people since they are more social and have more daily contacts. The second method assumes that susceptibility and transmissibility are highest for older individuals since the effects of COVID-19 worsen with age. Section 4.7 describes the process of re-estimating highly sensitive parameters from hospitalisation data and from literature for parameters relating to asymptomatic infection. Finally, Section 4.8 describes the process of fitting the models with re-estimated parameters to the second infection wave and assessing the degree of correlation between the observed and modelled infections using a χ^2 Goodness of Fit (GoF) test. Chapter 5 includes an analysis of the results.

2 Literature Review

The first case of the respiratory syndrome COVID-19 was observed in Wuhan China in December 2019. Since then, there has been a surge of interest in understanding the dynamics of COVID-19 and modelling its spread using statistical and mathematical models. A variety of method have been used to model the spread of COVID-19. These models include mathematical, logistic (Ahmed, Salam, Mohammad, Akgul, & Khoshnaw, 2020), Markov (Arumugam & Rajathi, 2020) and compartmental (Cooper et al., 2020) models amidst several others. Statistical and mathematical modelling of the spread of infectious diseases has been particularly of interest to governments globally to aid in adequately preparing their health systems for an influx of patients and meeting high demands for Personal Protective Equipment (PPE) and supportive medical supplies such as ventilators. These models have also been useful in making timely adjustments to fiscal and monetary policies to mitigate the secondary effects of COVID-19 which include business closures, suppressed financial sectors caused by lockdown restrictions, negative Gross Domestic Product (GDP) growth and rising unemployment rates (Ataguba, 2020).

The focus of this review is compartmental models used to understand the unique dynamics of COVID-19. This includes a history of compartmental modelling in epidemiology, applications of compartmental models, an exploration of multi-compartmental models and extensions of the basic SIR model, the effects of adjusting parameters and assumptions of the basic SIR model to account for the unique characteristics of COVID-19 and finally an exploration of stratified compartmental models used to capture population heterogeneity and the non-uniform attack of COVID-19 and different age, gender and income groups.

2.1 Compartmental Modelling in Epidemiology

Compartmental models have been used in Epidemiology since the early 1900s (Ross, 1916). These models have been used to understand the spread of several infectious diseases including measles (Bjørnstad, Finkenstädt, & Grenfell, 2002), HIV/AIDS (Silva & Torres, 2017), rubella (Thompson, 2016), swine flu (H1N1) (Coburn, Wagner, & Blower, 2009) and recently COVID-19. These models have been of particular interest because they accurately capture the dynamics of infectious diseases. This is because compartmental models are versatile and can be used to model various aspects of infectious

diseases by altering parameters or adjusting the number of compartments in the model. Compartmental models are also useful for estimating the duration of epidemics and forecasting the extent of the effects of infectious diseases (Brauer, 2008).

Compartmental models are also used to determine transitions between states or compartments. The basic compartmental model is the SIR model. Under the SIR model, individuals in a closed population are classified as Susceptible (S), Infected (I) or Recovered/Immune (R) (SIR). Probabilities of transitioning between these states are estimated from historical data and/or literature. Rarely does the basic SIR model fully capture the dynamics of an infectious disease due to having too few parameters, too few compartments or having a heterogeneous population. Another consideration is whether these models are deterministic or stochastic. As a result, extensions and adjustments have been made to the basic SIR model to capture the unique characteristics of an infectious disease.

2.2 Deterministic versus Stochastic Models

Both deterministic and stochastic compartmental models have been used in epidemiology to model the spread of infectious diseases. Deterministic models are useful for modelling the average behaviour of large populations while stochastic models are useful for modelling the behaviour of smaller populations and for assessing the variability of estimates and predictions (Ndiaye, Tendeng, & Seck, 2020).

Due to the novelty and volatility of COVID-19, deterministic and stochastic models have been used to model its spread. Stochastic compartmental models enable an understanding of the dynamics of COVID-19 over a continuous time interval. This means that in a closed population, the number of Susceptible (S(t)), Infected (I(t)) and Recovered/Immune (R(t)) for $t \in [0, \infty)$ individuals varies over time. This also means that the rates of transition between compartments also vary with time. Stochastic compartmental models allow an understanding of infectious diseases with dynamics that change over time (Simha, Prasad, & Narayana, 2020). Since COVID-19 is a novel epidemic, its dynamics have varied considerably since it first appeared in December 2019. These disease dynamic variations have been caused by improved treatments, scientific research, disease prevention measures (lockdowns, social distancing, sanitation, wearing masks etc.,) and the introduction of vaccines. The dynamics of COVID-19 varying with time allow for

time-dependent compartments and transition probabilities and are an argument for the viability of stochastic compartmental models in modelling the spread of COVID-19. A limitation of these models is that they have a high number of highly sensitive parameters. This means that continuous parameter re-estimation is necessary to yield accurate results and predictions. It also means that age structures, co-morbidities, spatial relationships, seasons, health care accessibility and other factors affecting the COVID-19 dynamics must be ignored to avoid having a parameter saturated model and hence over fitting and weak predictions (Wacker & Schlüter, 2020).

Deterministic compartmental models do not explain the dynamics of an infectious disease over time. Instead they model the average behaviour of infectious diseases in large populations. Since COVID-19 is a global pandemic, compartmental models are built using extensive datasets. Deterministic models therefore yield similar results, estimates and predictions to stochastic models and that the accuracy associated with stochastic models is maintained in a deterministic model (Wacker & Schlüter, 2020). Deterministic compartmental models have fewer parameters than stochastic models meaning that factors such as age structures and spatial relationships can be included in deterministic model without having a parameter saturated model (Balabdaoui & Mohr, 2020). This argument suggests that deterministic compartmental models are more viable than stochastic models for modelling the spread of COVID-19, especially since COVID-19 is new. Stochastic models may be preferable in the long run.

2.3 Multi-Compartmental Models

One of the limitations of the basic SIR model is its inability to fully capture the dynamics of infectious diseases due to having too few compartments (Balabdaoui & Mohr, 2020). Although most infectious diseases involve susceptibility, infection and recovery, considerations must be made to account for the unique attributes of a disease. Some adjusted models include the SIS (Susceptible-Infected-Susceptible) model used when there is no immunity upon recovery e.g., HIV/AIDS; the MSIR (Maternal antibody-Susceptible-Infected- Recovered) model used for infections where infants are born with immunity or become immune after being vaccinated e.g., Hepatitis B; the SIRS (Susceptible-Infected-Recovered-Susceptible) model used when the duration of recovery/ immunity is short e.g., COVID-19 (Ross, 1916) and the SEIS (Susceptible - Exposed-Infected-Susceptible) and SEIR (Susceptible-Exposed- Infected- Recovered) models used when there is a latent period

prior to infection e.g., cancer (Korobeinikov, 2004).

Due to the complex nature of COVID-19, several compartments were added to the basic SIR model to fully capture the dynamics of the disease. Adding compartments was necessary to avoid incorrect parameter estimates, to avoid under or over estimating the COVID-19 epidemic duration, to minimize model sensitivity and to avoid under or over estimating transition probabilities (Landaw & DiStefano 3rd, 1984). Having too many compartments can be disadvantageous as it may result in over fitting and hence incorrectly estimated model parameters and transition probabilities. The stochastic multi-compartmental model developed by South Africa's National Institute of Communicable Diseases (NICD) was an extension of the basic SEIR model. The model was modified to incorporate asymptomatic and isolated/hospitalised individuals which are unique features of the COVID-19 epidemic. The model therefore includes the following mutually exclusive compartments: Susceptible, Exposed, Asymptomatic Infectious, Symptomatic Infectious, Isolated/hospitalised, Recovered and Dead (Garba, Lubuma, & Tsanou, 2020). Sub-compartments were included to account for the extent of infection i.e., pre-infection, mild infection and severe infection.

2.4 Parameter and Assumption Adjustments

One of the advantages of compartmental models is that transition probabilities can be adjusted to account for different aspects of an infectious disease and to minimise model sensitivity. Another advantage is that assumptions can be relaxed to account for unique attributes of infectious diseases (Garba et al., 2020). The assumptions of classic compartmental models are as follows:

- The population is closed
- There is homogenous mixing of individuals
- Demographic and spatial features are ignored
- The population is homogenous
- Incubation is instantaneous
- No individual has initial immunity

- Only susceptible individuals can be infected and recovery/immunity is guaranteed after infection

As with the number of compartments, these model assumptions are not applicable to all infectious diseases. These assumptions enable ease of understanding and computation however they are not realistic in the context of COVID-19 (Adamu, Muhammad, Jingi, & Usman, 2019). The assumption that the population is closed is inaccurate due to international travel. It is however true during periods where borders are closed as a means of limiting the spread of the virus. Homogenous mixing of individuals is usually an accurate assumption however with lockdown restrictions this is not the case. Spatial and demographic features should be considered because COVID-19 spreads when people are in close proximity of one another. This means that transition probabilities vary depending on how densely populated an area is. Another assumption which needs to be adjusted is that incubation is instantaneous because COVID-19 symptoms only appear after 2-14 days or not at all if an individual is asymptomatic (Lauer et al., 2020).

Compartmental models have been built with relaxed assumptions to account for the unique nature of the COVID-19 epidemic. COVID-19 is a global epidemic and governments globally have been implementing measures such as lockdowns, border closures and curfews to limit the spread of COVID-19. The constant implementation and relaxation of measures has resulted in waves of infections (fluctuating numbers of infections) and has adversely affected the accuracy of statistical models.

Cooper proposed an extended SIR model to model the spread of COVID-19 (Cooper et al., 2020). The study was conducted based on data from January 2020 to June 2020 and the countries studied were China, South Korea, India, Australia, USA and Italy. The model relaxed the assumptions that populations are closed and that mixing is homogenous in a population (Cooper et al., 2020). An interesting attribute of the model was that the susceptible population was considered a variable which could be adjusted at various times to account for surges in infections. This was useful in understanding how the virus spreads over time, especially during infection surges.

The study showed that densely populated countries like India have notably higher transmission rates than sparsely populated countries. The study also showed that the increase in infections occurs quickly and exponentially hence harsh lockdowns must be implemented while infection numbers are low to

lower susceptibility and hence prevent surges in infections. Minimizing susceptibility results in fewer infections, fewer surges in infection and hence fewer deaths.
The advantage of having the susceptible population as a variable is that the model provides insights on the dynamics of COVID-19 during infection surges and declines. Another advantage of the study is that a variety of countries with different age structures, population densities and income distributions were studied to identify the effects of these factors on the behaviour of the virus. The disadvantage of the study was that the study was based on data over a short period of time which included a maximum of two waves of infection per country. To improve accuracy and reliability of conclusions and predictions, the study duration would have to be extended.

2.5 Stratified Models

An assumption of compartmental models is that the population is homogeneous and that demographics should be ignored. This is not realistic in the context of COVID-19 as its effects vary depending on age, gender, co-morbidity and population density amongst others. The effects of COVID-19 have been harshest for older people, males, people with underlying conditions and people who live in densely populated areas (Davies et al., 2020). A means of accounting for this population heterogeneity is model stratification. This is involves stratifying a population by age, gender, co-morbidity, income etc., and building a compartmental model for each stratum.

A unique feature of COVID-19 is its highly non-uniform attack on the different age strata of society. To account for this, Balabdaoui and Mohr proposed an age stratified discrete compartmental model (as an alternative to a differential equation based SIR model) to model the transmission dynamics of COVID-19 in Switzerland (Balabdaoui & Mohr, 2020). The population was stratified into 17 age groups (0-4, 5-9, ..., 75-79, ≥ 80) and the model included 9 compartments. The incubation period was assumed to be 5 days. The model included sub-compartments representing time spent in each compartment.

The study showed that the reproduction number for the most infected age group (15-20) was five times that of the least infected age group (≥ 80) because young people socialise more. The study also showed that people over 80 years account for 70% of deaths while people below 50 years account for 0.1% of deaths. The conclusion was that in order to avoid a second wave of

infections, the probability of transmission should be limited at schools. In Epidemiology, the reproduction number (R_0) represents the expected number of infections caused by a single infected individual. If $R_0 < 1$ then the disease dies out and if $R_0 > 1$ then a pandemic occurs as each infected individual infects more than 1 person causing the disease to spread exponentially. The conclusion was therefore that the reproduction number should be kept close to 1 to avoid silent multiplication of fatalities and the disease becoming epidemic.

The advantages of age stratified models are that they provide an age specific understanding of the dynamics of COVID-19. They also ensure timely decision making of when to implement or relax measures taken to limit the spread of COVID-19. The disadvantage of the model was that it was built only using data from Switzerland. This means that the results and conclusions are only applicable to countries with ageing populations. To improve accuracy and reliability of predictions, the model would have to be trained with data from other countries with both young and ageing population structures. The model could also be improved by stratifying with respect to co-morbidity, gender and standard of living as these factors also vary the effects of COVID-19.

3 The Dataset

The first dataset was obtained from (GitHub, 2021) which is a repository that collates data on the ongoing COVID-19 pandemic, specifically in Africa. The dataset comprises of 9 243 019 COVID-19 cases recorded between 5 March 2020 - 1 August 2020 i.e., during the first wave of COVID-19 cases in South Africa. The data was collected for males and females between ages 2-85 from all 9 provinces in South Africa. The dataset also included the underlying conditions and symptoms of the individuals in question. Other variables included in the dataset were date of death, date of hospital admission, travel history and date when symptoms began to show.

Since the COVID-19 pandemic is recent and ongoing, the data availability was limited to the first wave of cases. This is disadvantageous because models and predications are then dependent on data collected over a limited time frame. This is also disadvantageous because the dynamics of COVID-19 altered over time due to lockdown restrictions, behavioural changes (wearing masks, sanitizing, social distancing etc.,), variants and introduction of vaccines. This means that models built using this data do not fully capture the dynamics of COVID-19 and consequently the effects of lockdowns, behaviour changes, variants and vaccines are determined inferentially.

Figure 1 below shows the number of COVID-19 cases between 5 March 2020 - 1 August 2020. It is evident from the time series plot that the number of cases was volatile and exponentially increasing over time. The maximum number of recorded active cases was 693 936 on 20 July 2020. The average number of active cases per day was 51 066 cases with a standard deviation of 58 583. This high standard deviation was expected since the number of cases was volatile and exponentially increasing during the time period in question.

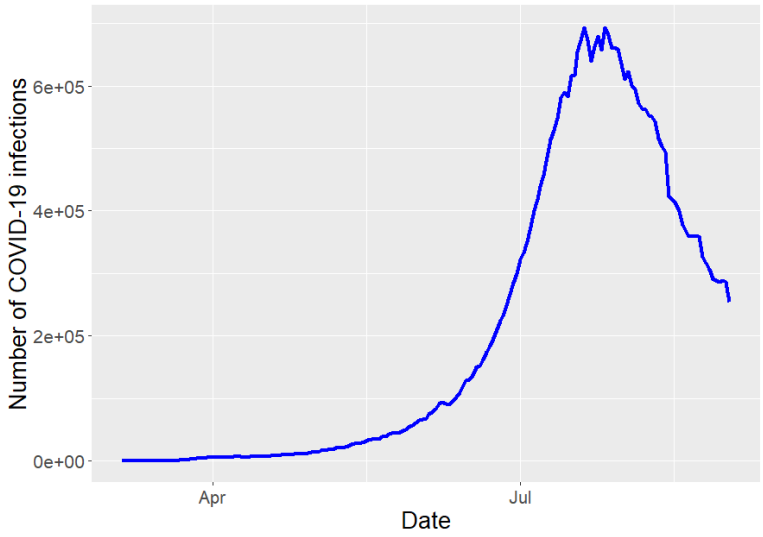


Figure 1: Observed COVID-19 cases during the first infection wave

The second dataset is the COVID-19 hospitalisation data obtained from the Gauteng DATCOV system. This data was used for parameter re-estimation. This dataset includes information about 125 244 COVID-19 related hospitalisations between 7 March 2020 - 7 October 2021. This specifically includes the age, sex, length of stay, facility type (public or private hospital), facility location, discharge status (discharged alive, transferred or died), ward upon admission (general ward or ICU), movements between wards and comorbidities of patients across Gauteng.

The average age of hospitalised patients was 53 with a standard deviation of 20 years. Hospitalisations were most prevalent amongst people aged 46-60. 47.1% of patients were male while 52.9% of patients were female. 42.3% of patients were hospitalised in public facilities and the remaining 57.7% in private facilities. Upon admission, 85.1% of people were admitted into General Ward, 14.5% into ICU and 0.4% into isolation wards indicating the the majority of cases were relatively mild. On average, people spent 11.6 days in General Ward and 11.6 days in ICU before being transferred, discharged or dying of COVID-19. The death rate amongst hospitalised patients was 21.9% and deaths were most prevalent amongst people aged 61-75. Figures 2 and 3 show the distribution of hospitalisations and deaths respectively.

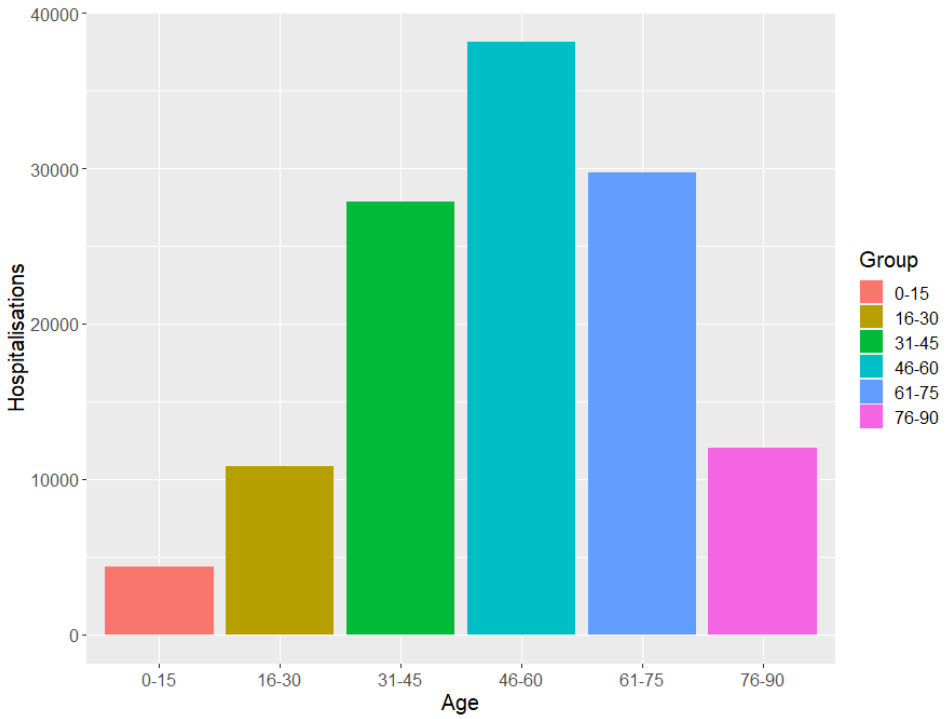


Figure 2: Hospitalisations by age

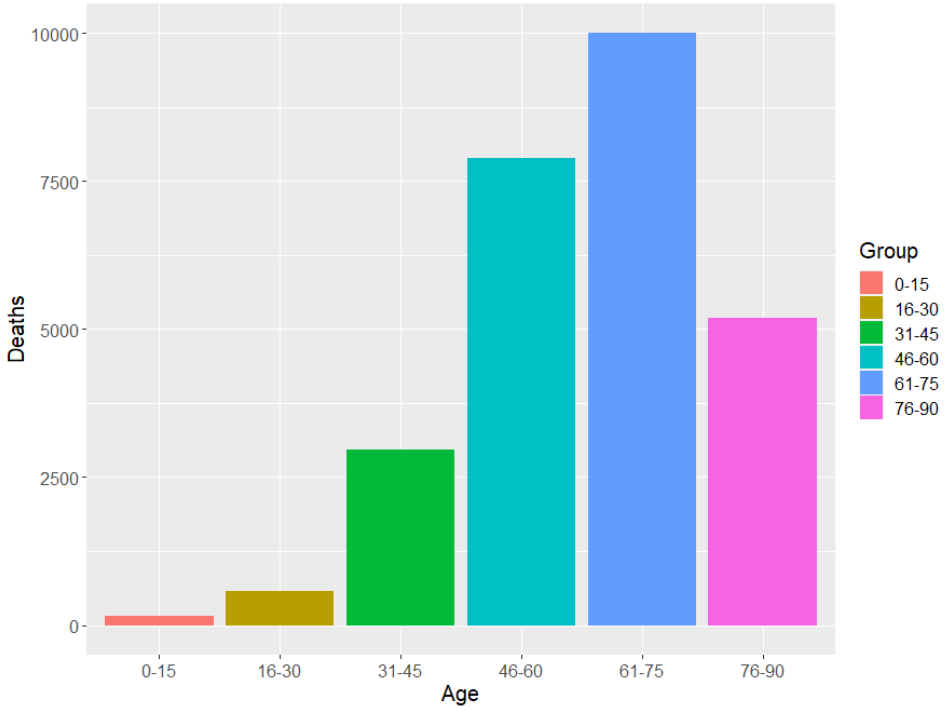


Figure 3: Deaths by age

From Figures 2 and 3, it is clear that hospitalisations and deaths were negatively skewed as the majority of hospitalisations and deaths were observed amongst older age groups. It is also evident that the modal age group for hospitalisations was 46-60 and 61-75 for deaths. The number of deaths was considerably lower than the number of hospitalisations indicating that the majority of hospitalised individuals were able to survive COVID-19.

Figure 4 is a pie chart showing the prevalence of various co-morbidities amongst patients who died of COVID-19. It is clear from this plot that the most common co-morbidities seen amongst patients who died of COVID-19 were hypertension, diabetes, obesity, cardiac disease and HIV. These are also the most common illnesses in South Africa (BusinessTech, 2021). With the exception of HIV, the most common co-morbidities are linked to diet and lifestyle indicating that South Africans have poor general health which is a common trait observed in developing countries.

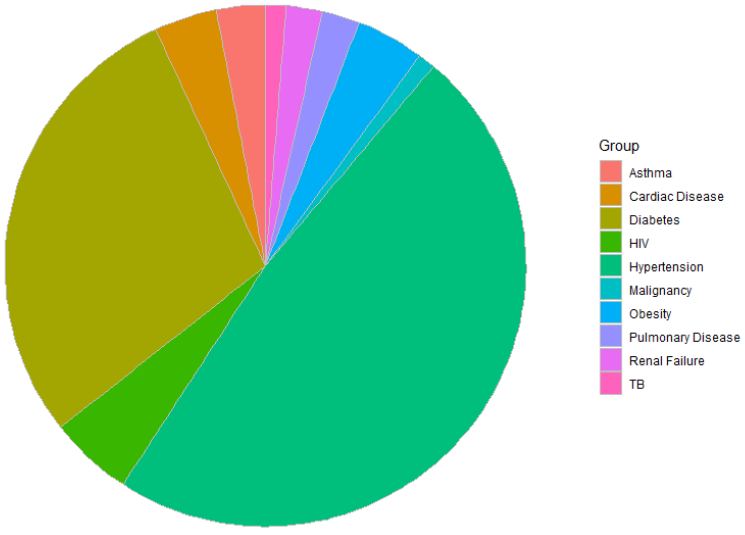


Figure 4: Co-morbidities of hospitalised COVID-19 patients

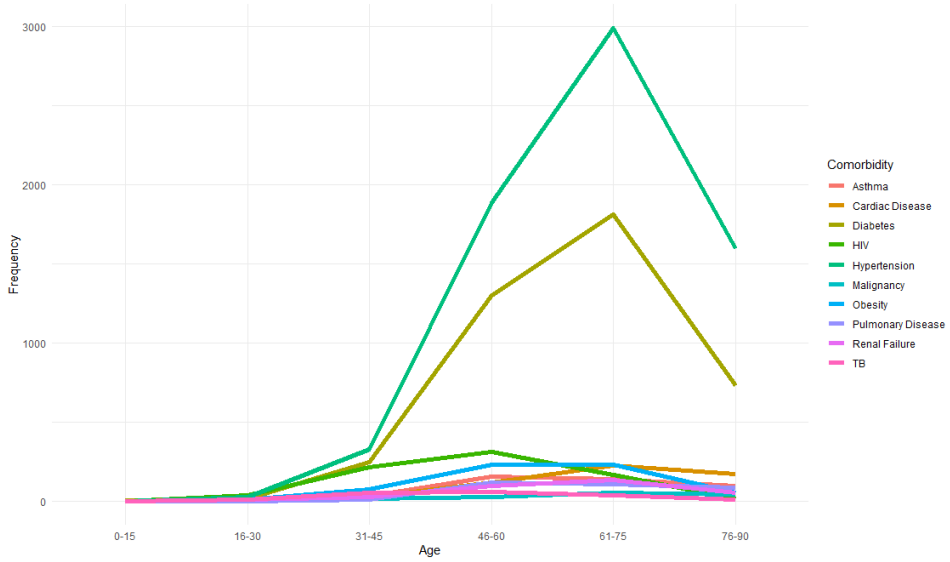


Figure 5: Co-morbidity specific death distributions

Figure 5 shows the age distribution of deaths according to co-morbidity. This plot again shows that the prevalence of deaths was highest amongst patients with hypertension, diabetes, cardiac disease, obesity and HIV. This plot also shows that the prevalence of co-morbidities was highest amongst older individuals as the majority of patients with underlying illnesses who succumbed to COVID-19 were aged 61-75. From Figure 5, it is clear that hypertension and diabetes are the most influential in determining the likelihood of COVID-19 related deaths. As previously mentioned, hypertension and diabetes are linked to poor diet and lifestyle and their high prevalence are indicative of poor general health amongst South Africans.

Figure 5 shows that COVID-19 infection numbers increase with age across all co-morbidities. Immunity weakens with age and the severity of co-morbidities also worsen with time. This graph therefore suggest that COVID-19 infections are not driven by age or co-morbidity independently. They are instead driven by the synergy between age and co-morbidity.

4 Methodology

4.1 Basic SIR Model

Compartmental models have been used in Epidemiology to model the spread of infectious diseases since 1927. The basic compartmental model is the SIR model. This model consists of three compartments namely Susceptible (S), Infected (I) and Recovered/Immune (R). The transition rates represent the probabilities of transition between compartments. β is the probability of transition between susceptibility and infection (transmission rate) and γ is the probability of transition between infection and recovery/immunity (recovery rate). The basic model is summarised in Figure 6 below:



Figure 6: Basic SIR Model

Compartmental models have characterisations and assumptions to ensure that the models accurately represent the data and to ensure accurate predictions. These characterisations and assumptions vary to account for the unique attributes of a specific infectious disease. The basic compartmental model assumptions are:

1. The population is fixed i.e., $S+I+R=N$ where N is fixed over time.
2. The population is closed.
3. There is homogenous mixing of individuals.
4. Demographic and spatial features are ignored.
5. No individual has initial immunity.
6. The incubation period is instantaneous.
7. The population is homogenous.
8. The duration of infectiousness of an individual is equal to the disease duration.

9. Only susceptible individuals can be infected and after infection, the individual recovers and becomes immune.

Let $S(t)$, $I(t)$ and $R(t)$ denote the number of Susceptible, Infected and Recovered individuals in a population at time t. Let N denote the population size. Since we assume that the population is closed, $S(t) + I(t) + R(t) = N$. The differential equations for the basic SIR model are then:

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N}, \tag{1}$$

$$\frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \gamma I(t), \tag{2}$$

$$\frac{dR(t)}{dt} = \gamma I(t), \tag{3}$$

with initial conditions $S(0) = s_0$, $I(0) = i_0$ and $R(0) = 0$ (Wang & Wang, 2016).

4.2 Basic Reproduction Number

The basic reproduction number (R_0) in Epidemiology is the expected number of infections generated by a single infected individual in a population where all individuals are susceptible to infection. It is a measure of the severity of an epidemic i.e., how contagious/infectious a given disease is. When $R_0 < 1$, the epidemic dies out as an infected individual infects less than one other individual. When $R_0 > 1$, the epidemic grows as each infected individual infects more than one person.

A disease is epidemic when the number of people infected increases over time i.e., when $\frac{dI(t)}{dt} > 0$. When the number of infected individuals decreases over time, $\frac{dI(t)}{dt} < 0$. Substituting $\frac{dI(t)}{dt} > 0$ into equation (2) gives:

$$\frac{\beta S(t)I(t)}{N} - \gamma I(t) > 0, \tag{4}$$

$$\implies \frac{\beta S(t)I(t)}{N\gamma} > I(t). \tag{5}$$

Since initially the entire population except the initially infected is susceptible, $S(t) \approxeq N$, thus

$$\frac{\beta}{\gamma} > 1.$$

The basic reproduction number is therefore

$$R_0 = \frac{\beta}{\gamma}. \tag{6}$$

This means that if $\beta > \gamma$ then $R_0 > 1$ and an epidemic occurs. Similarly, if $\beta < \gamma$ then $R_0 < 1$ and there is just a small outbreak and no epidemic.

4.3 Stratified Compartmental Model

Compartmental models are versatile and can be adjusted to capture the unique dynamics of an infectious disease. These adjustments include adding compartments and/or adjusting transition probabilities. Another adjustment, is changing the assumptions to account for the unique dynamics of a disease. The basic SIR model assumes that infectious diseases have a uniform attack on populations and that there is homogenous mixing of individuals in a population. This assumption does not hold true for all infectious diseases and hence the basic deterministic compartmental model may not be the most viable. A solution for this is to stratify the population into sub-populations (e.g. age, gender, co-morbidity, income etc.,) and identify the disease dynamics for each stratum.

Assume a population is stratified into m sub-populations. Let $S_j(t)$, $I_j(t)$ and $R_j(t)$ represent the number of susceptible, infected and recovered individuals respectively in the jth sub-group at time t, where $j = 1, 2, ..., m$ and $t \geq 0$. β_{ij} represents the probability of transmission between the ith and j^{th} subgroups at time $t \geq 0$, where $i = 1, 2, ..., m$ and $j = 1, 2, ..., m$.

The differential equations for the jth sub-group are then:

$$\frac{dS_j(t)}{dt} = -S_j(t)\left(\frac{\beta_{1j}I_1(t)}{N_1} + \frac{\beta_{2j}I_2(t)}{N_2} + ... + \frac{\beta_{mj}I_m(t)}{N_m}\right), \tag{7}$$

$$\frac{dI_j(t)}{dt} = S_j(t)\left(\frac{\beta_{1j}I_1(t)}{N_1} + \frac{\beta_{2j}I_2(t)}{N_2} + ... + \frac{\beta_{mj}I_m(t)}{N_m}\right) - \gamma_j I_j(t), \tag{8}$$

$$\frac{dR_j(t)}{dt} = \gamma_j I_j(t), \tag{9}$$

with initial conditions $S_j(0) = s_{j0}$, $I_j(0) = i_{j0}$ and $R_j(0) = 0$ for $j = 1, 2, ..., m$.

4.4 Age Stratified Model

Since the effects of COVID-19 are non-uniform across age bands, the population was stratified into 6 age groups. These age groups were in increments of 15 years i.e., 1-15, 16-30, ..., 76-90. These age groups were used to show how COVID-19 affects the youth who attend school and university, the working class and the retired individuals in a population.

The age stratified model had more compartments than the basic SIR model to capture the unique dynamics of COVID-19. The model compartments included: Susceptible (S), Exposed (E), Asymptomatic (A), Infected (I), Infectious and hospitalised (I_H), Infectious and in ICU (I_{ICU}), Recovered (R) and Dead (D). The model was built for each age group to determine the dynamics of COVID-19 for each age stratum.

The assumptions of the age-stratified model were then:

1. Individuals within a sub-population are homogenous and there is homogenous mixing of individuals,
2. Individuals in hospital and in ICU cannot spread the virus.
3. Only individuals in hospital or in ICU can die if their conditions become severe enough.
4. The death rate is homogenous within a sub-population but varies across the population.

The compartmental model to be built for each age stratum is summarised in Figure 7 below.

Assume the population is stratified into 6 sub-populations. Let $S_i(t)$, $E_i(t)$, $A_i(t)$, $I_i(t)$, $I_{H_i}(t)$, $I_{ICU_i}(t)$, $R_i(t)$ and $D_i(t)$ represent the number of Susceptible, Exposed, Asymptomatic, Infected, Infectious and hospitalised, Infectious in ICU, Recovered and Dead respectively individuals in the ith subgroup at time t, where $i = 1, 2, ...6$ and $t \geq 0$.

The differential equations, adapted from (South African COVID-19 Modelling Consortium, 2020), corresponding to the age-stratified model are then:

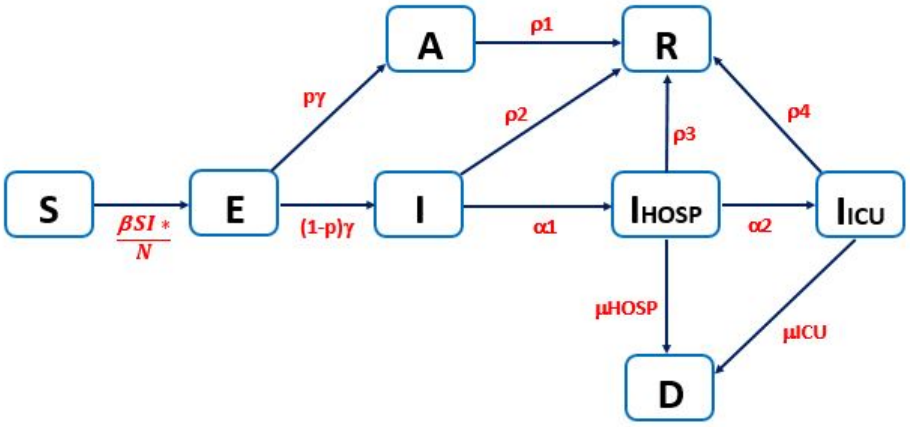


Figure 7: Age stratified COVID-19 Model

$$\frac{dS_i(t)}{dt} = -S_i(t)\sum_{i=1}^{6}\beta_{ij}\left[\sum_{j=1}^{6}\frac{(I_j(t)+0.75A_j(t))}{N_j}\right], \tag{10}$$

$$\frac{dE_i(t)}{dt} = -S_i(t)\sum_{i=1}^{6}\beta_{ij}\left[\sum_{j=1}^{6}\frac{(I_j(t)+0.75A_j(t))}{N_j}\right] - \gamma E_i(t), \tag{11}$$

$$\frac{dA_i(t)}{dt} = p\gamma E_i(t) - \rho_1 A_i(t), \tag{12}$$

$$\frac{dI_i(t)}{dt} = (1-p)\gamma E_i(t) - \alpha_1 I_i(t) - \rho_2 I_i(t), \tag{13}$$

$$\frac{dI_{H_i}(t)}{dt} = \alpha_1 I_i(t) - \alpha_2 I_{H_i}(t) - \rho_3 I_{H_i}(t) - \mu_{H_i} I_{H_i}(t), \tag{14}$$

$$\frac{dI_{ICU_i}(t)}{dt} = \alpha_2 I_{H_i}(t) - \rho_4 I_{ICU_i}(t) - \mu_{ICU_i} I_{ICU_i}(t), \tag{15}$$

$$\frac{dR_i(t)}{dt} = \rho_1 A_i(t) + \rho_2 I_i(t) + \rho_3 I_{H_i}(t) + \rho_4 I_{ICU_i}(t), \tag{16}$$

$$\frac{dD_i(t)}{dt} = \mu_{H_i} I_{H_i}(t) + \mu_{ICU_i} I_{ICU_i}(t), \tag{17}$$

where $i = 1, 2, ..., 6$ represents each age stratum.

The numerical parameters for the age-stratified model were obtained from

literature. The model parameters are summarized in Table 1 below:

Table 1: Age stratified model Parameters

Parameter	**Description**	**Value**	**Source**
γ	No. of days in incubation	14	(Silal et al., 2020)
p	Proportion of asymptomatic infections	0.75	(Davies et al., 2020)
ρ_1	No. of days with asymptomatic infection	7	(Gomes, 2020)
ρ_2	No. of days with symptomatic infection	5	(Silal et al., 2020)
ρ_3	No. of days in hospital	5	(Gaythorpe et al., 2020)
ρ_4	No. of days in ICU	9	(Gaythorpe et al., 2020)
α_1	No. of days before hospitalisation	7	(Silal et al., 2020)
α_2	No. of days before admission in ICU	4	(Silal et al., 2020)
μ_i	Death rate of age group i, i=1,2,...,6	0.002 - 0.148	(Veri, 2020)

The parameters can be explained and motivated as follows:

1. γ : This parameter represents the number of days before a person exposed to COVID-19 shows symptoms.

2. p : This represents the proportion of COVID-19 infected patients who do not show symptoms.

3. ρ_1 : A unique attribute of COVID-19 is that symptoms to do not show in all infected individuals. This parameter therefore represents the estimated infection duration for an asymptomatic individual. This parameter value was estimated from small samples since people rarely test for COVID-19 if they do not show any symptoms. Very little research has been conducted on asymptomatic individuals as the focus has primarily been on symptomatic patients due to the pandemic severity, hence 7 days proved to be a best estimate of asymptomatic infection duration due to limited data availability.

4. ρ_2 : This parameter represents the average number of days that infected individuals showed symptoms of COVID-19 infection. Common symptoms included shortness of breath, loss of smell and taste, fatigue and fever amongst others. This parameter was estimated from hospitalised individuals as it was easier to monitor them and determine the duration that they showed COVID-19 related symptoms.

5. $\rho 3$: This parameter represents the average length of stay in hospital for COVID-19 patients. Only patients with symptomatic infections were hospitalised hence this parameter estimate is similar to the estimated ρ_2 parameter.

6. ρ_4 : This parameter represents the average length of stay in ICU (Intensive Care Unit) for patients with severe COVID-19 infections. These were patients who required ventilators and to be monitored 24/7. Infection duration increased with severity hence the number of days in ICU was estimated to be almost double the number of days in hospital as infections were typically milder in hospital than in ICU.

7. α_1 : This parameter represents the number of days before hospitalisation is required i.e. number of days between infection and hospitalisation. This parameter was estimated by determining the average time taken for symptomatic infections to become severe enough for hospitalisation to be required.

8. α_2 : This parameter represents the number of days before a COVID-19 infected individual is admitted in ICU. Individuals were admitted into ICU either from home or after being transferred from the hospital. α_2 was therefore estimated by averaging the time taken for COVID-19 infections to progress to a point that they were severe enough for ICU admission to be required.

9. μ_i : This parameter is the age-specific death rate. It was calculated by determining the proportion of hospitalised individuals who died of COVID-19 by age group. This paper assumes that only individuals hospitalised and ICU individuals can transition to the absorbing death state hence death rates were estimated based on hospitalised individuals.

Table 2 shows the parameters that are expected to change after age stratification:
Contact network heterogeneity and population structure are highly influential in determining whether an infectious disease becomes epidemic or persists at epidemic levels (Prem et al., 2021). Since the effects of COVID-19 are non-uniform across different age groups of a population, understanding the population structure is necessary to understand the dynamics of COVID-19. Figure 8(a) shows the total population of South Africa in 2020 by age group. This includes the number of males and females in South Africa in age bands

Table 2: Expected age stratified model parameter changes

Parameter	**Description**	**Change**
p	Proportion of asymptomatic infections	Decrease with age
ρ_2	Number of days with symptomatic infection	Increase with age
ρ_3	Number of days in hospital	Increase with age
ρ_4	Number of days in ICU	Increase with age
μ_i	Death rate of age group i, i=1,2,...,6	Increase with age

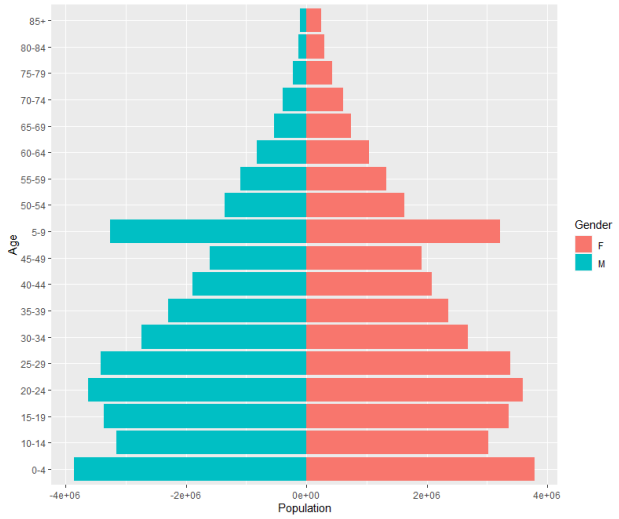
0-10, 10-20, 20-30, ..., 80-90. It is evident from Figure 8(a) that South Africa has a young population as the majority of the population is below the age of 35. It is also evident that the number of individuals decreases as the age bands increase. This is a common characteristic of developing countries and is attributable to developing countries having high birth and death rates as is the case in South Africa. The severity of COVID-19 increases with age as younger people have lower susceptibility to infection, low propensity to show clinical symptoms and a low likelihood of having a co-morbidity (Davies et al., 2020). Since South Africa has a young population, it is therefore expected that the COVID-19 death, infection and hospitalisation rates are considerably lower than the those of countries with ageing populations.

Contact network heterogeneity is another influential factor in infectious disease modelling. Contact matrices show the average number of interactions between age groups per day. They are necessary in modelling the spread of contact-transmissible infectious diseases like COVID-19 as they show the transmissibility of a disease across various age groups. The South African contact matrix below shows the average number of daily interactions between the six age bands of the age stratified model:

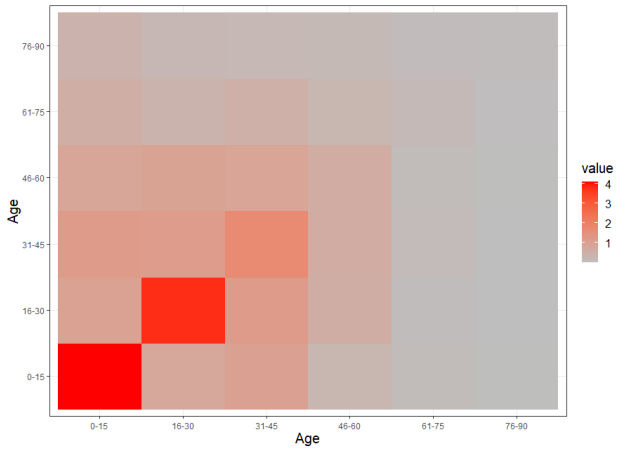
$$\mathbf{C} = \begin{pmatrix} 4.1044 & 0.8074 & 0.9901 & 0.3113 & 0.1001 & 0.0317 \\ 0.9837 & 3.8031 & 1.1860 & 0.5881 & 0.0871 & 0.0203 \\ 1.1921 & 1.1412 & 1.7481 & 0.6296 & 0.1236 & 0.0181 \\ 0.8357 & 0.9513 & 0.8797 & 0.6349 & 0.1122 & 0.0301 \\ 0.5623 & 0.4023 & 0.5079 & 0.3096 & 0.1843 & 0.0466 \\ 0.4203 & 0.2452 & 0.2274 & 0.2081 & 0.0920 & 0.0764 \end{pmatrix}.$$

It is evident from the contact matrix that younger individuals have the highest number of daily contacts which is expected since they attend school, university and social events. The opposite is true for older age groups who

have fewer daily contacts. This is also evident from the surface plot in Figure 8(b) which shows that the highest amount of contacts made were amongst individuals aged 0-45 who, as previously mentioned, attend school, work, university and social events (Prem et al., 2021).



(a)



(b)

Figure 8: The (a) Population pyramid and (b) Contact matrix for South Africa

4.5 Co-morbidity Stratified Model

The effects of COVID-19 vary depending on whether an individual has underlying illnesses or not. The population was therefore stratified into 2 groups: healthy individuals and individuals with co-morbidities. The base model was built based on healthy individuals and used the parameters specified in Table 1. The second model was built based on individuals with co-morbidities and the parameters in Table 1 were adjusted to account for co-morbidities.

Stratifying the population by co-morbidity allows for an understanding of the effect of COVID-19 on people depending on their pre-existing state of health. The model compartments included: Susceptible (S), Exposed (E), Asymptomatic (A), Infected (I), Infectious and hospitalised (I_H), Infectious and in ICU (I_{ICU}), Recovered (R) and Dead (D).

The assumptions for the co-morbidity stratified model are:

1. Individuals within a sub-population are homogenous i.e., the effects of COVID-19 are uniform across all co-morbidities.
2. Hospitalised individuals and individuals in ICU cannot spread the virus.
3. Only individuals in ICU or in hospital can die of COVID-19.
4. The transition probabilities are homogenous within a sub-population but vary across the population.

Figures 9 and 10 below show the co-morbidity stratified models for healthy individuals and individuals with co-morbidities respectively:
The differential equations for the co-morbidity stratified model are similar to those for the age-stratified model but are applicable for 2 strata rather than 6 strata. Assume the population is stratified into 2 sub-populations. Let $S_i(t)$, $E_i(t)$, $A_i(t)$, $I_i(t)$, $I_{H_i}(t)$, $I_{ICU_i}(t)$, $R_i(t)$ and $D_i(t)$ represent the number of Susceptible, Exposed, Asymptomatic, Infected, Infectious and hospitalised, Infectious in ICU, Recovered and Dead respectively individuals in the ith sub-group at time t, where $i = 1, 2$ and $t \geq 0$. The differential equations, adapted from (South African COVID-19 Modelling Consortium, 2020), are then:

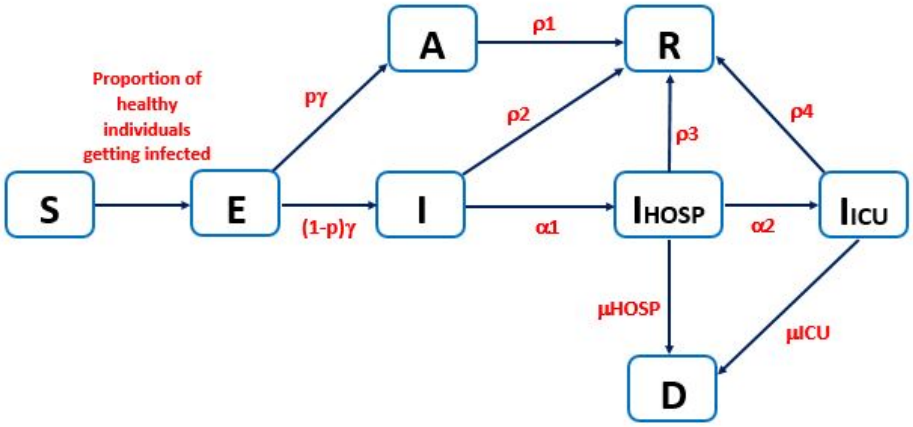


Figure 9: Co-morbidity stratified model for healthy stratum

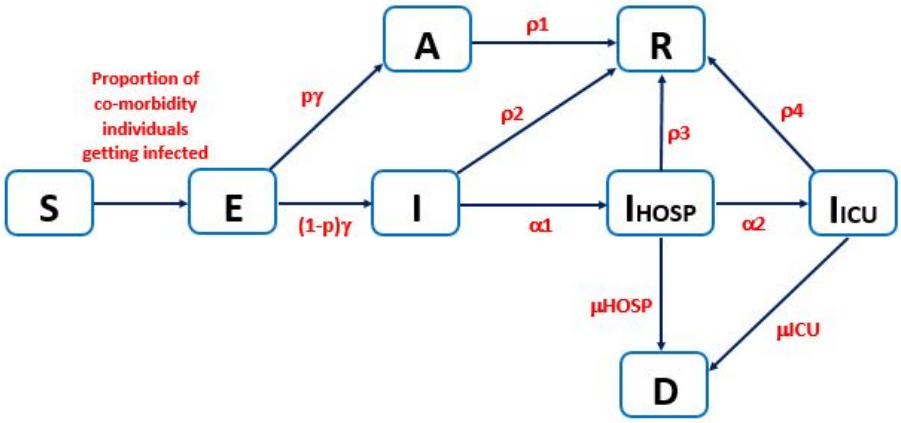


Figure 10: Co-morbidity stratified model for co-morbidity stratum

$$\frac{dS_i(t)}{dt} = -S_i(t)\sum_{i=1}^{2}\beta_{ij}\left[\sum_{j=1}^{2}\frac{(I_j(t)+0.75A_j(t))}{N_j}\right], \tag{18}$$

$$\frac{dE_i(t)}{dt} = -S_i(t)\sum_{i=1}^{2}\beta_{ij}\left[\sum_{j=1}^{2}\frac{(I_j(t)+0.75A_j(t))}{N_j}\right] - \gamma E_i(t), \tag{19}$$

$$\frac{dA_i(t)}{dt} = p\gamma E_i(t) - \rho_1 A_i(t), \tag{20}$$

$$\frac{dI_i(t)}{dt} = (1-p)\gamma E_i(t) - \alpha_1 I_i(t) - \rho_2 I_i(t), \tag{21}$$

$$\frac{dI_{H_i}(t)}{dt} = \alpha_1 I_i(t) - \alpha_2 I_{H_i}(t) - \rho_3 I_{H_i}(t) - \mu_{H_i} I_{H_i}(t), \tag{22}$$

$$\frac{dI_{ICU_i}(t)}{dt} = \alpha_2 I_{H_i}(t) - \rho_4 I_{ICU_i}(t) - \mu_{ICU_i} I_{ICU_i}(t), \tag{23}$$

$$\frac{dR_i(t)}{dt} = \rho_1 A_i(t) + \rho_2 I_i(t) + \rho_3 I_{H_i}(t) + \rho_4 I_{ICU_i}(t), \tag{24}$$

$$\frac{dD_i(t)}{dt} = \mu_{H_i} I_{H_i}(t) + \mu_{ICU_i} I_{ICU_i}(t), \tag{25}$$

where $i = 1, 2$ represents healthy individuals and individuals with co-morbidities respectively. The parameters are the same as those defined in Table 1.

Table 3 below shows the parameters that are expected to change after stratification by co-morbidity.

Table 3: Expected co-morbidity stratified model parameter changes

Parameter	**Description**	**Change**
p	Proportion of asymptomatic infections	Higher for healthy individuals
ρ_2	Number of days with symptomatic infection	Higher for co-morbidity individuals
ρ_3	Number of days in hospital	Higher for co-morbidity individuals
ρ_4	Number of days in ICU	Higher for co-morbidity individuals
μ_i	Death rate of population group i, i=1,2	Higher for co-morbidity individuals

An assumption of the co-morbidity stratified model is that there is homogenous mixing of individuals in a population. This means that an individual is equally likely to interact with a healthy individual as they are to interact with an individual with a co-morbidity. This means that contact patterns do not have a considerable impact on the dynamics of the co-morbidity stratified model.

4.6 Sensitivity Analysis

A sensitivity analysis is a means of determining the effect of parameter changes on model performance under a given set of assumptions (Davies et al., 2020). Sensitivity analyses are used in a variety of fields including Epidemiology, economics, engineering etc., to understand the relationships between model inputs and outputs and to test the robustness of mathematical/statistical models. Compartmental models are parametric models and in parametric modelling, outcomes are dependent on parameter changes. A sensitivity analysis is therefore necessary in compartmental modelling to ascertain the impact of parameters such as R_0, γ, β etc., on the model outcomes and performance.

Sensitivity analyses can be broadly classified as local or global. A local sensitivity analysis involves estimating the model's sensitivity to a single parameter by treating the output as a function of a single parameter and fixing all other parameters. Global sensitivity analysis involves varying an entire parameter space and identifying the effect on the output (Hamby, 1994). Within these broad classifications are several sensitivity analysis methods including One-by-one, Derivative Based, Regression Analysis and Multivariate amongst several others (Christopher Frey & Patil, 2002). This paper focuses primarily on One-by-one and Multivariate sensitivity analysis from a a parametric perspective. Due to the novelty and unique dynamics of COVID-19, this paper also assesses the models' sensitivity to initialisation date and lockdown level to determine the effect of these parameters on the models and to identify the optimal pandemic start date.

4.6.1 One-by-one sensitivity analysis

One-by one sensitivity analysis is the most common sensitivity analysis method due to its simplicity. It involves fixing all parameters except one and running the model to identify the effect of that variable on the model output. These effects include changes in death rate, number of infections, transmission rates etc. This process is repeated for all parameters and the parameter(s) that result in considerable model output changes are re-estimated until model sensitivity is minimised (Hamby, 1994). Model performance can be assessed graphically by assessing plots of a given variable e.g., R_0 against a given output e.g., death rate. Model performance can also be assessed using performance measures such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and R^2 (R-squared). Adequate plan-

ning for the effects of COVID-19 involves implementing strategies to minimise COVID-19 related infections, hospitalisations and deaths. Parameter sensitivity was therefore determined with respect to these model outputs (infections, hospitalisations and deaths). Individual parameters were varied according to a uniform distribution and the remaining parameters were kept constant. The parameters with the greatest effect on infections, hospitalisations and deaths were then identified and expressed graphically using violin plots. These highly sensitive parameters were then re-estimated to improve model accuracy.

4.6.2 Multivariate sensitivity analysis

Multivariate sensitivity analysis involves varying several variables simultaneously to determine their effect on the model output (Sobie, 2009). A multivariate sensitivity analysis involves selecting a model output e.g., death rate, infection rate, hospitalisation rate etc., and determining which variable the model output(s) are most sensitive to. This is done by identifying the model outputs with the highest standard coefficients. Multivariate sensitivity analysis quantifies the importance of model inputs and their interactions with respect to model output. It provides an overall view on the influence of inputs on outputs as opposed to a local view of partial derivatives as in one-by-one sensitivity analysis (Lamboni, Monod, & Makowski, 2011). As with the One-by-one sensitivity analysis, the multivariate analysis was done in a hospitalisation, infection and death context. This determined the parameters which needed to be re-estimated as well as the parameters most influenced by age and/or co-morbidity.

4.6.3 Sensitivity to initialisation date

Due to the novelty of COVID-19, there is a great deal of uncertainty regarding the start date of the epidemic. Determining the models' sensitivity to initialisation date is necessary to identify the optimal initialisation date that ensures that the height and time of the observed infection peak is matched by the models. This analysis involved initialising and re-initialising the models to reflect the changes in symptomatic and asymptomatic infections between 5 - 20 March 2020. This was done until the model peaks matched the height and timing of the observed infection peak. Early initialisation dates have fewer infections causing the models to underestimate the effects of a disease. As a result, models peak later and lower than observed infection peaks. The opposite is true for late initialisation dates which have higher numbers of

infections causing models to overestimate the effects of a given disease. As a result, models instead peak earlier and higher than observed infection peaks. It is therefore essential for the analysis to determine the optimal initialisation date to ensure the models are able to accurately predict COVID-19 infections.

4.6.4 Sensitivity to lockdown level

Lockdowns have been used globally as a means of limiting the spread of COVID-19 and reducing the pressure on health systems. This is due to the fact that lockdowns reduce transmissibility considerably as people have considerably fewer daily contacts. In South Africa, lockdown restrictions were implemented in levels with Level 5 having the harshest restrictions and Level 1 having the lightest restrictions. To account for the effects of lockdowns, the transmission matrices were adjusted according to lockdown level. Transmission probabilities were assumed to be low during harsh lockdowns and high during lighter lockdowns.

COVID-19 infections reduced considerably during higher levels of lockdown. To assess the effects of implementing harsh lockdowns for extended periods of time, transmission probabilities were reduced by approximately 50% between March - August 2020 for all models. The changes in the number of infections, hospitalisations and deaths were then identified. Since there is a trade off between saving lives and saving livelihoods, the feasibility of an extended lockdown was assessed by weighing the reduced pressure on the health system against the possibility of economic decline.

4.6.5 Sensitivity to transmission matrix β

The estimation of β is highly important in Epidemiology as β provides insight on the rate of spread of an infectious disease. Accurately estimating the transmission matrix, β, is especially important when dealing with stratified models as this matrix provides meaningful insight on how an infectious disease spreads across various strata.

Transmission matrices are generally estimated using contact matrices which show the number of successful contacts between sub-populations. Co-morbidity contact matrices are very rarely used in Epidemiology as it is generally assumed that there is an equal probability of contact between healthy and co-morbidity individuals. This assumption does not necessarily hold true

for COVID-19 due to the heightened risk aversion amongst individuals with co-morbidities. Individuals with co-morbidities are therefore more likely to stay at home to avoid severe infection. β is the rate of effective transmissions per unit time. It is calculated as:

$$\beta = \frac{R_0}{\gamma N_i}, \tag{26}$$

for $i = 1, 2$.

R_0 is generally estimated using seroprevalence data. Due to the novelty of COVID-19, this data is not yet available. It was therefore assumed that the rate of effective transmissions from healthy individuals was twice the rate of effective transmissions from individuals with co-morbidities. The transmission matrix was then:

$$\beta = \begin{pmatrix} 0.6 & 0.6 \\ 0.3 & 0.3 \end{pmatrix}.$$

As previously mentioned, the severity of COVID-19 infection increases with age. Also, younger individuals are generally more social than older individuals as they attend school, university and social events. To account for this population heterogeneity, the following method was proposed to determine the transmission matrix β. This approach uses contact matrices scaled by a ratio of reproduction number, R_0, to the average infection duration, D.

Method 1

This estimation method for the age-stratified model was based on active cases by age group data and a South Africa-level contact matrix. The combination of the data and the contact matrix was used to determine the age-stratified effective transmission matrix, β. Let C be the 6x6 contact matrix for South Africa given by:

$$C = \begin{pmatrix} 4.1044 & 0.8000 & 0.9901 & 0.3113 & 0.1001 & 0.0317 \\ 0.9837 & 3.8031 & 1.1860 & 0.5881 & 0.0871 & 0.0203 \\ 1.1921 & 1.1412 & 1.7481 & 0.6296 & 0.1236 & 0.0181 \\ 0.8357 & 0.9513 & 0.8797 & 0.6349 & 0.1122 & 0.0301 \\ 0.5623 & 0.4023 & 0.5079 & 0.3096 & 0.1843 & 0.0466 \\ 0.4203 & 0.2452 & 0.2274 & 0.2081 & 0.0920 & 0.0764 \end{pmatrix}.$$

As with the co-morbidity stratified model, β is the rate of effective transmissions per unit time. It is calculated as:

$$\beta = \frac{R_0 \alpha_{ij}}{D}, \tag{27}$$

for $i = 1, 2, ..., 6$ where D is the mean infectious period calculated as:

$$D = \left(\frac{p(0.7)}{\rho_1} + \frac{(1-p)}{\alpha_1} + \frac{(1-p)(0.25)}{\rho_2} + \frac{(1-p)(1-0.25)}{\alpha_2} \right), \tag{28}$$

and α_{ij} is a scaling factor (standardised contact matrix) with $i = 1, ..., 6$ $j = 1, ..., 6$.

This gives $D = 4.46$ and

$$\alpha_{ij} = \begin{pmatrix} 0.647 & 0.127 & 0.156 & 0.049 & 0.016 & 0.005 \\ 0.148 & 0.570 & 0.178 & 0.088 & 0.013 & 0.003 \\ 0.246 & 0.235 & 0.360 & 0.130 & 0.025 & 0.003 \\ 0.243 & 0.276 & 0.255 & 0.184 & 0.033 & 0.008 \\ 0.279 & 0.200 & 0.252 & 0.154 & 0.092 & 0.023 \\ 0.331 & 0.193 & 0.179 & 0.164 & 0.072 & 0.060 \end{pmatrix}.$$

According to the South African COVID-19 Modelling Consortium, the average R_0 during the 1st wave of COVID-19 infections was $\in [2.7; 2.8]$ (approx. 2.75). The transmission matrix β was then:

$$\beta = \begin{pmatrix} 0.399 & 0.078 & 0.096 & 0.030 & 0.009 & 0.003 \\ 0.091 & 0.351 & 0.110 & 0.054 & 0.008 & 0.002 \\ 0.152 & 0.145 & 0.222 & 0.080 & 0.015 & 0.002 \\ 0.150 & 0.170 & 0.157 & 0.113 & 0.020 & 0.004 \\ 0.172 & 0.123 & 0.155 & 0.095 & 0.057 & 0.014 \\ 0.204 & 0.119 & 0.110 & 0.101 & 0.044 & 0.037 \end{pmatrix}.$$

Aggregating across the diagonal to make the matrix symmetrical:

$$\beta = \begin{pmatrix} 0.399 & 0.085 & 0.124 & 0.090 & 0.091 & 0.104 \\ 0.085 & 0.351 & 0.128 & 0.112 & 0.066 & 0.061 \\ 0.124 & 0.128 & 0.222 & 0.119 & 0.085 & 0.056 \\ 0.090 & 0.112 & 0.119 & 0.113 & 0.058 & 0.053 \\ 0.091 & 0.066 & 0.085 & 0.058 & 0.057 & 0.029 \\ 0.104 & 0.061 & 0.056 & 0.053 & 0.029 & 0.037 \end{pmatrix}.$$

Method 1 assumed that COVID-19 susceptibility and transmissibility was higher amongst younger people since young people are more social and hence have more daily contacts. The second method proposed to estimate β assumes that COVID-19 susceptibility and transmissibility is higher amongst older individuals since the severity of the effects of COVID-19 increase with age.

Method 2

This method of estimation is based on a transmission model. This model assumes that relative susceptibility is not fixed but instead varies across age strata. This model also assumes that infections (A, I, I_{Hosp} and I_{ICU}) follow a Gamma distribution . The force of infection for an individual in age strata i at time t is:

$$\lambda_{i,t} = v_i \sum_j C(I_j + I_{Hosp_j} + I_{ICU_j} + \zeta A)/N_j, \tag{29}$$

where:
v_i is the susceptibility to infection for an individual in age strata i, $i = 1, 2, ..., 6$ and
ζ is the relative infectiousness of an asymptomatic individual (assumed to be 70%).

Due to the limited age-specific susceptibility data in South Africa, relative

susceptibility values observed in Israel were used as estimates for v_i. Israel was used as a proxy since it has a similar age structure to South Africa and hence was able to provide reasonable estimates of age-specific relative susceptibility to COVID-19 infection. Dattner et al., determined age-specific relative susceptibility to COVID-19 by stratifying the Israeli population into children (1-19) and adults (20+). They determined that children have a lower susceptibility to COVID-19 than adults with a relative susceptibility between 35%-43% compared to 57%-65% for adults. For the purposes of this study, the age-specific susceptibility v_i was assumed to be 35% for age group 1-15, 39% for age group 16-30, 43% for age group 31-45, 57% for age group 46-60, 61% for age group 61-75 and 65% for age group 76-90. That is $v_1 = 0.35, v_2 = 0.39, v_3 = 0.43, v_4 = 0.57, v_5 = 0.61$ and $v_6 = 0.65$.

As previously mentioned, infections were assumed to follow a Gamma distribution. According to Davies et al., the infections were distributed as follows:
$I \sim \text{Gamma}(\mu = 2.1, k = 4)$
$I_{Hosp}, I_{ICU} \sim \text{Gamma}(\mu = 2.9, k = 4)$
$A \sim \text{Gamma}(\mu = 5, k = 4)$

The Next Generation Matrix (NGM) is used as a means of determining the reproduction number R_0. It is defined as follows:

$$NGMij = v_i C(pE(I + I_{Hosp} + I_{ICU}) + (1 - p)\zeta E(A)), \tag{30}$$

where p is the proportion of symptomatic infections i.e., $p = 0.25$.

The dominant eigenvalues of the NGM gave estimates of the age-specific $\mathbf{R_0}$ values. $\mathbf{R_0}$ was then:

$$\mathbf{R_0} = \begin{pmatrix} 1.064 \\ 1.185 \\ 1.307 \\ 1.732 \\ 1.854 \\ 1.975 \end{pmatrix}.$$

As with Method 1, the rate of transmission was given by:

$$\beta = \frac{\mathbf{R_0}}{D}, \tag{31}$$

for $i = 1, ..., 6$ where D is the mean infection duration calculated as:

$$D = \left(\frac{p(0.7)}{\rho_1} + \frac{(1-p)}{\alpha_1} + \frac{(1-p)(0.25)}{\rho_2} + \frac{(1-p)(1-0.25)}{\alpha_2} \right). \tag{32}$$

To obtain a symmetrical square matrix, $\mathbf{R_0}$ was multiplied by its transpose $\mathbf{R_0^T}$. This gave:

$$\mathbf{R_0^*} = \mathbf{R_0}\mathbf{R_0^T} = \begin{pmatrix} 1.132 & 1.261 & 1.391 & 1.843 & 1.973 & 2.101 \\ 1.261 & 1.404 & 1.549 & 2.052 & 2.197 & 2.340 \\ 1.391 & 1.549 & 1.708 & 2.264 & 2.423 & 2.581 \\ 1.843 & 2.052 & 2.264 & 3.000 & 3.211 & 3.421 \\ 1.973 & 2.197 & 2.423 & 3.211 & 3.437 & 3.662 \\ 2.101 & 2.340 & 2.581 & 3.421 & 3.662 & 3.901 \end{pmatrix}.$$

From Method 1, $D = 4.46$. The transmission matrix β was then:

$$\begin{aligned} \beta &= \frac{\mathbf{R_0^*}}{D}, \\ &= \frac{\mathbf{R_0^*}}{4.46}, \\ &= \begin{pmatrix} 0.254 & 0.283 & 0.312 & 0.413 & 0.442 & 0.471 \\ 0.283 & 0.315 & 0.347 & 0.460 & 0.493 & 0.525 \\ 0.312 & 0.347 & 0.383 & 0.508 & 0.543 & 0.579 \\ 0.413 & 0.460 & 0.508 & 0.673 & 0.720 & 0.767 \\ 0.442 & 0.493 & 0.543 & 0.720 & 0.771 & 0.821 \\ 0.471 & 0.525 & 0.579 & 0.767 & 0.821 & 0.875 \end{pmatrix}. \end{aligned} \tag{33}$$

4.7 Re-estimation of Parameters

A sensitivity analysis determines which parameters have the greatest effect on model performance i.e., which parameters the model is most sensitive to. In order to minimize prediction error and maximize model performance, highly sensitive parameters need to be re-estimated (Brakenhoff, Roes, & Nikolakopoulos, 2019). Highly sensitive parameters were re-estimated using COVID-19 hospitalisation data obtained from the Gauteng DATCOV system. Although the objective of this study is to model the spread of COVID-19 across the whole of South Africa, the Gauteng hospitalisation data still provided reasonable estimates for the parameters in question.

Since the dataset only included COVID-19 cases where hospitalisation was required, parameters relating to asymptomatic infections were not able to be re-estimated from the data. The dataset was therefore only able to re-estimate ρ_2 (number of days with symptomatic infection), ρ_3 (Number of days in hospital), ρ_4 (number of days in ICU), α_2 (number of days before admission in ICU) and μ_i (age-specific death rate) since the models assume that only hospitalised individuals can die of COVID-19. The remaining parameters (p (proportion of asymptomatic infections) and ρ_1 (Number of days with asymptomatic infection)) were re-estimated from literature as the dataset only included data for symptomatic infections. The parameters relating to asymptomatic infection were re-estimated based on a study conducted in Italy that determined the association of age with likelihood of developing asymptomatic COVID-19 infection (Poletti et al., 2021). The incubation period (γ) was not re-estimated since extensive research has shown that the COVID-19 incubation period is between 10-14 days.

4.8 Comparison of Models and Prediction

The compartmental models built were compared visually and by means of a χ^2-Goodness of Fit test.

χ^2-Goodness of Fit

The χ^2-Goodness of Fit test is a statistical hypothesis test used to test how well a fitted model fits the data. It is conducted by testing:

H_0: There is no lack of fit
H_1: There is lack of fit

The value of the test-statistic is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \tag{34}$$

where O_i is the ith observed value, E_i is the ith estimated value and n is the sample size.

Since the χ^2 test is highly sensitive to large sample sizes, the test statistic was calculated using the observed and estimated COVID-19 cases as a proportion of the sample size. The test statistic was therefore be calculated

as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(\frac{O_i - E_i}{n})^2}{\frac{E_i}{n}}. \tag{35}$$

The χ^2 GoF test was conducted at a 5% level of significance with $9 - 1 = 8$ degrees of freedom since the models have 9 parameters. The critical value was then $\chi^2_{0.05,8} = 15.507$. If the test statistic exceeds the critical value then the null hypothesis is rejected in favour of the alternative hypothesis and a conclusion is made that the model does not fit the data. Conversely, if the test statistic is less than the critical value then we fail to reject the null hypothesis and conclude that the model fits the data. The χ^2 GoF test was conducted prior to and after parameter re-estimation to determine if there was an improvement in model fit after parameter re-estimation.

5 Analysis

5.1 Sensitivity Analysis

COVID-19 is a novel disease with unique dynamics. As a result, accurate estimation of parameters and initialisation values is essential to ensuring that compartmental models are representative and fully capture the unique dynamics of COVID-19. The sensitivity analysis was used as a means of investigating the model's sensitivity to initialisation date and lockdown level. The sensitivity analysis was also used to identify the parameters that needed to be re-estimated to improve model accuracy. The sensitivity to lockdown level, initialisation dates and paramters was investigated as these are the major determinants of the frequency and severity of COVID-19 infections, hospitalisations and deaths.

5.1.1 Sensitivity to initialisation date

The basic COVID-19 model was built as a means of determining the dynamics of COVID-19 in South Africa. The model was built prior to stratification in order to identify the effects of stratification by age and co-morbidity on the model dynamics. In order for the model to be built, an appropriate start date and β/R_0 had to be determined. Several start dates between 5 March 2020 - 20 March 2020 were selected and models were built for each date until the model peak matched the peak of the observed data. This process was repeated for various β values. β values between 0.4 - 0.7 were selected and models were built for each beta value until the model peak again matched the peak of the observed data.

The following assumptions were made while determining the optimal β value and start date for the base model:

1. Active cases not new cases were used
2. 75% of active cases were asymptomatic (A)
3. 25% of active cases were symptomatic (mild or severe) (I)
4. An inflation factor of 4 was applied to the observed active cases
5. $\beta = 0.55$ was assumed while determining the optimal start date
6. The observed data peak occurred on day 138

Table 4 shows the various beta values and their effects on the model peak:

Table 4: Model Parameters

β	γ	R_0	**Comment**
0.45	0.45	0.89	Late peak (day 170)
0.5	0.45	1	Late peak (day 150)
0.55	**0.45**	**1.11**	**Peak matches (day 138)**
0.6	0.45	1.22	Early peak (day 130)
0.65	0.45	1.33	Early peak (day 125)
0.7	0.45	1.44	Early peak (day 120)

It is evident from Table 4 above that the optimal β value was 0.55. Table 5 shows the various start dates and their effects on the model peak assuming that $\beta = 0.55$.

Table 5: Determining the optimal start date

Date	**Active**	**Cumulative**	**I**	**A**	**Comment**
05/03/20	4	4	1	3	Late peak (day 200)
06/03/20	4	8	2	6	Late peak (day 190)
07/03/20	8	16	4	12	Late peak
08/03/20	12	28	7	21	Late peak
09/03/20	28	56	14	42	Late peak (day 170)
10/03/20	28	84	21	63	Late peak
11/03/20	52	136	34	102	Late peak (day 155)
12/03/20	64	200	50	150	Late peak
13/03/20	96	296	74	222	Late peak (day 150)
14/03/20	152	448	112	336	Late peak
15/03/20	**244**	**692**	**173**	**519**	**Peak matches (day 138)**
16/03/20	256	948	237	711	Early peak (day 130)
17/03/20	340	1288	322	966	Early peak

From the Tables 4 and 5 above, it is clear that the optimal start date is 15 March 2020 and the optimal β value is 0.55.

This process was repeated for the age and co-morbidity stratified models. The optimal start date for the co-morbidity stratified model ended up being 10 March 2020 and 15 March 2020 for the age stratified model.

5.1.2 Sensitivity to lockdown restrictions

Nationwide lockdowns have been used globally as a means of limiting the spread of COVID-19 and consequently reducing the pressure on health systems. South Africa had one of the earliest and harshest lockdowns seen across the world and hence saw a lower and later peak in infections. Lockdown restrictions were implemented in levels with Level 5 having the harshest restrictions and Level 1 having the lightest restrictions. During level 5 lockdown, inter-provincial travel was prohibited, alcohol sales were prohibited as a means of lowering trauma cases in hospitals, borders were closed, all public gatherings except funerals were prohibited, funeral attendance was limited to 50 people, a curfew was created between 21:00 - 04:00, cigarette sales were banned and eating/drinking in bars and restaurants was prohibited. Other measures taken to limit the spread of COVID-19 included wearing masks in public spaces, sanitizing prior to entering indoor venues and social distancing. As cases began to subside, lockdown levels lowered and restrictions were lifted. This involved allowing inter-provincial travel, reducing curfew hours, allowing the sale of alcohol and increasing the maximum attendance at public gatherings.

The dates of the various lockdown levels during the first wave of infections were as follows:

- Level 5 : 26 March 2020 - 30 April 2020
- Level 4 : 1 May 2020 - 31 May 2020
- Level 3 : 1 June 2020 - 17 August 2020

To account for the various lockdown levels, the transmission probabilities were assumed to be low during harsh lockdowns (i.e., Level 5) and high during periods of lighter lockdown restrictions (i.e., Level 1). In order to determine the effect of extending lockdown Level 5, the transmission probability values (β) were reduced by approximately 50% between 26 March 2020 - 30 August 2020 (i.e., day 20 - 200) for the basic, age-stratified and co-morbidity stratified models.

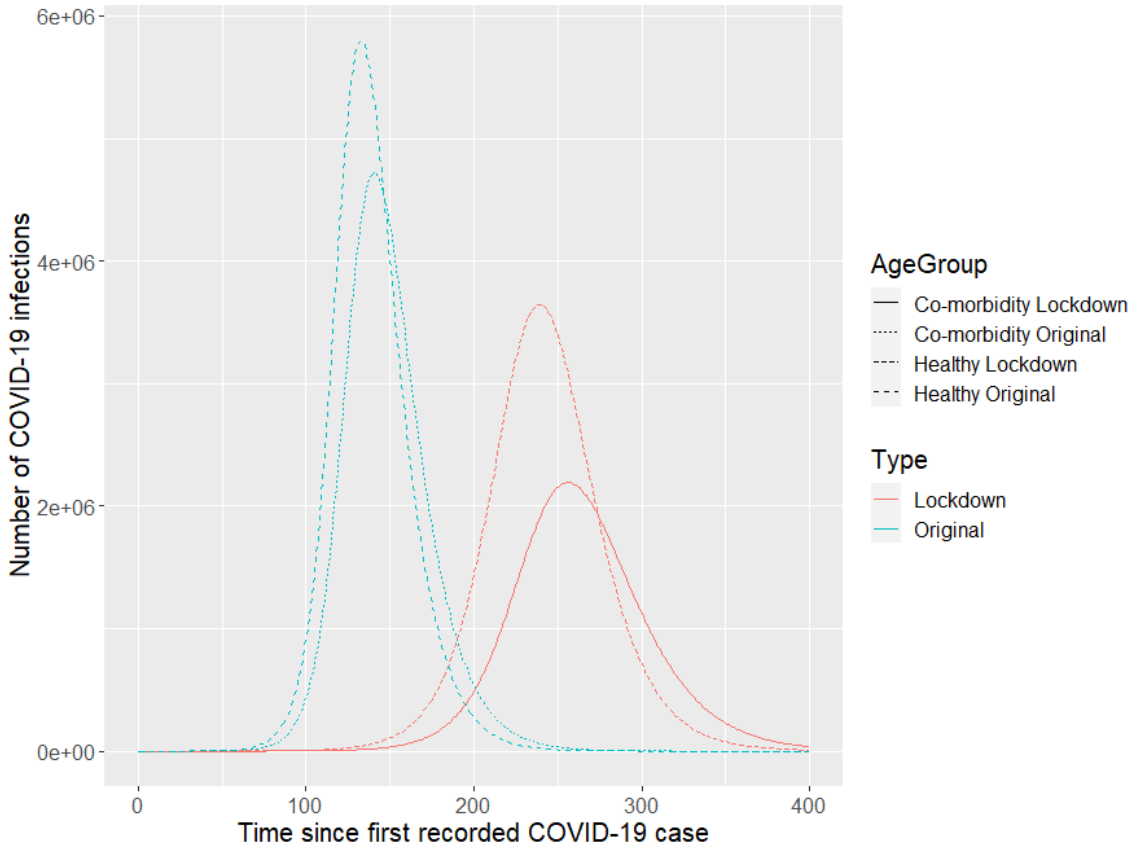


Figure 11: Co-morbidity models with and without an extended Level 5 lockdown

Figure 11 shows the total infections (asymptomatic, symptomatic, hospitalised and in ICU) for the healthy and co-morbidity populations with and without an extended Level 5 lockdown. Figure 12 then shows the total infections for each of the 6 age groups with and without an extended Level 5 lockdown. The purpose of these plots is to identify the impact of extending the duration of harsh lockdowns on the spread of COVID-19 in South Africa.

It is evident from Figure 11 that the total number of COVID-19 infections were higher amongst healthy individuals than individuals with co-morbidities. This was due to the greater risk aversion seen amongst individuals with co-morbidities. It is also evident from Figure 11 that extending the

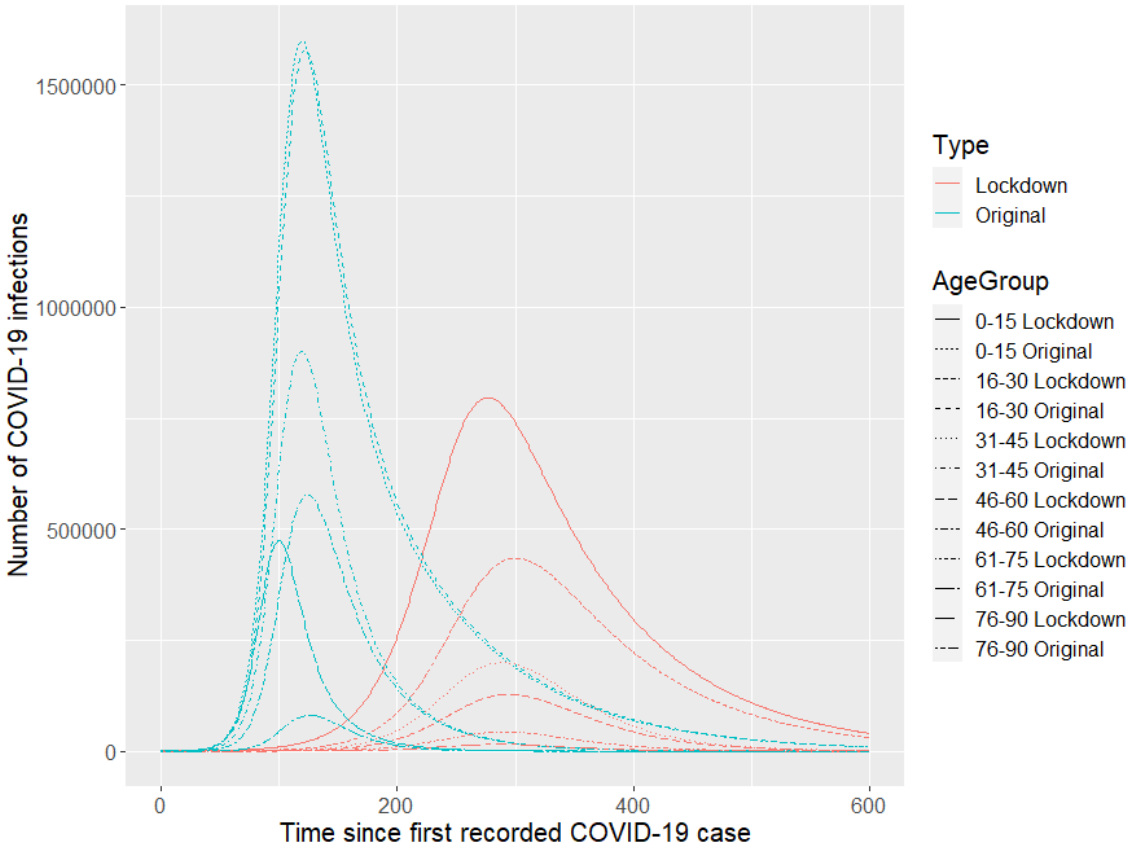


Figure 12: Age models with and without an extended Level 5 lockdown

duration of Level 5 lockdown delayed the infection peaks by approximately 130 days (4 months) for both the healthy and co-morbidity populations. The extended lockdown Level 5 also reduced the height of the infection peaks by over 100 000 infections for both populations. Since hospitalisations and ICU cases are included in the total infections, Figure 11 also shows that extending the Level 5 lockdown would have greatly reduced hospitalisations and hence lightened the pressure on the South African health system.

From Figure 12 it is evident that the total number of COVID-19 infections was higher among younger age groups than older age groups. This is due to the fact that South Africa has a young population i.e., the majority of South Africans are younger than age 35. It is also evident from Figure 12 that extending the duration of Level 5 lockdown delayed the infection peak by approximately 180 days (6 months) for all age groups. The extended lockdown also reduced the height of the infection peaks by over 500 000 infections for all age groups. Again, since hospitalisations and ICU cases are included in the total infections, Figure 12 shows that an extended lockdown would have greatly reduced hospitalisations and hence lightened the pressure on the health care system. The reduced and delayed hospitalisations would also have allowed hospitals to adequately prepare for high volumes of COVID-19 patients.

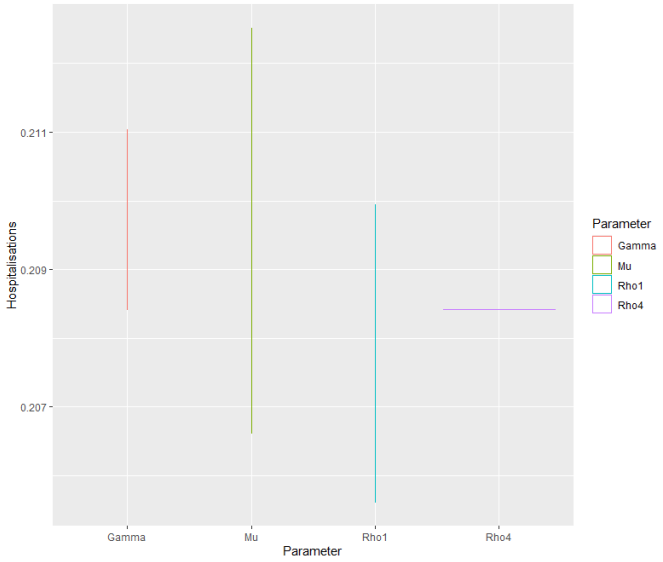
There is however a trade-off between saving lives and saving livelihoods. Although extending the Level 5 lockdown would have lowered the number of infections, this would not have been sustainable as businesses would have had to close permanently due to the prolonged lack of income which would have increased unemployment and consequently damaged the South African economy.

5.1.3 One-by one sensitivity analysis

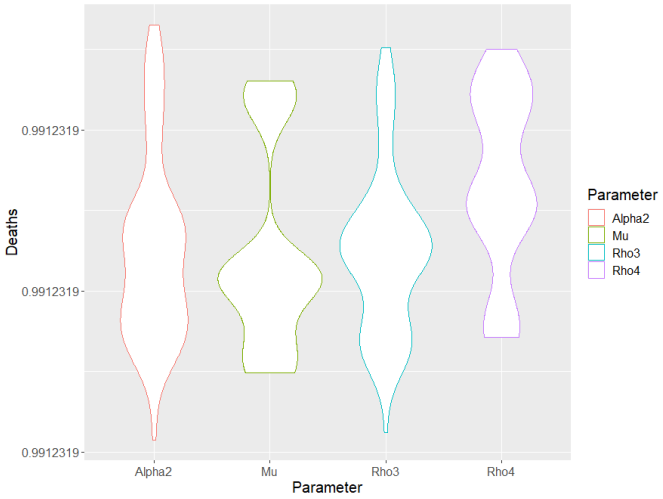
A one-by-one sensitivity analysis involves fixing all parameters except one and running the model to identify the effect of that variable on the model output. For the purposes of this study, parameter sensitivity was determined with respect to hospitalisation, infection and death. Individual parameters were varied according to a uniform distribution while the remaining parameters were kept constant. A one-by-one sensitivity analysis was only conduced for the basic COVID-19 model with a scalar β. A multivariate sensitivity analysis was conducted for the co-morbidity and age stratified models which had multi-dimensional βs.

After performing a one-by-one sensitivity analysis, it was determined that the three most sensitive parameters in the context of hospitalisations, infections and deaths were ρ_4 (number of days in ICU), ρ_1 (number of days with asymptomatic infection) and μ (death rate). This is illustrated by the violin plots in Figures 13-15.

The violin plots in Figure 13-15 show the most sensitive parameters in a hospitalisation, infection and death context. The parameters ρ_4 (number fo days in ICU) and μ (death rate) were sensitive in all contexts and therefore need to be re-estimated to improve model accuracy. A notable feature of these plots is that the plots are top-heavy for hospitalisations and total infections indicating that the above mentioned parameters are most sensitive during infection peaks in these contexts. In a death context however, the opposite is true as the plots are bottom-heavy and hence parameters are least sensitive during infection peaks. The one-by-one sensitivity analysis found that γ was a highly sensitive parameter in a hospitalisation and total infection context however, extensive research has shown that $\gamma \in (12, 14)$ days and so this parameter was not re-estimated.



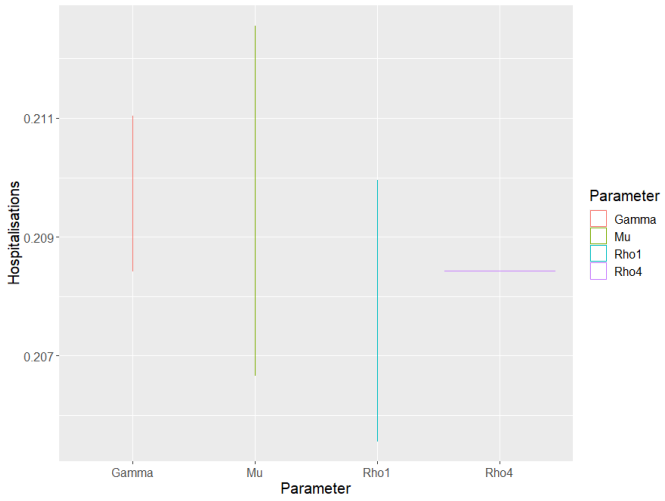
(a) Including ρ_4



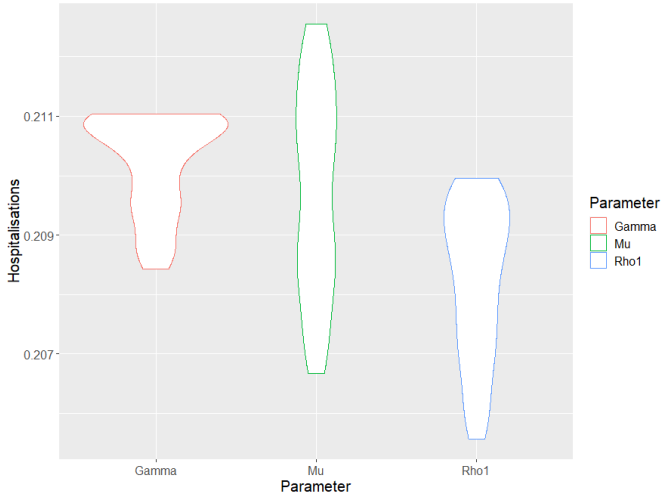
(b) Excluding ρ_4

Figure 13: Sensitivity of parameters to hospitalisations

.



(a) Including ρ_4



(b) Excluding ρ_4

Figure 14: Sensitivity of parameters to total infections

5.1.4 Multivariate sensitivity analysis

A multivariate sensitivity analysis involves varying several variables/parameters simultaneously to determine their effect on the model output. As with the one-by-one sensitivity analysis, this analysis was done in a hospitalisation, infection and death context. Table 6 shows the most sensitive param-

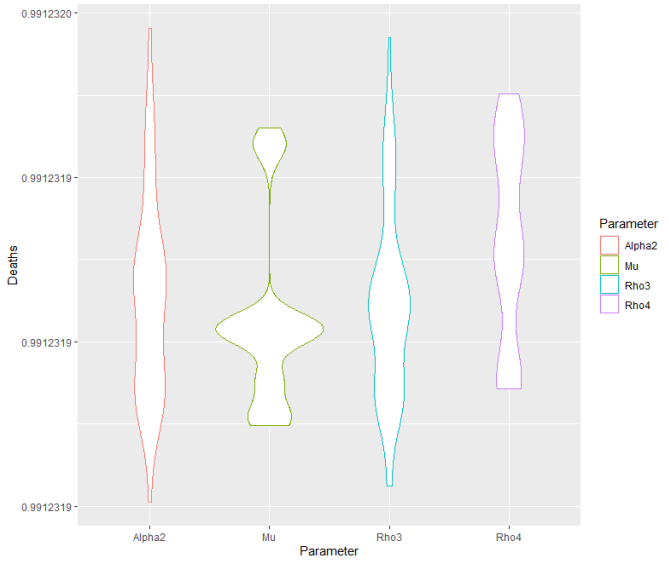


Figure 15: Sensitivity of parameters to total deaths

eters in all contexts for the basic, age stratified and co-morbidity stratified models:

Table 6: Multivariate sensitivity analysis results

Model	**Hospitalisation**	**Infection**	**Death**
Basic	γ, ρ_4, ρ_1	ρ_1, γ, p	ρ_1, γ, p
Age	p, ρ_2, α_1	p, ρ_2, α_1	μ_i, ρ_2, ρ_4
Co-morbidity	p, ρ_2, α_1	p, ρ_2, α_1	μ_i, ρ_4, ρ_1
Overall	p	ρ_2	μ_i

From Table 6, it is evident that the most sensitive parameters across all models and across all contexts were p (proportion of asymptomatic infections), ρ_2 (symptomatic infection duration) and μ_i (age/co-morbidity strata specific death rate). These parameters were highly sensitive, especially for the stratified models, because the proportion of asymptomatic cases, the infection duration and the death rate vary across age groups and between healthy and co-morbidity individuals. Asymptomatic COVID-19 cases are more prevalent in younger and/or healthy individuals. Infection duration

generally increases with infection severity. Since infection severity is higher for older and/or co-morbidity individuals, infection duration varies across the age and co-morbidity strata and hence ρ_2 is highly sensitive. Similarly, the death rate is higher amongst older individuals and individuals with co-morbidities and hence μ_i is highly sensitive. These parameters therefore need to be re-estimated to reflect the effects of age and co-morbidity in order to improve the accuracy of the various models.

5.1.5 Sensitivity of transmission matrix β

The multivariate sensitivity analysis showed that β in the co-morbidity stratified model was not significant in determining hospitalisations, infections or deaths. The sensitivity of β was therefore only determined for the age stratified model.

Standardised coefficients are used to show the strength of the effect of independent variables on the dependent variable. For this study, standard coefficients were used to identify the age groups driving hospitalisations, infections and deaths. This was done by identifying the age groups with the highest average standard coefficient values. The results are summarised in Table 7 below:

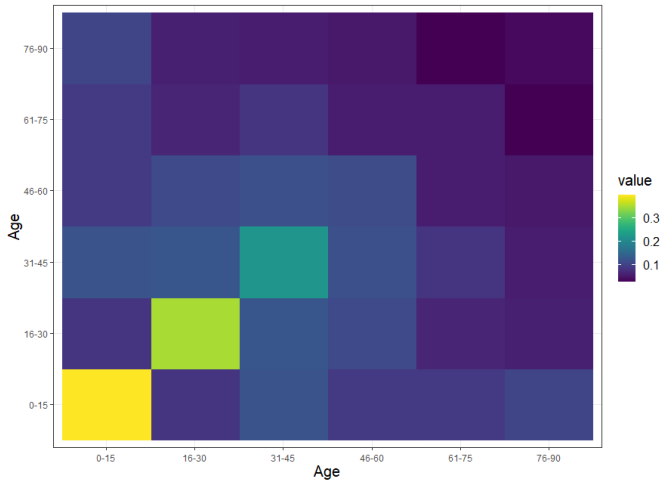
Table 7: Determining the age groups driving hospitalisations, infections and deaths

Age group	**Average standard coefficient value**		
	Hospitalisation	**Infection**	**Death**
0-15	0.1415	0.1482	0.2340
16-30	0.1512	0.1382	0.2795
31-45	0.1302	0.1377	0.2562
46-60	0.1260	0.1317	0.2553
61-75	**0.1572**	0.1502	0.2602
76-90	0.1485	**0.1508**	**0.3008**

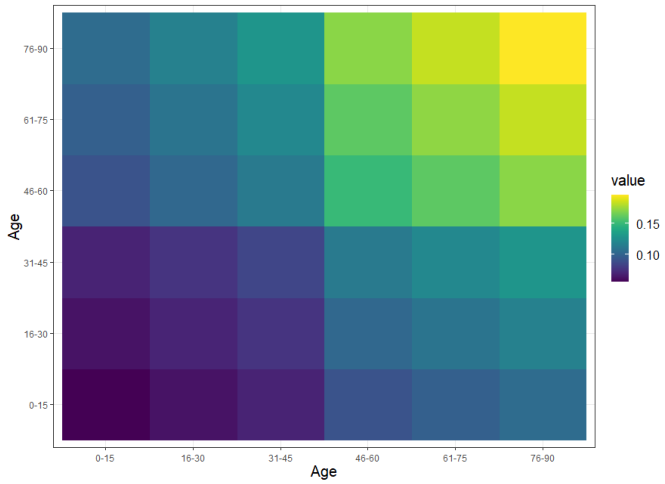
It is evident from Table 7 that hospitalisations were driven by the 61-75 age group while infections and deaths were driven by the 76-90 age group as these groups had the highest average standard coefficient values. This aligns with the observation that the effects of COVID-19 are harsher for older individuals as seen by the infection, hospitalisation and death rates

which increase with age. The β transmission matrix estimated in Method 1 assumed that susceptibility and transmissibility were highest for younger individuals as young people are more social and have more daily interactions. Since the average standard coefficients actually increase with age, susceptibility and transmissibility should also increase with age. This means that a more accurate β transmission matrix would have transmission probabilities which increase with age rather than decrease with age i.e., the β calculated in Method 2.

Figure 16 shows the transmission matrices from Method 1 and Method 2 of estimating β. Method 1 assumed higher transmissibility amongst younger age groups while Method 2 assumed higher transmissibility amongst older age groups. The sensitivity analysis showed that infections, hospitalisations and deaths were driven by older age groups (61-75 and 76-90) hence the transmission matrix in Figure 16(b) is a more accurate representation of COVID-19 transmissibility. This Figure is also a more accurate representation of the prevalence of infections, hospitalisations and deaths amongst different age groups.



(a) Method 1



(b) Method 2

Figure 16: Transmission matrices from Methods 1 and 2 of estimating β

5.2 Parameter re-estimation

5.2.1 Re-estimating ρ_2, ρ_4 and μ_i

The one-by-one and multivariate sensitivity analyses identified the most sensitive parameters to be p (proportion of asymptomatic infections), ρ_2 (symptomatic infection duration), ρ_4 (number of days in ICU) and μ_i (age/co-

morbidity strata specific death rate) across all models. These parameters were expected to be sensitive as they are the most dependent on age and/or co-morbidity. Parameters ρ_2, ρ_4 and μ_i were re-estimated using COVID-19 hospitalisation data obtained from the Gauteng DATCOV system. Due to limited data availability, p was re-estimated using literature in Section 5.2.2 below.

Table 8 below shows the age-specific estimates for ρ_2, ρ_4 and μ_i calculated from the hospitalisation data:

Table 8: Age specific ρ_2, ρ_4 and μ_i

Parameter	**0-15**	**16-30**	**31-45**	**46-60**	**61-75**	**76-90**
ρ_2	7	9	10	13	13	12
ρ_4	10.5	10.2	10.7	12.3	13.8	12.2
μ_i	0.04	0.05	0.11	0.21	0.34	0.44

It is evident from Table 8 above that the severity of the effects of COVID-19 worsen with age as the infection duration, days in ICU and COVID-19 death rates all increased with age. This table also suggests that ρ_2, ρ_4 and μ_i were highly sensitive parameters because they were assumed to be constant rather than age-dependent in the age-stratified model.

Table 9 below shows the number of COVID-19 deaths observed for various co-morbidities:
The hospitalisation data had 27 394 COVID-19 related deaths out of 125 244 hospitalisations (21.87% death rate). As shown in Table 9, 14 490 of the observed deaths (52.89%) were patients with underlying illnesses proving that the effects of COVID-19 are harsher amongst individuals with co-morbidities. The death rate for healthy individuals was then 10.3% and 11.6% for individuals with co-morbidities. The co-morbidity stratified model assumed that individuals with co-morbidities were 5 times more likely to die of COVID-19 than healthy individuals, however, the data shows that underlying illnesses only increased the likelihood of COVID-19 related death by 12.62% amongst mild and severe cases. Asymptomatic infections were not included due to limited data availability and the fact that people generally only visit the hospital if they present with mild or severe infection.

Table 9: Co-morbidity specific deaths

Co-morbidity	**Observed deaths**
Hypertension	6962
Diabetes	4157
HIV	752
Obesity	611
Cardiac Disease	564
Asthma	440
Pulmonary Disease	341
Renal Failure	325
Tuberculosis	183
Malignancy	155
Total	**14 490**

5.2.2 Re-estimating p

Another conclusion from the sensitivity analyses was that p was a very sensitive parameter for the age stratified and co-morbidity stratified models. This possibly indicates that the proportion of asymptomatic infections is not a constant $p = 0.75$ but instead varies with age and varies between healthy and co-morbidity individuals. Since the severity of COVID-19 infections worsens with age and co-morbidity, the proportion of asymptomatic infections is higher for younger and/or healthier individuals.

Determining the proportion of asymptomatic cases has been challenging as people generally only test for COVID-19 if they present with flu-like symptoms. As a result, accurate estimates of the proportion of asymptomatic cases is limited and so is data relating to this. The parameter p was therefore re-estimated based on a study conducted in Italy that determined the association of age with likelihood of developing symptomatic COVID-19 infection . The study saw the following results:

- People under age 20 : $p = 0.819 = 81.9\%$
- People under 60 : $p = 0.739 = 73.9\%$
- People over 80 : $p = 0.354 = 35.4\%$

Evidently, the likelihood of asymptomatic infection decreases considerably

with age. Taking this into account, the age-specific p was then re-estimated as follows:

Table 10: Age-specific p

	0-15	**16-30**	**31-45**	**46-60**	**61-75**	**76-90**
p	0.82	0.78	0.74	0.74	0.55	0.35

5.3 Fitting model to the second wave

To test accuracy and feasibility, the models were re-initialised and compared to the observed active cases for the second infection wave. This comparison was done by means of a χ^2 Goodness od Fit (GoF) test which is a statistical hypothesis test used to determine how well a fitted model fits the data.

Table 11 shows the χ^2 test statistics for the basic, age stratified and co-morbidity stratified models calculated before and after parameter re-estimation.

Table 11: Co-morbidity specific deaths

Model	χ^2 **Original**	χ^2 **Re-estimated**
Basic	4.067	2.384
Age stratified	26.564	12.970
Co-morbidity stratified	0.412	0.215

The χ^2 GoF test was conducted at a 5% significance level. The critical value was $\chi^2_{0.05,8} = 15.507$. Prior to parameter re-estimation, the test statistics for the basic and co-morbidity stratified models were less than the critical value indicating that there was insufficient evidence to reject the null hypothesis and hence these models were able to predict the second infection wave. This, however, was not the case for the age-stratified model which had a large test statistic value prior to re-estimation and hence was not able to accurately predict the second wave of infections.

After re-estimation, the test statistics of all models decreased significantly and were all less than the critical value. There was therefore insufficient evidence to reject the null hypothesis and it could be concluded that all models

were able to accurately predict the second wave of infections after parameter re-estimation. Table 11 also shows that the co-morbidity stratified model had the lowest χ^2 test statistics indicating that this model had the highest correlation with the observed data and hence was the best model for predicting the second infection waves. Another observation was that the age stratified model was only able to predict the second infection wave after accounting for the severity of the effects of COVID-19 which worsen with age. This further proves the high influence of age on the dynamics of COVID-19 and the importance of stratification in capturing population heterogeneity.

The sensitivity analyses and parameter re-estimation illustrated the importance of accounting for population heterogeneity when modelling the spread of COVID-19. Prior to re-estimation, all parameters were assumed to be constant across all age and co-morbidity strata however this analysis proved that this was inaccurate as there are variations in parameter values depending on age and co-morbidity. This analysis also proved that age was the most influential factor due to the transition probabilities having high standard coefficient values and the prediction accuracy of the age stratified model improving significantly after accounting for the effects of age on parameter values. This study therefore showed that age is a highly influential factor in understanding the unique dynamics of COVID-19 and should be accounted for in order to improve the accuracy of stratified compartmental models.

6 Conclusion

6.1 Discussion

This paper modelled the spread of COVID-19 in South Africa using a basic, age-stratified and co-morbidity stratified model to understand the unique dynamics of the novel corona virus. The population was stratified by age and co-morbidity as these are some of the factors which have the greatest influence on the transmission, hospitalisation and death rates for COVID-19.

Once the models were built, several sensitivity analyses were conducted to determine the models' sensitivity to initialisation date and lockdown restrictions and to identify highly sensitive parameters. Due to the novelty of COVID-19, there was a great deal of uncertainty regarding the start date of the pandemic, however the sensitivity analysis identified the optimal start date to be between 10-15 March 2020. Lockdown restrictions were implemented as a means of limiting the spread of COVID-19. The analysis determined that harsh lockdown restrictions for extended time periods would have delayed the infection peak by 4 - 6 months and reduced infections by approximately 46%. There is however a trade-off between saving lives and saving livelihoods. While a harsh lockdown would have significantly reduced COVID-19 cases, it also would have led to business closures, rising unemployment rates and a declining GDP which would have been detrimental for the long term health of the South African economy.

The one-by-one and multivariate sensitivity analyses identified the most sensitive parameters across all models to be transmission matrix β, p (proportion of asymptomatic cases), ρ_2 (symptomatic infection duration), ρ_4 (number of days in ICU) and μ_i (age/co-morbidity specific death rate). These parameters were identified to be highly sensitive specifically due to their high dependence on age. It was initially assumed that COVID-19 transmissibility was highest amongst younger age groups as young people attend school, university and social events and therefore have more daily interactions. The analysis however showed that older age groups (61-75 and 76-90) were in fact the drivers of hospitalisations, infections and deaths as the severity of the effects of COVID-19 worsen with age. These parameters were all re-estimated using hospitalisation data from the Gauteng DATCOV system. From this dataset it was established that since immunity weakens with age and the severity of co-morbidities also worsen with time, COVID-19 infections are not driven by age or co-morbidity independently. They are in-

stead driven by the synergy between age and co-morbidity. Parameters not related to hospitalisation were re-estimated from literature. These parameters were re-estimated accounting for the effects of age and co-morbidity and fit to the second wave of infections. The χ^2 GoF test showed that accounting for age and co-morbidity significantly improved the prediction accuracy of the models. The test also showed that prediction accuracy was most improved by accounting for the effects of COVID-19 which worsen with age and hence age was the most influential factor in understanding the dynamics of COVID-19 in South Africa.

6.2 Limitations

Modelling the spread of novel infectious diseases has several limitations as there is limited information and a great deal of misinformation during the emergence of an infectious disease. As a consequence, many variables are assumed or estimated, associations are established between unrelated events, parameters may be over or underestimated and there may be gaps in information since certain conclusions can only be made after several years of study and observation e.g. long term effects of COVID-19, long term effects of vaccines etc. The limitations for this study were mainly attributed to the novelty of COVID-19. Since COVID-19 only emerged in 2019, there was limited data availability and literature. There was also limited information related to asymptomatic infections since people only visit the hospital when they present with severe infection. Finally, age specific literature and data on susceptibility and transmissibility were limited hence proxies and estimates were frequently used which adversely affected accuracy of the models.

6.3 Further work

Age and co-morbidity are major contributors to the dynamics of COVID-19 however there are other factors which influence COVID-19 dynamics. These factors include spatial location, population density and vaccination. Further research can be conducted to identify the extent of the effects of these factors on the spread and dynamics of COVID-19. There is also limited data and literature available related to asymptomatic infection, age specific susceptibility and transmissibility and co-morbidity dependent contact matrices, specifically in a South African context, so further work in these areas would be worthwhile. Another consideration would be to add mobility to the age stratified model to include the effects of travel between provinces/cities/districts on the transmissibility of COVID-19 and determine the effects

of implementing travel restrictions on the rate of spread of COVID-19 in South Africa.

References

Adamu, H. A., Muhammad, M., Jingi, A., & Usman, M. (2019). Mathematical modelling using improved sir model with more realistic assumptions. *Int. J. Eng. Appl. Sci*, *6*, 64–69.

Ahmed, A., Salam, B., Mohammad, M., Akgul, A., & Khoshnaw, S. (2020). Analysis coronavirus disease (covid-19) model using numerical approaches and logistic model. *Aims Bioengineering*, *7*(3), 130–146.

Arumugam, R., & Rajathi, M. (2020). A markov model for prediction of corona virus covid-19 in india-a statistical study. *J Xidian Univ*, *14*(4), 1422–1426.

Ataguba, J. E. (2020). *Covid-19 pandemic, a war to be won: understanding its economic implications for africa.* Springer.

Balabdaoui, F., & Mohr, D. (2020). Age-stratified discrete compartment model of the covid-19 epidemic with application to switzerland. *Scientific reports*, *10*(1), 1–12.

Bjørnstad, O. N., Finkenstädt, B. F., & Grenfell, B. T. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. *Ecological monographs*, *72*(2), 169–184.

Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. (2020). *Unemployment in the time of covid-19: A research agenda.* Elsevier.

Brakenhoff, T. B., Roes, K. C., & Nikolakopoulos, S. (2019). Bayesian sample size re-estimation using power priors. *Statistical methods in medical research*, *28*(6), 1664–1675.

Brauer, F. (2008). Compartmental models in epidemiology. , 19–79.

BusinessTech. (2021). *Leading causes of death in South Africa.* Retrieved from `https://businesstech.co.za/news/lifestyle/498579/leading-causes-of-death-in-south-africa-2/` ((Accessed: 17 November 2021))

Christopher Frey, H., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk analysis*, *22*(3), 553–578.

Coburn, B. J., Wagner, B. G., & Blower, S. (2009). Modeling influenza epidemics and pandemics: insights into the future of swine flu (h1n1). *BMC medicine*, *7*(1), 1–8.

Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals*, *139*, 110057.

Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., & Eggo, R. M. (2020). Age-dependent effects in the transmission and control of covid-19 epi-

demics. *Nature medicine*, *26*(8), 1205–1211.

Garba, S. M., Lubuma, J. M.-S., & Tsanou, B. (2020). Modeling the transmission dynamics of the covid-19 pandemic in south africa. *Mathematical biosciences*, *328*, 108441.

Gaythorpe, K., Imai, N., Cuomo-Dannenburg, G., Baguelin, M., Bhatia, S., Boonyasiri, A., ... others (2020). Report 8: Symptom progression of covid-19. *Imperial College London*, *10*, 77344.

Gibson, G. J., & Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using markov chain methods. *Mathematical Medicine and Biology: A Journal of the IMA*, *15*(1), 19–40.

GitHub. (2021). *dsfsi/covid19africa.* Retrieved from `https://github.com/dsfsi/covid19africa` ((Accessed: 21 September 2021))

Gomes, C. (2020). Report of the who-china joint mission on coronavirus disease 2019 (covid-19). *Brazilian Journal of Implantology and Health Sciences*, *2*(3).

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, *32*(2), 135–154.

Korobeinikov, A. (2004). Lyapunov functions and global properties for seir and seis epidemic models. *Mathematical medicine and biology: a journal of the IMA*, *21*(2), 75–83.

Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, *96*(4), 450–459.

Landaw, E., & DiStefano 3rd, J. (1984). Multiexponential, multicompartmental, and noncompartmental modeling. ii. data analysis and statistical considerations. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *246*(5), R665–R677.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... Lessler, J. (2020). The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, *172*(9), 577–582.

McKibbin, W., & Fernando, R. (2020). The economic impact of covid-19. *Economics in the Time of COVID-19*, *45*.

Ndiaye, B. M., Tendeng, L., & Seck, D. (2020). Comparative prediction of confirmed cases with covid-19 pandemic by machine learning, deterministic and stochastic sir models. *arXiv preprint arXiv:2004.13489*.

Poletti, P., Tirani, M., Cereda, D., Trentini, F., Guzzetta, G., Sabatino, G., ... others (2021). Association of age with likelihood of developing symptoms and critical disease among close contacts exposed to

patients with confirmed sars-cov-2 infection in italy. *JAMA network open*, *4*(3), e211085–e211085.

Prem, K., Zandvoort, K. v., Klepac, P., Eggo, R. M., Davies, N. G., for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, C., ... Jit, M. (2021). Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the covid-19 era. *PLoS computational biology*, *17*(7), e1009098.

Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, *92*(638), 204–230.

Silal, S., Pulliam, J., Meyer-Rath, G., Nichols, B., Jamieson, L., Kimmie, Z., & Moultrie, H. (2020). Estimating cases for covid-19 in south africa update: 19 may 2020. *Update*.

Silva, C. J., & Torres, D. F. (2017). A sica compartmental model in epidemiology with application to hiv/aids in cape verde. *Ecological complexity*, *30*, 70–75.

Simha, A., Prasad, R. V., & Narayana, S. (2020). A simple stochastic sir model for covid 19 infection dynamics for karnataka: Learning from europe. *arXiv preprint arXiv:2003.11920*.

Sobie, E. A. (2009). Parameter sensitivity analysis in electrophysiological models using multivariable regression. *Biophysical journal*, *96*(4), 1264–1274.

South African COVID-19 Modelling Consortium. (2020). *National COVID Epi Model.* Retrieved from `https://sacovid19mc.github.io/supplementaryInformation` ((Accessed: 13 June 2021))

Thompson, K. M. (2016). Evolution and use of dynamic transmission models for measles and rubella risk and policy analysis. *Risk Analysis*, *36*(7), 1383–1403.

Wacker, B., & Schlüter, J. (2020). Time-continuous and time-discrete sir models revisited: theory and applications. *Advances in Difference Equations*, *2020*(1), 1–44.

Wang, H., & Wang, X.-S. (2016). Traveling wave phenomena in a kermack–mckendrick sir model. *Journal of Dynamics and Differential Equations*, *28*(1), 143–166.

World Health Organisation. (2020). *Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19.* Retrieved from `https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19?` ((Accessed: 21 September 2021))

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, *395*(10225), 689–697.

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M., Gill, H., Phan, L., ... others (2020). Impact of covid-19 pandemic on mental health in the general population: A systematic review. *Journal of affective disorders*.

A Understanding the parameters

Table 12: Effects of changes in parameter estimates

Parameter	**Value**	**Effect of increase**	**Effect of decrease**	**Comment**
β (Transmission probability)	0.55	• I and E peak early and higher • S drops faster • R peaks sooner	• I and E peak later and lower • S drops slower • R peaks later	Likely to be sensitive
γ (Incubation period)	14	• I and E peak later and higher • S drops slower • R peaks later and lower	• I and E peak sooner and lower • S drops faster • R peaks sooner and higher	Likely to be sensitive
p (Proportion of A infections)	0.75	• Significant drop in I and D • A peaks higher and sooner • E peaks slightly sooner • R peaks sooner	• Significant increase in I and D • A peaks lower and later • E peaks slightly later • R peaks later	Likely to be sensitive

B R Code

```
library(readxl)
library(ggplot2)
Active<- read_excel("Active cases.xlsx")
View(Active)

head(Active)
Date1=Active$Date
Inflated1=Active$Inflated
Dataf1=data.frame(Date1,Inflated1)
head(Dataf1)

obs<-ggplot(Dataf1,aes(x=Date1))+
  geom_line(aes(y=Inflated1),color="blue",size=1.5)+
  theme_grey(base_size = 12)+
  ggtitle("")+
  xlab("Date")+
  ylab("Number of COVID-19 infections")
obs+theme(text = element_text(size =20))+ theme(axis.title = element_text(




library(deSolve)
sir <- function(t, x, parms)  {
  with(as.list(c(parms, x)), {
    dS=-beta*I/59000000*S
    dI=beta*I/59000000*S-v*I-r*I
    dR=r*I
    dD=v*I
    output <- c(dS, dI, dR,dD)
    list(output)
  })
}
#SIRD Model (Basic COVID-19 Model)

#the Initial values
start<-c(S=58999999, I=1,R=0,D=0)
```

```
#Start with one infected person in the population

## The parameters
parms <- c(beta=0.68,v=0.02,r=0.55)
#r=recovery rate
#beta=infection rate
#v=death rate

## vector of timesteps
times <- seq(0, 200, 1)
#Time sequence of 200 days in 1 day increments

run_d<-ode(times=times, y=start, func=sir,parms=parms)
run_d.df<-as.data.frame(run_d)
head(run_d.df)
library(ggplot2)
title <- bquote("COVID-19 SIRD Model")
subtit<-bquote("Lets see")
res<-ggplot(run_d.df,aes(x=time))+
  ggtitle(bquote(atop(bold(.(title)),atop(bold(.(subtit))))))+
  geom_line(aes(y=S,colour="Susceptible"))+
  geom_line(aes(y=I,colour="Infected"))+
  geom_line(aes(y=R,colour="Recovered"))+
  geom_line(aes(y=D,colour="Dead"))+
  ylab(label="Frequency")+
  xlab(label="Time (days)")+
  theme(legend.justification=c(1,0), legend.position=c(0.3,0.05))+
  theme(legend.title=element_text(size=12,face="bold"),
  legend.background = element_rect(fill='#FFFFFF',
  size=0.5,linetype="solid"),
  legend.text=element_text(size=10),
  legend.key=element_rect(colour="#FFFFFF",
  fill='#C2C2C2',
  size=0.25,
  linetype="solid"))+
  scale_colour_manual("Compartments",
  breaks=c("Susceptible","Infected","Recovered","Dead"),
  values=c("blue","red","darkgreen","purple"))
print(res)
```

```
summary(run_d)
#Shows 5 number summary of S,I,R and D
run_d
#Shows the time, S, I, R and D over 70 days
plot(run_d[,3], col="blue",lwd=2,type = "l",main = "Obsevred and expected
lines(TimeSeries1,col="orange",lwd=2)
legend("topleft",legend=c("Expected Infections","Observed Infections"), co
#Plots the observed and expected infections over 150 days
#Both plots peak after approx. 66 days


#Multi-compartmental model
sirCov <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N=S+E+A+I+IHosp+IICU+R+D
    dS=-(beta*(I+0.75*A)/N)*S
    dE=(beta*(I+0.75*A)/N)*S-gamma*E
    dA=p*gamma*E-rho1*A
    dI=(1-p)*gamma*E-rho2*I-alpha1*I
    dIHosp=alpha1*I-alpha2*IHosp-rho3*IHosp-mu*IHosp
    dIICU=alpha2*IHosp-rho4*IICU-mu*IICU
    dR=rho1*A+rho2*I+rho3*IHosp+rho4*IICU
    dD=mu*IHosp+mu*IICU
    output <- c(dS,dE,dA,dI,dIHosp,dIICU,dR,dD)
    list(output)
  })
}

start1<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms1<-c(beta=0.58,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
times1 <- seq(0, 200, 1)
run_d<-ode(times=times1, y=start1, func=sirCov,parms=parms1)
summary(run_d)
head(run_d)


#VARYING PARAMETERS
#Assuming start date of 15 March 2020
start2<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms2<-c(beta=0.4,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9,
```

```
times2<- seq(0, 200, 1)
run_d2<-ode(times=times2, y=start2, func=sirCov,parms=parms2)
head(run_d2)
matplot(run_d2[,2:9], type="l", main="Beta=0.4")
lines(TimeSeries1,col="red",lwd=2)
#Late peak (day 180)

start2<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms3<-c(beta=0.45,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
times3<- seq(0, 200, 1)
run_d3<-ode(times=times3, y=start2, func=sirCov,parms=parms3)
matplot(run_d3[,2:9], type="l", main="Beta=0.45")
lines(TimeSeries1,col="red",lwd=2)
#Late peak (day 170)

start2<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms4<-c(beta=0.5,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9,
times4<- seq(0, 200, 1)
run_d4<-ode(times=times4, y=start2, func=sirCov,parms=parms4)
matplot(run_d4[,2:9], type="l", main="Beta=0.5")
lines(TimeSeries1,col="red",lwd=2)
#Late peak (day 150)

start2<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
times5<- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start2, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="Beta=0.55")
lines(TimeSeries1,col="red",lwd=2)
#Peak matches (day 138)

start2<-c(S=58998390, E=100, A=600, I=200, IHosp=10, IICU=5, R=50, D=10)
parms6<-c(beta=0.6,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9,
times6<- seq(0, 200, 1)
run_d6<-ode(times=times6, y=start2, func=sirCov,parms=parms6)
matplot(run_d6[,2:9], type="l", main="Beta=0.6")
lines(TimeSeries1,col="red",lwd=2)
#Early peak (day 130)
#Optimal beta is 0.55
```

```
#VARYING START DATES
par(mfrow=c(1,1))
start3<-c(S=58999989, E=5, A=3, I=3, IHosp=0, IICU=0, R=0, D=0)
#5 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 300, 1)
run_d5<-ode(times=times5, y=start3, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="5 March")
lines(TimeSeries1,col="red",lwd=2)
#Very late peak (day 190)

start4<-c(S=58999989, E=10, A=6, I=4, IHosp=0, IICU=0, R=0, D=0)
#6 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start4, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="6 March")
lines(TimeSeries1,col="red",lwd=2)
#Ver late peak (day 180)

start5<-c(S=58999887, E=50, A=42, I=21, IHosp=0, IICU=0, R=0, D=0)
#9 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start5, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="9 March")
lines(TimeSeries1,col="red",lwd=2)
#Ver late peak (day 160)

start6<-c(S=58999699, E=150, A=102, I=49, IHosp=0, IICU=0, R=0, D=0)
#11 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start6, func=sirCov,parms=parms5)
```

```
matplot(run_d5[,2:9], type="l", main="11 March")
lines(TimeSeries1,col="red",lwd=2)
#Ver late peak (day 155)

start7<-c(S=58999424, E=250, A=222, I=104, IHosp=0, IICU=0, R=0, D=0)
#13 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start7, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="13 March")
lines(TimeSeries1,col="red",lwd=2)
#Ver late peak (day 150)

start8<-c(S=58998678, E=550, A=519, I=223, IHosp=0, IICU=0, R=0, D=0)
#15 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start8, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="15 March")
lines(TimeSeries1,col="red",lwd=2)
#Peak matches (day 138)

start9<-c(S=58997512, E=1050, A=966, I=472, IHosp=0, IICU=0, R=0, D=0)
#17 March
parms5<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/7,rho3=1/5,rho4=1/9
#Beta=0.55
times5 <- seq(0, 200, 1)
run_d5<-ode(times=times5, y=start9, func=sirCov,parms=parms5)
matplot(run_d5[,2:9], type="l", main="17 March")
lines(TimeSeries1,col="red",lwd=2)
#Peak earyl (day 130)

##FINAL MODEL###
startf<-c(S=58997512, E=1050, A=966, I=472, IHosp=0, IICU=0, R=0, D=0)
#15 March
parmsf<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,rho4=1/9
#Beta=0.55
timesf <- seq(0, 200, 1)
```

```
run_df<-ode(times=timesf, y=startf, func=sirCov,parms=parmsf)
matplot(run_df[,2:9], type="l", main="Basic COVID-19 model")
lines(TimeSeries1,col="red",lwd=2)
#Peak matches (day 138)


### CO-MORBIDITY STRATIFIED MODEL ###
sirCoMorb <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2

    #Healthy equations
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I2+0.75*A2)/N2)-gamma*E1
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-rho2*I1-alpha1*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-alpha1*I1
    dIICU1=alpha2*IHosp1-rho4*IICU1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #Co-morbidity equations
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I2+0.75*A2)/N2)-gamma*E2
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-rho2*I2-alpha1*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-alpha1*I2
    dIICU2=alpha2*IHosp2-rho4*IICU2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2


    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

#48% of South Africans have a co-morbidity therefore:
#N1=30 680 000
```

```
#N2=28 320 000

#Look into seeding models (means of determining actual start date of pander
#Look into parameters (which ones need literature or data)
#How will you adjust data for parameters

starts<-c(S1=30679950 ,E1=22 ,A1=11, I1=20 ,IHosp1=2 ,IICU1=2 ,R1=4 ,D1=0
#Start date was 10 March 2020
parmss<-c(beta11=0.5,beta12=0.5,beta21=0.25,beta22=0.25,gamma=1/14,p=0.75,
timess <- seq(0, 200, 1)
run_ds<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmss)
head(run_ds)
population1<-run_ds[,2]+run_ds[,3]+run_ds[,4]+run_ds[,5]+run_ds[,6]+run_ds
population2<-run_ds[,10]+run_ds[,11]+run_ds[,12]+run_ds[,13]+run_ds[,14]+r


### AGE STRATIFIED MODEL ###
sirAge <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
    N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
    N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
    N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
    N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

    #1-15
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-alpha1*I1-rho2*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
    dIICU1=alpha2*IHosp1-rho4*IHosp1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #16-30
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
    dA2=p*gamma*E2-rho1*A2
```

```
dI2=(1-p)*gamma*E2-alpha1*I2-rho2*I2
dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
dIICU2=alpha2*IHosp2-rho4*IHosp2-mu2*IICU2
dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
dD2=mu2*IHosp2+mu2*IICU2

#31-45
dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
dA3=p*gamma*E3-rho1*A3
dI3=(1-p)*gamma*E3-alpha1*I3-rho2*I3
dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
dIICU3=alpha2*IHosp3-rho4*IHosp3-mu3*IICU3
dR3=rho1*A3+rho2*I3+rho3*IHosp3+rho4*IICU3
dD3=mu3*IHosp3+mu3*IICU3

#46-60
dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
dA4=p*gamma*E4-rho1*A4
dI4=(1-p)*gamma*E4-alpha1*I4-rho2*I4
dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
dIICU4=alpha2*IHosp4-rho4*IHosp4-mu4*IICU4
dR4=rho1*A4+rho2*I4+rho3*IHosp4+rho4*IICU4
dD4=mu4*IHosp4+mu4*IICU4

#61-75
dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
dA5=p*gamma*E5-rho1*A5
dI5=(1-p)*gamma*E5-alpha1*I5-rho2*I5
dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
dIICU5=alpha2*IHosp5-rho4*IHosp5-mu5*IICU5
dR5=rho1*A5+rho2*I5+rho3*IHosp5+rho4*IICU5
dD5=mu4*IHosp5+mu5*IICU5

#76-90
dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
dA6=p*gamma*E6-rho1*A6
```

```
    dI6=(1-p)*gamma*E6-alpha1*I6-rho2*I6
    dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
    dIICU6=alpha2*IHosp6-rho4*IHosp6-mu6*IICU6
    dR6=rho1*A6+rho2*I6+rho3*IHosp6+rho4*IICU6
    dD6=mu6*IHosp6+mu6*IICU6



    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

#SA Population by age
#N1=16 579 000
#N2=17 110 000
#N3=11 741 000
#N4= 8 024 000
#N5= 3 953 000
#N6= 1 593 000


#METHOD 1
library(deSolve)
R0=2.75
gamma=4.46
beta=R0/gamma
starts<-c(S1=16578939 ,E1=40 ,A1=8 ,I1=11 ,IHosp1=0 ,IICU1=0 ,R1=6 ,D1=0 ,S
#Start date was 15 March 2020
parmss<-c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15=0.091
#Check if parameters are per day or per year
timess <- seq(0, 200, 1)
run_ds<-ode(times=timess, y=starts, func=sirAge,parms=parmss)
head(run_ds)
tail(run_ds)
population1<-run_ds[,2]+run_ds[,3]+run_ds[,4]+run_ds[,5]+run_ds[,6]+run_ds
population2<-run_ds[,10]+run_ds[,11]+run_ds[,12]+run_ds[,13]+run_ds[,14]+ru
population3<-run_ds[,18]+run_ds[,19]+run_ds[,20]+run_ds[,21]+run_ds[,22]+ru
population4<-run_ds[,26]+run_ds[,27]+run_ds[,28]+run_ds[,29]+run_ds[,30]+ru
population5<-run_ds[,34]+run_ds[,35]+run_ds[,36]+run_ds[,37]+run_ds[,38]+ru
```

```
population6<-run_ds[,42]+run_ds[,43]+run_ds[,44]+run_ds[,45]+run_ds[,46]+r


#METHOD 2 (CHECK LATER)
#R0= (1.064 1.185 1.307 1.732 1.854 1.975)
starts2<-c(S1=16578939 ,E1=40 ,A1=8 ,Ip1=6 ,Im1=3 ,Is1=2 ,I31=0 ,I41=0 ,R1=
#Start date was 15 March 2020
parmss2<-c(beta11=0.254,beta12=0.283,beta13=0.312,beta14=0.413,beta15=0.44
#Check if parameters are per day or per year
timess2 <- seq(0, 200, 1)
run_ds<-ode(times=timess2, y=starts2, func=sirAge,parms=parmss2)
head(run_ds)
tail(run_ds)
population1<-run_ds[,2]+run_ds[,3]+run_ds[,4]+run_ds[,5]+run_ds[,6]+run_ds
population2<-run_ds[,12]+run_ds[,13]+run_ds[,14]+run_ds[,15]+run_ds[,16]+r
population3<-run_ds[,22]+run_ds[,23]+run_ds[,24]+run_ds[,25]+run_ds[,26]+r
population4<-run_ds[,32]+run_ds[,33]+run_ds[,34]+run_ds[,35]+run_ds[,36]+r
population5<-run_ds[,42]+run_ds[,43]+run_ds[,44]+run_ds[,45]+run_ds[,46]+r
population6<-run_ds[,52]+run_ds[,53]+run_ds[,54]+run_ds[,55]+run_ds[,56]+r



### ACCOUNTING FOR THE DIFFERENT LOCKDOWN LEVELS ###
#Level 5: 26 March 2020 - 30 April 2020 (Day 21 - Day 56)
#Level 4: 1 May 2020 - 31 May 2020 (Day 57 - Day 87)
#Level 3: 1 June 2020 - 17 August 2020 (Day 88 - Day 165)

#From NICD:
#R0_1 = 1.5
#R0_2 = 1.53
#R0_3 = 1.28

#Beta_1 (Level 5) = 0.45
#Beta_2 (Level 4) = 0.65
#Beta_3 (Level 3) = 0.7

sirCovLock <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N=S+E+A+I+IHosp+IICU+R+D
```

```
    beta=vector(length=200)
    for(t in 1:20){beta=0.8}
    for(t in 21:200){beta=0.45}

    dS=-(beta*(I+0.75*A)/N)*S
    dE=(beta*(I+0.75*A)/N)*S-gamma*E
    dA=p*gamma*E-rho1*A
    dI=(1-p)*gamma*E-rho2*I-alpha1*I
    dIHosp=alpha1*I-alpha2*IHosp-rho3*IHosp-mu*IHosp
    dIICU=alpha2*IHosp-rho4*IICU-mu*IICU
    dR=rho1*A+rho2*I+rho3*IHosp+rho4*IICU
    dD=mu*IHosp+mu*IICU
    output <- c(dS,dE,dA,dI,dIHosp,dIICU,dR,dD)
    list(output)
  })
}


startf<-c(S=58997512, E=1050, A=966, I=472, IHosp=0, IICU=0, R=0, D=0)
#15 March
parmsf<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,rho4=1/9
#Beta=0.55
timesf <- seq(0, 200, 1)
run_df<-ode(times=timesf, y=startf, func=sirCovLock,parms=parmsf)
matplot(run_df[,2:9], type="l", main="Basic COVID-19 model")
lines(TimeSeries1,col="red",lwd=2)
#Adjusting for lockdown levels shifts the plot right
#Infections peaking around day 150 (12 days later)


#Co-morbidity model
sirCoMorbL <- function(t, x, parms)  {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2


    beta11=vector(length=200)
    beta12=vector(length=200)
    beta21=vector(length=200)
```

```
    beta22=vector(length=200)
    for(t in 1:20){beta11=0.6}
    for(t in 21:200){beta11=0.3}
    for(t in 1:20){beta12=0.6}
    for(t in 21:200){beta12=0.3}
    for(t in 1:20){beta21=0.3}
    for(t in 21:200){beta21=0.1}
    for(t in 1:20){beta22=0.3}
    for(t in 21:200){beta22=0.1}

    #Healthy equations
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I2+0.75*A2)/N2)-gamma*E1
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-rho2*I1-alpha1*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-alpha1*I1
    dIICU1=alpha2*IHosp1-rho4*IICU1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #Co-morbidity equations
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I2+0.75*A2)/N2)-gamma*E2
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-rho2*I2-alpha1*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-alpha1*I2
    dIICU2=alpha2*IHosp2-rho4*IICU2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2


    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

starts<-c(S1=30679950 ,E1=22 ,A1=11, I1=20 ,IHosp1=2 ,IICU1=2 ,R1=4 ,D1=0
#Start date was 10 March 2020
parmss<-c(beta11=0.5,beta12=0.5,beta21=0.25,beta22=0.25,gamma=1/14,p=0.75,
timess <- seq(0, 400, 1)
```

```
run_ds<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmss)
run_dsL<-ode(times=timess, y=starts, func=sirCoMorbL,parms=parmss)

OGcomorb<-as.data.frame(run_ds)
LOCKcomorb<-as.data.frame(run_dsL)

dates2<-seq(as.Date("2020-03-15"),as.Date("2021-04-19"),by="days")
LOCKcomorb2<-cbind(dates2,LOCKcomorb)
head(LOCKcomorb2)

I1OG<-OGcomorb$I1+OGcomorb$A1+OGcomorb$IHosp1+OGcomorb$IICU1
I2OG<-OGcomorb$I2+OGcomorb$A2+OGcomorb$IHosp2+OGcomorb$IICU2
I1Lock<-LOCKcomorb$I1+LOCKcomorb$A1+LOCKcomorb$IHosp1+LOCKcomorb$IICU1
I2Lock<-LOCKcomorb$I2+LOCKcomorb$A2+LOCKcomorb$IHosp2+LOCKcomorb$IICU2

TIMEcomorb<-OGcomorb$time

OGH<-rep("Healthy Original",times=401)
OGC<-rep("Co-morbidity Original",times=401)
LockH<-rep("Healthy Lockdown",times=401)
LockC<-rep("Co-morbidity Lockdown",times=401)

O2<-rep("Original",times=401)
L2<-rep("Lockdown",times=401)

Infe1<-cbind(TIMEcomorb,OGH,I1OG,O2)
Infe2<-cbind(TIMEcomorb,OGC,I2OG,O2)
InfeL1<-cbind(TIMEcomorb,LockH,I1Lock,L2)
InfeL2<-cbind(TIMEcomorb,LockC,I2Lock,L2)

All2<-rbind(Infe1,Infe2,InfeL1,InfeL2)
head(All2)
B<-as.data.frame(All2)
colnames(B)<-c("Time","AgeGroup","N","Type")
head(B)
B$N<-as.integer(B$N)
B$Time<-as.integer(B$Time)
str(B)
BR<-round(B$N)
head(B)
```

```
colo<-ggplot(B,aes(x=Time, y=N, group=AgeGroup))+
  geom_line(aes(linetype=AgeGroup,colour=Type))+
  theme_grey()+
  scale_x_continuous(name="Time since first recorded COVID-19 case",limits=
  xlab("Time")+
  ylab('Number of COVID-19 infections')
colo+theme(text = element_text(size =15))+ theme(axis.title = element_text


#Age-model
sirAge <- function(t, x, parms)  {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
    N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
    N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
    N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
    N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

    #1-15
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-alpha1*I1-rho2*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
    dIICU1=alpha2*IHosp1-rho4*IHosp1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #16-30
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-alpha1*I2-rho2*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
    dIICU2=alpha2*IHosp2-rho4*IHosp2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2

    #31-45
```

```
dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
dA3=p*gamma*E3-rho1*A3
dI3=(1-p)*gamma*E3-alpha1*I3-rho2*I3
dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
dIICU3=alpha2*IHosp3-rho4*IHosp3-mu3*IICU3
dR3=rho1*A3+rho2*I3+rho3*IHosp3+rho4*IICU3
dD3=mu3*IHosp3+mu3*IICU3

#46-60
dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
dA4=p*gamma*E4-rho1*A4
dI4=(1-p)*gamma*E4-alpha1*I4-rho2*I4
dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
dIICU4=alpha2*IHosp4-rho4*IHosp4-mu4*IICU4
dR4=rho1*A4+rho2*I4+rho3*IHosp4+rho4*IICU4
dD4=mu4*IHosp4+mu4*IICU4

#61-75
dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
dA5=p*gamma*E5-rho1*A5
dI5=(1-p)*gamma*E5-alpha1*I5-rho2*I5
dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
dIICU5=alpha2*IHosp5-rho4*IHosp5-mu5*IICU5
dR5=rho1*A5+rho2*I5+rho3*IHosp5+rho4*IICU5
dD5=mu4*IHosp5+mu5*IICU5

#76-90
dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
dA6=p*gamma*E6-rho1*A6
dI6=(1-p)*gamma*E6-alpha1*I6-rho2*I6
dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
dIICU6=alpha2*IHosp6-rho4*IHosp6-mu6*IICU6
dR6=rho1*A6+rho2*I6+rho3*IHosp6+rho4*IICU6
dD6=mu6*IHosp6+mu6*IICU6
```

```
    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

#Standard model

sirAgeLock <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
    N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
    N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
    N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
    N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

    #1-15
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-alpha1*I1-rho2*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
    dIICU1=alpha2*IHosp1-rho4*IHosp1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    beta11=vector(length=200)
    beta12=vector(length=200)
    beta13=vector(length=200)
    beta14=vector(length=200)
    beta15=vector(length=200)
    beta16=vector(length=200)
    beta21=vector(length=200)
    beta22=vector(length=200)
    beta23=vector(length=200)
    beta24=vector(length=200)
    beta25=vector(length=200)
    beta26=vector(length=200)
    beta31=vector(length=200)
```

```
beta32=vector(length=200)
beta33=vector(length=200)
beta34=vector(length=200)
beta35=vector(length=200)
beta36=vector(length=200)
beta41=vector(length=200)
beta42=vector(length=200)
beta43=vector(length=200)
beta44=vector(length=200)
beta45=vector(length=200)
beta46=vector(length=200)
beta51=vector(length=200)
beta52=vector(length=200)
beta53=vector(length=200)
beta54=vector(length=200)
beta55=vector(length=200)
beta56=vector(length=200)
beta61=vector(length=200)
beta62=vector(length=200)
beta63=vector(length=200)
beta64=vector(length=200)
beta65=vector(length=200)
beta66=vector(length=200)
for(t in 1:20){beta11=0.399}
for(t in 21:200){beta11=0.198}
for(t in 1:20){beta12=0.085}
for(t in 21:200){beta12=0.042}
for(t in 1:20){beta13=0.124}
for(t in 21:200){beta13=0.062}
for(t in 1:20){beta14=0.090}
for(t in 21:200){beta14=0.045}
for(t in 1:20){beta15=0.091}
for(t in 21:200){beta15=0.045}
for(t in 1:20){beta16=0.104}
for(t in 21:200){beta16=0.052}
for(t in 1:20){beta21=0.085}
for(t in 21:200){beta21=0.042}
for(t in 1:20){beta22=0.351}
for(t in 21:200){beta22=0.175}
for(t in 1:20){beta23=0.128}
```

```
for(t in 21:200){beta23=0.064}
for(t in 1:20){beta24=0.112}
for(t in 21:200){beta24=0.066}
for(t in 1:20){beta25=0.066}
for(t in 21:200){beta25=0.033}
for(t in 1:20){beta26=0.061}
for(t in 21:200){beta26=0.030}
for(t in 1:20){beta31=0.124}
for(t in 21:200){beta31=0.062}
for(t in 1:20){beta32=0.128}
for(t in 21:200){beta32=0.064}
for(t in 1:20){beta33=0.222}
for(t in 21:200){beta33=0.111}
for(t in 1:20){beta34=0.119}
for(t in 21:200){beta34=0.059}
for(t in 1:20){beta35=0.085}
for(t in 21:200){beta35=0.042}
for(t in 1:20){beta36=0.056}
for(t in 21:200){beta36=0.028}
for(t in 1:20){beta41=0.090}
for(t in 21:200){beta41=0.045}
for(t in 1:20){beta42=0.112}
for(t in 21:200){beta42=0.066}
for(t in 1:20){beta43=0.119}
for(t in 21:200){beta43=0.059}
for(t in 1:20){beta44=0.113}
for(t in 21:200){beta44=0.056}
for(t in 1:20){beta45=0.058}
for(t in 21:200){beta45=0.029}
for(t in 1:20){beta46=0.053}
for(t in 21:200){beta46=0.026}
for(t in 1:20){beta51=0.291}
for(t in 21:200){beta51=0.045}
for(t in 1:20){beta52=0.066}
for(t in 21:200){beta52=0.033}
for(t in 1:20){beta53=0.085}
for(t in 21:200){beta53=0.042}
for(t in 1:20){beta54=0.058}
for(t in 21:200){beta54=0.029}
for(t in 1:20){beta55=0.057}
```

```
for(t in 21:200){beta55=0.029}
for(t in 1:20){beta56=0.029}
for(t in 21:200){beta56=0.014}
for(t in 1:20){beta61=0.104}
for(t in 21:200){beta61=0.052}
for(t in 1:20){beta62=0.061}
for(t in 21:200){beta62=0.030}
for(t in 1:20){beta63=0.056}
for(t in 21:200){beta63=0.028}
for(t in 1:20){beta64=0.053}
for(t in 21:200){beta64=0.026}
for(t in 1:20){beta65=0.029}
for(t in 21:200){beta65=0.014}
for(t in 1:20){beta66=0.037}
for(t in 21:200){beta66=0.018}

#16-30
dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
dA2=p*gamma*E2-rho1*A2
dI2=(1-p)*gamma*E2-alpha1*I2-rho2*I2
dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
dIICU2=alpha2*IHosp2-rho4*IHosp2-mu2*IICU2
dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
dD2=mu2*IHosp2+mu2*IICU2

#31-45
dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
dA3=p*gamma*E3-rho1*A3
dI3=(1-p)*gamma*E3-alpha1*I3-rho2*I3
dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
dIICU3=alpha2*IHosp3-rho4*IHosp3-mu3*IICU3
dR3=rho1*A3+rho2*I3+rho3*IHosp3+rho4*IICU3
dD3=mu3*IHosp3+mu3*IICU3

#46-60
dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
dA4=p*gamma*E4-rho1*A4
```

```
    dI4=(1-p)*gamma*E4-alpha1*I4-rho2*I4
    dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
    dIICU4=alpha2*IHosp4-rho4*IHosp4-mu4*IICU4
    dR4=rho1*A4+rho2*I4+rho3*IHosp4+rho4*IICU4
    dD4=mu4*IHosp4+mu4*IICU4

    #61-75
    dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
    dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
    dA5=p*gamma*E5-rho1*A5
    dI5=(1-p)*gamma*E5-alpha1*I5-rho2*I5
    dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
    dIICU5=alpha2*IHosp5-rho4*IHosp5-mu5*IICU5
    dR5=rho1*A5+rho2*I5+rho3*IHosp5+rho4*IICU5
    dD5=mu4*IHosp5+mu5*IICU5

    #76-90
    dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
    dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
    dA6=p*gamma*E6-rho1*A6
    dI6=(1-p)*gamma*E6-alpha1*I6-rho2*I6
    dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
    dIICU6=alpha2*IHosp6-rho4*IHosp6-mu6*IICU6
    dR6=rho1*A6+rho2*I6+rho3*IHosp6+rho4*IICU6
    dD6=mu6*IHosp6+mu6*IICU6



    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}


#METHOD 1
R0=2.75
gamma=4.46
beta=R0/gamma
starts<-c(S1=16578939 ,E1=40 ,A1=8 ,I1=11 ,IHosp1=0 ,IICU1=0 ,R1=6 ,D1=0 ,S
#Start date was 15 March 2020
```

```
parmss<-c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15=0.091
#Check if parameters are per day or per year
timess <- seq(0, 600, 1)
run_ds1<-ode(times=timess, y=starts, func=sirAge,parms=parmss)
run_ds2<-ode(times=timess, y=starts, func=sirAgeLock,parms=parmss)
run_ds1A<-as.data.frame(run_ds1)
run_ds2A<-as.data.frame(run_ds2)
I1AOG<-run_ds1A$A1+run_ds1A$I1+run_ds1A$IHosp1+run_ds1A$IICU1
I2AOG<-run_ds1A$A2+run_ds1A$I2+run_ds1A$IHosp2+run_ds1A$IICU2
I3AOG<-run_ds1A$A3+run_ds1A$I3+run_ds1A$IHosp3+run_ds1A$IICU3
I4AOG<-run_ds1A$A4+run_ds1A$I4+run_ds1A$IHosp4+run_ds1A$IICU4
I5AOG<-run_ds1A$A5+run_ds1A$I5+run_ds1A$IHosp5+run_ds1A$IICU5
I6AOG<-run_ds1A$A6+run_ds1A$I6+run_ds1A$IHosp6+run_ds1A$IICU6
I1AL<-run_ds2A$A1+run_ds2A$I1+run_ds2A$IHosp1+run_ds2A$IICU1
I2AL<-run_ds2A$A2+run_ds2A$I2+run_ds2A$IHosp2+run_ds2A$IICU2
I3AL<-run_ds2A$A3+run_ds2A$I3+run_ds2A$IHosp3+run_ds2A$IICU3
I4AL<-run_ds2A$A4+run_ds2A$I4+run_ds2A$IHosp4+run_ds2A$IICU4
I5AL<-run_ds2A$A5+run_ds2A$I5+run_ds2A$IHosp5+run_ds2A$IICU5
I6AL<-run_ds2A$A6+run_ds2A$I6+run_ds2A$IHosp6+run_ds2A$IICU6
TIME<-run_ds1A$time #601 values

OG015<-rep("0-15 Original",times=601)
OG1630<-rep("16-30 Original",times=601)
OG3145<-rep("31-45 Original",times=601)
OG4660<-rep("46-60 Original",times=601)
OG6175<-rep("61-75 Original",times=601)
OG7690<-rep("76-90 Original",times=601)

L015<-rep("0-15 Lockdown",times=601)
L1630<-rep("16-30 Lockdown",times=601)
L3145<-rep("31-45 Lockdown",times=601)
L4660<-rep("46-60 Lockdown",times=601)
L6175<-rep("61-75 Lockdown",times=601)
L7690<-rep("76-90 Lockdown",times=601)

O1<-rep("Original",times=601)
L1<-rep("Lockdown",times=601)

Inf1<-cbind(TIME,OG015,I1AOG,O1)
Inf2<-cbind(TIME,OG1630,I2AOG,O1)
```

```
Inf3<-cbind(TIME,OG3145,I3AOG,O1)
Inf4<-cbind(TIME,OG4660,I4AOG,O1)
Inf5<-cbind(TIME,OG6175,I5AOG,O1)
Inf6<-cbind(TIME,OG7690,I6AOG,O1)

InfL1<-cbind(TIME,L015,I1AL,L1)
InfL2<-cbind(TIME,L1630,I2AL,L1)
InfL3<-cbind(TIME,L3145,I3AL,L1)
InfL4<-cbind(TIME,L4660,I4AL,L1)
InfL5<-cbind(TIME,L6175,I5AL,L1)
InfL6<-cbind(TIME,L7690,I6AL,L1)

ALL<-rbind(Inf1,Inf2,Inf3,Inf4,Inf5,Inf6,InfL1,InfL2,InfL3,InfL4,InfL5,InfL
head(ALL)
A<-as.data.frame(ALL)
View(A)
colnames(A)<-c("Time","AgeGroup","N","Type")
head(A)
tail(A)
A$N<-as.integer(A$N)
A$Time<-as.integer(A$Time)
str(A)
AR<-round(A$N)
head(A)
View(A)

aglo<-ggplot(A,aes(x=Time, y=N, group=AgeGroup))+
  geom_line(aes(linetype=AgeGroup,color=Type))+
  theme_grey()+
  scale_x_continuous(name="Time since first recorded COVID-19 case",limits=
  xlab('Time since first reported COVID-19 case')+
  ylab('Number of COVID-19 infections')
aglo+theme(text = element_text(size =15))+ theme(axis.title = element_text



#METHOD 2 (CHECK LATER)
#R0= (1.064 1.185 1.307 1.732 1.854 1.975)
starts2<-c(S1=16578939 ,E1=40 ,A1=8 ,Ip1=6 ,Im1=3 ,Is1=2 ,I31=0 ,I41=0 ,R1=
#Start date was 15 March 2020
```

```
parmss2<-c(beta11=0.254,beta12=0.283,beta13=0.312,beta14=0.413,beta15=0.44
#Check if parameters are per day or per year
timess2 <- seq(0, 200, 1)
run_ds<-ode(times=timess2, y=starts2, func=sirAge,parms=parmss2)
head(run_ds)
tail(run_ds)
population1<-run_ds[,2]+run_ds[,3]+run_ds[,4]+run_ds[,5]+run_ds[,6]+run_ds
population2<-run_ds[,12]+run_ds[,13]+run_ds[,14]+run_ds[,15]+run_ds[,16]+r
population3<-run_ds[,22]+run_ds[,23]+run_ds[,24]+run_ds[,25]+run_ds[,26]+r
population4<-run_ds[,32]+run_ds[,33]+run_ds[,34]+run_ds[,35]+run_ds[,36]+r
population5<-run_ds[,42]+run_ds[,43]+run_ds[,44]+run_ds[,45]+run_ds[,46]+r
population6<-run_ds[,52]+run_ds[,53]+run_ds[,54]+run_ds[,55]+run_ds[,56]+r


#Beta estimation method 2
timetry<-0:1
IGam<-dgamma(timetry,2.1,4)
IGam
mean(IpGam)
IGam<-dgamma(timetry,2.9,4)
IHospGam<-dgamma(timetry,2.9,4)
IICUGam<-dgamma(timetry,2.9,4)
AGam<-dgamma(timetry,5,4)
mean(IGam+IHospGam+IICUGam) #1.277708
j<-(1-0.25)*0.7*mean(AGam)+0.25*mean(IGam+IHospGam+IICUGam)
v1=0.35
v2=0.39
v3=0.43
v4=0.57
v5=0.61
v6=0.65
v1*j
?matrix
C<-matrix(c(4.1044,0.8,0.9901,0.3113,0.1001,0.0317,0.9837,3.8031,1.1860,0.5
C #Contact matrix

NGM1<-v1*j*C
eigen(NGM1) #Dominant=1.064
#Dominant=0.8405
```

```
NGM2<-v2*j*C
eigen(NGM2) #Dominant=1.185
#Dominant=0.9366


NGM3<-v3*j*C
eigen(NGM3) #Dominant=1.307
#Dominant=1.0326

NGM4<-v4*j*C
eigen(NGM4) #Dominant=1.732
#Dominant=1.3688


NGM5<-v5*j*C
eigen(NGM5) #Dominant=1.854
#Dominant=1.4649

NGM6<-v6*j*C
eigen(NGM6) #Dominant=1.975
#Dominant=1.5610

T<-matrix(c(1.132,1.261,1.391,1.843,1.973,2.101,1.261,1.404,1.549,2.052,2.1
T
4.46^2
Tnew<-1/(4.46^2)*T
Tnew
Tinv<-solve(Tnew)

R0b<-matrix(c(0.8405,0.9366,1.0326,1.3688,1.4649,1.5610),nrow=6,ncol = 1,by
R0bT<-matrix(c(0.8405,0.9366,1.0326,1.3688,1.4649,1.5610),nrow=1,ncol = 6,b
R0TEST<-R0b%*%R0bT
Beta2<-1/5.5625^2*R0TEST
#This is the beta transmission matrix for Method 2 (re-estimation)
Beta2

library(deSolve)
library(ggplot2)

###Sensitivity Analysis ###
```

```
#Basic Model
library(deSolve)
sirCov <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N=S+E+A+I+IHosp+IICU+R+D
    dS=-(beta*(I+0.75*A)/N)*S
    dE=(beta*(I+0.75*A)/N)*S-gamma*E
    dA=p*gamma*E-rho1*A
    dI=(1-p)*gamma*E-rho2*I-alpha1*I
    dIHosp=alpha1*I-alpha2*IHosp-rho3*IHosp-mu*IHosp
    dIICU=alpha2*IHosp-rho4*IICU-mu*IICU
    dR=rho1*A+rho2*I+rho3*IHosp+rho4*IICU
    dD=mu*IHosp+mu*IICU
    dCumHosp=alpha1*I+alpha2*IHosp
    dCumInf=p*gamma*E+(1-p)*gamma*E+alpha1*I+alpha2*IHosp
    dCumExp=(beta*(I+0.75*A)/N)*S
    output <- c(dS,dE,dA,dI,dIHosp,dIICU,dR,dD,dCumHosp,dCumInf,dCumExp)
    list(output)
  })
}

startf<-c(S=58998450, E=100, A=600, I=200, IHosp=10, IICU=5, R=0, D=0,CumH
#15 March
parmsf<-c(beta=0.55,gamma=1/14,p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,rho4=1/9
#Beta=0.55
timesf <- seq(0, 200, 1)
run_df<-ode(times=timesf, y=startf, func=sirCov,parms=parmsf)
head(run_df)
#Time,S,E,A,I,IHosp,IICU,R,D,CumHosp,CumInf,CumExp


### HOSPITALISATIONS ###
N=59000000
BetaSensitive<-NULL
for (i in 1:100){
  parmsfsen <- c(beta=runif(1,0.2, 1),gamma=1/14, p=0.75,rho1=1/16,rho2=1/
  run_dsen<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen)
  BetaSensitive<-rbind(BetaSensitive, c(parmsfsen[1],run_dsen[length(run_ds
}
```

```
plot(BetaSensitive[,1],BetaSensitive[,2],xlab = "Beta",ylab = "Hospitaliza
#Hospitalisations increase as transmission rate increases
bs<-as.data.frame(BetaSensitive)
colnames(bs)<-c("Parm","Hospitalisations")
Parameter<-rep("Beta",times=100)
bsf<-cbind(bs,Parameter)


GammaSensitive<-NULL
for (i in 1:100){
  parmsfsen1 <- c(beta=0.5,gamma=1/runif(1,10,14), p=0.75,rho1=1/16,rho2=1
  run_dsen1<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen1)
  GammaSensitive<-rbind(GammaSensitive, c(parmsfsen1[2],run_dsen1[length(ru
}
plot(GammaSensitive[,1],GammaSensitive[,2],xlab = "Gamma",ylab = "Hospital
#Number of hospitalisations increase as incubation period increases
bsg<-as.data.frame(GammaSensitive)
colnames(bsg)<-c("Parm","Hospitalisations")
Parameter<-rep("Gamma",times=100)
bsgf<-cbind(bsg,Parameter)


PSensitive<-NULL
for (i in 1:100){
  parmsfsen2 <- c(beta=0.5,gamma=1/14, p=runif(1,0.6,0.9),rho1=1/16,rho2=1
  run_dsen2<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen2)
  PSensitive<-rbind(PSensitive, c(parmsfsen2[3],run_dsen2[length(run_dsen2
}
plot(PSensitive[,1],PSensitive[,2],xlab = "P",ylab = "Hospitalizations (I3
#Hosps decrease as proportion of A infections increase
bsp<-as.data.frame(PSensitive)
colnames(bsp)<-c("Parm","Hospitalisations")
Parameter<-rep("P",times=100)
bspf<-cbind(bsp,Parameter)


Rho1Sensitive<-NULL
for (i in 1:100){
  parmsfsen3 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/runif(1,14,18),rho2=1
  run_dsen3<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen3)
```

```
  Rho1Sensitive<-rbind(Rho1Sensitive, c(parmsfsen3[4],run_dsen3[length(run
}
plot(Rho1Sensitive[,1],Rho1Sensitive[,2],xlab = "Rho1",ylab = "Hospitaliza
#Hosps decrease as number of days with A infection increase
bsr1<-as.data.frame(Rho1Sensitive)
colnames(bsr1)<-c("Parm","Hospitalisations")
Parameter<-rep("Rho1",times=100)
bsr1f<-cbind(bsr1,Parameter)


Rho2Sensitive<-NULL
for (i in 1:100){
  parmsfsen4 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/runif(1,7,1
  run_dsen4<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen4)
  Rho2Sensitive<-rbind(Rho2Sensitive, c(parmsfsen4[5],run_dsen4[length(run
}
plot(Rho2Sensitive[,1],Rho2Sensitive[,2],xlab = "Rho2",ylab = "Hospitaliza
#Hosps decrease as number of days with symptomatic infection increase
bsr2<-as.data.frame(Rho2Sensitive)
colnames(bsr2)<-c("Parm","Hospitalisations")
Parameter<-rep("Rho2",times=100)
bsr2f<-cbind(bsr2,Parameter)


Rho3Sensitive<-NULL
for (i in 1:100){
  parmsfsen5 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/ru
  run_dsen5<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen5)
  Rho3Sensitive<-rbind(Rho3Sensitive, c(parmsfsen5[6],run_dsen5[length(run
}
plot(Rho3Sensitive[,1],Rho3Sensitive[,2],xlab = "Rho3",ylab = "Hospitaliza
#Hosps decrease as no.of days in hospital increase
bsr3<-as.data.frame(Rho3Sensitive)
colnames(bsr3)<-c("Parm","Hospitalisations")
Parameter<-rep("Rho3",times=100)
bsr3f<-cbind(bsr3,Parameter)


Rho4Sensitive<-NULL
for (i in 1:100){
```

```
  parmsfsen6 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen6<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen6)
  Rho4Sensitive<-rbind(Rho4Sensitive, c(parmsfsen6[7],run_dsen6[length(run
}
plot(Rho4Sensitive[,1],Rho4Sensitive[,2],xlab = "Rho4",ylab = "Hospitaliza
#Hosps increase as no. of days in ICU increase (no trend really)
bsr4<-as.data.frame(Rho4Sensitive)
colnames(bsr4)<-c("Parm","Hospitalisations")
Parameter<-rep("Rho4",times=100)
bsr4f<-cbind(bsr4,Parameter)


Alpha1Sensitive<-NULL
for (i in 1:100){
  parmsfsen7 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen7 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen7)
  Alpha1Sensitive<-rbind(Alpha1Sensitive, c(parmsfsen7[8],run_dsen7[length
}
plot(Alpha1Sensitive[,1],Alpha1Sensitive[,2],xlab = "Alpha1",ylab = "Hospi
#Hosps increase as the no. of days before hospitalisation is required incr
bsa1<-as.data.frame(Alpha1Sensitive)
colnames(bsa1)<-c("Parm","Hospitalisations")
Parameter<-rep("Alpha1",times=100)
bsa1f<-cbind(bsa1,Parameter)


Alpha2Sensitive<-NULL
for (i in 1:100){
  parmsfsen8 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen8 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen8)
  Alpha2Sensitive<-rbind(Alpha2Sensitive, c(parmsfsen8[9],run_dsen8[length
}
plot(Alpha2Sensitive[,1],Alpha2Sensitive[,2],xlab = "Alpha2",ylab = "Hospi
#Hosps increase as number of days before ICU admission increases
bsa2<-as.data.frame(Alpha2Sensitive)
colnames(bsa2)<-c("Parm","Hospitalisations")
Parameter<-rep("Alpha2",times=100)
bsa2f<-cbind(bsa2,Parameter)
```

```
MuSensitive<-NULL
for (i in 1:100){
  parmsfsen11 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5
  run_dsen11 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen11)
  MuSensitive<-rbind(MuSensitive, c(parmsfsen11[10],run_dsen11[length(run_
}
head(Alpha2Sensitive)
head(MuSensitive)
plot(MuSensitive[,1],MuSensitive[,2],xlab = "Mu",ylab = "Hospitalizations
#Hosps decrease as death rate increases
bsmu<-as.data.frame(MuSensitive)
colnames(bsmu)<-c("Parm","Hospitalisations")
Parameter<-rep("Mu",times=100)
bsmuf<-cbind(bsmu,Parameter)


#p,rho1,rho2,rho3,rho4,alpha1,alpha2,alpha3,alpha4,mu
letstry<-rbind(bsf,bsgf,bspf,bsr1f,bsr2f,bsr3f,bsa1f,bsa2f,bsmuf)
top3<-rbind(bsgf,bsr1f,bsmuf,bsr4f)
head(letstry)
violintry<-ggplot(top3,aes(x=Parameter,y=Hospitalisations,color=Parameter))
violintry+theme(text = element_text(size =15))+ theme(axis.title = element
#Top 3 most sensitive parameters were rho4,gamma, rho1 and mu
#i.e. No. of days in ICU, incubation period, number of days with A infectio
#and death rate



### INFECTIONS ###
N=59000000
BetaSensitive<-NULL
for (i in 1:100){
  parmsfsen <- c(beta=runif(1,0.2, 1),gamma=1/14, p=0.75,rho1=1/16,rho2=1/
  run_dsen<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen)
  BetaSensitive<-rbind(BetaSensitive, c(parmsfsen[1],run_dsen[length(run_ds
}
plot(BetaSensitive[,1],BetaSensitive[,2],xlab = "Beta",ylab = "Infections"
#Infections increase as transmission rate increases
bs<-as.data.frame(BetaSensitive)
colnames(bs)<-c("Parm","Infections")
```

```
Parameter<-rep("Beta",times=100)
bsf<-cbind(bs,Parameter)


GammaSensitive<-NULL
for (i in 1:100){
  parmsfsen1 <- c(beta=0.5,gamma=1/runif(1,10,14), p=0.75,rho1=1/16,rho2=1
  run_dsen1<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen1)
  GammaSensitive<-rbind(GammaSensitive, c(parmsfsen1[2],run_dsen1[length(ru
}
plot(GammaSensitive[,1],GammaSensitive[,2],xlab = "Gamma",ylab = "Infectior
#Number of Infections increase as incubation period increases
bsg<-as.data.frame(GammaSensitive)
colnames(bsg)<-c("Parm","Infections")
Parameter<-rep("Gamma",times=100)
bsgf<-cbind(bsg,Parameter)


PSensitive<-NULL
for (i in 1:100){
  parmsfsen2 <- c(beta=0.5,gamma=1/14, p=runif(1,0.6,0.9),rho1=1/16,rho2=1
  run_dsen2<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen2)
  PSensitive<-rbind(PSensitive, c(parmsfsen2[3],run_dsen2[length(run_dsen2
}
plot(PSensitive[,1],PSensitive[,2],xlab = "P",ylab = "Infections")
#Infections decrease as proportion of A infections increase
bsp<-as.data.frame(PSensitive)
colnames(bsp)<-c("Parm","Infections")
Parameter<-rep("P",times=100)
bspf<-cbind(bsp,Parameter)


Rho1Sensitive<-NULL
for (i in 1:100){
  parmsfsen3 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/runif(1,14,18),rho2=1
  run_dsen3<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen3)
  Rho1Sensitive<-rbind(Rho1Sensitive, c(parmsfsen3[4],run_dsen3[length(run
}
plot(Rho1Sensitive[,1],Rho1Sensitive[,2],xlab = "Rho1",ylab = "Infections"
#Infections decrease as number of days with A infection increase
```

```
bsr1<-as.data.frame(Rho1Sensitive)
colnames(bsr1)<-c("Parm","Infections")
Parameter<-rep("Rho1",times=100)
bsr1f<-cbind(bsr1,Parameter)


Rho2Sensitive<-NULL
for (i in 1:100){
  parmsfsen4 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/runif(1,7,1
  run_dsen4<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen4)
  Rho2Sensitive<-rbind(Rho2Sensitive, c(parmsfsen4[5],run_dsen4[length(run
}
plot(Rho2Sensitive[,1],Rho2Sensitive[,2],xlab = "Rho2",ylab = "Infections"
#Infections decrease as number of days with symptomatic infection increase
bsr2<-as.data.frame(Rho2Sensitive)
colnames(bsr2)<-c("Parm","Infections")
Parameter<-rep("Rho2",times=100)
bsr2f<-cbind(bsr2,Parameter)


Rho3Sensitive<-NULL
for (i in 1:100){
  parmsfsen5 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/ru
  run_dsen5<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen5)
  Rho3Sensitive<-rbind(Rho3Sensitive, c(parmsfsen5[6],run_dsen5[length(run
}
plot(Rho3Sensitive[,1],Rho3Sensitive[,2],xlab = "Rho3",ylab = "Infections"
#Infections decrease as no.of days in hospital increase
bsr3<-as.data.frame(Rho3Sensitive)
colnames(bsr3)<-c("Parm","Infections")
Parameter<-rep("Rho3",times=100)
bsr3f<-cbind(bsr3,Parameter)


Rho4Sensitive<-NULL
for (i in 1:100){
  parmsfsen6 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen6<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen6)
  Rho4Sensitive<-rbind(Rho4Sensitive, c(parmsfsen6[7],run_dsen6[length(run
}
```

```
plot(Rho4Sensitive[,1],Rho4Sensitive[,2],xlab = "Rho4",ylab = "Infections"
#Infections increase as no. of days in ICU increase (no trend really)
bsr4<-as.data.frame(Rho4Sensitive)
colnames(bsr4)<-c("Parm","Infections")
Parameter<-rep("Rho4",times=100)
bsr4f<-cbind(bsr4,Parameter)


Alpha1Sensitive<-NULL
for (i in 1:100){
  parmsfsen7 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen7 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen7)
  Alpha1Sensitive<-rbind(Alpha1Sensitive, c(parmsfsen7[8],run_dsen7[length
}
plot(Alpha1Sensitive[,1],Alpha1Sensitive[,2],xlab = "Alpha1",ylab = "Infec
#Infections increase as the number of days before hospitalisation is requi
bsa1<-as.data.frame(Alpha1Sensitive)
colnames(bsa1)<-c("Parm","Infections")
Parameter<-rep("Alpha1",times=100)
bsa1f<-cbind(bsa1,Parameter)


Alpha2Sensitive<-NULL
for (i in 1:100){
  parmsfsen8 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen8 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen8)
  Alpha2Sensitive<-rbind(Alpha2Sensitive, c(parmsfsen8[9],run_dsen8[length
}
plot(Alpha2Sensitive[,1],Alpha2Sensitive[,2],xlab = "Alpha2",ylab = "Infec
#Infections increase as time before ICU admission increases
bsa2<-as.data.frame(Alpha2Sensitive)
colnames(bsa2)<-c("Parm","Infections")
Parameter<-rep("Alpha2",times=100)
bsa2f<-cbind(bsa2,Parameter)


MuSensitive<-NULL
for (i in 1:100){
  parmsfsen11 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5
  run_dsen11 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen11)
```

```
  MuSensitive<-rbind(MuSensitive, c(parmsfsen11[10],run_dsen11[length(run_d
}
plot(MuSensitive[,1],MuSensitive[,2],xlab = "Mu",ylab = "Infections")
#Infections decrease as death rate increases
bsmu<-as.data.frame(MuSensitive)
colnames(bsmu)<-c("Parm","Infections")
Parameter<-rep("Mu",times=100)
bsmuf<-cbind(bsmu,Parameter)


#p,rho1,rho2,rho3,rho4,alpha1,alpha2,alpha3,alpha4,mu
letstry<-rbind(bsf,bsgf,bspf,bsr1f,bsr2f,bsr3f,bsa1f,bsa2f,bsmuf)
top3<-rbind(bsmuf,bsgf,bsr1f) #Rho4,alpha3,mu,gamma
head(letstry)
violintry<-ggplot(top3,aes(x=Parameter,y=Infections,color=Parameter))+geom
violintry +theme(text = element_text(size =15))+ theme(axis.title = elemen
#Top 4 most sensitive parameters were rho4,mu,gamma and rho1
#i.e. No. of days in ICU, death rate, incubation period and number of days
# an A infection



### DEATHS ###
library(deSolve)
N=59000000
BetaSensitive<-NULL
for (i in 1:100){
  parmsfsen <- c(beta=runif(1,0.2, 1),gamma=1/14, p=0.75,rho1=1/16,rho2=1/
  run_dsen<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen)
  BetaSensitive<-rbind(BetaSensitive, c(parmsfsen[1],run_dsen[length(run_ds
}
plot(BetaSensitive[,1],BetaSensitive[,2],xlab = "Beta",ylab = "Deaths")
#Deaths increase as transmission rate increases
bs<-as.data.frame(BetaSensitive)
colnames(bs)<-c("Parm","Deaths")
Parameter<-rep("Beta",times=100)
bsf<-cbind(bs,Parameter)


GammaSensitive<-NULL
```

```
for (i in 1:100){
  parmsfsen1 <- c(beta=0.5,gamma=1/runif(1,10,14), p=0.75,rho1=1/16,rho2=1
  run_dsen1<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen1)
  GammaSensitive<-rbind(GammaSensitive, c(parmsfsen1[2],run_dsen1[length(ru
}
plot(GammaSensitive[,1],GammaSensitive[,2],xlab = "Gamma",ylab = "Deaths")
#Deaths increase as incubation period increases
bsg<-as.data.frame(GammaSensitive)
colnames(bsg)<-c("Parm","Deaths")
Parameter<-rep("Gamma",times=100)
bsgf<-cbind(bsg,Parameter)


PSensitive<-NULL
for (i in 1:100){
  parmsfsen2 <- c(beta=0.5,gamma=1/14, p=runif(1,0.6,0.9),rho1=1/16,rho2=1
  run_dsen2<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen2)
  PSensitive<-rbind(PSensitive, c(parmsfsen2[3],run_dsen2[length(run_dsen2
}
plot(PSensitive[,1],PSensitive[,2],xlab = "P",ylab = "Deaths")
#Deaths decrease as proportion of A infections increase
#Shows that death rate is considerably lower amongst asymptomatic individu
bsp<-as.data.frame(PSensitive)
colnames(bsp)<-c("Parm","Deaths")
Parameter<-rep("P",times=100)
bspf<-cbind(bsp,Parameter)


Rho1Sensitive<-NULL
for (i in 1:100){
  parmsfsen3 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/runif(1,14,18),rho2=1
  run_dsen3<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen3)
  Rho1Sensitive<-rbind(Rho1Sensitive, c(parmsfsen3[4],run_dsen3[length(run
}
plot(Rho1Sensitive[,1],Rho1Sensitive[,2],xlab = "Rho1",ylab = "Deaths")
#Deaths decrease as number of days with A infection increase
bsr1<-as.data.frame(Rho1Sensitive)
colnames(bsr1)<-c("Parm","Deaths")
Parameter<-rep("Rho1",times=100)
bsr1f<-cbind(bsr1,Parameter)
```

```
Rho2Sensitive<-NULL
for (i in 1:100){
  parmsfsen4 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/runif(1,7,1
  run_dsen4<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen4)
  Rho2Sensitive<-rbind(Rho2Sensitive, c(parmsfsen4[5],run_dsen4[length(run
}
plot(Rho2Sensitive[,1],Rho2Sensitive[,2],xlab = "Rho2",ylab = "Deaths")
#Deaths decrease as number of days with symptomatic infection increase
#Fewer transitions into Is, I3 and I4
bsr2<-as.data.frame(Rho2Sensitive)
colnames(bsr2)<-c("Parm","Deaths")
Parameter<-rep("Rho2",times=100)
bsr2f<-cbind(bsr2,Parameter)


Rho3Sensitive<-NULL
for (i in 1:100){
  parmsfsen5 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/ru
  run_dsen5<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen5)
  Rho3Sensitive<-rbind(Rho3Sensitive, c(parmsfsen5[6],run_dsen5[length(run
}
plot(Rho3Sensitive[,1],Rho3Sensitive[,2],xlab = "Rho3",ylab = "Deaths")
#No trend
bsr3<-as.data.frame(Rho3Sensitive)
colnames(bsr3)<-c("Parm","Deaths")
Parameter<-rep("Rho3",times=100)
bsr3f<-cbind(bsr3,Parameter)


Rho4Sensitive<-NULL
for (i in 1:100){
  parmsfsen6 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen6<-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen6)
  Rho4Sensitive<-rbind(Rho4Sensitive, c(parmsfsen6[7],run_dsen6[length(run
}
plot(Rho4Sensitive[,1],Rho4Sensitive[,2],xlab = "Rho4",ylab = "Deaths")
#Deaths decrease as no. of days in ICU increase (no trend really)
bsr4<-as.data.frame(Rho4Sensitive)
```

```
colnames(bsr4)<-c("Parm","Deaths")
Parameter<-rep("Rho4",times=100)
bsr4f<-cbind(bsr4,Parameter)


Alpha1Sensitive<-NULL
for (i in 1:100){
  parmsfsen7 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen7 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen7)
  Alpha1Sensitive<-rbind(Alpha1Sensitive, c(parmsfsen7[8],run_dsen7[length
}
plot(Alpha1Sensitive[,1],Alpha1Sensitive[,2],xlab = "Alpha1",ylab = "Deaths
#Deaths decrease as time before hospitalisation required increases
bsa1<-as.data.frame(Alpha1Sensitive)
colnames(bsa1)<-c("Parm","Deaths")
Parameter<-rep("Alpha1",times=100)
bsa1f<-cbind(bsa1,Parameter)


Alpha2Sensitive<-NULL
for (i in 1:100){
  parmsfsen8 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,
  run_dsen8 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen8)
  Alpha2Sensitive<-rbind(Alpha2Sensitive, c(parmsfsen8[9],run_dsen8[length
}
plot(Alpha2Sensitive[,1],Alpha2Sensitive[,2],xlab = "Alpha2",ylab = "Deaths
#No trend for deaths and number of days before ICU admission
bsa2<-as.data.frame(Alpha2Sensitive)
colnames(bsa2)<-c("Parm","Deaths")
Parameter<-rep("Alpha2",times=100)
bsa2f<-cbind(bsa2,Parameter)


MuSensitive<-NULL
for (i in 1:100){
  parmsfsen11 <- c(beta=0.5,gamma=1/14, p=0.75,rho1=1/16,rho2=1/9,rho3=1/5
  run_dsen11 <-ode(times=timesf, y=startf, func=sirCov,parms=parmsfsen11)
  MuSensitive<-rbind(MuSensitive, c(parmsfsen11[10],run_dsen11[length(run_d
}
plot(MuSensitive[,1],MuSensitive[,2],xlab = "Mu",ylab = "Deaths")
```

```
#Deaths increase as death rate increases
bsmu<-as.data.frame(MuSensitive)
colnames(bsmu)<-c("Parm","Deaths")
Parameter<-rep("Mu",times=100)
bsmuf<-cbind(bsmu,Parameter)


#p,rho1,rho2,rho3,rho4,alpha1,alpha2,alpha3,alpha4,mu
library(ggplot2)
top3<-rbind(bsmuf,bsa2f,bsr3f,bsr4f) #Alpha3, gamma,rho1,rho2
head(letstry)
violintry<-ggplot(top3,aes(x=Parameter,y=Deaths,color=Parameter))+geom_vio
violintry+theme(text = element_text(size =15))+ theme(axis.title = element
#Top 4 most sensitive parameters were mu, rho3, alpha2 and rho4
#i.e. Death rate, number of days in hospital, number of days before ICU ad
#and number of days in ICU



#Multivariate sensitivity
#Hospitalisations
data<-NULL
for (i in 1:100){
  parmsmulti <- c(beta=runif(1,0.2, 1),gamma=1/runif(1,10,14), p=runif(1,0
  run_dmulti<-ode(times=timesf, y=startf, func=sirCov,parms=parmsmulti)
  data<-rbind(data, c(parmsmulti,run_dmulti[length(run_dmulti[,1]),10]/N))
}
head(data)
lm1<-lm(data[,11]~data[,1:10], data=as.data.frame(data))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#Most sensitive parameters are gamma, rho4, rho1, alpha1


#Total infections
data<-NULL
for (i in 1:100){
  parmsmulti <- c(beta=runif(1,0.2, 1),gamma=1/runif(1,10,14), p=runif(1,0
```

```
  run_dmulti<-ode(times=timesf, y=startf, func=sirCov,parms=parmsmulti)
  data<-rbind(data, c(parmsmulti,run_dmulti[length(run_dmulti[,1]),10]/N))
}
head(data)
lm1<-lm(data[,11]~data[,1:10], data=as.data.frame(data))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#rho1, gamma, mu, p


#Deaths
data<-NULL
for (i in 1:100){
  parmsmulti <- c(beta=runif(1,0.2, 1),gamma=1/runif(1,10,14), p=runif(1,0
  run_dmulti<-ode(times=timesf, y=startf, func=sirCov,parms=parmsmulti)
  data<-rbind(data, c(parmsmulti,run_dmulti[length(run_dmulti[,1]),10]/N))
}
head(data)
lm1<-lm(data[,11]~data[,1:10], data=as.data.frame(data))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#rho1,rho2,p,gamma


#CO-MORBIDITY STRATIFIED MODEL
#(Multivariate Analysis)
sirCoMorb <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2

    #Healthy equations
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I2+0.75*A2)/N2)-gamma*E1
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-rho2*I1-alpha1*I1
```

```
    dIHosp1=alpha1*I1-alpha2*IHosp1-alpha1*I1
    dIICU1=alpha2*IHosp1-rho4*IICU1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1
    dCumHosp1=alpha1*I1+alpha2*IHosp1
    dCumInf1=p*gamma*E1+(1-p)*gamma*E1+alpha1*I1+alpha2*IHosp1

    #Co-morbidity equations
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I2+0.75*A2)/N2)-gamma*E2
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-rho2*I2-alpha1*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-alpha1*I2
    dIICU2=alpha2*IHosp2-rho4*IICU2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2
    dCumHosp2=alpha1*I2+alpha2*IHosp2
    dCumInf2=p*gamma*E2+(1-p)*gamma*E2+alpha1*I2+alpha2*IHosp2


    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

#48% of South Africans have a co-morbidity therefore:
#N1=30 680 000
#N2=28 320 000

starts<-c(S1=30679950 ,E1=22 ,A1=11, I1=20 ,IHosp1=2 ,IICU1=2 ,R1=4 ,D1=0
#Start date was 10 March 2020
parmss<-c(beta11=0.5,beta12=0.5,beta21=0.25,beta22=0.25,gamma=1/14,p=0.75,
timess <- seq(0, 200, 1)
run_ds<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmss)
head(run_ds)
#D1->9, D2->17,CumHosp1->18, CumHosp2->19, CumInf1->20, CumInf2=21
population1<-run_ds[,2]+run_ds[,3]+run_ds[,4]+run_ds[,5]+run_ds[,6]+run_ds
population2<-run_ds[,10]+run_ds[,11]+run_ds[,12]+run_ds[,13]+run_ds[,14]+r
```

```
###HOSPITALISATIONS###
#Only doing a multivariate sensitivity analysis for models with matrix beta
N1=30680000
N2=28320000

#Healthy population
DataN1<-NULL
for (i in 1:100){
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,run_dsen[length(run_dsen[,1]),18]/N1))
}
head(DataN1)
lm1<-lm(DataN1[,15]~DataN1[,1:14], data=as.data.frame(DataN1))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#p,rho2,alpha1,gamma
#rho2,p,alpha1,rho1
#p, rho2,alpha1,rho1


#Co-morbidity population
DataN2<-NULL
for (i in 1:100){
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN2<-rbind(DataN2, c(parmsfsen,run_dsen[length(run_dsen[,1]),19]/N2))
}
head(DataN2)
lm1<-lm(DataN2[,15]~DataN2[,1:14], data=as.data.frame(DataN2))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#p,rho2,alpha1,rho1
#rho1,rho2,mu1,p
#rho2,p,alpha1,rho1
#rho2,p,alpha1,rho1
```

```
###TOTAL INFECTIONS###
DataN1<-NULL
for (i in 1:100){
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,run_dsen[length(run_dsen[,1]),20]/N1))
}
head(DataN1)
lm1<-lm(DataN1[,15]~DataN1[,1:14], data=as.data.frame(DataN1))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#p,rho2,alpha1,gamma
#p,rho2,alpha1,rho1
#p,rho2,alpha1,rho1

DataN2<-NULL
for (i in 1:100){
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN2<-rbind(DataN2, c(parmsfsen,run_dsen[length(run_dsen[,1]),21]/N2))
}
head(DataN2)
lm1<-lm(DataN2[,15]~DataN2[,1:14], data=as.data.frame(DataN2))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#rho1,rho2,p,alpha1
#p,rho2,alpha1,gamma
#gamma,p,rho2,rho1
#rho2,p,alpha1,rho1


###TOTAL DEATHS###
DataN1<-NULL
for (i in 1:100){
```

```
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,run_dsen[length(run_dsen[,1]),9]/N1))
}
head(DataN1)
lm1<-lm(DataN1[,15]~DataN1[,1:14], data=as.data.frame(DataN1))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#Mu1,rho4,alpha2,rho1
#mu1,rho4,rho1,alpha2
#mu1,rho4,rho1,alpha2


DataN2<-NULL
for (i in 1:100){
  parmsfsen <- c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta21=runi
  run_dsen<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmsfsen)
  DataN2<-rbind(DataN2, c(parmsfsen,run_dsen[length(run_dsen[,1]),17]/N2))
}
head(DataN2)
lm1<-lm(DataN2[,15]~DataN2[,1:14], data=as.data.frame(DataN2))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#Rho1,mu2,mu1,gamma
#mu2,alpha2,rho2,rho4
#mu2,rho1,rho4,alpha2
#mu2,gamma,rho1,alpha2, rho4



#AGE STRATIFIED MODEL
#Multivariate Analysis
sirAge <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
```

```
N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

#1-15
dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
dA1=p*gamma*E1-rho1*A1
dI1=(1-p)*gamma*E1-alpha1*I1-rho2*I1
dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
dIICU1=alpha2*IHosp1-rho4*IHosp1-mu1*IICU1
dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
dD1=mu1*IHosp1+mu1*IICU1
dCumHospAge1=alpha1*I1+alpha2*IHosp1
dCumInfAge1=p*gamma*E1+(1-p)*gamma*E1+alpha1*I1+alpha2*IHosp1

#16-30
dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
dA2=p*gamma*E2-rho1*A2
dI2=(1-p)*gamma*E2-alpha1*I2-rho2*I2
dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
dIICU2=alpha2*IHosp2-rho4*IHosp2-mu2*IICU2
dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
dD2=mu2*IHosp2+mu2*IICU2
dCumHospAge2=alpha1*I2+alpha2*IHosp2
dCumInfAge2=p*gamma*E2+(1-p)*gamma*E2+alpha1*I2+alpha2*IHosp2

#31-45
dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
dA3=p*gamma*E3-rho1*A3
dI3=(1-p)*gamma*E3-alpha1*I3-rho2*I3
dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
dIICU3=alpha2*IHosp3-rho4*IHosp3-mu3*IICU3
dR3=rho1*A3+rho2*I3+rho3*IHosp3+rho4*IICU3
dD3=mu3*IHosp3+mu3*IICU3
dCumHospAge3=alpha1*I3+alpha2*IHosp3
dCumInfAge3=p*gamma*E3+(1-p)*gamma*E3+alpha1*I3+alpha2*IHosp3
```

```
#46-60
dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
dA4=p*gamma*E4-rho1*A4
dI4=(1-p)*gamma*E4-alpha1*I4-rho2*I4
dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
dIICU4=alpha2*IHosp4-rho4*IHosp4-mu4*IICU4
dR4=rho1*A4+rho2*I4+rho3*IHosp4+rho4*IICU4
dD4=mu4*IHosp4+mu4*IICU4
dCumHospAge4=alpha1*I4+alpha2*IHosp4
dCumInfAge4=p*gamma*E4+(1-p)*gamma*E4+alpha1*I4+alpha2*IHosp4

#61-75
dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
dA5=p*gamma*E5-rho1*A5
dI5=(1-p)*gamma*E5-alpha1*I5-rho2*I5
dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
dIICU5=alpha2*IHosp5-rho4*IHosp5-mu5*IICU5
dR5=rho1*A5+rho2*I5+rho3*IHosp5+rho4*IICU5
dD5=mu4*IHosp5+mu5*IICU5
dCumHospAge5=alpha1*I5+alpha2*IHosp5
dCumInfAge5=p*gamma*E5+(1-p)*gamma*E5+alpha1*I5+alpha2*IHosp5

#76-90
dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
dA6=p*gamma*E6-rho1*A6
dI6=(1-p)*gamma*E6-alpha1*I6-rho2*I6
dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
dIICU6=alpha2*IHosp6-rho4*IHosp6-mu6*IICU6
dR6=rho1*A6+rho2*I6+rho3*IHosp6+rho4*IICU6
dD6=mu6*IHosp6+mu6*IICU6
dCumHospAge6=alpha1*I6+alpha2*IHosp6
dCumInfAge6=p*gamma*E6+(1-p)*gamma*E6+alpha1*I6+alpha2*IHosp6



output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
```

```
    list(output)
  })
}

#SA Population by age
#N1=16 579 000
#N2=17 110 000
#N3=11 741 000
#N4= 8 024 000
#N5= 3 953 000
#N6= 1 593 000


#METHOD 1
library(deSolve)
R0=2.75
gamma=4.46
beta=R0/gamma
starts<-c(S1=16578939 ,E1=40 ,A1=8 ,I1=11 ,IHosp1=0 ,IICU1=0 ,R1=6 ,D1=0 ,S
#Start date was 15 March 2020
parmss<-c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15=0.091
#Check if parameters are per day or per year
timess <- seq(0, 200, 1)
run_ds<-ode(times=timess, y=starts, func=sirAge,parms=parmss)
head(run_ds)
#D1->9, D2->17,D3->25,D4->33, D5->41, D6->49
#H1->50, H2->51,...,H6->55
#I1->56, I2->57,...,I6->61

###HOSPITALISATIONS###
#Only doing a multivariate sensitivity analysis for models with matrix beta

N=59000000
N1=16579000
N2=17110000
N3=11741000
N4=8024000
N5=3953000
N6=1593000
DataN1<-NULL
```

```
for (i in 1:100){
  parmsfsen <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15
  run_dsen<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,(run_dsen[length(run_dsen[,1]),55])/N6
}
head(DataN1)
lm1<-lm(DataN1[,51]~DataN1[,1:50], data=as.data.frame(DataN1))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#1-15: p,rho2,alpha1,mu1
#16-30: p,rho2,alpha1,rho1,mu2
#31-45: p,rho2,alpha1,rho3,mu3
#46-60: p,rho2,alpha1,rho1,rho3
#61-75: p,rho2,rho1,alpha1,mu5
#76-90: p,rho2,alpha1,rho1,rho3

#p (prop.of A infection) very sensitive acrosss all age strata (maybe it sl
#be age specific)
#Other sensitive parameters: rho2 (number of days with symptomatic infectio
#alpha1 (number of days before hospitalisation),
#rho1 (number of days with A infection),
#Age specific mu




###TOTAL INFECTIONS###
DataN1<-NULL
for (i in 1:100){
  parmsfsen <-c(beta11=runif(1,0.1, 1),beta12=runif(1,0.1, 1),beta13=runif
  run_dsen<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,(run_dsen[length(run_dsen[,1]),61])/N6
}
head(DataN1)
lm1<-lm(DataN1[,51]~DataN1[,1:50], data=as.data.frame(DataN1))
library(QuantPsyc)
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
```

```
stdcoeff
#1-15:  p,rho2,rho1,alpha1,mu4
#16-30: p,rho2,rho1,alpha1,mu2
#31-45: p,rho2,rho1,alpha1,rho3
#46-60: p,rho1,rho2,alpha1,mu4
#61-75: p,rho1,rho2,alpha1,rho3
#76-90: p,rho1,rho2,alpha1,rho3

#Most sensitive parameters in an infections context were p,rho1,rho2,alpha1
#p-proportion of asymptomatic
#rho1-Number of days with asymptomatic infection
#rho2-Number of days with symptomatic infection
#Sensitivity of rho1 increased with age with aligns with p being sensitive


###DEATHS###
#D1->9, D2->17,D3->25,D4->33, D5->41, D6->49
DataN1<-NULL
for (i in 1:100){
  parmsfsen <-c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15=
  run_dsen<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsen)
  DataN1<-rbind(DataN1, c(parmsfsen,(run_dsen[length(run_dsen[,1]),49])/N6
}
head(DataN1)
lm1<-lm(DataN1[,51]~DataN1[,1:50], data=as.data.frame(DataN1))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#1-15:  mu1,rho1,rho4,alpha1,p
#16-30: mu2,p,alpha1,rho2,rho4
#31-45: mu3,rho4,rho2,alpha2,rho1
#46-60: mu4,rho2,rho1,rho4,p
#61-75: mu5,rho2,rho4.p,mu4
#76-90: mu6,rho4,rho1,rho2,p



###TESTING SENSITIVITY OF BETA/R0###
N1=16579000
N2=17110000
```

```
N3=11741000
N4=8024000
N5=3953000
N6=1593000

#HOSPITALISATIONS (Driven by the 61-75 age group - largest population group
DataBet11<-NULL
for (i in 1:100){
  parmsfsenb1 <- c(beta11=runif(1,0.1,1),beta12=runif(1,0.1,1),beta13=runi
  run_dsenb1<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb1)
  DataBet11<-rbind(DataBet11, c(parmsfsenb1,(run_dsenb1[length(run_dsenb1[
}
head(DataBet11)
lm1<-lm(DataBet11[,51]~DataBet11[,1:50], data=as.data.frame(DataBet11))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b11=0.037, b12=0.778, b13=0.010, b14=0.012, b15=0.006, b16=0.006
#ave=0.1415


DataBet22<-NULL
for (i in 1:100){
  parmsfsenb2 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb2<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb2)
  DataBet22<-rbind(DataBet22, c(parmsfsenb2,(run_dsenb2[length(run_dsenb2[
}
head(DataBet22)
lm1<-lm(DataBet22[,51]~DataBet22[,1:50], data=as.data.frame(DataBet22))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b21=0.021, b22=0.833, b23=0.019, b24=0.015, b25=0.012, b26=0.007
#ave=0.1512


DataBet33<-NULL
for (i in 1:100){
  parmsfsenb3 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb3<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb3)
```

```
  DataBet33<-rbind(DataBet33, c(parmsfsenb3,(run_dsenb3[length(run_dsenb3[
}
head(DataBet33)
lm1<-lm(DataBet33[,51]~DataBet33[,1:50], data=as.data.frame(DataBet33))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b31=0.003, b32=0.295, b33=0.301, b34=0.009, b35=0.066, b36=0.107
#ave=0.1302


DataBet44<-NULL
for (i in 1:100){
  parmsfsenb4 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb4<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb4)
  DataBet44<-rbind(DataBet44, c(parmsfsenb4,(run_dsenb4[length(run_dsenb4[
}
head(DataBet44)
lm1<-lm(DataBet44[,51]~DataBet44[,1:50], data=as.data.frame(DataBet44))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b41=0.178, b42=0.298, b43=0.207, b44=0.037, b45=0.001, b46=0.035
#ave=0.126


DataBet55<-NULL
for (i in 1:100){
  parmsfsenb5 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb5<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb5)
  DataBet55<-rbind(DataBet55, c(parmsfsenb5,(run_dsenb5[length(run_dsenb5[
}
head(DataBet55)
lm1<-lm(DataBet55[,51]~DataBet55[,1:50], data=as.data.frame(DataBet55))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b51=0.051, b52=0.694, b53=0.004, b54=0.051, b55=0.052, b56=0.091
#ave=0.1572
```

```
DataBet66<-NULL
for (i in 1:100){
  parmsfsenb6 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb6<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb6)
  DataBet66<-rbind(DataBet66, c(parmsfsenb6,(run_dsenb6[length(run_dsenb6[
}
head(DataBet66)
lm1<-lm(DataBet66[,51]~DataBet66[,1:50], data=as.data.frame(DataBet66))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b61=0.042, b62=0.783, b63=0.013, b64=0.017, b65=0.019, b66=0.017
#ave=0.1485



###DEATHS###
#Deaths driven by 76-90 age group.
DataBet11<-NULL
for (i in 1:100){
  parmsfsenb1 <- c(beta11=runif(1,0.1,1),beta12=runif(1,0.1,1),beta13=runi
  run_dsenb1<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb1)
  DataBet11<-rbind(DataBet11, c(parmsfsenb1,(run_dsenb1[length(run_dsenb1[
}
head(DataBet11)
lm1<-lm(DataBet11[,51]~DataBet11[,1:50], data=as.data.frame(DataBet11))
lm2<-lm.beta(lm1)
summary(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b11=0.414, b12=0.512, b13=0.170, b14=0.146, b15=0.065, b16=0.097
#ave=0.2340


DataBet22<-NULL
for (i in 1:100){
  parmsfsenb2 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb2<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb2)
  DataBet22<-rbind(DataBet22, c(parmsfsenb2,(run_dsenb2[length(run_dsenb2[
```

```
}
head(DataBet22)
lm1<-lm(DataBet22[,51]~DataBet22[,1:50], data=as.data.frame(DataBet22))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b21=0.326, b22=0.249, b23=0.424, b24=0.296, b25=0.220, b26=0.162
#ave=0.2795


DataBet33<-NULL
for (i in 1:100){
  parmsfsenb3 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb3<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb3)
  DataBet33<-rbind(DataBet33, c(parmsfsenb3,(run_dsenb3[length(run_dsenb3[
}
head(DataBet33)
lm1<-lm(DataBet33[,51]~DataBet33[,1:50], data=as.data.frame(DataBet33))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b31=0.273, b32=0.255, b33=0.555, b34=0.140, b35=0.155, b36=0.159
#ave=0.2562


DataBet44<-NULL
for (i in 1:100){
  parmsfsenb4 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb4<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb4)
  DataBet44<-rbind(DataBet44, c(parmsfsenb4,(run_dsenb4[length(run_dsenb4[
}
head(DataBet44)
lm1<-lm(DataBet44[,51]~DataBet44[,1:50], data=as.data.frame(DataBet44))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b41=0.169, b42=0.163, b43=0.228, b44=0.660, b45=0.146, b46=0.106
#ave=0.2453
```

```
DataBet55<-NULL
for (i in 1:100){
  parmsfsenb5 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb5<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb5)
  DataBet55<-rbind(DataBet55, c(parmsfsenb5,(run_dsenb5[length(run_dsenb5[
}
head(DataBet55)
lm1<-lm(DataBet55[,51]~DataBet55[,1:50], data=as.data.frame(DataBet55))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b51=0.193, b52=0.281, b53=0.163, b54=0.165, b55=0.660, b56=0.099
#ave=0.2602


DataBet66<-NULL
for (i in 1:100){
  parmsfsenb6 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb6<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb6)
  DataBet66<-rbind(DataBet66, c(parmsfsenb6,(run_dsenb6[length(run_dsenb6[
}
head(DataBet66)
lm1<-lm(DataBet66[,51]~DataBet66[,1:50], data=as.data.frame(DataBet66))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b61=0.181, b62=0.501, b63=0.216, b64=0.182, b65=0.115, b66=0.610
#ave=0.3008



###TOTAL INFECTIONS###
#Infections driven by 76-90 age group (1-15 also has a big impact - check v
DataBet11<-NULL
for (i in 1:100){
  parmsfsenb1 <- c(beta11=runif(1,0.1,1),beta12=runif(1,0.1,1),beta13=runi
  run_dsenb1<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb1)
  DataBet11<-rbind(DataBet11, c(parmsfsenb1,(run_dsenb1[length(run_dsenb1[
}
head(DataBet11)
```

```
lm1<-lm(DataBet11[,51]~DataBet11[,1:50], data=as.data.frame(DataBet11))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b11=0.061, b12=0.739, b13=0.025, b14=0.030, b15=0.031, b16=0.003
#ave=0.1482


DataBet22<-NULL
for (i in 1:100){
  parmsfsenb2 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb2<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb2)
  DataBet22<-rbind(DataBet22, c(parmsfsenb2,(run_dsenb2[length(run_dsenb2[
}
head(DataBet22)
lm1<-lm(DataBet22[,51]~DataBet22[,1:50], data=as.data.frame(DataBet22))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b21=0.017, b22=0.770, b23=0.015, b24=0.014, b25=0.009, b26=0.004
#ave=0.1382


DataBet33<-NULL
for (i in 1:100){
  parmsfsenb3 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb3<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb3)
  DataBet33<-rbind(DataBet33, c(parmsfsenb3,(run_dsenb3[length(run_dsenb3[
}
head(DataBet33)
lm1<-lm(DataBet33[,51]~DataBet33[,1:50], data=as.data.frame(DataBet33))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b31=0.081, b32=0.331, b33=0.109, b34=0.151, b35=0.107, b36=0.047
#ave=0.1377


DataBet44<-NULL
for (i in 1:100){
```

```
  parmsfsenb4 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb4<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb4)
  DataBet44<-rbind(DataBet44, c(parmsfsenb4,(run_dsenb4[length(run_dsenb4[
}
head(DataBet44)
lm1<-lm(DataBet44[,51]~DataBet44[,1:50], data=as.data.frame(DataBet44))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b41=0.221, b42=0.104, b43=0.170, b44=0.103, b45=0.106, b46=0.086
#ave=0.1317


DataBet55<-NULL
for (i in 1:100){
  parmsfsenb5 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb5<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb5)
  DataBet55<-rbind(DataBet55, c(parmsfsenb5,(run_dsenb5[length(run_dsenb5[
}
head(DataBet55)
lm1<-lm(DataBet55[,51]~DataBet55[,1:50], data=as.data.frame(DataBet55))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b51=0.011, b52=0.708, b53=0.011, b54=0.035, b55=0.109, b56=0.027
#ave=0.1502


DataBet66<-NULL
for (i in 1:100){
  parmsfsenb6 <- c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta
  run_dsenb6<-ode(times=timess, y=starts, func=sirAge,parms=parmsfsenb6)
  DataBet66<-rbind(DataBet66, c(parmsfsenb6,(run_dsenb6[length(run_dsenb6[
}
head(DataBet66)
lm1<-lm(DataBet66[,51]~DataBet66[,1:50], data=as.data.frame(DataBet66))
lm2<-lm.beta(lm1)
stdcoeff<-cbind(sort(abs(lm2),decreasing = TRUE))
stdcoeff
#b61=0.017, b62=0.731, b63=0.001, b64=0.024, b65=0.023, b66=0.109
```

```
#ave=0.1508



#RE-initialising:

sirCov <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N=S+E+A+I+IHosp+IICU+R+D
    dS=-(beta*(I+0.75*A)/N)*S
    dE=(beta*(I+0.75*A)/N)*S-gamma*E
    dA=p*gamma*E-rho1*A
    dI=(1-p)*gamma*E-rho2*I-alpha1*I
    dIHosp=alpha1*I-alpha2*IHosp-rho3*IHosp-mu*IHosp
    dIICU=alpha2*IHosp-rho4*IICU-mu*IICU
    dR=rho1*A+rho2*I+rho3*IHosp+rho4*IICU
    dD=mu*IHosp+mu*IICU
    output <- c(dS,dE,dA,dI,dIHosp,dIICU,dR,dD)
    list(output)
  })
}

startf<-c(S=58262285, E=254824, A=191118, I=63706, IHosp=12741, IICU=6371,
#1 September 2020
parmsf<-c(beta=0.35,gamma=1/14,p=0.75,rho1=1/16,rho2=1/9,rho3=1/5,rho4=1/9
#Beta=0.55
timesf <- seq(0, 211, 1)
run_df<-ode(times=timesf, y=startf, func=sirCov,parms=parmsf)
BasicDF<-as.data.frame(run_df)
BasicInf<-BasicDF$I+BasicDF$IHosp+BasicDF$IICU

library(deSolve)
### CO-MORBIDITY STRATIFIED MODEL ###
sirCoMorb <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2

    #Healthy equations
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)
```

```
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I2+0.75*A2)/N2)-gamma*E1
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-rho2*I1-alpha1*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-alpha1*I1
    dIICU1=alpha2*IHosp1-rho4*IICU1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #Co-morbidity equations
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I2+0.75*A2)/N2)-gamma*E2
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-rho2*I2-alpha1*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-alpha1*I2
    dIICU2=alpha2*IHosp2-rho4*IICU2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2


    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

#48% of South Africans have a co-morbidity therefore:
#N1=30 680 000
#N2=28 320 000

starts<-c(S1=30272539 ,E1=132508 ,A1=99381, I1=33127 ,IHosp1=6625 ,IICU1=3
#Start date was 10 March 2020
parmss<-c(beta11=0.2,beta12=0.2,beta21=0.1,beta22=0.1,gamma=1/14,p=0.75,rho
timess <- seq(0, 211, 1)
run_ds<-ode(times=timess, y=starts, func=sirCoMorb,parms=parmss)
CoDF<-as.data.frame(run_ds)
CoInf1<-CoDF$I1+CoDF$IHosp1+CoDF$IICU1
CoInf2<-CoDF$I2+CoDF$IHosp2+CoDF$IICU2
CoInfoverall<-CoInf1+CoInf2
```

```
sirAge <- function(t, x, parms)  {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
    N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
    N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
    N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
    N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

    #1-15
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
    dA1=p*gamma*E1-rho1*A1
    dI1=(1-p)*gamma*E1-alpha1*I1-rho2*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
    dIICU1=alpha2*IHosp1-rho4*IHosp1-mu1*IICU1
    dR1=rho1*A1+rho2*I1+rho3*IHosp1+rho4*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #16-30
    dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
    dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
    dA2=p*gamma*E2-rho1*A2
    dI2=(1-p)*gamma*E2-alpha1*I2-rho2*I2
    dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
    dIICU2=alpha2*IHosp2-rho4*IHosp2-mu2*IICU2
    dR2=rho1*A2+rho2*I2+rho3*IHosp2+rho4*IICU2
    dD2=mu2*IHosp2+mu2*IICU2

    #31-45
    dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
    dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
    dA3=p*gamma*E3-rho1*A3
    dI3=(1-p)*gamma*E3-alpha1*I3-rho2*I3
    dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
    dIICU3=alpha2*IHosp3-rho4*IHosp3-mu3*IICU3
    dR3=rho1*A3+rho2*I3+rho3*IHosp3+rho4*IICU3
    dD3=mu3*IHosp3+mu3*IICU3

    #46-60
```

```
    dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
    dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
    dA4=p*gamma*E4-rho1*A4
    dI4=(1-p)*gamma*E4-alpha1*I4-rho2*I4
    dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
    dIICU4=alpha2*IHosp4-rho4*IHosp4-mu4*IICU4
    dR4=rho1*A4+rho2*I4+rho3*IHosp4+rho4*IICU4
    dD4=mu4*IHosp4+mu4*IICU4

    #61-75
    dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
    dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
    dA5=p*gamma*E5-rho1*A5
    dI5=(1-p)*gamma*E5-alpha1*I5-rho2*I5
    dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
    dIICU5=alpha2*IHosp5-rho4*IHosp5-mu5*IICU5
    dR5=rho1*A5+rho2*I5+rho3*IHosp5+rho4*IICU5
    dD5=mu4*IHosp5+mu5*IICU5

    #76-90
    dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
    dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
    dA6=p*gamma*E6-rho1*A6
    dI6=(1-p)*gamma*E6-alpha1*I6-rho2*I6
    dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
    dIICU6=alpha2*IHosp6-rho4*IHosp6-mu6*IICU6
    dR6=rho1*A6+rho2*I6+rho3*IHosp6+rho4*IICU6
    dD6=mu6*IHosp6+mu6*IICU6



    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

starts<-c(S1=16358815 ,E1=71606 ,A1=53704 ,I1=17901 ,IHosp1=3580 ,IICU1=17
#Start date was 15 March 2020
parmss<-c(beta11=0.399,beta12=0.085,beta13=0.124,beta14=0.090,beta15=0.091
#Check if parameters are per day or per year
```

```
timess <- seq(0, 211, 1)
run_dsA<-ode(times=timess, y=starts, func=sirAge,parms=parmss)
ADF<-as.data.frame(run_dsA)
I1A<-ADF$I1+ADF$IHosp1+ADF$IICU1
I2A<-ADF$I2+ADF$IHosp2+ADF$IICU2
I3A<-ADF$I3+ADF$IHosp3+ADF$IICU3
I4A<-ADF$I4+ADF$IHosp4+ADF$IICU4
I5A<-ADF$I5+ADF$IHosp5+ADF$IICU5
I6A<-ADF$I6+ADF$IHosp6+ADF$IICU6
Ioverall<-I1A+I2A+I3A+I4A+I5A+I6A


###FITTING TO 2ND WAVE####
library(readxl)
Active2<- read_excel("Active Cases 2.xlsx")
View(Active2)

head(Active2)
Date2=Active2$Date
Inflated2=Active2$Inflated
Dataf2=data.frame(Date2,Inflated2,BasicInf,CoInf1,CoInf2,I1A,I2A,I3A,I4A,I5

head(Dataf2)
plot(TimeSeries2,lty=1,lwd=2,col="blue",main="Number of active COVID-19 ca
max(Inflated2)
library(ggplot2)
ggplot(Dataf2,aes(x=Date2))+
  geom_line(aes(y=Inflated2),color="magenta",size=1)+
  geom_line(aes(y=BasicInf),color="purple",size=1)+
  geom_line(aes(y=5*CoInfoverall),color="red",size=1)+
  geom_line(aes(y=Ioverall),color="limegreen",size=1)+
  ggtitle("Observed and estimated active COVID-19 cases during the 2nd wav
  xlab("Time")+
  ylab("Frequency")

#GOF
#X^2 GOF
#X^2=sum[(O-E)^2/E]

#Basic
```

```
Xsquare<-((Dataf2$Inflated2/59000000)-(Dataf2$BasicInf/59000000))^2/(Dataf
head(Xsquare)
sum(Xsquare) #4.067011
chisq.test(Dataf2$Inflated2,Dataf2$BasicInf)
#p-value=0.2394>0.05 therefore observed infections are not significantly d
#from expected (modelled) infections using the basic model.

#Co-morbidity
Xsquare2<-((Dataf2$Inflated2/59000000)-(CoInfoverall/59000000))^2/(Dataf2$
sum(Xsquare2) #0.4123091
chisq.test(Dataf2$Inflated2,CoInfoverall)

#Age
Xsquare2<-((Dataf2$Inflated2/59000000)-(Ioverall/59000000))^2/(Dataf2$Infla
sum(Xsquare2) #25.48048
chisq.test(Dataf2$Inflated2,Ioverall)



### AGE MODEL WITH RE-ESTIMATED PARAMETERS ###
sirAgeRE <- function(t, x, parms) {
  with(as.list(c(parms, x)), {
    N1=S1+E1+A1+I1+IHosp1+IICU1+R1+D1
    N2=S2+E2+A2+I2+IHosp2+IICU2+R2+D2
    N3=S3+E3+A3+I3+IHosp3+IICU3+R3+D3
    N4=S4+E4+A4+I4+IHosp4+IICU4+R4+D4
    N5=S5+E5+A5+I5+IHosp5+IICU5+R5+D5
    N6=S6+E6+A6+I6+IHosp6+IICU6+R6+D6

    #1-15
    dS1=-beta11*S1*((I1+0.75*A1)/N1)-beta12*S1*((I2+0.75*A2)/N2)-beta13*S1
    dE1=beta11*S1*((I1+0.75*A1)/N1)+beta12*S1*((I1+0.75*A2)/N2)+beta13*S1*
    dA1=p1*gamma*E1-rho1*A1
    dI1=(1-p1)*gamma*E1-alpha1*I1-rho21*I1
    dIHosp1=alpha1*I1-alpha2*IHosp1-rho3*IHosp1-mu1*IHosp1
    dIICU1=alpha2*IHosp1-rho41*IHosp1-mu1*IICU1
    dR1=rho1*A1+rho21*I1+rho3*IHosp1+rho41*IICU1
    dD1=mu1*IHosp1+mu1*IICU1

    #16-30
```

```
dS2=-beta21*S2*((I1+0.75*A1)/N1)-beta22*S2*((I2+0.75*A2)/N2)-beta23*S2
dE2=beta21*S2*((I1+0.75*A1)/N1)+beta22*S2*((I1+0.75*A2)/N2)+beta23*S2*
dA2=p2*gamma*E2-rho1*A2
dI2=(1-p2)*gamma*E2-alpha1*I2-rho22*I2
dIHosp2=alpha1*I2-alpha2*IHosp2-rho3*IHosp2-mu2*IHosp2
dIICU2=alpha2*IHosp2-rho42*IHosp2-mu2*IICU2
dR2=rho1*A2+rho22*I2+rho3*IHosp2+rho42*IICU2
dD2=mu2*IHosp2+mu2*IICU2

#31-45
dS3=-beta31*S3*((I1+0.75*A1)/N1)-beta32*S3*((I2+0.75*A2)/N2)-beta33*S3
dE3=beta31*S3*((I1+0.75*A1)/N1)+beta32*S3*((I1+0.75*A2)/N2)+beta33*S3*
dA3=p3*gamma*E3-rho1*A3
dI3=(1-p3)*gamma*E3-alpha1*I3-rho23*I3
dIHosp3=alpha1*I3-alpha2*IHosp3-rho3*IHosp2-mu3*IHosp3
dIICU3=alpha2*IHosp3-rho43*IHosp3-mu3*IICU3
dR3=rho1*A3+rho23*I3+rho3*IHosp3+rho43*IICU3
dD3=mu3*IHosp3+mu3*IICU3

#46-60
dS4=-beta41*S4*((I1+0.75*A1)/N1)-beta42*S4*((I2+0.75*A2)/N2)-beta43*S4
dE4=beta41*S4*((I1+0.75*A1)/N1)+beta42*S4*((I1+0.75*A2)/N2)+beta43*S4*
dA4=p4*gamma*E4-rho1*A4
dI4=(1-p4)*gamma*E4-alpha1*I4-rho24*I4
dIHosp4=alpha1*I4-alpha2*IHosp4-rho3*IHosp4-mu4*IHosp4
dIICU4=alpha2*IHosp4-rho44*IHosp4-mu4*IICU4
dR4=rho1*A4+rho24*I4+rho3*IHosp4+rho44*IICU4
dD4=mu4*IHosp4+mu4*IICU4

#61-75
dS5=-beta51*S5*((I1+0.75*A1)/N1)-beta52*S5*((I2+0.75*A2)/N2)-beta53*S5
dE5=beta51*S5*((I1+0.75*A1)/N1)+beta52*S5*((I1+0.75*A2)/N2)+beta53*S5*
dA5=p5*gamma*E5-rho1*A5
dI5=(1-p5)*gamma*E5-alpha1*I5-rho25*I5
dIHosp5=alpha1*I5-alpha2*IHosp5-rho3*IHosp5-mu5*IHosp5
dIICU5=alpha2*IHosp5-rho45*IHosp5-mu5*IICU5
dR5=rho1*A5+rho25*I5+rho3*IHosp5+rho45*IICU5
dD5=mu4*IHosp5+mu5*IICU5

#76-90
```

```
    dS6=-beta61*S6*((I1+0.75*A1)/N1)-beta62*S6*((I2+0.75*A2)/N2)-beta63*S6
    dE6=beta61*S6*((I1+0.75*A1)/N1)+beta62*S6*((I1+0.75*A2)/N2)+beta63*S6*
    dA6=p6*gamma*E6-rho1*A6
    dI6=(1-p6)*gamma*E6-alpha1*I6-rho26*I6
    dIHosp6=alpha1*I6-alpha2*IHosp6-rho3*IHosp6-mu6*IHosp6
    dIICU6=alpha2*IHosp6-rho46*IHosp6-mu6*IICU6
    dR6=rho1*A6+rho26*I6+rho3*IHosp6+rho46*IICU6
    dD6=mu6*IHosp6+mu6*IICU6



    output <- c(dS1,dE1,dA1,dI1,dIHosp1,dIICU1,dR1,dD1,dS2,dE2,dA2,dI2,dIH
    list(output)
  })
}

starts<-c(S1=16358815 ,E1=71606 ,A1=53704 ,I1=17901 ,IHosp1=3580 ,IICU1=17
#Start date was 15 March 2020
parmss<-c(beta11=0.191,beta12=0.212,beta13=0.234,beta14=0.310,beta15=0.332
#Check if parameters are per day or per year
timess <- seq(0, 211, 1)
run_dsARe<-ode(times=timess, y=starts, func=sirAgeRE,parms=parmss)
ADF<-as.data.frame(run_dsARe)
I1A<-ADF$I1+ADF$IHosp1+ADF$IICU1
I2A<-ADF$I2+ADF$IHosp2+ADF$IICU2
I3A<-ADF$I3+ADF$IHosp3+ADF$IICU3
I4A<-ADF$I4+ADF$IHosp4+ADF$IICU4
I5A<-ADF$I5+ADF$IHosp5+ADF$IICU5
I6A<-ADF$I6+ADF$IHosp6+ADF$IICU6
Ioverall2<-I1A+I2A+I3A+I4A+I5A+I6A


#Age
Xsquare2<-((Dataf2$Inflated2/59000000)-(Ioverall2/59000000))^2/(Dataf2$Inf
sum(Xsquare2) #12.1735
chisq.test(Dataf2$Inflated2,Ioverall)

ggplot(Dataf2,aes(x=Date2))+
  geom_line(aes(y=Inflated2),color="magenta",size=1)+
  geom_line(aes(y=BasicInf),color="purple",size=1)+
```

```
geom_line(aes(y=5*CoInfoverall),color="red",size=1)+
geom_line(aes(y=Ioverall2/2),color="limegreen",size=1)+
ggtitle("Observed and estimated active COVID-19 cases during the 2nd wav
xlab("Time")+
ylab("Frequency")
```