2022

# COVID-19: A Review of Random Testing as an Augmentation of a Close-Contact Testing Regime in Response to the COVID-19 Pandemic

John McDonnell
*Technological University Dublin, Ireland*

# COVID-19: A review of random testing as an augmentation of a close-contact testing regime in response to the COVID-19 Pandemic


**John McDonnell**

A dissertation submitted in partial fulfilment of the requirements of
Technological University Dublin for the degree of
M.Sc. in Computing (Data Science)

**2nd March 2022**

# Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Science), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Technological University Dublin and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

***Signed: John McDonnell***

***Date: 2nd March 2022***

# Abstract

The COVID-19 pandemic has presented the world with a huge challenge to balance the health of the population with the scale of stringent response measures. Policy makers need evidence based recommendations to inform their response to the pandemic and the ever changing landscape it creates.

Testing and isolation of infected individuals plays a critical role in containing the spread of the disease, but is aimed primarily at detecting symptomatic individuals. However, asymptomatic individuals are much more difficult to detect. This study presents the augmentation of a close contact testing regime with random testing to examine the effect on the number of COVID-19 related deaths.

An extended Susceptible, Exposed, Infected, Removed (SEIR) model was built using the agent based modelling software, NetLogo. This model included several non-pharmaceutical interventions such as testing, tracing, public health measures, and level 5 lockdown. A specific outbreak in county Carlow between December 1$^{st}$ 2020 and February 28$^{th}$ 2021 was modeled. Population data based on the Ireland Census 2016 and COVID-19 statistics were used as inputs into the model and for model validation.

The results showed a statistically significant difference between scenarios where no random testing was performed (base model) and scenarios where 40% and 50% daily testing capacity was reserved for random testing. This result demonstrates that both testing regimes work best when used together, and not with one replacing the other.

**Keywords:** SEIR, Agent Based Model, Epidemiology, COVID-19, random testing

# Acknowledgments

I would like to thank my supervisor, Patricia O'Byrne for all of her guidance, support, and patience over the past 6 months. Her insight and honesty were invaluable to me, particularly in the early stages of the study but also throughout. My Friday mornings won't be the same without our check-ins!

I would also like to thank my wife Dee for her unwavering support, encouragement, guidance, and motivation to just keep going! She has felt like a single parent at times during this course, in the final weeks of each semester as well as the months leading up to the submission of this dissertation. Now, it's time for her to take a well-earned break! I would also like to thank my beautiful three daughters Lola, Lucy, and Ciara for their hugs and kisses, and many cards and posters they've created for me. Having them stuck to the walls around my desk has kept me going late at night when the rest of the house is asleep!

Finally, I would like to thank my extended family and friends for their support and encouragement over the past few years.

# Contents

**Declaration** **I**

**Abstract** **II**

**Acknowledgments** **III**

**Contents** **IV**

**List of Figures** **VIII**

**List of Tables** **IX**

**List of Acronyms** **X**

**1 Introduction** **1**

1.1 Background . . . . . 1

1.2 Research Project and Problem . . . . . 2

1.3 Research Objectives . . . . . 3

1.3.1 Design and build an Agent-based model . . . . . 4

1.3.2 Initialise the model . . . . . 12

1.3.3 Validate the model . . . . . 14

1.3.4 Experiment design . . . . . 14

1.4 Research Methodologies . . . . . 15

1.4.1 Evaluation of designed solution with statistical tests . . . . . 16

1.5 Scope and Limitations . . . . . 16

1.6 Document Outline . . . . . 16

**2 Review of existing literature** **18**

2.1 Background . . . . . 18

2.1.1 COVID-19 . . . . . 18

2.1.2 Government responses . . . . . 23

2.2 Approaches to solve the problem . . . . . 27

2.2.1 Epidemiology Modelling . . . . . 28

2.2.2 Equation based models . . . . . 29

2.2.3 Agent based models . . . . . 33

2.2.4 Network based models . . . . . 35

2.2.5 Swarm Intelligence . . . . . 36

2.3 Research Question . . . . . 37

2.3.1 Gaps in the research . . . . . 38

2.3.2 Research question . . . . . 38

2.3.3 Research hypothesis . . . . . 38

2.4 Summary . . . . . 39

**3 Experiment design and methodology** **40**

3.1 Design and Build an Agent Based Model . . . . . 40

3.1.1 Agent Based Modelling Software . . . . . 40

3.1.2 SEIR Model . . . . . 42

3.1.3 Modelling dynamics of society . . . . . 43

3.1.4 Modelling disease dynamics . . . . . 45

3.1.5 Modelling agent movement . . . . . 48

3.1.6 Modelling time . . . . . 53

3.1.7 Modelling government responses . . . . . 54

3.2 Data used to build the model . . . . . 57

3.2.1 General Data Protection Regulation (GDPR) . . . . . 57

3.2.2 Small Areas . . . . . 58

3.2.3 Household Data . . . . . 58

3.2.4 Population Data . . . 59
3.2.5 Build sample population . . . 60
3.2.6 COVID-19 Statistics . . . 62
3.2.7 Disease dynamics data . . . 62
3.3 Initialise the model . . . 63
3.3.1 Global variables . . . 63
3.3.2 Import small areas . . . 65
3.3.3 Import households . . . 65
3.3.4 Import people . . . 66
3.3.5 Create workplaces and hospitals . . . 66
3.3.6 Assign people to households . . . 67
3.3.7 Set initial number of infected individuals . . . 67
3.4 Validate the model . . . 68
3.4.1 Behaviour Space in NetLogo . . . 68
3.4.2 Selecting the number of runs . . . 68
3.4.3 Cross Validation . . . 69
3.4.4 Sensitivity Analysis . . . 70
3.4.5 Comparison to real world data . . . 72
3.5 Design the experiment . . . 73
3.5.1 Testing Capacity . . . 73
3.5.2 Experiment setup . . . 73
3.5.3 Evaluation of experiment results . . . 75

**4 Results, evaluation and discussion 78**

4.1 Results of Validation . . . 78
4.1.1 Validate the number of runs . . . 78
4.1.2 Cross validation with equation based model . . . 82
4.1.3 Sensitivity analysis . . . 85
4.1.4 Validation compared to real world data . . . 92
4.2 Experiment Results . . . 97

4.2.1 Testing the assumptions of a parametric test . . . . . 97
4.2.2 Test for difference in means . . . . . 98
4.3 Evaluation and discussion . . . . . 100

**5 Conclusion** **104**
5.1 Research Overview . . . . . 104
5.2 Problem Definition . . . . . 105
5.3 Design of Experimentation, Evaluation and Results . . . . . 106
5.4 Contributions and impact . . . . . 110
5.5 Future Work and recommendations . . . . . 111

**References** **113**

**A Code used in this study** **125**

**B Screenshots of NetLogo Model** **126**

# List of Figures


1.1 Agent activity cycle . . . . . 11
1.2 Epidemiological State . . . . . 15

3.1 Epidemiological Process . . . . . 47
3.2 Agent activity cycle . . . . . 51
3.3 Agent movement process at every tick . . . . . 52

4.1 Number of confirmed cases reaches at least 5% of the sample population 79
4.2 Number of confirmed cases reaches at least 5% of the sample population 81
4.3 Cross validation of Agent and Equation based models . . . . . 83
4.4 Cross validation of each compartment in Agent and Equation based models . . . . . 84
4.5 Sensitivity analysis of level 5 lockdown parameter . . . . . 86
4.6 Sensitivity analysis of testing and tracing functions . . . . . 89
4.7 Sensitivity analysis of transmissibility . . . . . 91
4.8 Comparison of simulated outbreak to real world outbreak . . . . . 93
4.9 Comparison of individually simulated outbreaks to real world . . . . . 96
4.10 Comparison of number of deaths between scenarios . . . . . 99
4.11 Trend of combined testing carried out during each scenario . . . . . 101

B.1 Simulation Environment . . . . . 127
B.2 Simulation Interface . . . . . 128
B.3 Epidemiology parameters in the simulation . . . . . 129
B.4 Behaviour space used to generate experiments . . . . . 130

# List of Tables


1.1 Person Agent attributes . . . . . 7
1.2 Household Agent attributes . . . . . 8
1.3 Agent and model parameters, $P$ . . . . . 10
1.4 Epidemiology parameter values by age . . . . . 10
1.5 Agent actions . . . . . 13
1.6 Model response variables . . . . . 15

3.1 Person agent attributes . . . . . 44
3.2 Household Agent attributes . . . . . 44
3.3 Epidemiology parameters and their sources . . . . . 46
3.4 Person agent attributes . . . . . 49
3.5 Variables used to count time or flag certain periods of time in NetLogo 53
3.6 Epidemiology parameter values by age . . . . . 63
3.7 Global variables initialised before each simulation run in NetLogo . . . 64
3.8 Parameters used in the experiment . . . . . 74
3.9 Response variables returned by the experiment . . . . . 75

4.1 Percentage of runs considered large enough for the simulation . . . . . 80
4.2 Percentage of runs with an outbreak long enough for the simulation . . 81
4.3 Correlations between agent based and equation based models . . . . . 82
4.4 Results of statistical tests used on experiment data . . . . . 94
4.5 Results of statistical tests used on experiment data . . . . . 98
4.6 Descriptive statistics from experiment . . . . . 100

# List of Acronyms

| | |
|---|---|
| **ABM** | Agent Based Model |
| **AI** | Artificial Intelligence |
| **ANOVA** | Analysis Of Variance |
| **CSO** | Central Statistics Office |
| **EMA** | European Medicines Agency |
| **EU** | European Union |
| **FDA** | Food and Drug Administration |
| **FPA** | Flower Pollination Algorithm |
| **GDPR** | General Data Protection Regulation |
| **ICU** | Intensive Care Unit |
| **NPHET** | National Public Health Emergency Team |
| **PCR** | Polymerase Chain Reaction |
| **PSO** | Particle Swarm Optimisation |
| **SAPS** | Small Area Population Statistics |
| **SEIR** | Susceptible Exposed Infected Removed |
| **SI** | Swarm Intelligence |
| **SNA** | Social Network Analysis |
| **WHO** | World Health Organisation |

# Chapter 1

# Introduction

## 1.1 Background

In December 2019, a newly identified coronavirus SARS-CoV-19 (COVID-19) was discovered in Wuhan, China, and has since then spread across most countries of the world, becoming a worldwide pandemic. As of August 26th 2021, the official figures from the WHO website show that the number of confirmed cases has reached 213,050,725 across 220 countries with 4,448,352 confirmed deaths associated with the disease[1].

Governments have implemented a range of non-pharmaceutical interventions in response to the pandemic and the effectiveness of these responses have varied across countries and territories. Such responses include social distancing, restriction of travel, closure of schools and business and isolation of confirmed cases (Hale et al., 2020). There is a need for policy makers to balance the scale of these and other responses in order to minimise the mortality rate while keeping disruption to the lives and livelihoods of the population to a minimum.

A vast array of statistical methods and simulation techniques have been used in epidemiology to model the dynamics and spread of infectious diseases including the effectiveness of interventions deployed to contain them. Policy makers can use these

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019

techniques to monitor and predict the prevalence of COVID-19, plan resources effectively, and control the spread of this infectious disease.

## 1.2 Research Project and Problem

Testing and isolation of infected individuals is vital to slow down the spread of the COVID-19 pandemic. During a pandemic, most of the reported cases will present to healthcare professionals with symptoms of the disease. However, asymptomatic individuals may also be a key factor in the spread of the disease, and secondary infections could have an impact on the more vulnerable members of society. Relying on the tracing of close contacts of confirmed cases may only identify a fraction of the number of asymptomatic cases and estimating the impact of these cases on the dynamics of epidemics is very difficult since the true number will never be known. Many countries have different levels of maturity in relation to testing and tracing of close contacts which also could have a significant impact on the level of control they can impose on the spread of the disease (Hale et al., 2020; Ivorra et al., 2020).

As of August 2021, many countries have already experienced multiple waves of the pandemic as cases and deaths rise and fall. With these rises, come government responses which place a heavy burden on society in terms of economic, social, and mental health aspects. Public health policies such as mask wearing, travel restrictions, and school and business closures, as well as vaccine roll-out have made a significant contribution to keeping the number of cases and deaths at a manageable level in some countries. However, there are still many countries whose policies have proven to be less effective and have suffered as a result (Silva et al., 2020; Murphy et al., 2021).
If testing could be improved, more asymptomatic infectious individuals could be detected, potentially impacting the number of COVID-19 related deaths. This may also reduce the severity of waves of cases and associated painful but necessary response policies.
As a result, the following question will be addressed in this research.

**Research Question:** Can the efficiency of a COVID-19 close-contact tracing regime in an Irish county be improved through the use of random testing of potentially asymptomatic infected individuals?

## 1.3 Research Objectives

The research question intends to examine whether introducing random testing to augment a close-contact tracing regime can detect and remove additional potentially asymptomatic infected individuals, resulting in a statistically significantly lower number of COVID-19 related deaths. The aim of this research is to examine whether random testing could play a significant role and form part of a strategy to reduce the need for stringent social isolation and travel restriction policies. This does not suggest that close-contact tracing is not effective.

Epidemiology models can be used to simulate the dynamics of infectious diseases and the interventions used to control them. An agent-based model uses a bottom-up approach where micro-level rules are applied to a heterogeneous population of agents. These agents share a closed simulated environment where they follow specific rules of behaviour. Agent based models can exhibit complex behaviour and provide valuable information at a population level which will be used to answer the research question described in this project (El-Sayed et al., 2012; Carpenter & Sattenspiel, 2009).

The following sub-sections will discuss the various elements of how an agent-based model will be built including the software, compartmental method, society attributes, population attributes, disease attributes, government responses, and the data sources used. In addition, sub-sections describe how the model will be initialised and validated, as well as how the experiment will be designed to test the hypothesis.

### 1.3.1 Design and build an Agent-based model

The first objective will be to design and build an Agent-based model on a modified version of the SEIR compartmental method.
The purpose of this agent based model will be to simulate the dynamics of a specific outbreak of COVID-19 between the dates December 1st 2020 and February 28th 2021. The location of the outbreak will be in a county in Ireland. To accurately represent this outbreak, the dynamics of the COVID-19 disease, Irish society, and the Irish Government response policies will be incorporated.

#### NetLogo

The model will be built using the multi-agent simulation environment and programming language, NetLogo. This easy-to-use environment can simulate complex natural and social phenomena and has been used for many years in the areas of research and education (Tisue & Wilensky, 2004). The NetLogo model will be built using agents, known as turtles, spatial units in the environment, known as patches, and functions and variables which can be programmed using the NetLogo language.

A unit of time in the NetLogo environment is called a tick. Each tick will represent one hour and time is determined as the modulus of the number of ticks and 24 (Hunter & Kelleher, 2021).

#### SEIR Model

A Susceptible Exposed Infected Removed (SEIR) model is a compartmental method used to simulate the effects of infectious diseases. Agents are separated into mutually exclusive groups or compartments and can therefore only reside in one compartment at any one time. The agents move between compartments depending on the parameters of the disease being modelled. These compartments are not physical spaces in the simulation environment but are listed as an attribute of an agent to indicate at which stage of infection an agent is at. Susceptible agents do not have the disease but can

become infected (i.e. they are not immune). Exposed agents have become infected but cannot yet pass the disease to another agent. Infected agents have the disease and can pass it on to other agents. Removed agents have either recovered from or have died as a result of the disease.

Epidemiological status is used as a parameter to track the compartment that an agent currently resides in and is used in many examples in the literature (Silva et al., 2020; Saravanan et al., 2013; Singh et al., 2018; El-Sayed et al., 2012; Hoertel et al., 2020). In the model initialisation process, each agent will be assigned a value of Susceptible or Infected in Section 1.3.2.

This model will also include a number of additional compartments in the standard SEIR model to account for dynamics specific to COVID-19 (Silva et al., 2020; Yang et al., 2020). The Infected compartment will be split into a number of separate compartments: Infected asymptomatic, Infected symptomatic, Infected severe, and Infected critical. These additional compartments represent the number of agents that are asymptomatic, symptomatic but have a mild infection, have been admitted to hospital, and have been admitted to an Intensive Care Unit (ICU) respectively. In addition, a compartment to track the number of dead agents will also be included. Therefore the model becomes $SEI^AI^MI^SI^CRD$. The number of agents in each compartments at the end of the simulation become the set of response variables and are listed in Table 1.6.

In order to accurately simulate the specific effects of COVID-19 on a population of agents, an age-structured table of probability values will be used to determine whether an agent becomes infected ($E$), becomes symptomatic ($I^A$, $I^M$), becomes hospitalised ($I^S$), becomes critically ill ($I^C$), dies or recovers ($D$, $R$). The probability values and associated parameter values are listed in Table 1.4 (Kerr et al., 2021; Zhang et al., 2020; Verity et al., 2020; Ferguson et al., 2020; Brazeau et al., 2020; O'Driscoll et al., 2021).

In addition to the probability of agents moving from one compartment to another,

several attribute values are assigned to each agent during the initialisation process to determine how many days each agent should remain in each compartment. These values are listed in Table 1.1 (Kerr et al., 2021; Lauer et al., 2020; Du et al., 2020; Linton et al., 2020; Wang et al., 2020; Chen et al., 2020; Verity et al., 2020).

**Elements of Society**

Each person agent in the model represents either a member of the population or a household and will have a number of attributes associated with them. Since the severity of COVID-19 varies depending on age, each agent will have an age attribute (Silva et al., 2020; Hunter et al., 2020; Barek et al., 2020). The Ireland Central Statistics Office (CSO) provides a detailed breakdown of the population age based on the latest Ireland Census data from 2016, which is broken down by county, city, and sex[7]. While this model will not differentiate the severity of COVID-19 based on gender, the sex of the agent is also included as an attribute since this will be used to assign agents to households.

Hunter et al. (2020) have shown how distributing agents according to real world data allows results to be more easily applied to real life scenarios and also offers reproducibility in cases where this level of data is openly available. Household data available from the Central Statistics Office (CSO) is based on the Ireland Census from 2016 and provides population distribution based on household size and age[8].

[2]https://data.cso.ie/table/E3003

[3]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[4]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[5]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[6]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[7]https://data.cso.ie/table/E3003

[8]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

| Person Agent attributes | | | |
|---|---|---|---|
| **Attribute** | **Description** | **Value** | **Source** |
| Age | Age of agent | 0 - 95 | CSO Ireland[2] |
| Epidemiological status | Status as set out in table 1.6 | All agents initialised as susceptible | |
| IsEssentialWorker | Is this agent an essential worker | Yes (20%) / No (80%) | (Redmond et al., 2020) |
| EconomicStatus | Economic status of the agent | Employed / unemployed / student / retired | CSO Ireland[3] |
| Days exposed | No of days the agent was infected but not infectious. | 4.5 mean, 1.5 std | (Lauer et al., 2020) (Du et al., 2020) |
| Days infectious to symptomatic | Number of days the agent was infectious | 1.1 mean, 0.9 std | (Linton et al., 2020) |
| Days severely ill | Number of days the agent was severely ill in hospital | 6.6 mean, 4.9 std | (Linton et al., 2020) (Wang et al., 2020) |
| Days critical | Number of days the agent was critically ill in ICU | 1.5 mean, 2.0 std | (Chen et al., 2020) (Wang et al., 2020) |
| Days critical to death | Number of days the agent was critically ill before death | 10.7 mean, 4.8 std | (Verity et al., 2020) |
| Days infectious to recovery mild | Number of days the agent was infectious before recovery (asymptomatic and mild cases) | 8.0 mean, 2.0 std | (Verity et al., 2020) |
| Days infectious to recovery severe | Number of days the agent was infectious before recovery (severe and critical cases) | 18.1 mean, 6.3 std | (Verity et al., 2020) |
| Home patch | Area where the agent lives | Patches are assigned households by household type and agents are assigned to households according to census household data | CSO Ireland[4] |
| Work patch | Area where the agent works or goes to school | Agents are assigned a patch not assigned as a household. | CSO Ireland[5] |

Table 1.1: Person Agent attributes

| Household agent attributes | | | |
|---|---|---|---|
| **Attribute** | **Description** | **Value** | **Source** |
| HouseholdType | Category of household | One person, married/co-habiting couple, married/co-habiting couple with children, one parent family with children, two or more non-related persons | CSO Ireland[6] |
| Location | Coordinates of household within patch | x, y | (Hunter et al., 2020) |

Table 1.2: Household Agent attributes

Hunter et al. (2020) and Hunter & Kelleher (2021) have proposed a method to create a society in Ireland that can represent a town or county. The CSO provides detailed population statistics across a number of so-called themes at different levels of geographic granularity. These statistics include the number of people per household type by age group, and the principle economic status of agents such as student, retired, working, unemployed etc. The lowest level of geographic granularity is called a small area and contains between 80 and 120 dwellings[9]. These small areas will be used as the basis to assign agents to households and since small areas can be aggregated up to county level, this aligns with statistics on COVID-19 cases also[10].
Redmond et al. (2020) describe essential workers as healthcare professionals, armed forces, retail sales, and transportation operatives. These essential workers will account for 20% of all workers and are flagged as such using an agent attribute in Table 1.1.

Each patch in the NetLogo environment will correspond to a small area from the Ireland CSO data, and household agents will be generated within the patch according to that data. Household agents will have a small number of attributes[11] such as location in the environment and household type as outlined in Table 1.2. Person agents will then be assigned to these households in the small area with the locations stored as an agent attribute called Home Patch (see Table 1.1). Household agents

[9]https://www.cso.ie/

[10]https://data.gov.ie/dataset/covid-19-hpsc-detailed-statistics-profile?package_type=dataset

[11]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

within a patch not assigned as a household will be assigned as a workplace or school instead.

### Government Responses

Government response policies to COVID-19 have evolved as the pandemic has progressed but there are a number of measures that have been implemented across many countries including physical distancing, business and school closures, travel restrictions, mask wearing, and personal hygiene advice. In order to respond to the spread of the disease, testing and tracing capabilities are also implemented by many countries, including Ireland.

To model measures such as mask wearing, physical distancing, and hygiene, a fractional reduction of 80% in transmissibility is applied to $\beta_1$ in Table 1.3, the probability of becoming infected by age (Chu et al., 2020). This fractional reduction is stored as parameter $\beta_9$ in Table 1.3.

The Ireland government testing and tracing regime involves the testing of symptomatic people using polymerase chain reaction (PCR) test and tracing people that were in close contact with a person that had tested positive for COVID-19. Mardani et al. (2020) have shown the accuracy of a PCR test ranging from 83.5% to 87.9%. However, Khatami et al. (2020) have leveraged 60 studies and have presented overall recall (sensitivity) as 87% which is used in this project as parameter $\beta_8$ (see Table 1.3).

As agents move around the environment, they will keep track of other agents they have been in close contact with. If an agent has tested positive for COVID-19, then their close contacts will also be tested.

### Agent Activity Cycle

The model will run through an iterative process where at each tick, every agent will follow the agent activity cycle laid out in Figure 1.1. The model design and iterative algorithm are based on a simplified version of the models produced by Silva et al.

[12]https://data.cso.ie/table/E3003

[13]https://data.cso.ie/table/E3003

| Demographic parameters | | |
|---|---|---|
| **Variable** | **Description** | **Source** |
| $\alpha_1$ | Population size | CSO Ireland[12] |
| $\alpha_2$ | Age distribution of agents | CSO Ireland[13] |
| Epidemiological parameters | | |
| Variable | Description | Source |
| $\beta_1$ | Probability of becoming infected by age | (Zhang et al., 2020) |
| $\beta_2$ | Probability of symptoms appearing by age | (Verity et al., 2020; Ferguson et al., 2020) |
| $\beta_3$ | Probability of developing severe symptoms by age | (Verity et al., 2020; Ferguson et al., 2020) |
| $\beta_4$ | Probability of developing into a critical case by age | Verity et al. (2020); Ferguson et al. (2020) |
| $\beta_5$ | Probability of death by age | (O'Driscoll et al., 2021; Brazeau et al., 2020) |
| $\beta_6$ | Number of initially infected agents | 1 |
| $\beta_7$ | Accuracy of PCR test (recall) | Recall 87% (Khatami et al., 2020) |
| $\beta_8$ | Random test probability | Applied as part of the experiment design |
| $\beta_9$ | Fractional reduction in becoming infected | 80% (Chu et al., 2020) |

Table 1.3: Agent and model parameters, $P$

| Epidemiology parameter values by age | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-9 | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 | 90+ |
| $\beta_1$ | 0.34 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.24 | 1.47 | 1.47 |
| $\beta_2$ | 0.5 | 0.55 | 0.6 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.90 |
| $\beta_3$ | 0.00050 | 0.00165 | 0.00720 | 0.02080 | 0.03430 | 0.07650 | 0.13280 | 0.20655 | 0.24570 | 0.24570 |
| $\beta_4$ | 0.00003 | 0.00008 | 0.00036 | 0.00104 | 0.00216 | 0.00933 | 0.03639 | 0.08923 | 0.17420 | 0.17420 |
| $\beta_5$ | 0.00002 | 0.00002 | 0.00010 | 0.00032 | 0.00098 | 0.00265 | 0.00766 | 0.02439 | 0.08292 | 0.16190 |

Table 1.4: Epidemiology parameter values by age

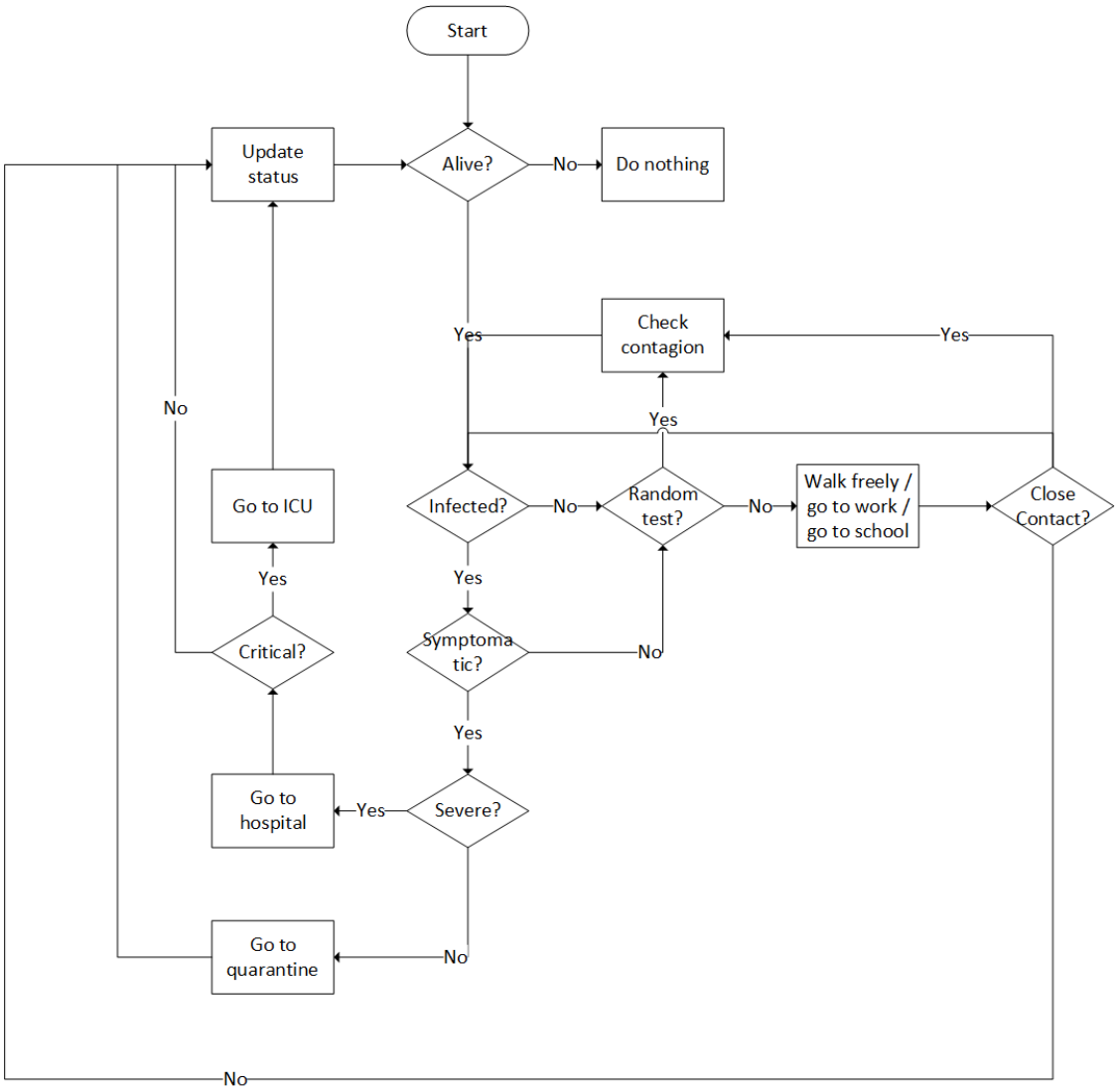



Figure 1.1: Agent activity cycle

(2020), Venkatramanan et al. (2018), and Hunter & Kelleher (2021).

At each iteration, every agent will check if they are alive. If so, then they will check if they have been infected. If the agent has been infected, then they will travel home to quarantine for 14 days if they have not already done so. A PCR test is assumed to be used to assess whether an agent is infected. The recall is 87% as shown in Table 1.3 (Khatami et al., 2020). If a PCR test is positive, then the agent stays in quarantine for 14 days. If the test is negative, then the agent returns to normal patterns of behaviour. If an agent has been infected, probabilities $\beta_3$, $\beta_4$, and $\beta_5$ are used to determine whether

the agent has a severe case ($I^S$), a critical case ($I^C$), and whether or not the agent dies or recovers ($D$ or $R$). The duration of time the agent spends in each of these compartments are determined by agent attributes as described in Table 1.1.
If the agent is not confirmed as infected, then they may be selected for a random test. The probability of being selected for a random test is determined by $\beta_8$ whose value varies according to the experiment scenarios. This random test is assumed to also be a PCR test and assumes the same recall accuracy of 87% as shown in Table 1.3 (Khatami et al., 2020). If this random test is negative, then they return to normal patterns of behaviour. If the random test is positive, then the probabilities and duration of time in each compartment are the same as outlined above.
During evenings and weekends, the agents will move randomly. During work/school hours, they will remain at work (assuming the agent is an essential worker) or school.

A predefined number of iterations are run, with each iteration representing one hour. The model follows the behaviour rules set out in the agent activity cycle for the simulated duration of 12 weeks and produces a set of response variables, $Y$, as outlined in Table 1.6.
If two agents occupy the same space in the environment, then a close contact is flagged and the agent may become infected according to $\beta_1$, the probability of becoming infected.
At the end of each tick, all response variables are updated. At the end of each simulation, the final values of the response variables are the results of a single run of the simulation.

### 1.3.2 Initialise the model

Before the simulation is run, the model must first be initialised. During this initialisation process, a number of components are built.
Firstly, the geographic area is populated. Each patch in the NetLogo environment will correspond to a small area. A representative sample of dwellings according to

| Agent actions | |
|---|---|
| Action | Description |
| Walk freely | Walk within the environment |
| Personal contact | Close contact with another agent |
| Go to hospital | For severe and critical cases |
| Go to quarantine | For less severe cases |
| Go to work | Go to work or school |

Table 1.5: Agent actions

the private households data from the CSO is generated in each small area[14]. Each dwelling is technically an agent on a patch but does not interact with person agents.

Next, person agents are generated according to the age profile statistics from Census 2016 and populated in the dwellings generated above[15]. As the agents are being created in the NetLogo environment, their agent attributes are also being generated including age, gender, as well as the epidemiology attributes in Table 1.1 These epidemiology attributes determine how long each agent will spend in each compartment. Whether an agent actually will spend time in some compartments at all will still be determined by the epidemiological parameters in Table 1.3.

Finally, the number of infected agents will also be initialised. Since on December 1st 2020, the pandemic will have been progressing for approximately 8 months, there will already be a sizeable amount of infected people in the community. A naive approach will be taken to estimate the number of infected agents. Kerr et al. (2021) have estimated that the number of days an agent will be infectious is 1.1. The 7 day average number of cases will be used to account for random variation in the data[16]. Since the population will be sampled, the percentage of confirmed cases will be used to account for this sampling. The infected agents will be selected at random.

[14]https://data.cso.ie/table/E3003

[15]https://data.cso.ie/table/E3003

[16]https://data.gov.ie/dataset/covid-19-hpsc-detailed-statistics-profile?package_type=dataset

### 1.3.3 Validate the model

Before accurate predictions can be made, the model will need to be validated. While no standard technique is employed for agent based modelling validation, the model can be validated using two methods.
Firstly, the agent based model will be validated to show that it behaves like an SEIR model by comparing its output to an equations based SEIR model using specified parameters (Silva et al., 2020). If the patterns and data match, this model can be shown to behave as expected.

The agent based model will also be validated against real world data[17] [18] by comparing the number and trend of confirmed cases against the number of confirmed cases in the model (Hunter et al., 2020).

### 1.3.4 Experiment design

A number of simulation scenarios will be executed to reflect different rates of random testing. Each scenario will simulate the impact of random testing at a rate of 0%, 10%, 20%, 30%, 40%, and 50%. The same testing capacity will be used for each scenario, with a different percentage being allocated to random testing. Each scenario will be simulated 300 times (Hunter & Kelleher, 2021) and will focus on the outbreak of cases between 1st December 2020 and 28th February 2021.

All other model parameters in Table 1.1 and Table 1.2 will be initialised at the same level for each scenario. These parameters will be set to simulate Level 2 of the response policy "Resilience and Recovery 2020-2021: Plan for Living with COVID-19"[19] in Ireland for the first 26 days. Level 2 of this policy assumes physical distancing is in place, and non-essential worker agents will not be traveling to work but are free to move around the environment. Since small events indoor and outdoor are permitted,

[17] https://data.gov.ie/dataset/covid-19-hpsc-detailed-statistics-profile?package_type=dataset
[18] https://covid-19.geohive.ie/
[19] https://www.gov.ie/en/campaigns/resilience-recovery-2020-2021-plan-for-living-with-covid-19/

| Response Variables, $Y$ | |
|---|---|
| **Variable** | **Description** |
| $S$ | Number of susceptible agents |
| $E$ | Number of exposed agents |
| $I^M$ | Number of infected symptomatic agents |
| $I^A$ | Number of infected asymptomatic agents |
| $I^S$ | Number of infected severe agents |
| $I^C$ | Number of infected critical agents |
| $R$ | Number of recovered agents |
| $D$ | Number of dead agents |

Table 1.6: Model response variables

house visits are permitted, and travel restrictions are not as stringent. From day 27 until the remainder of the simulation, the parameters will be set to simulate Level 5 of the government response policy. Level 5 assumes travel restrictions are in place, schools and non-essential workplaces are closed, and person agents movements around the environment will be reduced by 80% (Chu et al., 2020).

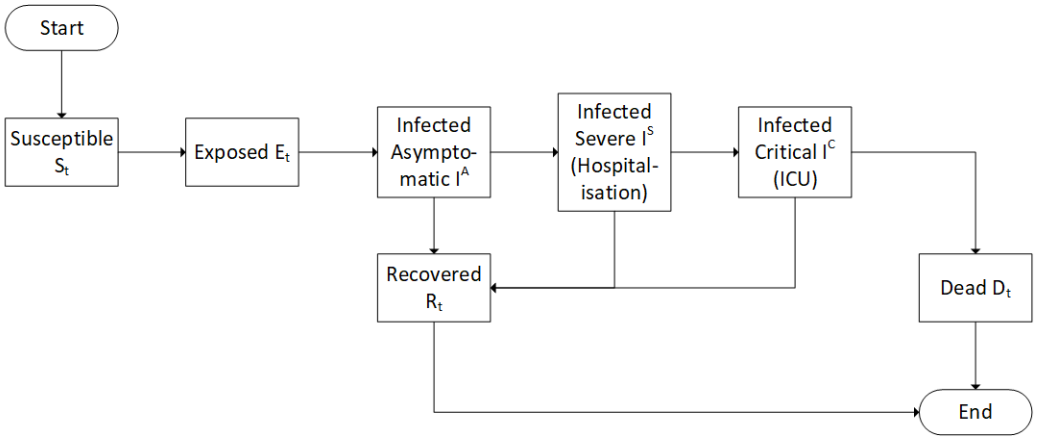



Figure 1.2: Epidemiological State

## 1.4 Research Methodologies

This research will be quantitative in order to identify any relationship between the detection and removal of asymptomatic agents infected with COVID-19 and the number of COVID related deaths.

### 1.4.1 Evaluation of designed solution with statistical tests

The output of each simulation will be a set of response variables as set out in Table 1.6. The response variable, the number of dead agents, $D$, will be returned at the end of each simulation. The results of the scenario with 0% random testing will correspond to close-contact tracing only and will be compared to each of the other scenarios in a two sample t-test for equal means. If the difference in means is statistically significant at a confidence level of 95% ($\alpha = 0.05$), then we can reject the Null hypothesis in favour of the Alternate. Rejecting the Null hypothesis will show that the number of COVID-19 related deaths is lower in a population using a regime of close-contact tracing augmented with random testing. This finding could help policy makers form strategies that involve a combination of testing approaches to reduce the need for stringent and painful policy measures in order to control the spread of infectious diseases.

## 1.5 Scope and Limitations

The domain of this research will be to investigate the use of decision support through modelling and simulation. The scope will be on the use of these techniques to simulate the effectiveness of government response policies on the spread of COVID-19. In terms of context, this research will only focus on a specific outbreak of the COVID-19 pandemic in a specific region in Ireland.

## 1.6 Document Outline

This dissertation is organised into five chapters, including this chapter. Chapter Two provides background on the origins and dynamics of the COVID-19 disease and the government response policies employed. This chapter also examines different epidemiology models used to predict the impact of various interventions and identifies gaps in the literature to use as the basis for this study. Chapter Three presents the model that this study will use to answer the research question including the model itself, the

data used as input into the model, how the model will be initialised and validated, and finally how the experiment will be designed and run. Chapter Four examines and discusses the results of the validation and the main experiment used to test the hypothesis. Finally, chapter Five presents the conclusions interpreted from the results and provides a critical analysis of the work undertaken in this study.

# Chapter 2

# Review of existing literature

## 2.1 Background

### 2.1.1 COVID-19

#### History of COVID-19

In December 2019, a series of unexplained cases of pneumonia were reported in the city of Wuhan, Hubei Province, in China. A novel virus from a family of viruses known as Coronavirus was identified and named as COVID-19 (Al-qaness et al., 2020; Zeng et al., 2020).

The Chinese government implemented a series of public health measures in an attempt to contain the virus such as surveillance of the population, epidemiological investigations, and the closure of the suspected source of the virus outbreak, the Human seafood wholesale market. These measures were not found to be sufficient by the Chinese government and additional, much more severe, measures were then introduced. These included the sealing off of the city of Wuhan, restrictions on movement and public transportation, and a stay at home notice (Al-qaness et al., 2020; Liu et al., 2020). While these measures did help to reduce the spread of the disease in China, COVID-19 has since spread across the world and has been classified as a pandemic by the World Health Organisation[1].

[1]https://www.who.int/emergencies/diseases/novel-coronavirus-2019

At the time of writing, several other members of this coronavirus family have been seen such as the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). By 2004, SARS-CoV had infected more than 3000 people, killing 774 and from 2012 to 2019, SARS-CoV had been reported to have had 2465 confirmed cases, with 850 deaths (Hui et al., 2020). However, by November 16, 2021, COVID-19 has become a worldwide pandemic, spreading to over 220 countries, infecting 253,640,693 people, and unfortunately causing the deaths of 5,104,899 individuals[2]. The following sections discuss the dynamics of the disease across a number of areas.

### Spread of the Disease

Infectious diseases are spread from an infected individual to a susceptible individual through direct or indirect contact. COVID-19 has proven to be an incredibly infectious disease when compared to SARS-CoV and MERS-CoV, two of its closest relatives.
The transmission of COVID-19 is accepted to be through one of two modes. The first is through respiratory droplets which can be exhaled by an infectious individual through breathing, singing, exercising, coughing, or sneezing. These droplets can fall close to where they were exhaled. The second mode is through direct contact with contaminated surfaces (Morawska et al., 2020).
The transmission can be exacerbated in indoor settings such as homes, workplaces, public transport, or entertainment and hospitality venues (Shim et al., 2020). Morawska et al. (2020) have also presented small airborne droplets which can be transmitted to susceptible individuals as a probable additional route of infection which would also have a higher risk indoors, particularly in busy areas and areas with with poor ventilation.

### Symptoms

Symptoms of COVID-19 can vary significantly from person to person. Some individuals do not experience any symptoms at all and are therefore asymptomatic. The

[2]https://www.who.int/emergencies/diseases/novel-coronavirus-2019

majority of infected individuals experience common symptoms such as cough, fever, breathing difficulties, joint pain, fatigue, gastrointestinal symptoms, and a loss of taste or smell (Carfì et al., 2020).
More severe cases require hospitalisation and patients suffer with low levels of oxygen in their blood (hypoxaemia), pneumonia, and a need for ventilation, mechanical or non-invasive (Bastos et al., 2021).

**Variants of the disease**

Viruses are continuously mutating which is an expected process. Sometimes these mutations result in a new variant of the virus. As these mutations evolve, some variants emerge but do not spread and die out quite quickly, while others persist. In the case of a global pandemic such as COVID-19, some of these variants can spread very quickly and become the dominant strain across whole regions and even the world. The properties of each COVID-19 variant has differed but a number of variants have evolved and spread, usurping the previous dominant strain. Countries use genome sequencing on a sample of confirmed cases in order to understand which variants exist in the population and also to identify new variants. These new variants are investigated and assessed by the World Health Organisation in terms of transmissibility, disease severity, and performance of vaccines and medicines, and identified as a Variant of Interest or Variant of Concern. The World Health Organisation can then advise countries on any recommended changes to public health responses[3].

A number of notable Variants of Concern were detected in Ireland. The Alpha variant (B.1.1.7), became the prominent strain between December 2020 and June 2021 with the Delta variant (B.1.617.2) quickly dominating the pandemic between July 2021 and November 2021[4]. The increased transmissibility of the Delta variant was significant cause for concern, however vaccines have been shown to be effective against severe cases (Lopez Bernal et al., 2021). From December 2021, the Omicron variant

[3]https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

[4]https://www.hpsc.ie/a-z/respiratory/coronavirus/novelcoronavirus/surveillance/summaryofcovid-19virusvariantsinireland/Virus%20Variant%20report.pdf

(B.1.1.529) has taken over as the dominant strain in Ireland[5]. Early indications suggest that the severity of the disease from this variant does not seem to be as significant[6], however further research is needed to confirm this hypothesis (Abdullah et al., 2022).

### Diagnosis

The diagnosis of COVID-19 can be confirmed using a number of diagnostic tests. The gold standard diagnostic test is called quantitative reverse transcription-PCR (RT-qPCR) which uses samples taken from the upper and lower respiratory tract (nose and throat). This and close variations such as direct rapid RNA extraction-free RT-qPCR and high-throughput RT-qPCR have been used extensively by many countries as part of their detection strategy (Khatami et al., 2020). Some studies have shown the sensitivity of PCR testing to ranges between 83.5% to 87.9% and 65.6% to 93.4% (Mardani et al., 2020; Nagura-Ikeda et al., 2020). However, Khatami et al. (2020) have leveraged 60 studies and have presented overall sensitivity as 87%. Other lesser known laboratory tests are the real-time RT-PCR test which suffers from a higher risk of false-negative and false-positive results and the reverse transcription-loop-mediated isothermal amplification (RT-LAMP) test, which is preferred in point-of-care settings (Tahamtan & Ardebili, 2020; Nagura-Ikeda et al., 2020; Khatami et al., 2020).
Another popular option is called the rapid antigen test (RAT) which is being used in both clinical settings and, at the time of writing, by members of the public since it does not require special equipment, returns a quick result, and does not require any specialist skills (Nagura-Ikeda et al., 2020). However, the sensitivity of these tests can vary significantly depending on whether the sample is taken from saliva or a nasal swab, but sensitivity has been presented as 66.7% (Nagura-Ikeda et al., 2020).

[5]https://www.hpsc.ie/a-z/respiratory/coronavirus/novelcoronavirus/surveillance/summaryofcovid-19virusvariantsinireland/Virus%20Variant%20report.pdf

[6]https://www.research.ed.ac.uk/en/publications/severity-of-omicron-variant-of-concern-and-vaccine-effectiveness-?fbclidĪwAR3OTgg5KBwFRPlYplpRI0uI3i_zF8t8zKfrQvlJRhgpw_Rp35UxHMFvfqc

### Prevention

Non-pharmaceutical interventions have been adopted worldwide since the pandemic began. These include personal hygiene measures such as hand-washing, use of hand-sanitiser, and the use of surface disinfectant. Additional measures such as physical distancing and the use of personal protective equipment (suitable goggles, face masks or respirators) have also played a prominent role in the prevention of the disease (Pradhan et al., 2020).

Despite the adoption of non-pharmaceutical interventions, the vast majority of the world's population is still susceptible to COVID-19. Over 100 vaccines have been in development with several vaccines from manufacturers such as Pfizer, Johnson & Johnson, Astra Zenica and Moderna being approved for use[7]. These vaccines have been approved by both the Foods and Drugs Association (FDA) in the USA and the European Medicines Agency (EMA) as well as other regulatory agencies across the world. These and other vaccines are being used as the main pillar of long term response to the COVID-19 disease. The effectiveness of these vaccines have been shown in studies to be very high with rates of over 90% being reporting in initial studies. However with the emergence of new variants, the transmission resistance of the virus has reduced, resulting in the need for additional booster doses in some cohorts of patients. However, the effectiveness against severe or critical disease has remained very high (Polack et al., 2020; Evans & Jewell, 2021).

### Treatment

Initial treatment of COVID-19 is similar to the non-pharmaceutical prevention of the disease. If an infected individual has been confirmed to have COVID-19, they are advised to isolate at home in a well ventilated room, keeping distance from others in the house, wearing a mask, and ensuring that no travel is undertaken such as to work, school, or any public place. These non-pharmaceutical responses are sufficient for asymptomatic and mild cases.

[7]https://covid19.trackvaccines.org/vaccines/

However, in severe cases, hospitalisation may be required with some patients requiring anti-viral agents, steroids, and ventilation (Lam et al., 2020). Unfortunately, the disease progresses with some patients requiring treatment in an Intensive Care Unit (ICU) with more invasive treatments being provided.

### 2.1.2 Government responses

Governments across the world have taken a range of both pharmaceutical and non-pharmaceutical measures in response to the COVID-19 pandemic. The scale of the responses have varied across countries which has depended on the timing of waves of infection and the economic and containment strategy of the government (Hale et al., 2020). In many countries, the pandemic has not only caused a public health emergency but has also created economic and political crises. The capacity of healthcare systems in countries has also reflected countries' ability to respond effectively to the pandemic and can inform the responses governments are in a position to make (Greer et al., 2020).

The responses outlined in the following sections are not an exhaustive list but are some of the most common responses used across the world, including by the Irish government.

#### Restriction of movement

The restriction of movement can take many forms including restricting the number of passengers on public transport, limiting the ability of travelers to arrive from or depart to countries, and stay-at-home guidance for the public, where travel outside a particular radius is prohibited except for individuals deemed to be essential workers. This policy has been used heavily by the Irish government and for extended periods of time. The Irish government have implemented a number of variations of movement restrictions at a national, regional, and county level. Visitors from designated countries have been required to quarantine for periods of up to 14 days, and public transport capacity has been severely reduced in an attempt to lower the transmission of the disease (Hale et al., 2020; Kennelly et al., 2020).

### Physical distancing and mask wearing

Measures such as physical distancing (initially known as social distancing) and mask wearing have been used heavily in many countries including Ireland due to their relative ease of implementation and low cost. Maintaining a distance of at least 1m between people has been shown to have a lower rate of transmission of the disease with higher distances associated with lower risk of infection. Transmission has also been shown to be significantly reduced by wearing masks and/or eye protection (Chu et al., 2020; Kennelly et al., 2020). However, the literature does not necessarily account for the types of masks used, or the correct usage of them, for example, wearing a mask without the nose covered or wearing disposable masks multiple times.

### School and business closures

School and business closures have been used across the world with most governments issuing work-from-home guidance for workers that have the option to do so. This was one of the first measures introduced along with restriction of movement and physical distancing. This measure has disproportionately affected some industries, particularly in the hospitality and entertainment industries (Kaushal & Srivastava, 2021).

This measure has forced many businesses to trade for reduced hours, operate at reduced capacity, or close down for extended periods of time. This has also forced many governments to introduce emergency economic stimulus measures, such as enhanced unemployment benefits, in an attempt to prevent mass unemployment and maintain a health economy (Hale et al., 2020; Kennelly et al., 2020; Silva et al., 2020). Within days of the Irish government announcing these measures in March 2020, RTE reported that 140,000 workers in the restaurant, bar, and night club industries were laid off[8]. Many of these workers have had little job security over the course of the pandemic as waves of the virus have forced these businesses to open and close multiple times or to restrict their services. Indoor and outdoor gatherings were also prohibited worldwide

[8]https://www.rte.ie/news/coronavirus/2020/0316/1123480-coronavirus-ireland/

with music events being cancelled, sports events being held behind closed doors, and the meeting of crowds of people being prohibited.

Schools, childcare facilities, and universities have also been affected across the world and in Ireland. For periods of weeks in October 2020 and from January 2021 until March 2021, primary and secondary schools in Ireland adopted a policy of home schooling where parents teach younger children at home while older children and teenagers attend remote classes. State examinations were also reduced in scope, such as the Junior Certificate, or cancelled, with estimates of marks being assigned to Leaving Certificate students. University students have also been impacted with classes being delivered online, and examinations being administered remotely as open book format (Kennelly et al., 2020).

**Testing and contact tracing**

From the beginning of the pandemic, a key government response was the identification and isolation of individuals with COVID-19. There are a number of approaches of identifying individuals infected by the virus. If an individual suffers from common symptoms such as fever, prolonged cough, tiredness, or loss of sense of taste or smell, then a PCR test, as described in section 2.1.1 is used as the gold standard to diagnose the disease. However, the criteria used by countries to determine who qualifies for a test has differed and these criteria have changed for Ireland over the course of the pandemic many times. For example, during the early stages of the pandemic, specific symptoms were required before a PCR test could be booked. Over the course of 2021, walk-in centres were introduced that allowed people to book a PCR test without presenting with any symptoms. However, more recently as demand has soared, the requirement for symptoms has returned before a PCR test can be booked[9].
In addition, newly diagnosed individuals have been advised to isolate at home and remain at home for a period of days unless their symptoms become more severe[10]. If

[9]https://www2.hse.ie/conditions/covid19/testing/get-tested/

[10]https://www2.hse.ie/conditions/covid19/restricted-movements/how-to-self-isolate/

the symptoms become more severe, then further treatment and hospitalisation may be required. Since the transmissibility of the virus is extremely high, contact tracing is also used to contact people that were in close contact with an infected individual within the previous few days. Depending on the criteria for testing, these individuals would also have a PCR test booked and the process would repeat if a positive result was produced.

Panovska-Griffiths et al. (2020) investigate the impact of testing and tracing in the UK, and suggest that large scale testing of symptomatic individuals and effective tracing of their close contacts would be required in order to limit the spread of COVID-19. In their modelling, a very small percentage (0.75%) of asymptomatic individuals were tested through the close contact regime which suggests another approach to identifying asymptomatic individuals is required. While asymptomatic individuals will not suffer from any adverse effects, they have the potential to unknowingly spread the disease to many other people, some of which may become quite ill. In Germany, for example, close contact tracing was accompanied by walk-in test centres where individuals could attend for a PCR test without a booking or without having shown any symptoms. This formed part of a successful strategy which helped to greatly reduce the number of COVID-19 deaths during the early phase of the pandemic in 2020 (Hale et al., 2020; Barbarossa et al., 2020).

**Use of technology**

Countries that have managed to maintain relatively lower rates of death have been successful in the areas of testing, close contact tracing, and surveillance. In order to manage the scale of coordination required at a national level, countries have relied on the use of digital technology to provide public health guidance, track disease activity in real time, and identify and track individuals that have come in contact with an infected person (Whitelaw et al., 2020).

In the UK and USA, a study used a smartphone app to track COVID symptoms and generated mathematical models to predict geographical hot-spots of incidence 5 to

7 days in advance of official public health reports (Drew et al., 2020). Whitelaw et al. (2020) describe how China and Taiwan used high performance infrared thermal cameras at airports to detect individuals with a fever. In Ireland, a collaboration between the public and private sectors and health authorities developed a smartphone app used to track and trace COVID-19 infections (Kennelly et al., 2020).

The ability for countries to implement such technology depends on a number of factors such as local and regional data privacy law, high costs, regulation, and the potential for violation of civil liberties (Whitelaw et al., 2020). As a result, the rate and scale of adoption across countries varies widely. This also raises ethical questions around the level of surveillance and control a government can have over its population and whether effective governance and regulation can be maintained. China has shown its extraordinary level of control over its population through the use of technology can be very effective to control the spread of an infectious disease by restricting movement and mass surveillance. Kozyreva et al. (2021) describe the use of contact tracing technology in Germany and the psychological factors influencing the public's perception of their use. Trust in the security of the technology and the perception of effectiveness of such technology were key to acceptance.
If these ethical concerns can be addressed, the use of the data generated through this technology can be leveraged to build effective models which can be used to inform government responses and minimise their impact on wider society.

## 2.2 Approaches to solve the problem

In order for governments to assess the severity of the COVID-19 pandemic and plan the best proportion of responses outlined in Section 2.1.2, they need to be able to analyse the infection rates in the population, and to predict possible contagion scenarios as well as model potential responses. Epidemiology models can be used to build such scenarios and these will be discussed in the following sections.

### 2.2.1 Epidemiology Modelling

According to the WHO, an epidemic disease is one "affecting many persons at the same time, and spreading from person to person in a locality where the disease is not permanently prevalent" and a pandemic as "a worldwide spread of a new disease"[11]. Epidemiology is the study of infectious diseases and in order to understand how epidemiology models can be used to inform policymakers to assess the effectiveness of various measures, it is important to understand a number of concepts.

#### Epidemiology concepts

**Basic Reproductive Number,** $R0$, is a measure used in epidemiology to assess the transmissibility of viruses, and estimate the severity of infectious disease outbreaks. $R0$ is an intuitive concept to understand and is defined as the average number of secondary infections generated by a single infectious individual in a susceptible population (Ajelli et al., 2010; Breban et al., 2007). If the value of $R0$ is less than 1, then an outbreak will become extinct. If the value of $R0$ is greater than 1, then the outbreak is expected to become an epidemic and continue to spread. The modelling for $R0$ however can be quite complex. One method uses data from a close contact tracing regime where secondary infections are calculated for each infected individual and then an average is taken, but a more common method uses differential equations on cumulative incidence data (Breban et al., 2007). The first option does not scale well, especially when attempting to do this on a scale of the COVID-19 pandemic. While the second option does make some assumptions and cannot include individual level modelling or time independent infection rates, it provides an estimate that can scale well and be used for assessing the status of an outbreak.

**Stages of infectious disease** vary depending on the dynamics of the disease, but there are a number of common concepts around the stages of infection. First, a susceptible individual will become exposed to the disease, usually through a micro-organism. This may not necessarily be from person to person since the transmission

[11] https://www.who.int/

method will vary. Next, the stage at which pathological changes occur within the infected individual is called the incubation period and will not show any symptoms. After this phase, the infected individual may become symptomatic and the range of severity of these symptoms will vary significantly. The infectious disease will progress from here and will end in recovery, death, or disability[12].

**Epidemiology Models**

A Susceptible, Exposed, Infected, Removed (SEIR) model is a compartmental method used to simulate the effects of infectious diseases. In this model, a population is split into four compartments; Susceptible, Exposed, Infected, and Removed, and the progression of an outbreak can be modelled as the population moves through each stage. Susceptible ($S$) members of the population do not have the disease but can become infected. Exposed ($E$) members of the population have become infected but may not be infectious yet. This simulates the incubation period of an infectious disease. The Infectious ($I$) compartment contains members of the population who are now infected by the disease and can transmit the disease to others and the Removed ($R$) compartment contains both members the of the population that have died as well as those that have recovered.

### 2.2.2 Equation based models

Equation based models can be described by a series of ordinary differential equations. A basic assumption is that most, if not all of the individuals in the model will progress through every stage of infection. Godio et al. (2020) describe the classic SEIR differential equations as follows:

$$\frac{dS(t)}{dt} = -\beta I(t).\frac{S(t)}{N} \tag{2.1}$$

The susceptible part of equation 2.1 is represented by the whole population. $\beta$ represents the probability of infection and $N$ is the total population, which is used

[12]https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section9.html

to normalise the system. The minus sign exists here since the number of susceptible individuals will reduce once they become exposed through contact at probability $\beta$.

$$\frac{dE(t)}{dt} = \beta I(t).\frac{S(t)}{N} - \gamma E(t) \tag{2.2}$$

Equation 2.2 describes the Exposed compartment and contains the same term as Susceptible, except with a plus sign since the individuals move into this compartment. Here, $\gamma$ denotes the inverse average latent time and how fast individuals start to show symptoms.

$$\frac{dI(t)}{dt} = \gamma E(t) - (\lambda + \kappa)I(t) \tag{2.3}$$

$$\frac{dR(t)}{dt} = (\lambda + \kappa)I(t) \tag{2.4}$$

The Infected compartment, denoted by equation 2.3, contains the individuals that are now infected with the disease after the latent period. $\lambda$ and $\kappa$ denote the recovery rate and the death rate respectively and correspond to how fast individuals will recover from the disease and how many of them will die as described in equation 2.4 (Godio et al., 2020; Yang et al., 2020; S. He et al., 2020; Prem et al., 2020).
A significant task performed while using these models is estimating the parameters of the model. In the classical example above, this means estimating the parameters $\gamma$, $\lambda$, and $\kappa$.

In the literature, there are many examples of research using equation based SEIR models for the modelling of infectious diseases including COVID-19. In most cases, the authors have adopted the classical model to add features that are specific to the disease dynamics or specific to the research topic. Godio et al. (2020) have added a time dependent infection rate as well as stochastic solver to incorporate different scenarios during the simulation such as mask wearing or lock-down policies. Yang et al. (2020) have added additional move-in move-out parameters which included the ability to incorporate the migration of individuals during the outbreak. These parameter values were based on publicly available migration index data. S. He et al. (2020)

added additional compartments, $I_1$, $I_2$, $H$, and $Q$ to account for individuals infectious without intervention, infectious with intervention, in hospital, and in quarantine respectively. Prem et al. (2020) also split the Infected compartment into two separate compartments, where different infection rates, and contributions of asymptomatic and sub-clinical cases are applied to two sets of age groups. The authors also included an age and location specific matrix of social mixing patterns to account for different scenarios in the simulation. Choi & Ki (2020) also included an additional compartment for hospitalised individuals. Ivorra et al. (2020) include additional compartments for hospitalised and removed individuals, but have a number of versions of each to account for different control measures.

While many of the examples above and others in the literature do introduce various novelties to account for different dynamics relating to COVID-19, they suffer from a number of limitations due to the nature of differential equation modelling. Equation based models are unable to take into account external influences, such as the actions containing the spread of the infection that may occur at different times during the development of the infection itself. Although Godio et al. (2020) did add time dependent infection rate, this was an estimate and was not based on the behaviours of individuals in the simulation. Since differential equations are deterministic, they will produce the same result for every simulation. Godio et al. (2020) did include a swarm intelligence component called Particle Swarm Optimisation to add a stochastic element to the simulations. However, approaches from other authors produced a wholly deterministic result (Yang et al., 2020; S. He et al., 2020; Prem et al., 2020; Choi & Ki, 2020). Ivorra et al. (2020) do raise an interesting point where deterministic models are more useful to use as a first approach since they are easier to use, easier to analyse, and require less data for the calibration of the model. In some examples, asymptomatic individuals are not taken into account which could be excluding a key contributor to the transmission and spread of the disease (Ibarra-Vega, 2020; Smirnova et al., 2019). Interestingly, Ivorra et al. (2020) estimated that as much as 37% of infections in the first wave of COVID-19 in China were due to undetected cases, the majority of which were

thought to be asymptomatic. Testing and diagnostic capacity are difficult elements of an epidemic to incorporate into differential equations and, while it was incorporated by Godio et al. (2020); Choi & Ki (2020), the majority of researchers did not include this aspect of the epidemic (S. He et al., 2020; Prem et al., 2020; Ibarra-Vega, 2020; Smirnova et al., 2019; Flaxman et al., 2020; Yang et al., 2020). Prem et al. (2020) and Smirnova et al. (2019) did not take quarantine into consideration in their equation based models, but these were included in other examples (Godio et al., 2020; S. He et al., 2020; Ibarra-Vega, 2020; Choi & Ki, 2020; Flaxman et al., 2020). An assumption of equation based models is also that almost every individual in the model moves through every compartment. This means that every individual will be exposed to, and be infected with the disease. While this does not reflect the real world in which only a small percentage of individuals may be infected with an infectious disease, it fulfils the purpose of these models which is to test the effect of various interventions on the spread and severity of infections. Results are quite often returned as an acceptable range of values rather than a specific value so as to account for limitations in input data as well as the deterministic nature of these models.

Other researchers offer an interesting insight into the COVID-19 epidemic through the lens of the 1918 influenza pandemic in the US and UK (Bootsma & Ferguson, 2007; D. He et al., 2013). Using deterministic SEIR modelling, the researchers investigated non-pharmaceutical responses to the 1918 epidemic which are quite similar to the COVID-19 pandemic. There are obvious differences in the world we live in today in terms of family size, travel patterns, levels of education, availability of information, and the huge difference in the level of healthcare. As a result, disease-mortality was much higher then than at the time of writing. However, it was interesting to note that cities that better implemented non-pharmaceutical interventions such as social distancing and isolation had a much milder first wave but much more severe second wave.

### 2.2.3 Agent based models

Deterministic approaches to epidemiology modelling, such as with ordinary differential equations as described in the previous section, are relatively simple models to create, notwithstanding the novelties discussed. Another approach to modelling epidemics such as COVID-19 using the SEIR compartmental method is called Agent Based Modelling. Agent Based Models (ABMs), which are stochastic in nature, are computer simulations of individuals or agents in a simulated space that takes place over simulated time (El-Sayed et al., 2012). This bottom-up approach simulates how a heterogeneous population of agents move and interact in ways that resemble human behaviour and are ideal for modelling macro-level patterns that arise from micro-level behaviour (El-Sayed et al., 2012; Carpenter & Sattenspiel, 2009).

Agent based models are ideal in scenarios where agent behaviour is complex and interactions between agents and the environment is sophisticated. Agents and the environment can be constructed to include social factors and behaviour, for example, agents can belong to a social class, contain attributes such as age, gender, diet, race, and income, and can also exhibit behaviours that can influence other agents and the environment such as mobility patterns and social networks. In the case of epidemiology, attributes such as risk factors can be used to determine the susceptibility to an infectious disease, and behaviours can include interactions that affect the transmission or progress of an infectious disease, such as infecting another agent or isolating in quarantine. Such complexity is very difficult to capture in an equations based model (E. Frias-Martinez et al., 2011).

Agent based models also allow multiple simulations to be run under different conditions in order to simulate the effectiveness of various interventions. In the example of epidemiology, these models are ideal to simulate the effectiveness of various non-pharmaceutical interventions on an infectious disease such as COVID-19. This ability to examine specific interventions and how they relate to outcomes at a population level offers a better understanding of the dynamics of an epidemic (El-Sayed et al., 2012; Carpenter & Sattenspiel, 2009).

Agent based models are commonly used to model the dynamics of the COVID-19 pandemic. Flaxman et al. (2020) study non-pharmaceutical interventions such as lockdowns, banning of public events, school closures, self-isolation, and social distancing and compare their effectiveness across 11 European countries. Silva et al. (2020) focus on social distancing, mask wearing, and isolation spread across a number of scenarios. Interestingly, the economic effects of the pandemic are also modeled alongside the epidemiological effects. Attributes including social, epidemiological, and economic factors are taken into consideration and additional compartments have also been introduced to account for asymptomatic, mild, hospitalised, and critically ill individuals in an Intensive Care Unit (ICU). While testing was not simulated in the environment, several quarantine scenarios were simulated during the experiment. Shuvo et al. (2020) examine the impact of hospital capacity and social distancing to model their effect on the length of an outbreak and COVID-19 related deaths. Additional compartments for hospitalisation and dead agents were implemented and while only one lockdown scenario was considered, a number of scenarios for differing levels of social isolation were explored. Shuvo et al. (2020) also acknowledge that their model has not been validated against real epidemic data or information related to social isolation since it was not available at the time the research was being conducted. Kerr et al. (2021) built an open source agent based model solution called Covasim specifically to model COVID-19 dynamics, as well as various interventions. This solution can be customised with country specific demographic information, disease transmission networks, and a set of pharmaceutical and non-pharmaceutical interventions. The authors presented a case study from Seattle, Washington that predicted a small change in mobility patterns ensuring the prevention of a rises in COVID-19 cases which was shown to match real world observations. While this case study was a small example, the solution has been used in many countries for local modelling of interventions including the UK. Panovska-Griffiths et al. (2020) have used Covasim to determine the impact of testing and tracing as a strategy for reopening schools in the UK, and the results of their modelling estimated that 75% of symptomatic infected people would need to be tested to avoid additional waves of infection. However, due to limited availability of input

data, a significant number of assumptions were made but were acknowledged by the authors. Carpenter & Sattenspiel (2009) used agent based models to simulate the effects of seasonal movement patterns on the 1918 influenza epidemic to investigate disease transmission and spread. Since the data availability was limited, the model was quite rudimentary.

An interesting perspective on the comparison of equation based models and agent based models in epidemiology consists of the combining of both approaches into a hybrid model, where an agent based model can be used to simulate the effects of a disease at a detailed local level, and the equations based model can be used to scale this up to a population level where high confidence datasets may not be available (Ajelli et al., 2010; Hunter et al., 2020; Hunter & Kelleher, 2021).

### 2.2.4 Network based models

Social Network Analysis (SNA) is an area of Data Science focused on the characterisation of structures of social networks. The analysis of such structures can be used to learn how people in a social network can be reached quickly, what communities exist in a network, how resilient a network is to attack, and to understand how a process can diffuse over a network. A network consists of a set of points called nodes which can be connected by lines called edges. The nodes and edges can have characteristics to represent attributes of people as well as the relationships between them. In the area of epidemiology, this translates to the use of networks to understand the influences of social factors on disease distribution in populations and how social phenomena influence health as they spread through social networks (El-Sayed et al., 2012).

Saravanan et al. (2013) make use of mobile phone data to generate social networks which are used to model mobility patterns on the spread of a disease across regions. The authors noted that specific individuals that communicate with lots of other users tend to be influencers in terms of disease spread, and this individualistic nature of people is something that agent based models do not take into account. Since this is

not specific to COVID-19, testing and tracing does not seem to be modeled here.

Social Network Analysis is quite often used as part of a larger agent based model. Venkatramanan et al. (2018) use social network analysis to simulate a social contact network of agents to forecast the emergence of the Ebola epidemic in Liberia. This network, called a bi-partite network, represents the relationships between two different types of agents - people and locations. Venkatramanan et al. (2018) acknowledge that the model does not have the granularity of data to simulate individual level behavioural interventions. Hoertel et al. (2020) examined the effects of non-pharmaceutical interventions after a lockdown has been lifted in an effort to reduce the risk of multiple waves of infections overwhelming ICU beds. The agent based model did not use the SEIR compartmental methodology and did not take into account asymptomatic individuals. However, social networks were generated for the agent based model to simulate different types of social contacts experienced during the day, for example, interfamilial contacts, contacts at work, or contacts with friends. While the research aimed to measure the effect of non-pharmaceutical intervention on ICU beds within a health system, it did not account for any potential change in fatality due to COVID-19 infections overwhelming this system. A number of studies used social network analysis, within a Python developed agent based modelling solution called Covasim, to simulate transmission networks across households, schools, workplaces and care facilities, while simulating the effectiveness of various interventions on the spread of the disease (Kerr et al., 2021; Panovska-Griffiths et al., 2020).

### 2.2.5 Swarm Intelligence

Swarm Intelligence (SI) is an area of Artificial Intelligence (AI) and is a collective term used to describe a group of algorithms that have been inspired by nature. These algorithms use a set of decentralised agents that follow a set of simple rules in order to solve a problem (Inje et al., 2019), for example, bees in nature need to collectively scout and decide on a location for a new colony and insects need to find the shortest path between their nest and a food source as well as organise their nest. Individually

bees or other insects are unsophisticated creatures but as a collective, they show a remarkable ability to make the correct decisions to solve these problems.

In computer science, swarm intelligence mimics these systems by creating a population of self-organising agents that follow a simple set of rules to solve common problems such as optimisation problems. Muthukaruppan & Er (2012) use an algorithm called Particle Swarm Optimisation (PSO) to interpret decisions made by a fuzzy expert system in the diagnosis of coronary heart disease. Al-qaness et al. (2020) use the Flower Pollination Algorithm (FPA) to optimise an early COVID-19 time-series forecast model. This algorithm simulates the transfer of pollen between flowers by pollinators in nature. The authors produced promising results for short term forecast within 10 days but this study was produced in early 2020 when dynamics of the disease were not widely known. The forecast model, while interesting for simpler use cases, does not lend itself to predicting more complex disease dynamics. Godio et al. (2020) uses Particle Swarm Optimisation (PSO) to improve the reliability of COVID-19 medium term predictions, and adds a stochastic element to a differential equations model. However, similar to the previous example, these models simulate less complexity than is possible with agent based models.

## 2.3 Research Question

This section will use the gaps identified in the literature discussed in Sections 2.1 and 2.2 and use them as the basis for this research dissertation. The rapid and prolonged spread of the COVID-19 pandemic has shown the need for social and economic scenarios to be assessed and incorporated into response policies. The literature has shown many approaches to solve this problem and while there are many examples of excellent research, the disease is still quite novel and not fully understood which is being exacerbated by the constantly changing dynamics of the disease through emerging variants. The following summary discusses themes that have emerged as gaps in the literature.

### 2.3.1 Gaps in the research

While deterministic equation based models can be extended to include stochastic behaviour, examples in the literature can become quite complex when a number of external interventions are included. Since individual level behaviour can not be easily modelled using differential equations, assumptions are made across a homogeneous population, which requires prior knowledge of disease dynamics, and can be difficult with such a novel virus. Agent based models do also rely on prior knowledge of disease dynamics but can apply simpler rules at an individual level and still simulate much more complex systems, and generate population level inferences.

Diagnostic capacity is not taken into account in some examples in the literature while testing and tracing of infected individuals has been shown to have a significant effect on the spread of the disease.

Asymptomatic individuals are a key cohort in the spread of COVID-19. However, some examples in the literature do not distinguish between symptomatic and asymptomatic individuals.

### 2.3.2 Research question

Based on the gaps identified in the research, the following question will be addressed in this research dissertation.

Can the efficiency of a COVID-19 close-contact tracing regime in an Irish county be improved through the use of random testing of potentially asymptomatic infected individuals?

### 2.3.3 Research hypothesis

**Alternate Hypothesis** The number of COVID-19 related deaths in an Irish county is significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing.

**Null Hypothesis** The number of COVID-19 related deaths in an Irish county is not significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing.

## 2.4 Summary

In order to answer the research question, a number of objectives will be set in order to build an agent based SEIR model. These objectives will be to design and build an agent based model, to utilise data sources to build the model, to initialise the model, to validate the model, and to design an experiment to test the hypothesis. These objectives will raise a number of questions, such as which compartments of the SEIR model will best represent COVID-19, what elements of society are critical to capture in the model, what disease dynamics should be included, which response strategies should be included in the model to complement the two testing regimes, how will individual level behaviour be designed, and how can symptomatic and asymptomatic individuals be differentiated? While this model will generate data in the experiment, many different data sources will be required to incorporate the dynamics of society, the COVID-19 disease, and as well individual behaviour - what data exactly is required, what level of granularity is required for the model, how will this data be sampled if required? Once the model has been built, how will it be validated before an experiment can be run, how can the results be trusted, especially if they are surprising, and what data should be generated to test the hypothesis?

Network based models and Swarm Intelligence have been used in the literature as components of larger models, forming contact networks and communities or as optimisation algorithms but won't be included in this project due to limitations in scope.

# Chapter 3

# Experiment design and methodology

A modified Agent Based SEIR Model was used to study the effects of random testing augmenting a regime of close contact testing and tracing on the spread of COVID-19. This study focused on a specific outbreak in county Carlow during the third wave of the pandemic between the dates December 1$^{st}$ 2020 and February 28$^{th}$ 2021. The following sections describe how the study was undertaken including how the model was designed, built, initialised, and validated. The experiment is also described in terms of design and execution.

## 3.1 Design and Build an Agent Based Model

### 3.1.1 Agent Based Modelling Software

The model was built using Netlogo, a multi-agent simulation environment and programming language (Tisue & Wilensky, 2004). The software is popular in the domains of research and education and it quite easy to use. The software also has an active community and there are many sample models that are built and included with the software. Agents, known as turtles, move across the environment which consists of a grid of so-called patches, which are also programmable agents. These turtles and

patches can interact with each other in complex ways and can perform actions (Tisue & Wilensky, 2004). Agents can also be split into different types known as breeds. Each breed of agent can have its own set of variables and behaviour in the simulation. As well as agents and patches, networks of nodes, connected by edges, can also be built. Network models can be analysed using Social Network Analysis (SNA) concepts such as centrality and processes such as diffusion can be simulated.

The software uses a graphical user interface which can be used to build a user interface for the model, including the environment itself as well as controls such as buttons and sliders which can be used as actions or parameters for the model. The interface can also include outputs such as plots or monitors that can give real time feedback during a simulation.

The software uses a language derived from Logo but extended to support multiple agents. The language can be used to instruct agents to perform very complex behaviour through its library of native or user defined procedures called commands (to perform actions) and reporters (to return a value). The language allows both global level variables plus agent or patch level variables so that each agent or patch can store it's own set of values during the simulation.

The software also includes additional tools. One such tool is called BehaviourSpace which is used to conduct experiments with the models. Models can be run multiple times with variables being set with different values and the results being recorded for analysis[1]. For example, a model may have 4 parameters with ranges of values for each. Behaviour space can run an experiment and systematically set the value of each combination of variable values, and run a set number of iterations of each setting. The resulting dataset can be analysed outside of BehaviourSpace, for example using Python.

[1]https://ccl.northwestern.edu/netlogo/docs/behaviorspace.html

Finally, Netlogo can interact with external data files to import input data that can be leveraged by the model, and can also incorporate external software capabilities using extensions. This allows the software to run R or Python code, for example, within a model or experiment.

### 3.1.2 SEIR Model

As described in Section 1.3.1, SEIR models are compartmental models used to simulate the effects of infectious diseases. The four standard compartments are Susceptible, Exposed, Infected, and Removed. To model the effects of COVID-19, this basic model was extended in this study to incorporate a number of additional compartments: Susceptible, Exposed, Infected Asymptomatic, Infected Mild, Infected Severe, Infected Critical, Dead, and Recovered. The following rules govern the behaviour of individuals in each compartment in this model. Susceptible individuals, $S$, could become infected by COVID-19 but did not yet. This model assumed that individuals previously infected with COVID-19 could not be re-infected. Exposed individuals, $E$, were infected but not yet infectious. Infected Asymptomatic individuals, $I^A$, were infectious but did not show any symptoms and therefore did not know to alter their behaviour unless they had been identified as a close contact. Infected Mild individuals, $I^M$, were infectious and were experiencing symptoms but did not require hospital admission in this model. This group self isolate and could be selected to complete a PCR test at the next available slot if there was capacity. Infected Severe, $I^S$, contained infected individuals that had been admitted to hospital while Infected Critical, $I^C$, were individuals admitted to an Intensive Care Unit. The Removed compartment was split into Recovered, $R$, and Dead, $D$. Since the Infected and Removed compartments from a standard SEIR model were split into a number of sub-compartments, the model became $SEI^AI^MI^SI^CRD$.

### 3.1.3 Modelling dynamics of society

In order to model a real world outbreak of COVID-19, it was essential to ensure that the simulation was representative of county Carlow. To achieve this, a number of entities were created in the NetLogo environment. Agents represented the population of people and an agent breed called People was created in NetLogo. An additional breed of agent was used to represent households. Households agents were further split into household type, hospitals, and workplaces. The Ireland Central Statistics Office (CSO) produces statistics on many areas of the Irish society and aggregates this data at different levels of geographic granularity. The lowest level of granularity is known as a Small Area and each patch in the NetLogo simulation represented a Small Area as defined by the CSO.

Since the severity of the disease depends on age, the age distribution of individuals in the simulation matched county Carlow based on the 2016 Census Data available from the CSO[2] (Silva et al., 2020; Hunter et al., 2020; Barek et al., 2020). Details of the data used and the pre-processing of that data are discussed in Section 3.2.4.
To model different patterns of agent behaviour, economic status data was incorporated to organise individuals into the following groups: students, retired, unemployed, employed and working from home, and essential workers. The proportion of employed people that were deemed to be essential workers was set to 20% (Redmond et al., 2020). The 2016 Census data from the CSO[3] was also used to incorporate this data and is described in Section 3.2.4.
The CSO has aggregated the number of household types in each small area. These household types include: one person, couple without children, couple with children, one-parent family, and two or more non-related persons[4]. Since each patch in NetLogo represents a single small area, the equivalent density of each household type in each small area is simulated. The population of people are then assigned to households as

[2]https://data.cso.ie/table/E3003

[3]https://data.cso.ie/table/E3003

[4]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

| Person Agent attributes | | | |
|---|---|---|---|
| **Attribute** | **Description** | **Value** | **Source** |
| Age | Age of agent | 0 - 95 | CSO Ireland[5] |
| Gender | Gender of agent | M or F | CSO Ireland[6] |
| IsEssentialWorker | Is this agent an essential worker | Yes (20%) / No (80%) | (Redmond et al., 2020) |
| EconomicStatus | Economic status of the agent | At_work, unemployed student, retired | CSO Ireland[7] |
| person_id | Unique identifier for the person agent | Generated by NetLogo | |
| family_id | Unique identifier for the household the agent will be assigned to | See Section 3.2.4 | CSO Ireland[8] |

Table 3.1: Person agent attributes

| Household agent attributes | | | |
|---|---|---|---|
| **Attribute** | **Description** | **Value** | **Source** |
| HouseholdType | Category of household | One person, married/co-habiting couple, married/co-habiting couple with children, one parent family with children, two or more non-related persons | CSO Ireland[9] |
| small_area_id | ID of small area | ID | CSO Ireland[10] |
| small_area_name | Label of small area | ID | CSO Ireland[11] |
| family_id | ID of household | ID | |
| isWorkplace? | Flag to indicate this household agent is a workplace | ID | |
| isHospital? | Flag to indicate this household agent is a hospital | ID | |

Table 3.2: Household Agent attributes

described in Section 3.2.4.

Table 3.1 shows the list of demographic parameters stored by each person agent in order to model the demographic elements of society.

[5]https://data.cso.ie/table/E3003

[6]https://data.cso.ie/table/E3003

[7]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[8]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

### 3.1.4 Modelling disease dynamics

Modelling the dynamics of this infectious disease required incorporating several features that interact with each other. Each compartment in the $SEI^AI^MI^SI^CRD$ model represents a different stage in the progression of an individual through a COVID-19 infection. However, each individual does not spend time in every compartment. Since the severity of the disease depends on the age of the individual, which acts like a proxy for underlying health conditions, a table of probabilities was used to determine how an individual progresses through the disease. The assumption made is that the alpha variant is used throughout the experiment and that no individuals have been vaccinated. Although vaccination had begun in Ireland from January 2021, the number of vaccinated people was relatively low during January and February, which was during the period of this experiment. The alpha variant was also the dominant variant during this time period. This is relevant for the parameters used to describe the disease dynamics. Table 3.3 lists out the epidemiology parameters and their sources, and Table 3.6 lists out their values (Kerr et al., 2021; Zhang et al., 2020; Verity et al., 2020; Ferguson et al., 2020; Brazeau et al., 2020; O'Driscoll et al., 2021).

The length of time individuals spend in each compartment is also taken from the literature and is listed in Table 3.4. The set of probabilities and epidemiological attributes are combined to form a process outlined in Figure 3.1.

Susceptible individuals had a probability of $\beta_1$ of becoming exposed to the disease only if they were within a close radius of an infected individual. As a result, not every individual became exposed. If an individual did become exposed to the disease, the days_since_infected agent attribute started counting and the disease progressed according to this process. After 4.5 days, the individual had a probability of $\beta_2$

[9]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[10]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

[11]https://census2016.geohive.ie/datasets/private-households-by-size-small-areas-census-2016-theme-5-2-ireland-2016-cso-osi/explore?location=53.433245%2C-8.329629%2C7.59

| Age Structured Epidemiology parameters | | |
|---|---|---|
| **Attribute** | **Description** | **Source** |
| $\beta_1$ | Probability of infection | (Zhang et al., 2020) |
| $\beta_2$ | Probability of developing symptoms | (Verity et al., 2020; Ferguson et al., 2020) |
| $\beta_3$ | Probability of developing a severe case | (Verity et al., 2020; Ferguson et al., 2020) |
| $\beta_4$ | Probability of developing a critical case | (Verity et al., 2020; Ferguson et al., 2020) |
| $\beta_5$ | Probability of death | (O'Driscoll et al., 2021; Brazeau et al., 2020) |

Table 3.3: Epidemiology parameters and their sources

of having an asymptomatic infection or a probability of $(1 - \beta_2)$ of having a mild infection. Asymptomatic individuals recovered after 8 days. Mild infections progressed to a severe infection at a probability of $\beta_3$ after 1.1 days. Severe infections required hospitalisation and the individual was transferred to a local hospital. If the individual recovered from a mild case, this was at a probability of $(1 - \beta_3)$ and full recovery was made after 8 days. Severe cases lasted for 6.6 days, at which point, the individual progressed to a critical infection at a probability of $\beta_4$. Recovering from a severe infection took 18.1 days. Critical cases represented the individual moving into an ICU in hospital where they remained critical for a period of 1.5 days. After this period, critical cases progressed to death at a probability of $\beta_5$ and death took place after 10.7 days. Critical cases recovered with a probability of $(1 - \beta_5)$ and recovery took 18.1 days.

An individual infected another if they were in any one of the infected asymptomatic ($I^A$), infected mild ($I^M$), infected severe($I^S$), or infected critical ($I^C$) compartments, and they were within a close radius of Susceptible individuals ($S$). This radius depended on the size of the agents in the simulation. The radius was set to the size of the agents so that if the agent size was very small, so too is the radius.

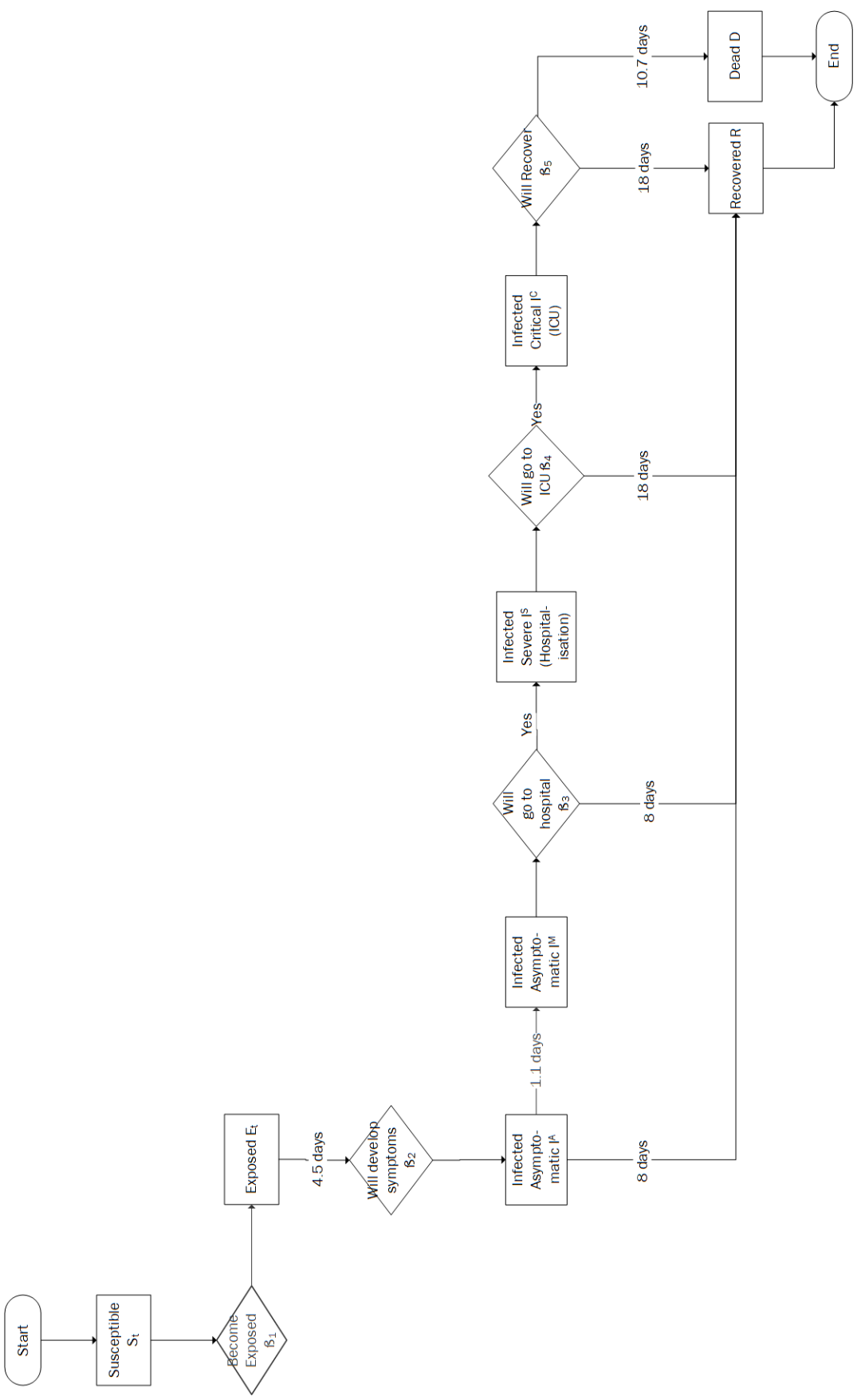



Figure 3.1: Epidemiological Process

### 3.1.5 Modelling agent movement

Two different types, or "breeds", of agents are used in the model. Household agents do have some attributes but do not move in the environment. Person agents represent individual people in the simulation and can move around the environment according to different rules. These rules are time dependent and align to the National Framework for living with COVID-19[12], which summarises the restrictions that apply in Ireland at different infection levels. While most of the restrictions in this policy focus on public gatherings, hospitality, and domestic travel[13], information is also provided on workplace and leisure time which is useful for modelling basic behaviour. The rules of behaviour in the simulation focus on Level 2 and Level 5 of this policy to align with the restrictions in place in county Carlow between December 1$^{st}$ 2020 and February 28$^{th}$ 2021.

Individuals with an economic status of "at_work" were assumed to work Monday to Friday 09:00 to 17:00. 80% of these individuals were assumed to work from home, while 20% were assumed to be essential workers and were randomly assigned to a workplace at the beginning of the simulation (Redmond et al., 2020). Essential workers traveled to their workplace at 09:00 where they remained until 17:00. At 17:00, they traveled back to their household and could roam freely around the environment from there according to the level of restrictions in place. During level 2, they can roam around the environment from 17:00 to 21:00. During Level 5, they can randomly roam around the environment from 18:00 to 21:00. Non-essential workers worked from home and so stayed in their assigned households until 17:00, at which point, they randomly roamed around the environment until 21:00 (and from 18:00 to 21:00 during level 5). Individuals with an economic status of retired or unemployed were free to randomly roam around the environment from 07:00 until 21:00 during level 2 and from 18:00 until 21:00 during level 5. Students were assumed to study from home but could roam

[12]https://assets.gov.ie/87604/405b1065-055a-4ca8-9513-390ce5298b10.pdf

[13]https://whatsnew.citizensinformation.ie/2020/09/15/the-new-framework-for-living-with-covid-19/

| Person agent epidemiological attributes | | | |
|---|---|---|---|
| **Attribute name** | **Description** | **Value** | **Source** |
| Epi_Status | Epidemiological status as set out in table 1.6 | All agents initialised as susceptible | - |
| DaysExposed | No of days the agent was infected but not infectious. | 4.5 | (Lauer et al., 2020) (Du et al., 2020) |
| DaysInfectious-ToSymptomatic | Number of days the agent was infectious | 1.1 | (Linton et al., 2020) |
| DaysSevere | Number of days the agent was severely ill in hospital | 6.6 | (Linton et al., 2020) (Wang et al., 2020) |
| DaysCritical | Number of days the agent was critically ill in ICU | 1.5 | (Chen et al., 2020) (Wang et al., 2020) |
| DaysCritical-ToDeath | Number of days the agent was critically ill before death | 10.7 | (Verity et al., 2020) |
| DaysInfectious-ToRecoveryAsympt-omatic | Number of days the agent was infectious before recovery (asympt-omatic cases) | 8.0 | (Verity et al., 2020) |
| DaysInfectious-ToRecoveryMild | Number of days the agent was infectious before recovery (mild case) | 8.0 | (Verity et al., 2020) |
| DaysInfectious-ToRecoverySevere | Number of days the agent was infectious before recovery (severe cases) | 18.1 | (Verity et al., 2020) |
| DaysInfectious-ToRecoveryCritical | Number of days the agent was infectious before recovery (critical cases) | 18.1 | (Verity et al., 2020) |
| Days_since_infected | Number of days since the agent was infected | Generated by NetLogo | - |
| will_develop-_symptoms? | Flag to indicate whether an agent will develop symptoms if infected | $\beta_2$ in tables 3.3 and 3.6 | (Verity et al., 2020) (Ferguson et al., 2020) |
| will_goto-_hospital? | Flag to indicate whether an agent will develop severe symptoms if infected | $\beta_3$ in tables 3.3 and 3.6 | (Verity et al., 2020) (Ferguson et al., 2020) |
| will_goto_ICU? | Flag to indicate whether an agent will develop critical symptoms if infected | $\beta_4$ in tables 3.3 and 3.6 | (Verity et al., 2020) (Ferguson et al., 2020) |
| will_recover? | Flag to indicate whether an agent will die if infected | $\beta_5$ in tables 3.3 and 3.6 | (O'Driscoll et al., 2021) (Brazeau et al., 2020) |

Table 3.4: Person agent attributes

around the environment from 12:00 until 21:00 during level 2 and from 18:00 until 21:00 during level 5. Since school and third level institution closures were widespread during this period and home-schooling was prevalent, students hours were assumed to be ad-hoc and shortened. All individuals traveled back to their assigned households at 22:00 every day where they stayed until at least 07:00 the next day. During weekends, the behaviour was different. All individuals were assumed to be not working and not studying. During weekends, all individuals were free to randomly roam around the environment from 07:00 until 21:00 during level 2 and from 18:00 until 21:00 during level 5.

If an individual had become infected and was asymptomatic, their behaviour did not change from the above. However, if such individuals were notified as being a close contact, they traveled back to their household and isolated while waiting for a PCR test. Individuals were assumed to not infect any susceptible individuals if they were isolating, and their compliance to isolation was assumed to be 100%. If an individual had become infected and was symptomatic, they traveled back to their household and isolated until a PCR test was taken. If an individual progressed to a severe case, they were transported to a local hospital where they were assumed to be isolating, and therefore could not infect another susceptible individual. Their disease progressed there until they recovered or died. Individuals that tested positive continued to isolate at home until they recovered or their case developed into a severe infection.

Any travel the individuals undertook in the simulations such as transport to/from work or to/from hospital is assumed to be instant. Therefore the individuals do not interact with other agents while traveling unless randomly roaming around the environment. Susceptible individuals can still become infected at work or at home if the infected person is not isolating.

This behaviour is described in an Agent Activity Cycle in Figure 3.2 and the movement of agents are described in more detail in Figure 3.3.

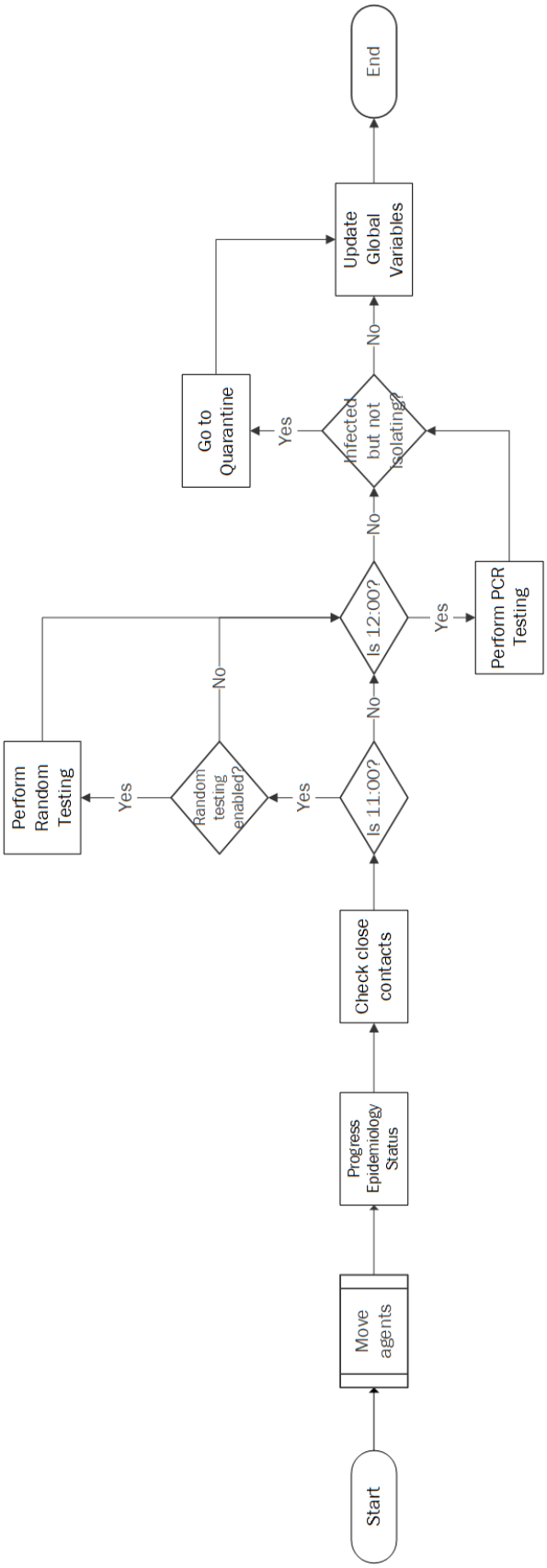



Figure 3.2: Agent activity cycle

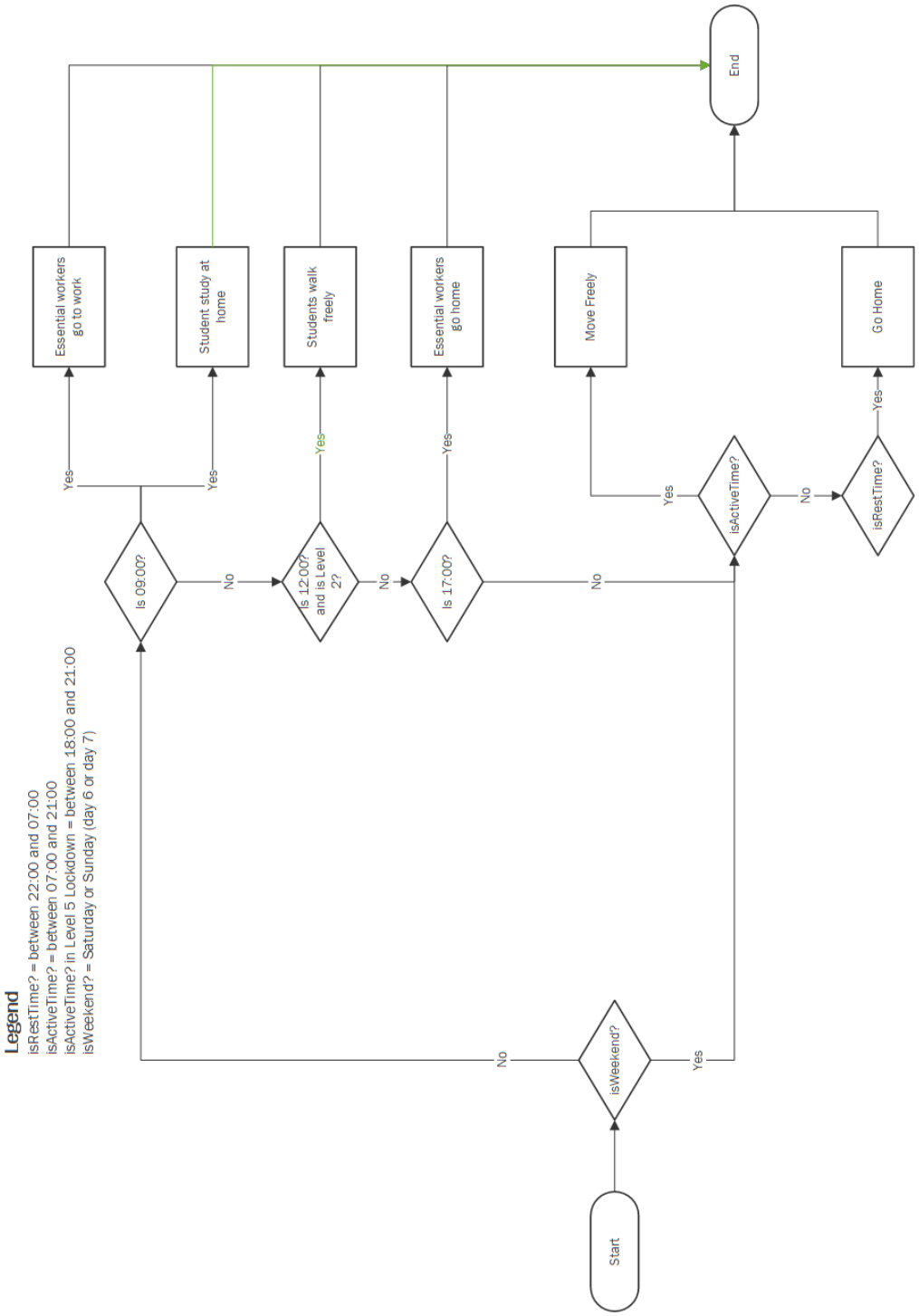



Figure 3.3: Agent movement process at every tick

| Modelling time in NetLogo | | |
|---|---|---|
| **Variable** | **Description** | **Values** |
| hours | Hour of the day (9 means 09:00) | 0 to 23 |
| days | Day number (1 means November 25th, 96 means February 28th) | 0 to 96 |
| isWeekend? | Sixth and seventh day of each week are assumed to be the weekend. This is calculated using the modulus | True or False |
| isWorkTime? | Work time is assumed to be between 09:00 and 17:00 for everyone | True or False |
| isActiveTime? | Agents can roam freely between 07:00 and 21:00 during level 2 restrictions except if they are working or isolating | True or False |
| isLevel5ActiveTime? | Agents can roam freely between 18:00 and 21:00 during level 5 restrictions except if they are working or isolating | True or False |
| isRestTime? | Agents are all at home to rest or sleep between 22:00 and 07:00 | True or False |
| lockdown_from_day_n | This parameter is used to trigger level 5 restrictions in the model and can be updated in the interface. When level 5 is triggered, active time gets severely reduced and agents spend much more time at home | 0 to 96 |

Table 3.5: Variables used to count time or flag certain periods of time in NetLogo

### 3.1.6 Modelling time

The NetLogo software counts the passing of time using "ticks" as a standardised measure, independent of all models or computer speeds (Tisue & Wilensky, 2004). Since the simulation was a series of iterations, a tick could be counted after or during each iteration. Each tick in the NetLogo environment was assumed to represent 1 hour, which was then rolled up to days using variables defined in the model. Day 1 was assumed to be Monday up to day 7, which was assumed to be Sunday. Weekdays were then defined as day number 1 to 5 and weekends were assumed to be days 6 and 7. Additional variables were used as flags to represent work time (09:00 to 17:00), rest time (22:00 to 07:00), active time (07:00 until 21:00), and level 5 active time (18:00 to 21:00). Some of these periods overlapped, for example, one individual may have been working from 09:00 to 17:00 and then had active time for leisure from then until 21:00, while a retired individual would have been free to enjoy active time from 07:00 until 21:00. These variables are listed in more detail in Table 3.5.

At the end of each iteration, a number of variables were updated which were subsequently used for reporting the results of the model. These are discussed in more

detail in Section 3.5.2. A number of model performance improvements were made by updating a series of agent sets at the end of each iteration. An agent set in NetLogo is a filtered subset of agents that can be generated during a simulation[14], for example, a list of all agents currently infected with COVID-19. However, it can be computationally expensive if every agent in the environment individually generates the same agent set for every tick. By updating a series of agent sets after each tick, agents can refer to them without having to generate them. This helped reduce a single run of a simulation from 12 hours down to 4 minutes (Railsback et al., 2017).

### 3.1.7 Modelling government responses

A number of Government responses policies were included in the model which could be either disabled or configured in the model interface. This allowed the model to be validated without any responses active as well as fine tune the parameters of the model. Each of the responses are described in more detail in the following sub sections.

#### Public health guidelines

Public health advice has been consistent across the world and was the first response to the emerging COVID-19 epidemic. This response included mask wearing, physical distancing, personal hygiene, and restriction of movement and gathering (Chu et al., 2020; Kennelly et al., 2020). To model this response the probability of infection, which is age stratified, was reduced to 20%. This response can be modified or disabled using the parameter, reduce_transmissibility_to.

#### Contact tracing

The close contact regime in Ireland was modelled by each infected individual, including asymptomatic cases, keeping track of every other individual that moved within a specific radius since they became infected. While this approach is a little naive in terms of people remembering every person they came into contact with in recent days,

[14]https://ccl.northwestern.edu/netlogo/docs/programming.html

this assumption was made since there did not seem to be reliable data in the literature to estimate what percentage of close contacts should be remembered.
Individuals listed as close contacts were not flagged as such until the infected individual was a confirmed case through a PCR test. These close contacts would, in turn, take a PCR test after at least one day, depending on testing capacity. If any of these close contact individuals tested positive, then their close contacts were flagged for a PCR test and so on. Contact tracing could be disabled in the interface using the binary variable, close_contact_tracing.

### PCR testing

Individuals can be flagged for a PCR test in one of two ways. Firstly, if an individual is infected with a symptomatic case, then they will be included in PCR testing at the next available time. Secondly, if an individual was flagged as a close contact, they will also be included for PCR testing at the next available time.

PCR testing is completed once per day at 12:00. This ensures that only a limited number of tests can be performed on any day, and also ensures that any individual flagged for testing will be tested at least one day later, which is aligned to the testing regime in Ireland during the time of this experiment. Testing capacity was limited to 180,000 PCR tests per week in Ireland at the time of writing[15], which is 25,000 per day for a population of 4,761,865 people, according to the 2016 Census[16]. The simulation model took the same proportion of tests and applied it to the sample of 4,843 agents in the model. This amounted to a total testing capacity of 26 tests per day. Individuals flagged for PCR tests were selected at random each day if the numbers to be tested exceeded capacity.
When random testing was introduced, the overall capacity did not increase, so as the capacity for random testing increased, the capacity for PCR testing decreased. This is discussed further in the following section.

[15]https://www.hse.ie/eng/services/news/media/pressrel/symptom-free-walk-in-covid-19-testing.html

[16]https://data.cso.ie/table/E3003

### Random testing

Random testing was not introduced as a response policy in Ireland since the start of the pandemic. However, walk-in testing centres did open from March 2021 and depending on available testing capacity do remain open to date. However, as of January 20$^{th}$ 2022, there are restrictions on who can avail of a walk-in test[17]. It is not feasible to constitute these walk-in centres as random tests since the decision to get tested remains with the individual. The proposal of this study is to test the hypothesis of whether truly random testing can have an effect on the spread of COVID-19 and ultimately reduce the number of associated deaths. The random testing regime in this model was built to randomly select a predefined number of individuals from the population every day excluding individuals that were already infected with a symptomatic case, recovered, or dead. The number of random tests per day are described as part of the experiment design in Section 3.5. The random tests were performed each day at 11:00. This ensured that the random tests were carried out before the standard testing regime already in place. The assumption made was that the random testing took preference over standard PCR testing. In reality, it is likely that this policy would shift if testing capacity was under strain from increased demand.

If a run of random testing identified an individual as positive for COVID-19, then they would travel to their home to isolate and their disease would progress as normal. That person's close contacts were also flagged for testing, which took place at least one day later depending on testing capacity. These close contacts would be tested as part of the standard PCR testing regime.

### Level 5 restrictions

Since elements of society such as public gatherings, events, visiting of households, and hospitality were not included in the model, the shift from Level 2 restrictions to Level 5 restrictions focused on the restriction of movement of individuals in the environment. During Level 2 restrictions, individuals were free to roam randomly

[17]https://www.citizensinformation.ie/en/health/covid19/testing_for_covid19.html

around the environment between 07:00 and 21:00 except for workers, who worked from 09:00 to 17:00, either from home or a designated workplace if they were essential workers. During Level 5 restrictions, individuals were free to roam randomly around the environment only from 18:00 until 21:00, thus significantly reducing the movement of individuals and also ensuring that individuals stayed reasonably local and did not travel far from their home. This also is aligned with travel restrictions during this outbreak.

**Working from home**

As described in Section 3.1.5, the economic status of individuals identified which individuals were employed. 80% of workers were assumed to work from home and 20% of workers were assumed to be essential workers and work from designated workplaces (Redmond et al., 2020). Their behaviour is also described in Section 3.1.5.

## 3.2 Data used to build the model

This section describes the datasets used in the building of the agent based model and how these datasets were pre-processed and loaded into the model. The sampling methodology is also described. The data pre-processing was completed in a Jupyter Notebook using Python, and has been generalised so that the county and sample size can be specified at the start of the script, and then a sample of that approximate size from that county will be generated. Details of the pre-processing are discussed further in the following sections.

### 3.2.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) is a European Law, in effect since May 25$^{th}$ 2018, that requires organisations to safeguard the personal data of EU citizens[18]. Personal data is defined as "any information that relates to an individual

[18]https://gdpr.eu/

who can be directly or indirectly identified"[19].

Information concerning health is classified as a special category and is subject to additional protection from the GDPR.

This study does not process any personal data of any individual. All data gathered and processed, including population, demographic, economic status, as well as data related to COVID-19 has been aggregated by the relevant data providers and cannot be used to identify any EU citizen, directly or indirectly.

### 3.2.2 Small Areas

The Central Statistics Office (CSO) lists many statistics on its website, several of which can be broken down by geographical area such as province, county, electoral division, town, and towns-land. The smallest of these geographical regions are called "small areas". When statistics are aggregated by small area, the county is not included in these datasets. Therefore, a separate dataset is required to map small areas to counties. The CSO provides this mapping dataset which was used to filter the list of small areas by county[20]. This filtered list of small areas was then joined onto the household dataset as described in the following section.

### 3.2.3 Household Data

The Small Area Population Statistics (SAPS) dataset is a set of population statistics produced from the Ireland Census 2016 and is aggregated at Small Area level[21]. The statistics produced are split across 14 themes from education to migration to disability to occupations. One such theme listed in this dataset was Theme 5 - Private Households. These statistics listed the density of households in each small area by household type. 13 different household types in the dataset were aggregated to five distinct types for this study in order to simplify the dataset. One person households

[19]https://gdpr.eu/what-is-gdpr/

[20]https://data-osi.opendata.arcgis.com/datasets/small-areas-generalised-20m-osi-national-statistical-boundaries-2015

[21]https://www.cso.ie/en/media/csoie/census/census2016/census2016boundaryfiles/SAPS2016_SA2017.csv

in the dataset were mapped to one person households for the study. Married couple households, cohabiting households, and couple with others in the dataset were mapped to couple households for this study. Married couple with children, cohabiting couple with children, and couple with children and others in the dataset were mapped to couple with children for this study. One parent family (father) with children, one parent family (mother) with children, one parent family (father) with children and others, and one parent family (mother) with children and others in the dataset were mapped to one-parent family for this study. Finally, non family households and relation, and two or more non-related person households in the dataset were mapped to non-related for this study.

The small areas dataset in Section 3.2.2 was then used in an inner join to filter the list of households by small area for a specific county. In the case of this study, county Carlow.
A sample of households was then created where each row represented one household. This sample matched the density of household types in each small area according to a proportion of the density in the SAPS dataset. In the case of this study, the sample size was set to 4,843 and the population of county Carlow is 56,932. Therefore, the sample of households in each small area was 8.5% of the actual density according to the SAPS dataset.

### 3.2.4 Population Data

Since the severity of COVID-19 depends on the age profile of the population, it was critical to ensure that this was representative as possible. The SAPS dataset was also leveraged to extract population statistics at small area level. Theme 1 - Sex, Age and Marital Status contains the number of people by age, by gender, by small area. Between the ages of 0 and 19, the number of people are listed for each age. From age 20 and upwards, the number of people are listed for age ranges. For example, 20-24, 25-29, 30-34 and so on.

A full scale sample population was generated for county Carlow where each row represented one person. This sample matched the population and gender attributes of the SAPS dataset, for example, if the SAPS dataset reported that 15 female people aged 54 were listed in small area abc, then 15 records were generated with attributes for gender, age, and small area populated as female, 54, and abc respectively. If the age statistic was listed for an age range, a random age was chosen between the upper and lower bounds of the age range. At the end, this full scale sample, listed 56,932 records for people with the age and gender distribution in each small area according to the SAPS dataset.

In order to simulate the behaviour of agents in the model, applying representative numbers of people that are working, retired, or were students was important. For this reason, the SAPS dataset was used again. This time, Theme 8 - Principle Status was used but not at small area level. The principle status was available in a separate dataset at an aggregate level which was then applied to the sample dataset[22]. In order to simplify the dataset used in this study, people age 5 and under are assumed to have a status of preschool and people and between 6 and 15 are assumed to have a status of student. Using the Theme 8 dataset a gender and age specific probability was used to determine whether people aged between 16 and 64 were students, unemployed, or at work. Any person older than 65 was assumed to have a status of retired.
Once the economic status was populated for the sample population, 20% of people at work were designated as essential workers (Redmond et al., 2020). All other people at work will be assumed to be working from home in the model.

### 3.2.5 Build sample population

In order to complete the population dataset, the full scale sample of the population was used to populate the sample of households. A sample population dataset was generated where each member was associated with a household according to each household type. One person households could be populated by any person aged be-

[22]https://census.cso.ie/sapmap2016/Results.aspx?Geog_Type=S&Geog_Code=S#SAPMAP_T8_801

tween 18 and 95. Couple households were populated with 2 people aged between 18 and 95 and within 10 years of each other. The dataset did assume that all couples were male and female. While this is not truly representative of the population, this bias does not affect the epidemiological modelling since the severity and spread of COVID-19 is not dependent on gender. One point to note here though is that there is a difference in economic status between males and females where a higher proportion of males are employed compared to females. While this does not directly affect the epidemiological status, it may affect the behaviour and movement of people in the model. Couples with children households were populated with 2 adults aged between 18 and 50 and within 10 years of each other. In addition, two children aged between 0 and 17 were added to each household. One parent family households were populated with one adult aged between 18 and 50 and one child aged between 0 and 17. Non-related households were populated with 3 adults aged between 18 and 95.

The performance of the NetLogo software is dependent on the number of agents in the simulation. During initial simulations, the performance of this model began to significantly deteriorate once the number of person agents in the model increased above 5000. County Carlow was chosen as the location of the outbreak to be modeled due to its relatively small size, and because a smaller sample of people would be more representative in terms of the distribution of attributes such as age, gender, and economic status, compared to a small sample from a more populous county such as Dublin or Cork. The target of 5000 agents was approximate since the pre-processing stage populated different types of households with different numbers of individuals. For this reason, the number of person agents actually imported into the model was 4,843.

The output of this script is three flat files. The first file contains information on the small areas which were used by NetLogo to import information for patches in the model. Secondly, a households file was generated in the script which was then used to import household agent information in the model. Finally, an agents file was generated by the script and used to import person agent information in the model.

### 3.2.6 COVID-19 Statistics

In order to validate that the model could create an accurate representation of a real world outbreak of COVID-19, real world data was required for comparison. County level statistics are available from the Ireland COVID-19 Data Hub and include confirmed cases of COVID-19 and population proportion of confirmed cases[23]. The methodology for validating the model is described in Section 3.4.

#### Outbreak

The outbreak chosen for this study is county Carlow between the dates December 1st 2020 and February 28th 2021. County Carlow is a relatively small county which would not require such a large sample to remain representative. The performance of the NetLogo software declines as more agents are used, and since the model was populated with approximately 5,000 person agents, this remains a reasonable percentage of the actual population.

### 3.2.7 Disease dynamics data

Section 3.1.4 describes how the disease is modeled in the NetLogo software and Table 3.3 lists the epidemiological parameters used to determine which path an agent takes through the progression of the disease, as well as the sources from the literature used. Table 3.6 lists the actual values from these sources which are used by the NetLogo software to progress the disease. These sources are taken from studies throughout 2020 but may not be guaranteed to derived from the alpha variant. The virus was still quite new during this time and the availability of variant-specific dynamics data was limited. When the model is initialised, as described in Section 3.3, this table of probabilities is used to set the agent attribute flags that control the progression of the disease. These agent attribute flags are described in Table 3.4.

[23]https://data.gov.ie/en_GB/dataset/covid-19-hpsc-county-statistics-historic-data?package_type=dataset

| Epidemiology parameter values by age | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-9 | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 | 90+ |
| $\beta_1$ | 0.34 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.24 | 1.47 | 1.47 |
| $\beta_2$ | 0.5 | 0.55 | 0.6 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.90 |
| $\beta_3$ | 0.00050 | 0.00165 | 0.00720 | 0.02080 | 0.03430 | 0.07650 | 0.13280 | 0.20655 | 0.24570 | 0.24570 |
| $\beta_4$ | 0.00003 | 0.00008 | 0.00036 | 0.00104 | 0.00216 | 0.00933 | 0.03639 | 0.08923 | 0.17420 | 0.17420 |
| $\beta_5$ | 0.00002 | 0.00002 | 0.00010 | 0.00032 | 0.00098 | 0.00265 | 0.00766 | 0.02439 | 0.08292 | 0.16190 |

Table 3.6: Epidemiology parameter values by age

## 3.3 Initialise the model

Once the data were pre-processed and exported to flat files, they were imported into the NetLogo environment during an initialisation process before each simulation run was executed. The details of the initialisation process are discussed in the following sections.

### 3.3.1 Global variables

A procedure in the NetLogo environment was used to initialise several groups of global variables. Firstly, the age-stratified epidemiology probabilities were imported into lists. These included the variables $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$ from Table 3.6.
A group of reporting variables were also initialised. These reporting variables were used to count various aspects of a simulation while it was being run and report these back to the user during execution, as well as being used to generate experiment results. Additional variables, described in Table 3.7, were also initialised and used in the modelling process.

Performance improvements in the simulation were also initialised here by generating a series of so-called agent sets. These agent sets are simply lists of agents and are initialised at the beginning of the simulation as well as after each tick. These pre-calculated agent sets remove the need for every agent to calculate the same list of agents for each tick, thus vastly improving the efficiency and performance of the model. The following agent sets were initialised. Firstly, a set of agents who are currently infected with COVID-19 and are not currently isolating. This can be used by

| **Global Reporting variables** | | |
|---|---|---|
| **Variable** | **Description** | **Initialisation** |
| dailyConfirmedCasesList | List of confirmed cases each day | Set to empty list |
| dailyActualCasesList | Number of new exposed individuals each day | Set to empty list |
| dailyConfirmedCases | Number of confirmed cases today | Set to 0 |
| dailyActualCases | Number of new exposed individuals today | Set to 0 |
| dailyPCRTestsConducted | Number of PCR test conducted today | Set to 0 |
| cumulativeConfirmedCases | Number of cumulative confirmed cases since the simulation started | Set to 0 |
| cumulativeConfirmedRandomCases | Number of cumulative confirmed cases identified using random testing | Set to 0 |
| cumulativeNegativeRandomCases | Number of cumulative random tests performed with a negative result | Set to 0 |
| cumulativeNegativeRandomCases | Number of cumulative random tests performed with a negative result | Set to 0 |
| **Variables used in the simulation** | | |
| PCR_sensitivity | Probability of PCR test correctly identifying positive case of COVID-19 | Set to 0.87 |
| ireland_population | Used to calculate testing capacity as proportion of population | Set to 4,761,865 |
| ireland_testing_capacity | Used to calculate testing capacity as proportion of population | Set to 25,000 |
| testing_capacity | Testing capacity in the simulation | Proportion of simulation population |
| daily_random_tests_percentage | Used to reserve a proportion of testing capacity to random testing | Set by parameter random_tests_percentage |
| daily_CC_tests_remaining | Number of PCR tests remaining today | Set to testing_capacity |
| in_Lockdown? | Flag used to indication Level 5 restrictions are in place | Set to False |

Table 3.7: Global variables initialised before each simulation run in NetLogo

each agent to determine if any infected agents are within a close radius which may cause them to become exposed to the virus. Secondly, an agent set is created containing asymptomatic agents and agents infected with a mild case that are not isolating. This agent set is used during contact tracing to track contact with other agents. Finally, an agent set is created which contains all susceptible agents in the environment and is also used in the close contact tracing function.

### 3.3.2 Import small areas

Each patch in the NetLogo corresponded to a small area as defined by the CSO and discussed in Section 3.2.2. A procedure was created to import the small areas text file generated in pre-processing and populate each patch specific attribute, patch_small_area, with a small area id. Since the small areas were sorted before being exported to the text file, this ensured that small areas geographically close to each other in the CSO data were mapped close to each other in the NetLogo environment. In reality, the small areas vary by size and shape but these were not implemented in this model due to scope constraint. Each patch in the model was the same size and shape but adjacent small areas from the CSO data were also adjacent in the model.

### 3.3.3 Import households

Next the households were imported from the text file from pre-processing. The file contained the attributes of each household as per Table 3.2. A household agent was created for each row in the file and the following household attributes were populated: small_area_id, small_area_name, householdtype, and familyid. In addition, attributes specific to the NetLogo environment were also initialised including the shape, colour, and size of the agent as well as the is_hospital flag. Finally, the households were distributed randomly within each patch.

### 3.3.4 Import people

The final file imported was the agents text file. This file contained many of the attributes of each agent in the simulation. A person agent was created for each row in the file and the following person agent attributes were populated. The gender, age, economic status, person_id, family_id, and whether the person was an essential worker or not. Other NetLogo specific attributes were also initialised which included the shape, colour, size, and x and y coordinates.

Finally, epidemiological attributes and flags were initialised according to the values in Table 3.4. While the number of days for each stage remained static for each run of the simulation, the flags used to control how an agent progressed through the disease (will_develop_symptoms, will_goto_hospital, will_goto_icu, and will_recover) were initialised according to the probabilities $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$ from Table 3.6, which ensured the disease dynamics remained stochastic.

### 3.3.5 Create workplaces and hospitals

Workplaces and hospitals were not created in the pre-processing stage and imported into the model. Instead, they were created by NetLogo in the initialisation process as follows. Since workplaces and hospitals are simply buildings in the model, 100 household agents were created as workplaces and 1 household agent was created as a hospital. Since only 20% of individuals at work were classified as essential workers, this amounted to 356 individuals. The number of workplaces was set to 100 as a rough estimate of employees per workplace according to a report on business in Ireland by the CSO[24]. Only industry, construction, and distribution categories of workers were taken into account. Healthcare workers were not identified in the model due to scope constraints and a lack of reliable data. Healthcare workers would all have been classified as essential workers and would also have been at a higher risk of becoming exposed to the disease. Due to the small size of county Carlow, there is only one

[24] https://www.cso.ie/en/releasesandpublications/ep/p-syi/statisticalyearbookofireland2019/bus/businessinireland/

hospital. These households were populated with the is_workplace and is_hospital flags respectively. In addition, appropriate values for NetLogo agent specific attributes such as shape, colour, and size were set. Workplaces were set as a blue factory shape and the hospital was set as a white $H$ symbol.

### 3.3.6 Assign people to households

The pre-processed data included a person attribute, family_id, which represented the household_id which the person was assigned to. This acted like a primary key / foreign key relationship in a relational database, and was used to associate person agents with household agents. Even after these attributes were populated, the NetLogo software still required this association to be made since it automatically generated its own ids for both sets of agents. Therefore, a procedure was written to associate the NetLogo id of each household agent with the corresponding person agents according to the imported ids.

In addition, person agents that were identified as essential workers were randomly assigned to workplaces.

### 3.3.7 Set initial number of infected individuals

In terms of setting the number of infected agents, historical data on COVID-19 was used to determine the 7 day average of confirmed cases as 4.7[25]. The exposed period of COVID-19 set to 4.5, so the outbreak required at least this many days to get started. Therefore a period of 7 days was set to allow the exposed individuals to progress their infections, so that by the time that December 1$^{st}$ was reached the outbreak was already in progress. Since this study is attempting to model an outbreak while the disease is already in the community, it was necessary to have this run in period. Therefore the initial number of infected individuals was set to 20.

[25]https://data.gov.ie/enGB/dataset/covid-19-hpsc-county-statistics-historic-data?packagetype=dataset

## 3.4 Validate the model

Before any experiments could be conducted, a number of validation steps were performed in order to ensure that the model was an accurate representation of a COVID-19 outbreak in county Carlow between the dates December 1st 2020 and February 28th 2021. The model validation was performed by running experiments in the BehaviourSpace module in NetLogo and analysing the results using Python in a Jupyter notebook. The literature did not present a standard methodology for validating agent based models. However, Hunter & Kelleher (2020) describe a framework for validating agent based models, using epidemiology models in case studies, which was leveraged for the validation of the agent based model used in this study. These validation steps are detailed in the following sections.

### 3.4.1 Behaviour Space in NetLogo

Behaviour Space is a tool within the NetLogo software designed to run experiments. The tool runs the model simulation a predefined number of times, systematically varying parameter values, and recording results along the way[26]. These results were exported to a csv flat file and were then read by a Python script for further analysis.

### 3.4.2 Selecting the number of runs

The number of runs during the experiments varied in the literature. Hunter & Kelleher (2020) propose the use of confidence intervals to determine how many runs of the experiment would be sufficient to produce a representative estimation of a calculated statistic. An assumption was made where only one scenario was simulated but the result was applied to every scenario in the experiment. This scenario was the base scenario in the model where all interventions were enabled except for the random testing element, which was the subject of the experiment. The model was run a pre-defined number of times which was much more than would be expected to calculate the statistic. Once the results were generated, samples were taken, where the

[26]https://ccl.northwestern.edu/netlogo/docs/behaviorspace.html

first sample contained the first 5 rows, the second sample contained the first 10 rows, the next sample contained the first 15 rows and so on until the final sample contained all rows from the results dataset. For each sample, the 95% confidence interval was calculated and plotted. A decision was made on the number of runs which provided sufficient confidence that the calculated statistic was close to the true value.

For this study, two statistics were estimated using this approach. Firstly the percentage of runs where the wave of infection was large enough to last the duration of the period (90 days). The second statistic was the percentage of runs where the wave of infection was large enough that approximately 5% of the population was confirmed as testing positive for COVID-19. In both cases, Formula 3.1 was used to calculate the confidence interval.

$$CI = \hat{p} \pm z * \sqrt{\frac{\hat{p} * (1 - \hat{p})}{n}} \tag{3.1}$$

Where $\hat{p}$ represents the proportion of runs for each statistic, $n$ is the sample size, and $z$ is the statistic for the 95% confidence interval (1.96). The results of this analysis are discussed in Section 4.1.1.

### 3.4.3 Cross Validation

The first validation of the model was cross validation against another previously validated model. The validated model was an equations based model which used the same selected parameters from the literature. The equation based model was considered to be validated because the standard equations described in Section 2.2.2 were used. This validation was used to test whether a basic version of the agent based model would behave as an SEIR model was expected to, using the selected parameters. For this reason, the government responses in the agent based model were disabled when being validated against the equation based model. If the outputs of the two models were similar, this would provide evidence that the basic version of the agent based model did behave as an SEIR model should (Hunter & Kelleher, 2020).

To generate an equations based model, a Python script was used to leverage the differential equations described in Section 2.2.2[27] [28]. The equations were populated with the same parameters used in the agent based model which produced a set of response variables, $SEIR$. Each time step represented one day.

An experiment was designed for the agent based model which ran 10 simulations with all response parameters disabled. This included the testing function, contact tracing function, no lockdown, no reduction in transmissibility, and no random testing. The experiment recorded the number of agents in each compartment at the end of every tick, with each tick representing one hour. The results were exported to a csv file.

Before the two models could be compared, time had to be standardised between the two. The equations based model was based on days but the agent based model was based on hours. A days column was created in the agent based model results dataset by integer dividing the number of ticks by 24. The results were then aggregated using the mean for each compartment. Finally some compartments were combined to aggregate up to the same level as the equations based model, i.e. $I^A$, $I^M$, $I^S$, and $I^C$ were aggregated to $I$, while $R$ and $D$ were aggregated to $R$, to correspond to the equations based model. To evaluate the similarity, both the Pearson and Spearman correlations were calculated and a number of plots were generated. The results of this analysis are discussed in Section 4.1.2.

### 3.4.4 Sensitivity Analysis

The cross validation against an equations based model was the first step in validating the agent based model by testing its fundamental behaviour. Next, sensitivity analysis was used to determine if the response from the agent based model to parameters was appropriate. The parameter values may be taken to extremes which may not reflect the real world, but unexpected behaviour identified can be used to investigate

[27] https://github.com/henrifroese/infectious_disease_modelling/blob/master/part_two.ipynb

[28] https://towardsdatascience.com/infectious-disease-modelling-beyond-the-basic-sir-model-216369c584c4

potential issues in the model (Hunter & Kelleher, 2020).
The parameters used in the sensitivity analysis were those which related to the government response policies. i.e. the rate of reduction in transmissibility, the day in which lockdown took effect, and the enabling and disabling of close contact tracing and PCR testing.

To perform the sensitivity analysis, three separate experiments were conducted using the Behaviour Space module in NetLogo. The response variables for each experiment were the number of agents in each compartment at the end of each step in the simulation, and the results of each experiment were saved as a csv file, which a Python script used to perform the analysis. The first experiment focused on when Level 5 lockdown took effect. The values for this parameter were systematically varied through the set of integers 10, 30, 50, 70, and 90 days with all other parameters remaining unchanged. 100 simulation runs were conducted for each value of the lockdown_from_day_n parameter, which was 500 runs in total. The second experiment toggled the PCR testing parameter on and off as well as the close contact tracing parameter on and off, with all other variables remaining unchanged. 100 runs of the simulation were conducted for each variable value, which amounted to 400 runs in total. The final experiment for the sensitivity analysis focused on the reduction of transmissibility to simulate public health measures. The value of the parameter reduce_transmissibility_to was varied using the values 20%, 40%, 60%, 80%, and 100%, where 100% was no reduction in transmissibility. All other parameter values were held constant.

To evaluate the impact of the parameter values from each experiment, the values of some compartments were aggregated together, i.e. $I^A$, $I^M$, $I^S$, and $I^C$ were aggregated to $I$, while $R$ and $D$ were aggregated to $R$, to correspond to a standard SEIR model. The values for each compartment were then grouped by the parameter being tested, and the mean of agents in each compartment across the 100 runs was calculated. The results were then plotted and summarised in tables. The results of this validation are discussed in Section 4.1.3.

### 3.4.5 Comparison to real world data

Once the sensitivity analysis was completed, the final step in the validation process was a comparison to real world data. Due to the complex nature of a stochastic model, each run of a simulation will produce a different result, the majority of which will not exactly match the real world data. Hunter & Kelleher (2020) propose that when comparing to real data, the comparison should be to a distribution of simulated outcomes, such that the real data could be a likely sample from that distribution.
In the case of this study, an experiment was created in the Behaviour Space module in NetLogo which ran 300 simulations with the model parameter values set to the base model, which is described in Section 3.5. The response variable for the experiment was the cumulative number of confirmed cases over the course of the 90 day outbreak, and the results were saved to a csv file which was used by a Python script to perform the analysis.

The real world data used was the number of confirmed cases of COVID-19 per day according to the dataset described in Section 3.2.6[29]. This data was specific to county Carlow between December 1$^{st}$ 2020 and February 28$^{th}$ 2021.

Since the simulated data were recorded at an hour level (i.e. per step) and the real world data were at a day level, the simulated data were aggregated up to day level by integer dividing the steps column by 24. The number of confirmed cases per day were created by calculating the difference between the current day's cumulative confirmed cases and the previous day. Finally, since the simulations were a sample from the population of county Carlow, the cases per day were scaled up according to the sample size in the model.

To compare the simulated data to the real world data, three comparisons were made. Firstly, a test was performed to check whether the two sets of data were from the same

[29]https://data.gov.ie/en_GB/dataset/covid-19-hpsc-county-statistics-historic-data?package_type=dataset

distribution. Secondly, Pearson and Spearman correlations were calculated between the two sets of data, and finally, the distribution of daily cases were compared for each run of the simulation to check whether the real data could be a likely sample from the distribution of simulations. These results are discussed in Section 4.1.4.

## 3.5 Design the experiment

Once the model had been validated, it was ready to be used in an experiment to generate reliable, trustworthy results. The purpose of the experiment outlined here was to test the hypothesis that the number of COVID-19 related deaths in an Irish county is significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing. The following sections describe how this hypothesis was tested.

### 3.5.1 Testing Capacity

As described in Section 3.1.7, the total testing capacity for the model was calculated to be 26 tests per day. This number was calculated as a proportion of the actual testing capacity, 25,000 PCR tests per day, based on the sample size of person agents in the model. To avoid the alternate hypothesis becoming trivially accepted, the introduction of random testing into the experiment did not simply add additional testing capacity to the model. Instead, each scenario varied the percentage of testing capacity which was reserved for random testing. During the simulations, the random testing was performed each day before the standard PCR testing based on the close contact testing regime, which gave preference to the random testing if the testing capacity was strained due to high demand.

### 3.5.2 Experiment setup

The Behaviour Space module within the NetLogo software was used to design the experiment. The parameters used in the experiment and their values are listed in

| Model Parameters | |
|---|---|
| **Parameter** | **Value** |
| close_contact_tracing | True |
| PCR_testing | True |
| reduce_transmissibility_to | 20% |
| lockdown_from_day_n | 31 |
| agent_size | 0.4 |
| random_tests_percentage | 0%, 10%, 20%, 30%, 40%, 50% |

Table 3.8: Parameters used in the experiment

Table 3.8. Close contact tracing and PCR testing were enabled to model the standard testing regime in the model. The transmissibility was reduced to 20% to model public health measures such as physical distancing, hand-washing, and mask wearing. Level 5 restrictions began after 31 days which represented December $26^{\text{th}}$ 2020, which was set by the lockdown_from_day_n parameter. The agent_size parameter controlled the physical size of the agents, households, hospitals, and workplaces in the model, as well as the radius between agents that triggered exposure to the disease and close contacts. This parameter was set to 0.4. Finally, the random_tests_percentage parameter was varied systematically through the values 0, 0.1, 0.2, 0.3, 0.4, and 0.5 which represented 0%, 10%, 20%, 30%, 40%, 50% respectively. When the random_tests_percentage was set to 0%, this was considered to be the base model. This represented the model with all modelled government interventions enabled using the standard test and trace regime, and this is the model which all subsequent scenarios would be compared to.

A number of scenarios were modeled for each value of the random_tests_percentage parameter, while all other parameters in the model were held constant. Each scenario was run 300 times, which amounted to 1800 runs in total.
The response variables recorded during the experiment are listed in Table 3.9. These response variables were measured at every step in every simulation. The results of the experiment were exported to a csv file and a Python script was used to generate the analysis. The results are discussed in detail in Section 4.2.

| Model Parameters | |
|---|---|
| **Response Variable** | **Description** |
| SusceptibleAgents | The number of susceptible agents |
| ExposedAgents | The number of exposed agents |
| InfectedAAgents | The number of infected asymptomatic agents |
| InfectedMAgents | The number of infected agents with a mild infection |
| InfectedSAgents | The number of infected agents with a severe infection |
| InfectedCAgents | The number of infected agents with a critical infection |
| RecoveredAgents | The number of agents recovered from infection |
| DeadAgents | The number of agents that have died from their infection |
| CumulativeConfirmedCases | The number of cumulative confirmed cases (PCR + random testing) |
| CumulativeConfirmedRandomCases | The number of cumulative confirmed cases (random testing only) |
| CumulativeNegativeRandomCases | The number of cumulative negative test results from random testing |
| DailyPCRTestsCompleted | The number of PCR tests conducted during the current day so far |

Table 3.9: Response variables returned by the experiment

### 3.5.3 Evaluation of experiment results

In order to evaluate whether a difference in the number of deaths exists between each scenario, a statistical test is required. The type of statistical test used depends on the measurement of the variable and the shape of the data. In the case of this study, the sample of the base model, where no random PCR tests were performed, was compared to each of the scenarios in which the percentage of random PCR tests was varied. In other words, the base model was compared to each of the models where 10%, 20%, 30%, 40%, and 50% of the testing capacity was allocated to random testing. While Analysis Of Variance (ANOVA) can be used to test the difference between 3 or more means, this study is not focused on comparing the samples with differing percentages of random testing to each other, but rather comparing each of them to the base model.

The test chosen was the difference between two means. An independent samples test was used since these are two different groups of agents. While the same agents are in the environment in each scenario, their epidemiological attributes will differ for each run of the simulation. In addition, the data are not reported at an agent level with a value for each agent for each scenario, but instead, the data are reported as

aggregated totals. There are two types of tests for comparing independent samples. An independent samples t-test is a parametric test, which is based on the assumptions of normality and homogeneity of variance. These assumptions can be tested and if the two assumptions do not hold, the Mann-Whitney U-Test is used since it is a non-parametric test and does not rely on these assumptions. Therefore, before the test was performed, tests for normality and equal variance were performed for all samples before the difference in means was compared.

**Test for normality**

To test if the sample data were normally distributed, the data from each sample were plotted in a histogram for a visual check, followed by the Shapiro-Wilk statistical test for normality, which is appropriate for samples containing thousands of observations or fewer. The Null Hypothesis of the Shapiro-Wilks test is that the sample is from a normal distribution and the alternate hypothesis is that the sample is not from a normal distribution[30]. If the p-value was less than than $\alpha$, where $\alpha = 0.05$, then the null hypothesis was rejected and the sample was not considered to be normally distributed.

**Test for equal variance**

To test for equal variance, Levene's test was used for all samples. The Null Hypothesis of the Levene test is that the samples have equal variance[31]. The alternate hypothesis is that the samples do not have equal variance. If the p-value was less than $\alpha$, where $\alpha = 0.05$, then the null hypothesis was rejected and the samples were shown to not have equal variances.

**Test for difference of means**

The Alternate Hypothesis of this study is that the number of COVID-19 related deaths in an Irish county is significantly higher in a population using a regime of close-

[30]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html
[31]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levene.html

contact tracing than in a population augmented with random testing. Therefore, this required a one-sided test, $D_C > D_R$, where $D_C$ is the number of deaths in the regime of close contact tracing and $D_R$ is the number of deaths from the regime augmented with random testing.

If the sample data were normally distributed and had equal variance, then an independent t-test was used[32]. The Null Hypothesis of a t-test is that the means of the two samples are equal and the Alternate Hypothesis is that the means are not equal. If the p-value is less than $\alpha$, where $\alpha = 0.05$, then the null hypothesis is rejected in favour of the alternate and the means are assumed to be not equal.

If the sample data were not normally distributed or did not have equal variance, then a Mann-Whitney U-test was performed[33]. The assumptions of the Mann-Whitney U-test of random and independent samples were held. The Null Hypothesis of a U-test is that the means of the two samples are equal and the Alternate Hypothesis is that the means are not equal. If the p-value is less than $\alpha$, where $\alpha = 0.05$, then the null hypothesis is rejected in favour of the alternate and the means are assumed to be not equal.

[32]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mstats.ttest_ind.html
[33]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html

# Chapter 4

# Results, evaluation and discussion

This chapter describes the actual work carried out in this study and focuses on the presentation and discussion of the findings from the validation of the model as well as the main experiment. The validation consisted of a number of smaller experiments which started at a basic level and built up to a validation of a fully functional model compared to real world data. The chapter also discusses the research hypotheses and how the experiment was used to test them.

## 4.1 Results of Validation

The following section discusses the results of the validation of the model before the experiment was run.

### 4.1.1 Validate the number of runs

Two statistics were estimated in order to validate the number of runs in the experiment. The first statistic was the percentage of runs where the number of confirmed cases reaches at least 5% of the sample population. The second statistic was the percentage of runs where the wave of infections continues throughout the complete period of the experiment. In each case, the base scenario was run 700 times, which is more than what would be expected to calculate each statistic.

### Outbreak size

The statistic was calculated at the last step of each run by populating a full_outbreak value for each each run with a value of 1 in cases where the number of confirmed cases of COVID-19 as a percentage of initial susceptible agents was greater than or equal to 0.05 (i.e. 5%). Otherwise, each run was populated with a value of 0. Samples of runs were then taken at intervals of 5, where the first sample takes the first 5 runs, then the second sample take the first 10 runs and so on until all runs are included in the final sample. The statistic was then calculated by taking the mean of each sample, with upper and lower confidence intervals using Equation 3.1. The samples are shown in Figure 4.1 and Table 4.1.

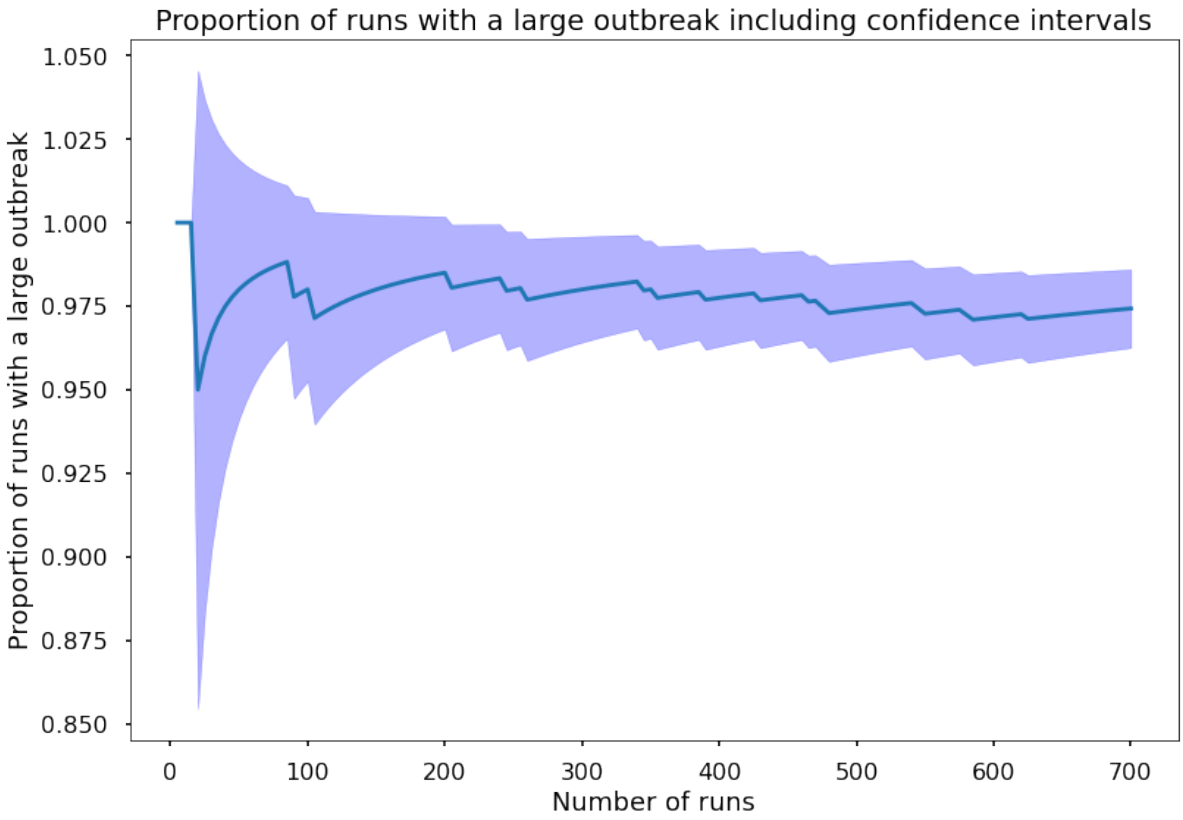



Figure 4.1: Number of confirmed cases reaches at least 5% of the sample population

The calculated statistic remained at approximately 97.5% of runs where the percentage of confirmed cases was greater than or equal to 5%. During the first 100 runs,

| Percentage of runs where outbreak infected at least 5% of Susceptible agents | | | |
|---|---|---|---|
| **Runs** | **Lower CI** | **Statistic** | **Upper CI** |
| 50 | 0.941 | 0.980 | 1.019 |
| 100 | 0.953 | 0.980 | 1.007 |
| 200 | 0.968 | 0.985 | 1.002 |
| 300 | 0.964 | 0.980 | 0.996 |
| 400 | 0.963 | 0.976 | 0.992 |
| 500 | 0.960 | 0.974 | 0.988 |
| 600 | 0.958 | 0.972 | 0.985 |
| 700 | 0.963 | 0.974 | 0.986 |

Table 4.1: Percentage of runs considered large enough for the simulation

the statistic increased from 0.98 with confidence intervals of 0.94 and 1.02, to a statistic of 0.98 with confidence intervals of 0.95 and 1.01. At 700 runs, this had reduced to 0.96 with confidence intervals of 0.974 and 0.986. At 300 runs, the statistic was 0.96 with confidence intervals of 0.98 and 0.99 which seemed to achieve a very close result to 700 runs without the additional computational overhead of running an additional 400 runs.

### Outbreak length

The second statistic was calculated at the last step of each run by populating each run with a value of 1 in cases where the step number was greater than or equal to 2300. This step number represented the number of hours since the experiment began, which was 95.8 when converted to days. Similarly to the first statistic, samples of runs were taken at 5 run intervals and the statistic was calculated by taking the mean of each sample with upper and lower confidence intervals, as shown in Figure 4.2.

The value of this statistic started very high at 100% for the first 100 runs and then seemed to stabilise at 0.988 with confidence intervals of 0.981 and 0.996 at 700 runs. However, similarly with the first statistic, at 300 runs, the statistic settled very close to these numbers with a statistic value of 0.987 with confidence intervals of 0.974 and 0.999. Table 4.2 and Figure 4.2 show these results.

Since both statistics had accurate estimations of the statistic at 300 runs, the experiment was run 300 times for each scenario. These results gave a high degree of

| Percentage of runs where outbreak lasted 95 days or more | | | |
|---|---|---|---|
| **Runs** | **Lower CI** | **Statistic** | **Upper CI** |
| 50 | 1.000 | 1.000 | 1.000 |
| 100 | 1.000 | 1.000 | 1.000 |
| 200 | 0.976 | 0.990 | 1.004 |
| 300 | 0.974 | 0.987 | 0.999 |
| 400 | 0.977 | 0.987 | 0.998 |
| 500 | 0.976 | 0.986 | 0.996 |
| 600 | 0.980 | 0.988 | 0.997 |
| 700 | 0.989 | 0.989 | 0.996 |

Table 4.2: Percentage of runs with an outbreak long enough for the simulation

confidence that the model accounted for stochasticity in the model. An assumption made in this methodology is that the other similar scenarios produced similar results within 300 runs also.

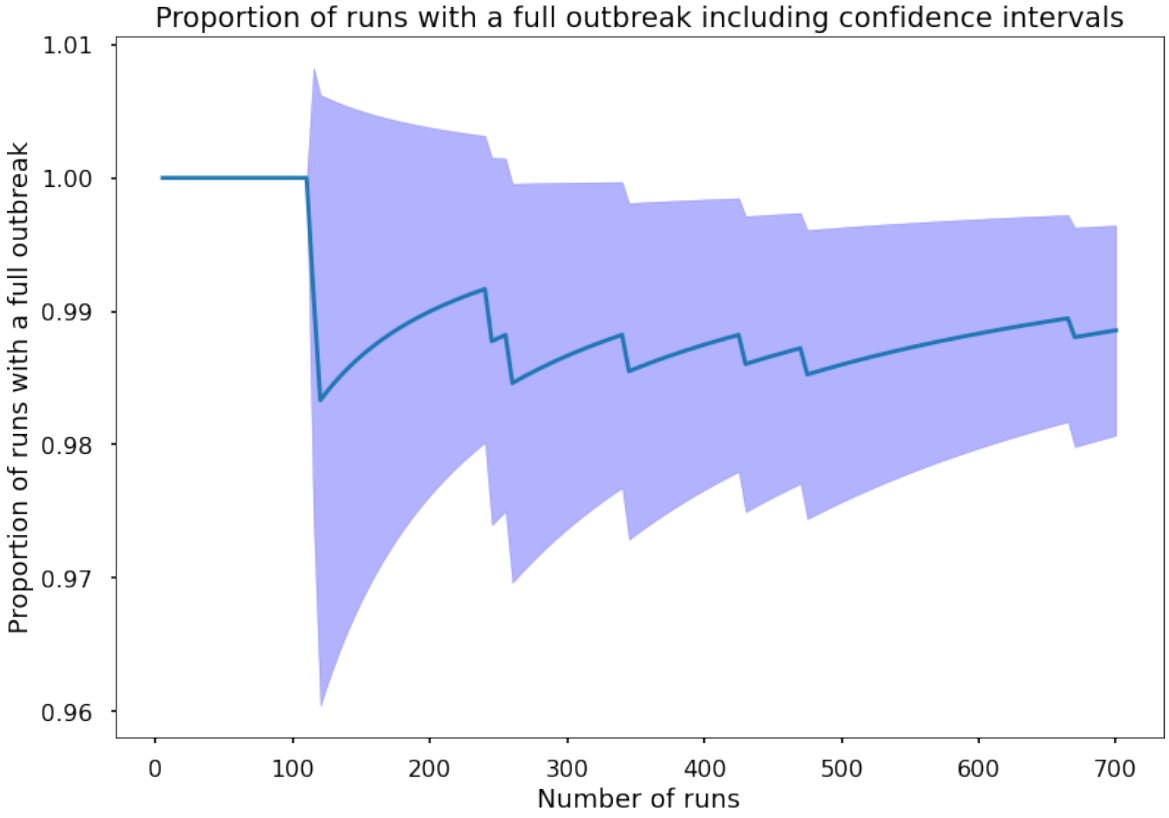



Figure 4.2: Number of confirmed cases reaches at least 5% of the sample population

| Spearman correlation between agent based and equation based models | | |
|---|---|---|
| **Compartment** | **Correlation** | **P-value** |
| Susceptible | 0.9983 | 3.376e-32 |
| Exposed | 0.6591 | 8.936e-14 |
| Infected | 0.5465 | 4.067e-09 |
| Removed | 0.9995 | 5.893e-150 |

Table 4.3: Correlations between agent based and equation based models

### 4.1.2 Cross validation with equation based model

The results of both models were plotted and shown in Figure 4.3 and separated into individual compartments in Figure 4.4. 10 runs of the agent based model with no interventions enabled, including testing and tracing, level 5 lockdown, reduction in transmissibility, and random testing. This resulted in the majority of susceptible agents in the model becoming infected during the outbreak for each run of the simulation. While this does not yet reflect the real world outbreak, it did allow the basic dynamics of the model to be compared to a validated model, which in this case was an equations based model. Figure 4.3 shows that while both models do not produce exactly the same results, the dynamics are very similar. The Susceptible and Removed compartments follow the same path almost exactly and the Exposed and Infected compartments follow a similar path with the peaks being slightly earlier in the agent based models. In both models, the peak of the Exposed compartment was earlier than the peak of the Infected compartment and the size of the Infected peak is approximately twice as large as the Exposed peak. The same variables were used as input to both models which were taken from the literature as shown in Table 3.4 and discussed in Section 3.1.4.

In addition to plotting both models for comparison, the output of both models was compared using the Spearman correlation. Spearman was preferred over Pearson because the Pearson correlation measures the linear relationship between two variables, while Spearman measures the monotonic relationship. In other words, Spearman measures the relationship, but allows for different rates of change. Figure 4.3 shows

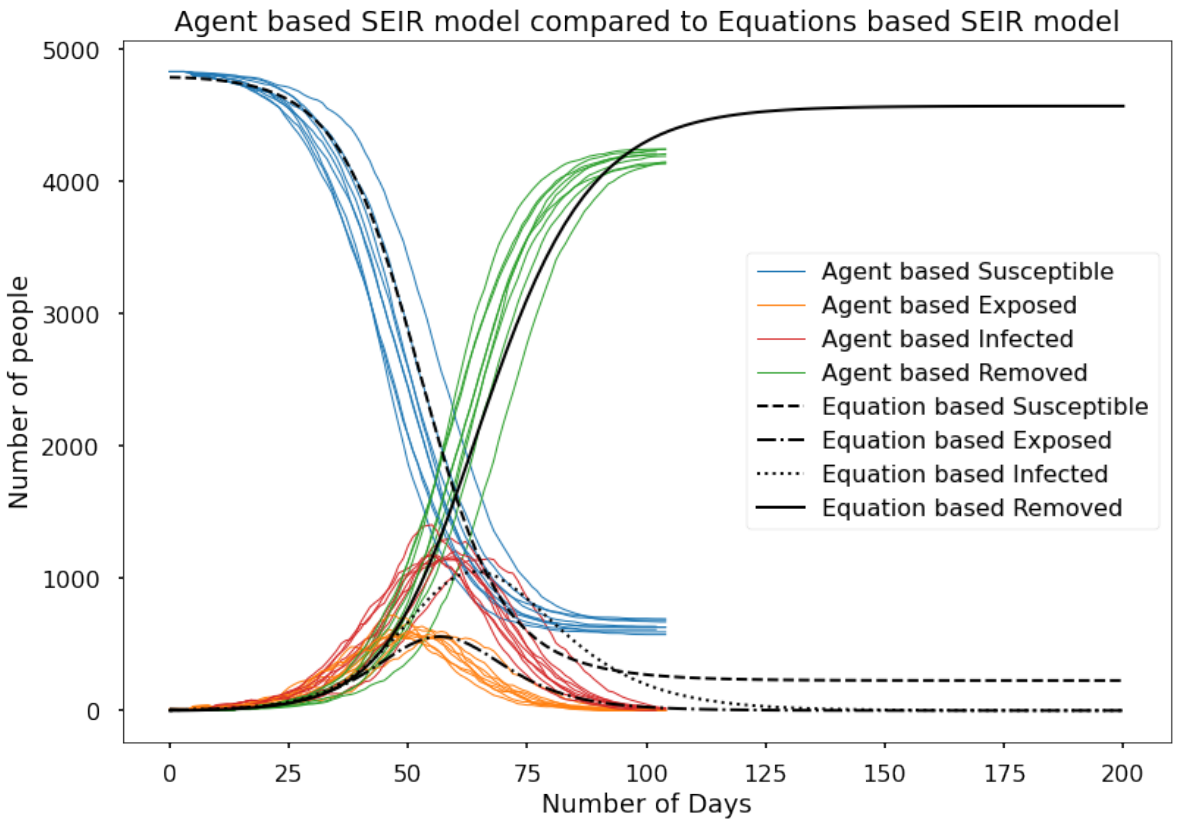
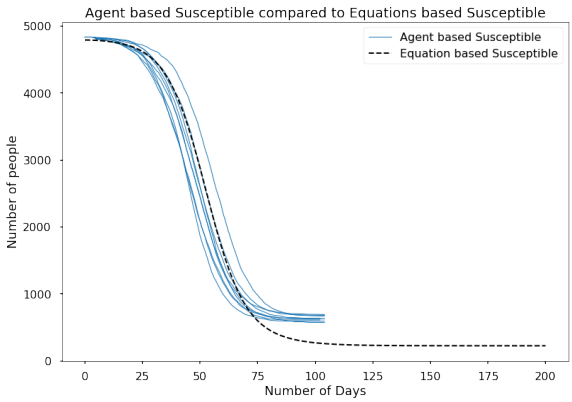


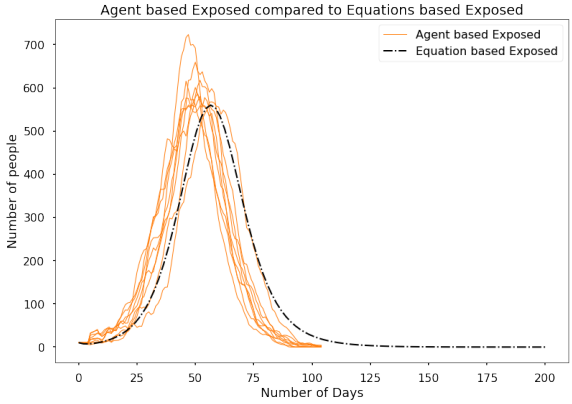

Figure 4.3: Cross validation of Agent and Equation based models

that the shape of each compartment over time is not linear and so Pearson is not appropriate to use.

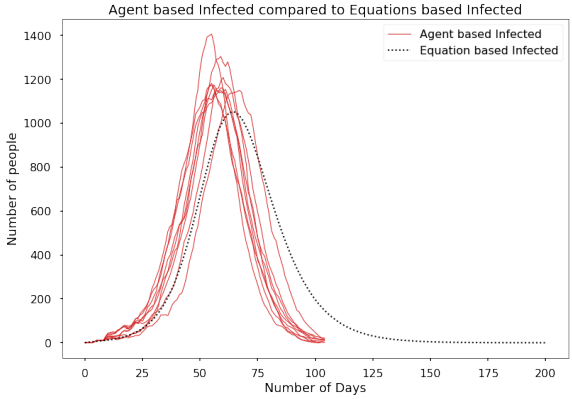
The response variables in the agent based model were combined in order to ensure they corresponded to the SEIR compartments from the equation based model. Each of the asymptomatic, mild, severe, and critical infected compartments were combined into an Infected compartment and the recovered and dead compartments were combined into a Removed compartment. The agent based model also returned values for each compartment at an hour level which needed to be transformed to a day level by integer dividing each step by 24. Once at a day level, the results were grouped by day and the mean was taken for each compartment. These mean values were then used in the correlation test and the results are shown in Table 4.3. There was a strong positive correlation between the agent based model and the equation based model for the Susceptible and Removed compartments with values of 0.99 and 0.99



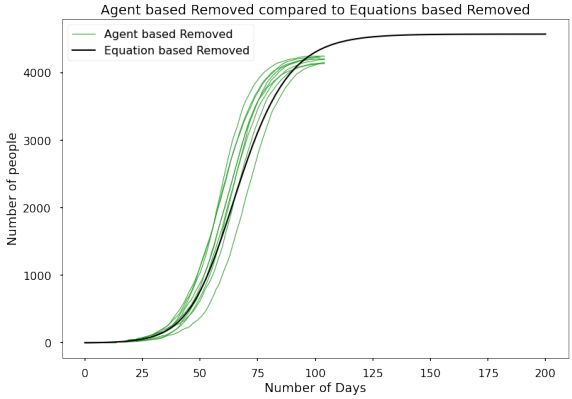

(a) Agent based Susceptible vs Equations based Susceptible




(b) Agent based Exposed vs Equations based Exposed




(c) Agent based Infected vs Equations based Infected




(d) Agent based Removed vs Equations based Removed

Figure 4.4: Cross validation of each compartment in Agent and Equation based models

respectively. The Exposed and Infected compartments had weaker correlations of 0.66 and 0.55 respectively but these values would still considered to be relatively high. This difference in correlation values could be explained by the offset of the peaks of the Exposed and Infected compartment between the two models. Figure 4.3 shows this offset is more prominent with the Infected compartment compared to the Exposed compartment, and this is reflected in the lower correlation values in Table 4.3.

Overall however, it can be concluded that the agent based model behaves as an SEIR model should, when compared to a validated model using the same variable values to simulate the disease dynamics of COVID-19. While this was a basic validation, it was

the first important step in validating the model before the experiment was run.

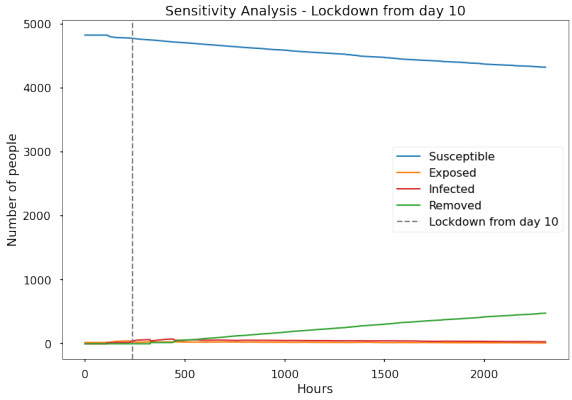
### 4.1.3 Sensitivity analysis

The sensitivity analysis undertaken in this study focused on testing three separate variables using three independent experiments managed by the Behaviour Space module in NetLogo. The response variables for each experiment were the number of agents in each compartment at the end of each step in the simulation. All other variables were held constant in each experiment. Similarly to the cross validation, some of the response variables were combined in order to simplify the analysis. The Infected compartments, $I^A$, $I^M$, $I^S$, and $I^C$, were combined into a single $I$ compartment and the $R$ and $D$ compartments were combined into an $R$ compartment. 100 runs of the simulation were performed for each value of the variable being analysed in each experiment. The mean value of the response variables were then taken across all 100 runs at each step in each experiment. The results of each experiment are discussed next.

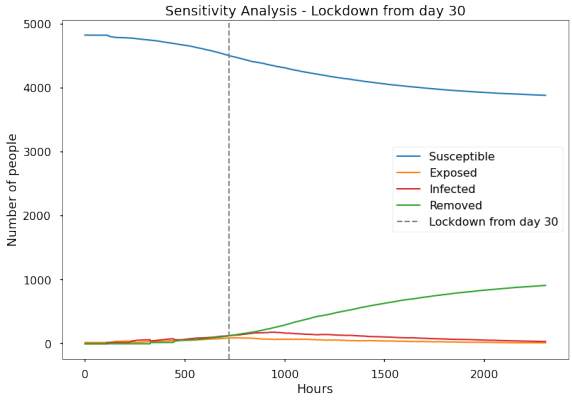
#### Level 5 lockdown

The first variables tested in the sensitivity analysis was the day from which level 5 lockdown was initiated. 100 runs of the simulation were performed for each of the values 10, 30, 50, 70, and 90 with all other variables being held constant. In total, this amounted to 500 runs. For each value of the lockdown_from_day_n variable, the mean value of each compartment at each step across all 100 runs was taken.

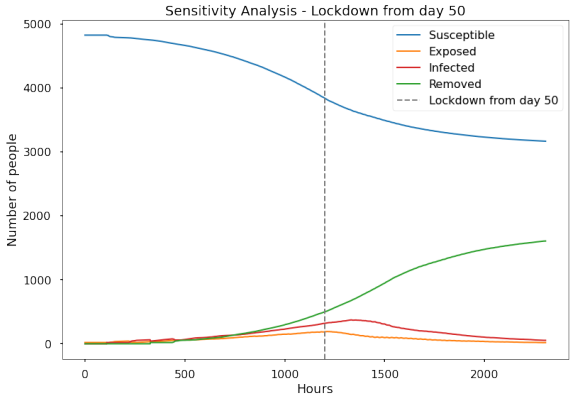
Intuitively, it was expected that as level 5 lockdown is started later and later during the outbreak simulations, that the rate of infection would increase. This would be due to the agents in the simulation spending more time mixing in the environment and increasing the opportunity to become exposed to the disease. Figure 4.5 shows the dynamics of the outbreak for each scenario where n, the day in which level 5 lockdown started, is equal to 10, 30, 50, 70, and 90. As expected, the dynamics of the outbreak are quite different in each case. A pattern emerges where the size of the outbreak increases as the day in which level 5 lockdown begins is delayed. From day 10, the number of susceptible agents steadily decreases throughout the outbreak



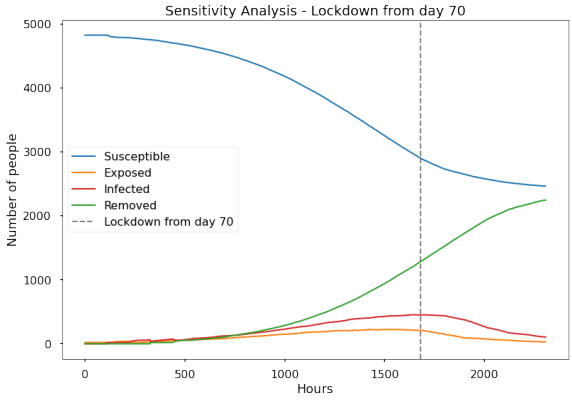

(a) Level 5 lockdown from day 10



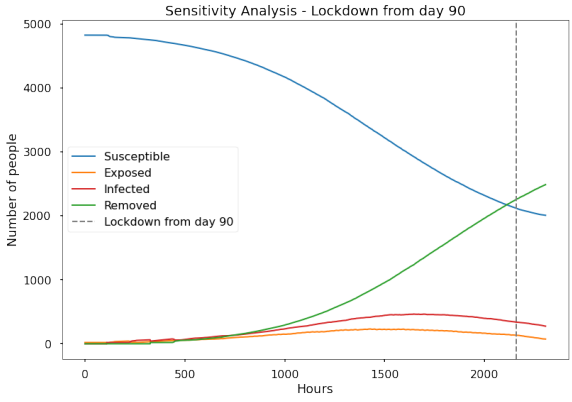

(b) Level 5 lockdown from day 30




(c) Level 5 lockdown from day 50




(d) Level 5 lockdown from day 70




(e) Level 5 lockdown from day 90

Figure 4.5: Sensitivity analysis of level 5 lockdown parameter

and is followed by the removed compartment increasing at a steady, linear rate. The exposed and infected groups remain very low. Starting level 5 lockdown from day 10 seems to control the outbreak in this simulation quite well and the outbreak looks as if it will continue at a steady pace but does not appear to taper off. This appearance is estimated by looking at the linear nature of the susceptible agents variable. However, starting level lockdown from day 30, the dynamics change significantly. The susceptible and removed compartments now start to develop a clear curve reminiscent of a classic SEIR model but taper off. This tapering off is quite slow and takes until the end of the simulation to level off. The infected and exposed curves start to become visible as the number of infections increases. Starting level 5 lockdown from day 30 did seem to control the outbreak in this simulation, but since the momentum of the disease took time to dissipate, the effect was slow. Starting level 5 lockdown from day 50 or day 70 were interesting. The spread of the disease can clearly be seen increasing as the start day of level 5 lockdown begins later, and the infected and exposed curves were now clearly visible. Interestingly in both scenarios, the exposed curve noticeably decreased soon after level 5 lockdown began. This suggested that the number of new infections in the model decreased sharply soon after level 5 lockdown, followed by a sharp decrease in the infected curve a short time later. This may also explain how the momentum of the outbreak was maintained. Even though the number of new infections (i.e. exposed individuals) decreased noticeably, the incubation time of 4.5 days ensures that the infected curve does not necessarily peak immediately. While the level 5 lockdown interventions did have an effect on the spread of the disease in this simulation after 50 and 70 days, the momentum gathered by the outbreak at this stage masked the effect of the intervention immediately after its implementation. Effective communication to the public would be required to reinforce and reassure the effectiveness of the intervention. Finally, introducing level 5 lockdown after day 90 had very little impact on the spread of the disease. By this point, the majority of the population in the simulation had been exposed to the disease and, while the intervention should decrease the number of new infections, the peak of exposed and infection compartments had already passed and the outbreak was naturally running

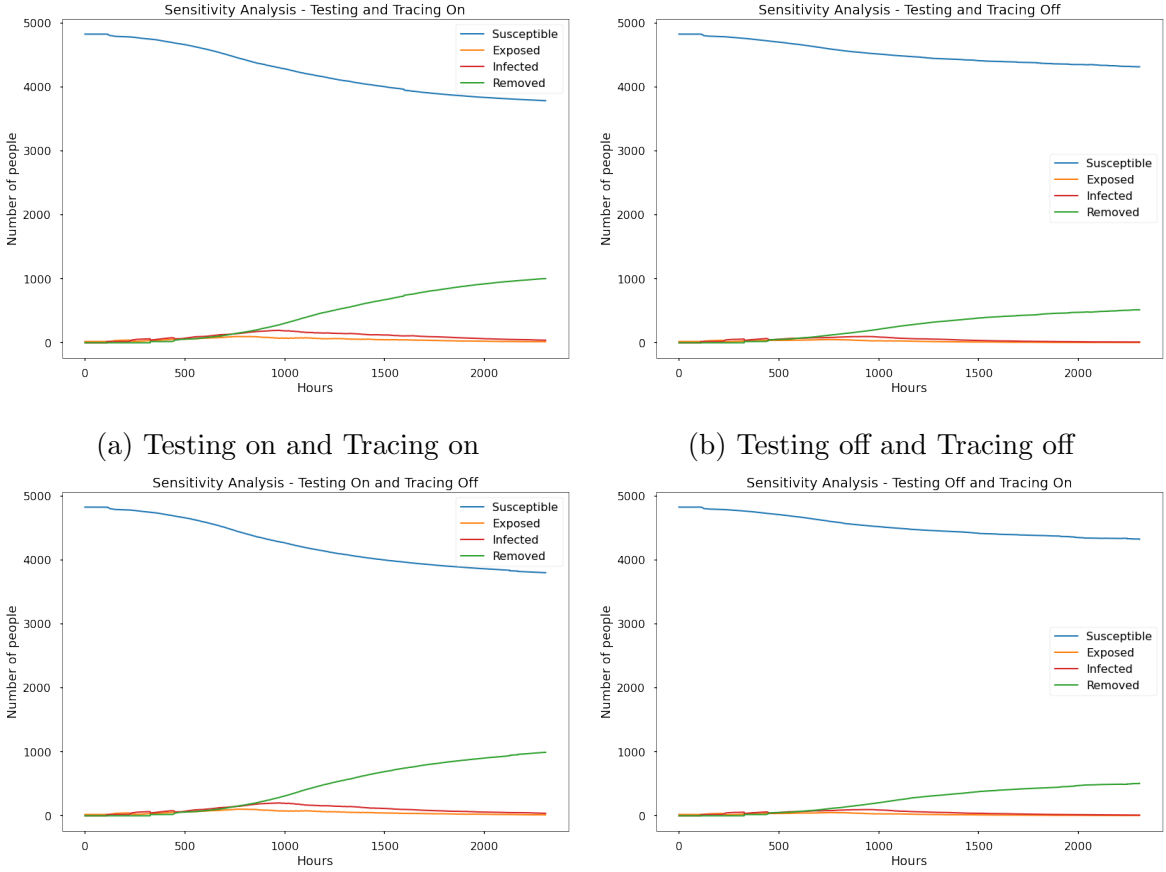
its course. The benefit of this was that the outbreak in the model was short, and recovery is quick. However, due to the huge scale of the outbreak, the consequence was an increased risk of a large number of disease related deaths.

### Testing and tracing

The second experiment in the sensitivity analysis varied the testing function and contact tracing function. Since each of these are binary values (on or off), 4 scenarios were simulated, totaling 400 runs.

The expectation before running the experiment was that the outbreak would have a larger rate of infection when testing and/or tracing is disabled, compared to the rate of infection when either of these interventions were switched on. However, Figure 4.6 shows a surprising result observed in the experiment. Regardless of whether the contract tracing intervention was enabled or not, a larger outbreak was observed when the testing intervention was switched off compared to when the testing intervention was switched on. The reason is caused by a limitation in the model. When testing is switched off, all agents with a mild infection are symptomatic and therefore will isolate at home. The compliance level of agents will be 100%, which will not be representative of the real world, where a small percentage of individuals may not be compliant with this guidance. In addition, the model assumes that agents isolating cannot infect other agents that come into close contact with them. Therefore, when testing is disabled, agents infected will isolate at home, or will isolate in hospital if the infection progresses to a severe or critical case until recovery or death. However, when testing is enabled, symptomatic cases and/or cases identified through close contacts, are periodically tested with a rate accuracy of 87%. Therefore, 13% of positive cases will test negative, stop isolating and go back into the community where they can infect other agents in the environment unless their case progresses to a severe or critical case when they will isolate in hospital.

While this limitation is not ideal in the simulation, it reflects the complexity of modelling human behaviour in complex circumstances. This will not adversely affect




(a) Testing on and Tracing on

(b) Testing off and Tracing off

(c) Testing on and Tracing off

(d) Testing off and Tracing on

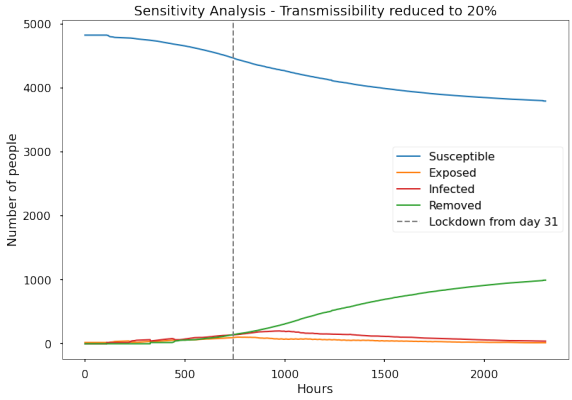
Figure 4.6: Sensitivity analysis of testing and tracing functions

the main experiment in this study since this will assume that both testing and contact tracing interventions are both enabled for the course of the outbreak. Having both interventions enabled does give realistic behaviour and the recommendation would be to not run this model with testing disabled. This also highlights the value of sensitivity analysis, and pushing the values of variables beyond the limits of what would be expected in a real world scenario to test the thresholds of the model's capability. For the purpose of this experiment and study, and to operate within the constraints of time and scope, the decision was made to acknowledge the limitation and continue with the experiment.

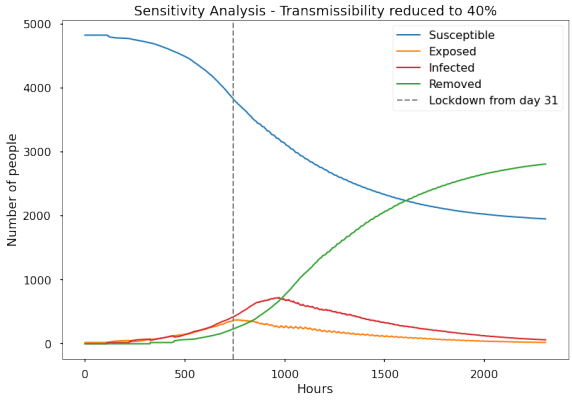
**Reduction of transmissibility**

The final experiment in the sensitivity analysis focused on the reduction of transmissibility of the disease. This represented the public health advice of handwashing, mask wearing, and physical distancing. A value of 100% represented no reduction of transmissibility and the probability of infection remained at the values of $\beta_1$ from Table 3.6. Whereas, a value of 50% represented the values of $\beta_1$ from Table 3.6 to be reduced by 50%, and so on. The value of this parameter in the model was set to 20% (Kennelly et al., 2020; Chu et al., 2020). However the sensitivity analysis varies this parameter through the values 20%, 40%, 60%, 80%, and 100% with all other parameters held constant.

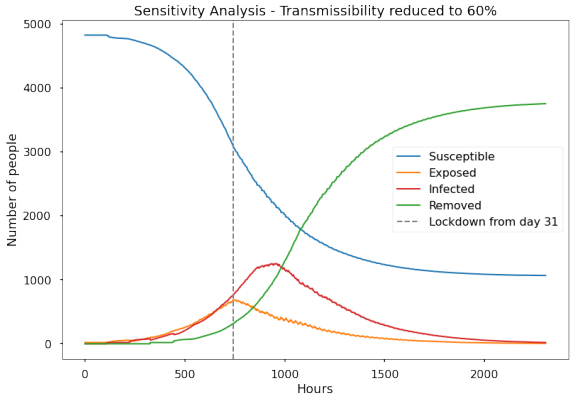
The expectation of the experiment was that as the transmissibility increased, so too would the size and speed of the outbreak. Figure 4.7 shows the results of the experiment for each of the five values of transmissibility. At 20%, the rate of infection is moderate with an initial wave of infection during the first half of the outbreak and exposure tapering off slowly during the course of the outbreak. Approximately 20% of the susceptible population become exposed to the disease. However this behaviour changed quickly when transmissibility was set to 40%. The wave of infection significantly increased, even after level 5 lockdown had started and new exposures reduced. By the end of the outbreak, approximately 60% of the population had become exposed to the disease, and the wave of infection peaked at approximately 30% of the



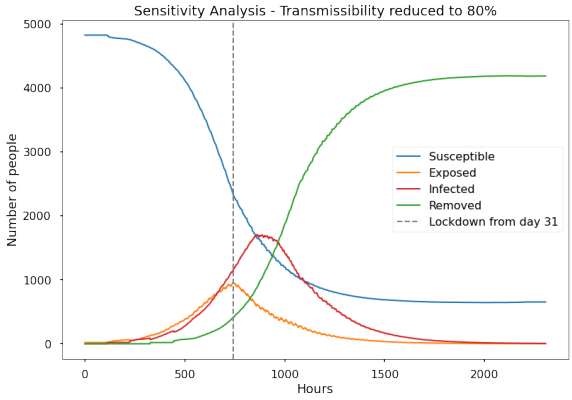

(a) Transmissibility of 20%



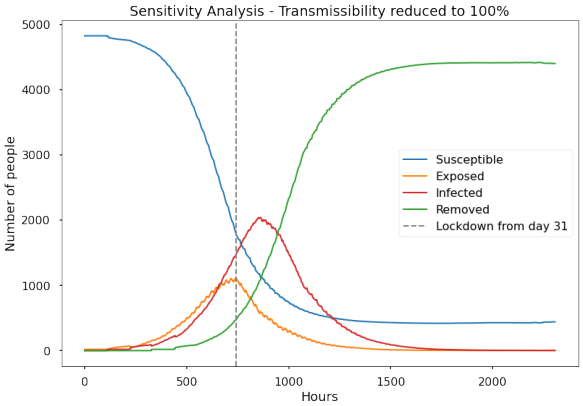

(b) Transmissibility of 40%




(c) Transmissibility of 60%




(d) Transmissibility of 80%




(e) Transmissibility of 100%

Figure 4.7: Sensitivity analysis of transmissibility

population (including both exposed and infected compartments). With transmissibility increased to 60% and 80%, the pattern continued with the size of the outbreak becoming larger and the wave of infection becoming steeper. The curve of the removed compartment also became steeper as transmissibility increased and the outbreak dissipated in the final third of the outbreak. Finally, at 100% transmissibility, the output of the model resembled SEIR curves where no interventions at all had been in place. A short, sharp wave of infection peaked at 60% of the population (including exposed and infected compartments), followed by a short, sharp recovery where the curve of the removed compartment tapers off for the final third of the outbreak. By the end of the outbreak, over 90% of the population has become exposed to the disease.

The experiment demonstrated that the model was highly sensitive to the transmissibility parameter. Small changes to the parameter could have a significant impact on the outbreak of the disease. The interpretation of this characteristic to a real world outbreak of COVID-19 intuitively suggests that close contact with other potentially infectious individuals without the use of public health practices can have a significant impact on the progression of the outbreak.

### 4.1.4 Validation compared to real world data

The final stage of the validation process outlined by Hunter & Kelleher (2020) is a comparison of the data from the agent based model to real world data for the outbreak in county Carlow. The validation was assessed by comparing the real world data, in terms of confirmed cases of COVID-19, to a distribution of 300 simulated runs from the base version of the agent based model. This base model, described in section 3.5, included all interventions enabled in the model except for random testing, and the response variable was the number of confirmed cases each day.

A number of pre-processing steps were required. Firstly, since the agent based model was calculated at hour level, the results were aggregated up to day level by integer dividing by 24, and taking the mean. Secondly, the agent based model calculated a running total of the number of confirmed cases, so a daily total was calculated

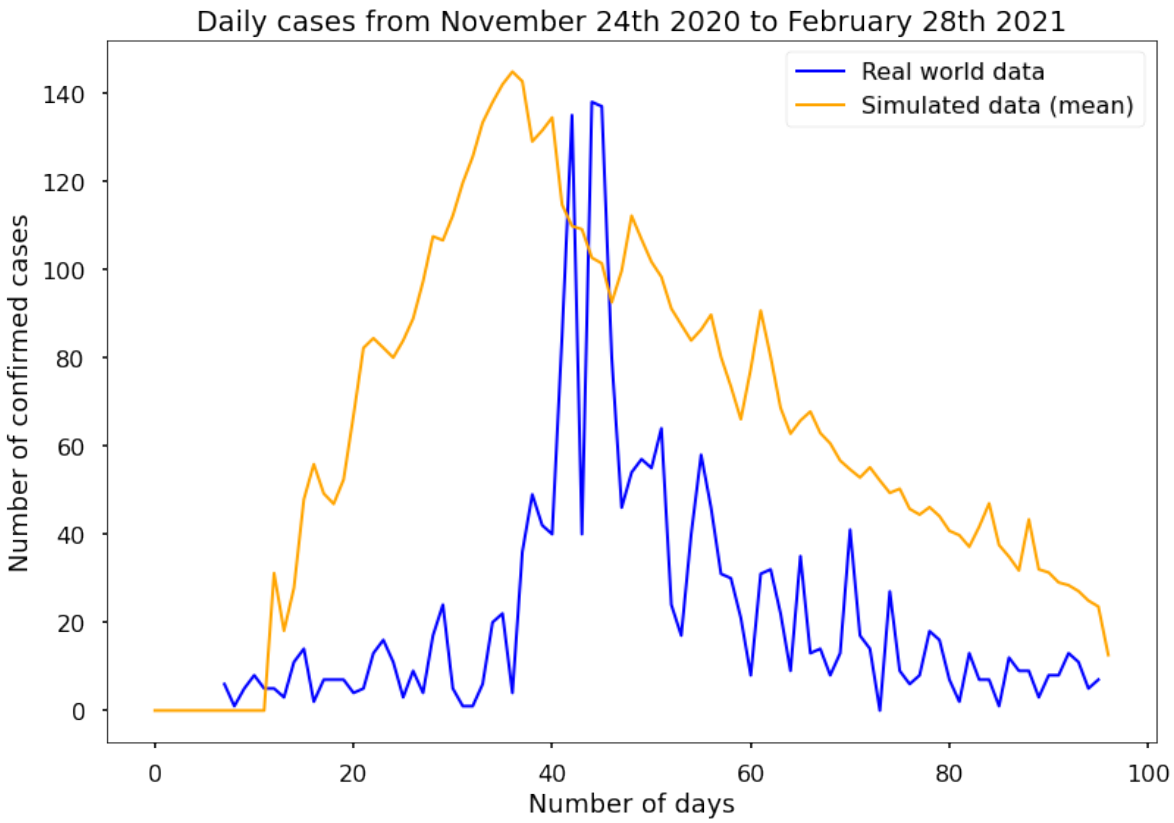



Figure 4.8: Comparison of simulated outbreak to real world outbreak

by taking the difference from the previous day. Since the model was a sample of the real world data, the results from each simulation were scaled up to match the real world data by dividing the daily cases from each simulation by 4843 (sample size) and multiplying by 56932 (population of county Carlow). Finally, the simulations all had a spike of cases at the beginning, which was as a result of the outbreak starting. To account for this a value of 0 was set for the daily cases for the first 12 days. This included the 6 days at the end of November and 6 days at the beginning of December.

Three comparisons were made between the agent based model and the real world data. Firstly, a test was performed to determine if the samples were from the same distribution. The agent based model was aggregated across all 300 runs using the mean of each day. Before the test was performed, the assumptions of an independent samples t-test were assessed, i.e. distribution normality and homogeneity of variance. Each sample was tested for normality using the Shapiro-Wilk statistical test and the

| Results of statistical tests from validation | | |
|---|---|---|
| **Test** | **Real world data** | **Simulated data** |
| Shapiro test (p-value) | 2.24206 | 0.00961 |
| Levene test (p-value) | - | 0.04547 |
| Mann Whitney test (p-value) | - | 0.02310 |
| Spearman correlation | - | 0.72782 (p-value: 6.48636e-16) |
| Pearson correlation | - | 0.68755 (p-value: 9.92688e-14) |

Table 4.4: Results of statistical tests used on experiment data

results are shown in Table 4.4. The Null hypothesis of the Shapiro-Wilk test is that the sample is from a normal distribution. Neither the real world data nor the simulated data from the agent based models were shown to be normally distributed since the p-values were less than $\alpha$, where $\alpha = 0.05$. To test for equal variance, Levene's test was used and the results are shown in Table 4.4. The Null hypothesis of Levene's test is that the samples have equal variance. Since the p-value was 0.04547, the Null hypothesis was rejected in favour of the alternate and the result was assumed to be that the samples did not have equal variance at a confidence level of 95% ($\alpha = 0.05$). Given that both samples were normally distributed but their variances were not equal, a two-sided Mann-Whitney U-test test was used in favour of an independent t-test. The Null hypothesis of the Mann-Whitney U-test is that the two samples are from the same distribution. The results in Table 4.4 show that the U statistic and associated p-value from the Mann-Whitney test had values of 3585 and 0.02310 respectively, which would assume that the Null hypothesis should be rejected in favour of the alternate, and conclude that the distributions were not equal. While this result was not as hoped, the agent based model was based on the assumptions that every person with symptoms of COVID-19 applied for and attended PCR testing, and that the reporting of testing was accurate and immediate. In addition, the agent based model did not account for differences in human behaviour over the Christmas period due to time and scope constraints. Figure 4.8 shows the distribution of daily cases, and while the distribution of the daily cases do not match exactly, the approximate pattern and scale was in line with expectations. With this in mind, a decision was made to scale up the simulated data differently. Instead of scaling up based on population size, the simulated data was

scaled up according to the size of the outbreak. Scaling up each simulated value by a factor of 4 produced a set of results that was a much closer match. With non-normal distributions but equal variance, an independent t-test had a p-value of 0.755 and a Mann-Whitney U-test had a p-value of 0.310, which showed that the samples could be from the same distribution using a different scaling factor. While these are not ideal results, the validation of the model continued as-is.

The second comparison to the real world data was a correlation test. No additional pre-processing was required to the two samples for the correlation test. Both Spearman and Pearson tests were performed with the results shown in Table 4.4. Both Pearson and Spearman correlation tests showed high correlations of 0.687 and 0.728 respectively and both sets of p-values were less than $\alpha$, where $\alpha = 0.05$. As a result, the assumption was made that the outbreaks from the agent based model and the real world were linearly related.

Finally, since every run of the simulation produced a different number of infected agents in the outbreak, a distribution of the cumulative number of confirmed cases of COVID-19 was produced by summing the number of daily cases for each simulation. The real data was then checked whether the cumulative number of confirmed cases during the real outbreak fell into the distribution of simulated cases. Since the simulation was based on a sample of the population, the simulation values were scaled up by a factor of 4 as per the first comparison in this validation.

To check whether the cumulative number of cases of COVID-19 from the real world outbreak fell into the distribution of simulated outcomes, a number of steps were followed. First, a Shapiro-Wilk test was used to produce a p-value of 0.0998 which showed that the distribution of simulated outcomes were normally distributed. Next, a z-score was calculated for the real world value of 2034 cases by using formula 4.1, where $x$ is 2034, $\bar{x}$ is the mean of simulated outcomes, and $\sigma$ is the standard deviation of simulated outcomes.

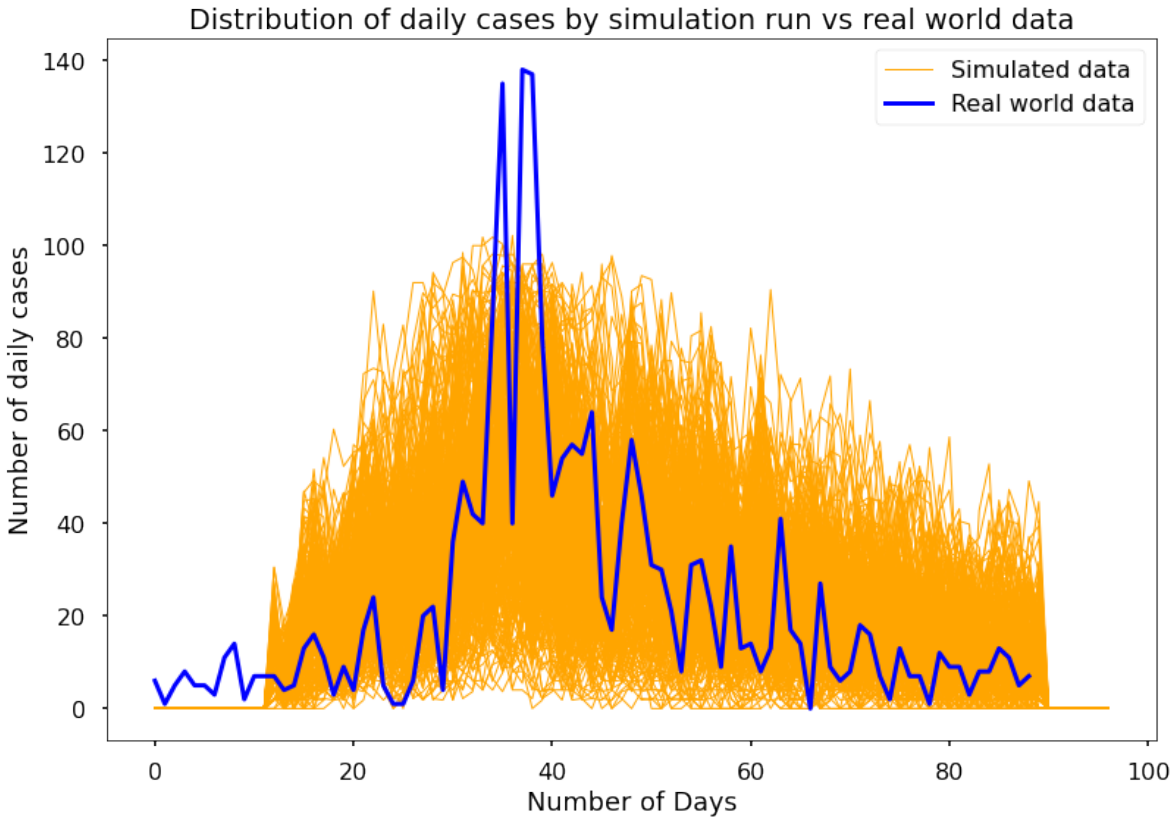



Figure 4.9: Comparison of individually simulated outbreaks to real world

$$z = \frac{x - \bar{x}}{\sigma} \tag{4.1}$$

This produced a z-score of 0.008811 which was tested to determine if it fell into a normal distribution by calculating a probability density function. This test produced a value of 0.9929699, which showed that this z-score was indeed from the normal distribution of simulated outcomes using a scaling factor of 4.

The validation of this model produced some mixed results, some of which worked well and some which were not ideal. In terms of the number of runs, a value of 300 seemed appropriate and was similar to the number of runs from examples in the literature (Hunter et al., 2020; Silva et al., 2020). Cross validating the agent based model compared to the equations based model produced excellent results with very high correlations in the results and near perfect match in terms of behaviour of each

compartment in the simplified model. The sensitivity analysis showed how the model responded when varying different attributes. This also highlighted one weakness which does not affect the outcome of this model but it useful for consideration in future work. Finally, the comparison to real world data showed the general pattern of behaviour in the outbreak did match the real world data, and while the simulation did not match perfectly, using a different sample scaling factor produced results that were acceptable to continue with the experiment.

## 4.2 Experiment Results

Once the model had been validated, the experiment produced 300 runs of the simulation for each of the following scenarios. The base scenario, where the percentage of testing capacity reserved for random testing was set to 0%, followed by scenarios where the percentage reserved for random testing was set to 10%, 20%, 30%, 40%, and 50%. All other parameters in the simulation were held constant. This amounted to 1800 runs of the simulation in total. The response variables generated by the simulation are listed in Table 3.9.

The results of the experiment were analysed by comparing the base model to each of the models where a percentage of testing capacity was reserved for random testing. The statistic used in the comparison of the base model and each of these scenarios was the cumulative number of deaths at the end of each simulation. This comparison was made by calculating the difference between two means. Before a parametric independent samples t-test could be used, the assumptions of such a test, distribution normality and homogeneity of variance, were tested first. If the assumptions were not met, then a non-parametric test, Mann-Whitney U-test was used instead.

### 4.2.1 Testing the assumptions of a parametric test

The Shapiro-Wilk statistical test was used to assess whether the samples were normally distributed. The results for each of the samples were shown in Table 4.5. In all

| Results of statistical tests in experiment | | | |
|---|---|---|---|
| **Scenario** | **Shapiro test (p-value)** | **Levene test (vs base model) (p-value)** | **Mann-Whitney U-test (p-value)** |
| Base model | 4.959744073773642e-11 | - | - |
| 10% random testing | 2.264694725154226e-12 | 0.2808275144057998 | 0.34582545466108683 |
| 20% random testing | 2.4450958618915664e-11 | 0.9063451553730657 | 0.05666387868892259 |
| 30% random testing | 3.486663735507989e-11 | 0.5591615905929357 | 0.07881522995053614 |
| 40% random testing | 8.980246017295523e-13 | 0.9688079291689023 | 0.007734610300772811 |
| 50% random testing | 1.054944976860983e-12 | 0.9703003908034008 | 0.019524229851639353 |

Table 4.5: Results of statistical tests used on experiment data

samples, the p-value was less than $\alpha$, where $\alpha = 0.05$, so the null hypothesis of normality was rejected for the alternate and all samples were assumed to be non-normal. Since these results violated one of the assumptions of the parametric independent samples t-test, the non-parametric test, Mann-Whitney U-test was used. However, for completeness, the results from the test for equal variance are described below.
Levene's test for equal variance was used between the base model and each of the random testing scenarios. The results of these tests are shown in Table 4.5. The results show that the base scenario and each of the random testing scenarios all have equal variance. However, as previously discussed, the test for normality had failed so the non-parametric test was used instead.

### 4.2.2 Test for difference in means

The Mann-Whitney U-test was used to test for difference in means. The alternate hypothesis in this study is that the number of COVID-19 related deaths in an Irish county is significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing. Therefore, a one-sided U-test was used where the results of the base model were tested to assess whether the number of deaths in this scenario were statistically significantly higher than the number of deaths in each of the scenarios with random testing. The results of this test are shown in Table 4.5.

The base model shows a difference in means in two scenarios with a very close result for a potential third scenario. The p-values in scenarios with 40% and 50% random testing are 0.0077 and 0.0195 respectively, which are both less than $\alpha$, where $\alpha = 0.05$. Therefore, the null hypothesis of equal means was rejected in favour of the alternate. It was then inferred that, in this simulation, scenarios with 40% and 50% random testing resulted in a statistically significant reduction of deaths compared to a testing regime of close contact tracing only, at a confidence level of 95%. There is also an argument for the scenario with 20% random testing to be included in this inference since the p-value of 0.0567 was very close to $\alpha$ (0.05).

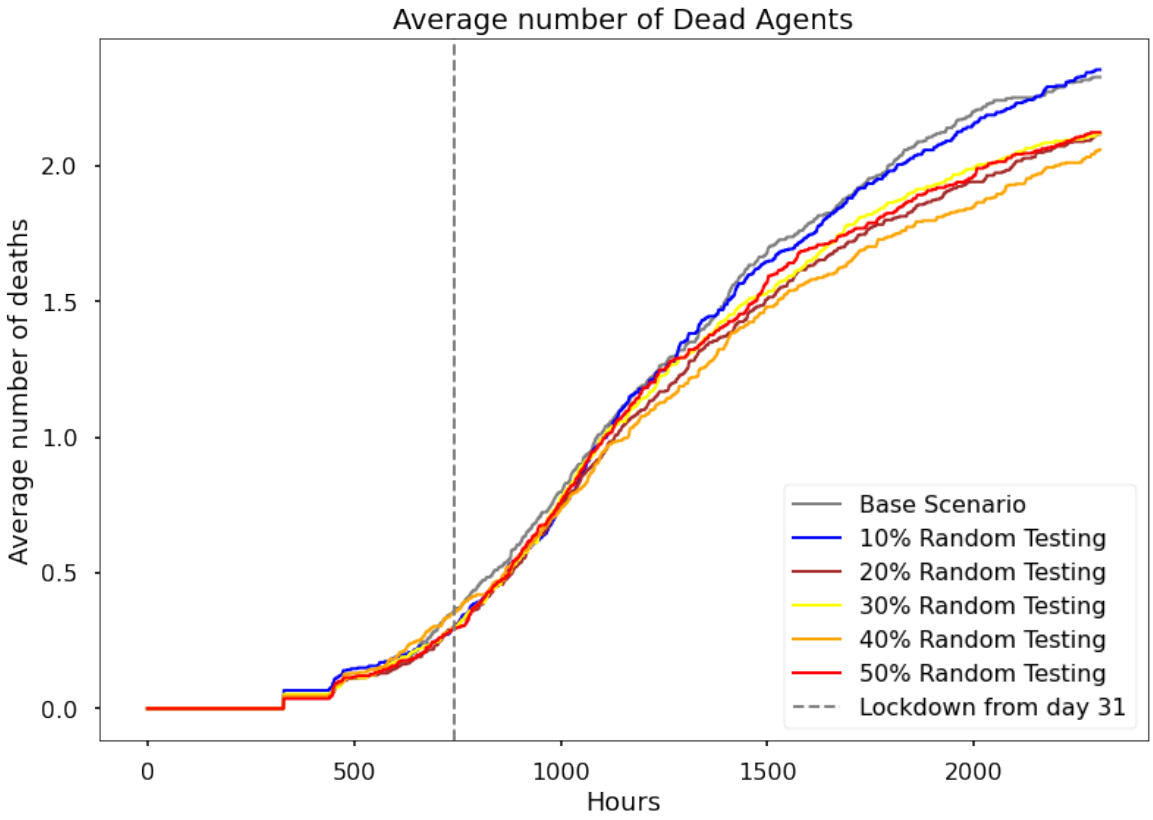



Figure 4.10: Comparison of number of deaths between scenarios

Figure 4.10 and Table 4.6 show the average number of deaths across all runs for each scenario. The same trend as described by the Mann-Whitney test results are shown. There seems to be two groups of results. One group contains the base model

| Descriptive statistics from experiment results | | | | |
|---|---|---|---|---|
| **Scenario** | **Mean** | **Standard deviation** | **CI Lower** | **CI Upper** |
| Base model | 2.31 | 1.61739 | 2.12697 | 2.49303 |
| 10% random testing | 2.30334 | 1.73050 | 2.10751 | 2.49916 |
| 20% random testing | 2.09334 | 1.56186 | 1.91659 | 2.27007 |
| 30% random testing | 2.10667 | 1.52415 | 1.93419 | 2.27914 |
| 40% random testing | 2.02 | 1.55812 | 1.84368 | 2.19632 |
| 50% random testing | 2.06667 | 1.63231 | 1.88195 | 2.25138 |

Table 4.6: Descriptive statistics from experiment

and the model with 10% random testing and the second group contains the models with random testing of 20%, 30$, 40%, and 50%.

The results generated by the experiment show in that, in this simulation, the introduction of random testing to augment a regime of close contact tracing and testing does have a significant impact on the number of deaths compared to contact tracing and testing alone. The number of deaths at the end of each simulation is still very low but since this experiment is a sample of the population, it can be scaled back up to real world levels using a factor of 4 as per the model validation.

## 4.3 Evaluation and discussion

The aim of this study was to test whether the introduction of random testing, as an augmentation to a testing regime that consisted of close contact tracing and PCR testing, could have a statistically significant impact on the number of COVID-19 related deaths. An agent based model was used to simulate a number of scenarios where the percentage of testing capacity reserved for random testing was varied from 0% to 50% in increments of 10%. Additional interventions were included in the simulations including reduced disease transmissibility to account for public health measures such as mask wearing and physical distancing, working from home and level 5 lockdown to lower mobility within the model, as well as contact tracing and PCR testing. Sensitivity analysis showed that this model was very sensitive to transmissibility reduction

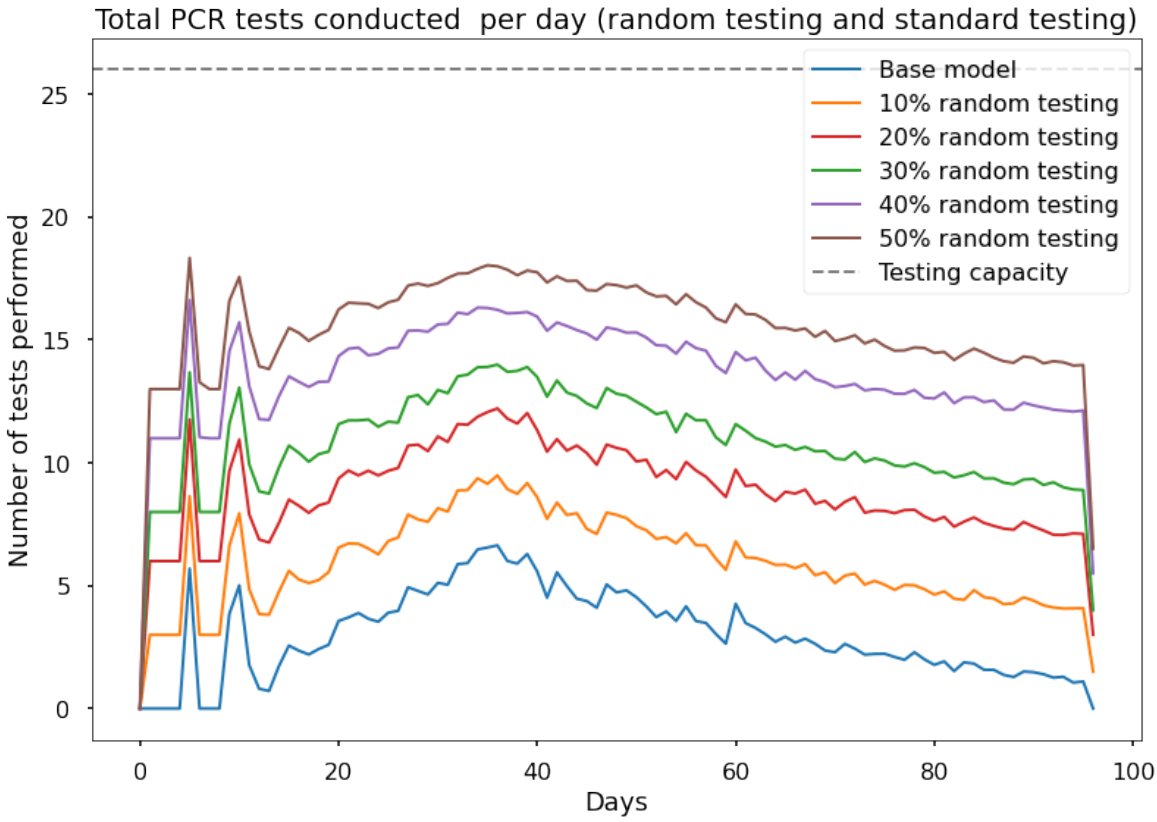



Figure 4.11: Trend of combined testing carried out during each scenario

(20%, 40%, 60%, 80%, and 100%) and the day on which level 5 lockdown started (10, 30, 50, 70, and 90), both of which would have been expected. However, due to a limitation, disabling PCR testing reduced the size of the outbreak in the model, since it assumed 100% isolation compliance from mildly infected individuals, which outperformed the detection and isolation of individuals identified through PCR testing. Each of these interventions were held constant while the percentage of random testing was varied during each scenario, and 300 runs of each scenario were simulated in the experiment. The base scenario was assumed to the the scenario where 0% of the testing capacity was reserved for random testing. Each of the other scenarios were compared to the base scenario using the Mann-Whitney U-test of difference in means. The statistic tested was the number of COVID-19 related deaths at the end of each simulation. The results of the comparison showed that the difference in means was statistically significant in cases where random testing levels of 40% and 50% were

used, at a confidence level of 95%. Two other scenarios were very close to being statistically significant where the percentage of random testing was 20% and 30% with p-values of 0.056 and 0.078 respectively. While the expectation of some of scenarios generating a statistically significant result was realistic, it was surprising that four out of five were either significant or very close to being significant. Equation based models discussed in this study did include a number of interventions to limit the spread of the disease, but testing and diagnostic capacity were only included in a small number of studies (Godio et al., 2020; Choi & Ki, 2020). It is difficult to compare the results of this model to such equation based models, since an assumption of these models is that the majority of the population will become infected. Even though equation based models are often not a realistic representation of real world outbreaks, these models are useful to test the effectiveness of various interventions and detect macro trends. Godio et al. (2020) did conclude that extensive testing in the early stages of an epidemic has a positive effect on the number of COVID-19 related deaths. While random testing was not considered, this did increase the number of tests performed in the early stages of the outbreak modeled in the agent based model from this study. Similarly, Choi & Ki (2020) concluded that the transmission period and rate both decreased with active testing, which is aligned with the proactive approach to testing that this study explores. Agent based models can simulate far more complex phenomena, including diagnostic testing, but these models do require significantly more input data and computational power. Panovska-Griffiths et al. (2020) used the Covasim software to generate an agent based model with the aim of reopening schools, and concluded that significant levels of testing would be required to prevent future waves of infection. While the random testing from this study does not constitute an increase in overall capacity, the use of random testing did increase testing activity throughout the outbreak.

Reserving a proportion of testing capacity each day for random testing implies that a small number of tests are attempting to detect positive cases of COVID-19 within a relatively large population of individuals. In the case of reserving 50% of testing

capacity for random testing, this amounts to 13 tests per day to detect positive cases within a susceptible population of over 4800 individuals. Although a low detection rate would be expected, this study has shown that early detection of asymptomatic individuals in the simulated outbreak significantly reduced the rate of infection, which in turn reduced the risk of COVID-19 related deaths. The detection rate of infected agents across all random testing scenarios was approximately 1.6% (number of detected cases from random testing / total random tests performed). There is a trade off here between reserving the testing capacity for random testing and retaining the ability to increase standard close contact testing during periods of high demand. While a real world regime may adjust the percentage of random testing depending on demand, this study retained the same percentage throughout each scenario regardless of demand. However, Figure 4.11 shows that in this model, the introduction of random testing was not affecting the overall testing capacity and therefore the testing of symptomatic agents.

In the following chapter, the results presented here will be discussed and critically evaluated in order to determine whether the aims of this study have been met.

# Chapter 5

# Conclusion

This chapter summarises the research undertaken in this study and discusses the results and their impact on COVID-19. The strengths and limitations of the research are also discussed as well as ideas for future work in this domain.

## 5.1 Research Overview

In December 2019, a novel coronavirus was identified in Wuhan, China and named as COVID-19. The Chinese government implemented a series of measures in an attempt to contain the outbreak, but instead, this quickly developed into a global pandemic. As of February 12$^{th}$ 2022, this has resulted in 404,910,528 confirmed cases of COVID-19 and 5,783,776 deaths[1]. Multiple variants of the disease have also emerged, with the two most notable being named as Delta and Omicron. Early non-pharmaceutical responses included physical distancing, mask wearing, working from home, school and business closures, testing and close contact tracing, and the use of technology. During 2021, vaccination was included as an additional and the most effective response, particularly from a long-term perspective. Policy makers need evidence based recommendations to inform their response to pandemic and the ever changing landscape it creates. Epidemiology models such as SEIR can be used to predict effectiveness of interventions. These models can be built (individually or in combination) using differential equations,

[1]https://covid19.who.int/

social network analysis, swarm intelligence, and agent based systems.

## 5.2 Problem Definition

Testing plays a critical role in the detection and isolation of infected individuals, which in turn slows down the spread of COVID-19. Symptomatic individuals can be found and isolated relatively easily, but asymptomatic individuals are more challenging to discover. Close contact tracing is aimed primarily at testing symptomatic individuals plus any others that they have been in contact with. However, this may only detect a fraction of the asymptomatic cases. There is an unmet need to detect more asymptomatic cases, which will further slow the spread of the disease and the associated deaths that unfortunately accompany it. Additional government responses, which are necessary, place a heavy burden on society in terms of economic, social, and mental health aspects. Any measure which eases this burden could have significant benefits for wider society.

This study proposed the following research question: Can the efficiency of a COVID-19 close-contact tracing regime in an Irish county be improved through the use of random testing of potentially asymptomatic infected individuals? The following hypothesis was derived from the research question:

**Alternate Hypothesis** The number of COVID-19 related deaths in an Irish county is significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing.

**Null Hypothesis** The number of COVID-19 related deaths in an Irish county is not significantly higher in a population using a regime of close-contact tracing than in a population augmented with random testing.

## 5.3 Design of Experimentation, Evaluation and Results

This study proposed the use of random testing to augment an existing regime of close contact tracing and testing. Due to the complexity of modelling disease, societal, and response dynamics, an agent based model was built. A specific outbreak was chosen in county Carlow between December 1st 2020 and February 28th 2021. This county was relatively small, so a small sample in absolute terms would represent a higher proportion of the population compared to a larger, more populous county. The input data was taken from the CSO website and included population demographic data such as household, age distribution and economic status as well as COVID-19 data on cumulative daily cases[2].

The agent based model contained households, people, and small areas, which were based on data from the CSO. This ensured that the population in the simulated environment was representative of county Carlow in terms of age distribution, population density, and economic status which affected daily behaviour. The CSO data was based on Census 2016, and while this data is now up to 6 years old, this was the most recent validated dataset available. A Python script used this data to build a sample population of households and individuals within each small area. This dataset was then automatically imported into the NetLogo environment during the initialisation process. A number of limitations of the model included several assumptions required for the sample population. Firstly individuals in the population were assumed to be retired if they were aged 65 or over. 20% of workers were assumed to be essential workers but no distinction was made for healthcare workers, which would have a much higher risk of infection, particularly in a hospital setting. Although the age distribution and economic status of the individuals were each representative of the wider population, they came from distinct datasets, and the economic status data was at a national level rather than small area level. However, it was felt that this would not

[2]https://www.cso.ie/

have a significant impact on the progression of the simulated outbreak since individuals aged 65 and older were assumed to be retired, and only 20% of workers commuted to workplaces.

The epidemiology attributes controlled the dynamics of the disease and were taken from the literature as discussed in Section 3.1.4 and Table 3.6. These values did produce an outbreak which was comparable to real world data and the incorporation of several additional compartments for infected asymptomatic, infected symptomatic, infected severe, and infected critical individuals allowed a much more granular view of what was happening within different aspects of the model. Future work here could include fine tuning the epidemiology parameter values given that the literature did not have values specific to Ireland. A grid search type function could work well here to fit the model to real world data. However care would need to be taken to avoid overfitting.

Government responses in the model included the reduction of transmissibility, contact tracing, PCR testing, level 5 lockdown, and random testing. Reduction in transmissibility worked well as a proxy for mask wearing, hand washing, and physical distancing. The value of 20% was taken from the literature, however, it was not possible to validate this number based on available data in Ireland. The contact tracing function enabled infected individuals to remember which other agents they came into contact with since they became infected. Flagging these close contacts for PCR testing in the case of a confirmed case of COVID-19 gave realistic behaviour in terms of how the response worked in the real world. A limitation in this model was every agent remembered every close contact since they became infected, which would not likely reflect the real world. It was a difficult decision to determine what percentage of close contacts were remembered since no data was available. The assumption was then made that all close contacts were remembered. The testing was assumed to be PCR testing, which is the gold standard and was used in Ireland at the time of the simulated outbreak. Ensuring the testing capacity was representative, given the sample size, and performing the testing once per day produced a realistic testing process.

The introduction of random testing into the process was seamless since the random tests were performed earlier each day. The random test was also assumed to be a PCR test. Future work could be to switch the random test to antigen-based. While the accuracy of the antigen is accepted to be lower, the volume of testing could be much higher, and with a higher reach could potentially detect a much higher proportion of asymptomatic individuals (Nagura-Ikeda et al., 2020). However, one concern voiced by the National Public Health Emergency Team (NPHET) was inappropriate use of antigen testing where a negative test is produced for infected symptomatic individuals, and they proceed to stop isolating and go back into the community[3]. Such scenarios demonstrate the complexity of modelling agent behaviour. Introducing two levels of restrictions focused on movement of individuals in the environment. Level 5 restrictions involved much less free movement around the environment in order to reduce the risk of agents having close contact. The reduced free movement time also represented travel restrictions since agents stayed reasonable close to their own households. A limitation of the model was that it did not differentiate between a normal working week and the Christmas period where people's behaviour would have been quite different. Also, the Level 5 restrictions came into effect from December 24$^{\text{th}}$ 2020 and no data exists on adherence to these rules for the Christmas period in 2020. The assumption in this model was that the population of agents all adhered to the Level 5 restrictions from December 24$^{\text{th}}$.

Three experiments were conducted in order to validate the model before the main experiment was performed in this study. A limitation of agent based models in general is the difficulty of validating them due to their stochastic nature. The danger with over-fitting an agent based model to real world data is that there may be a number of configurations of parameters that predict the same observations. The validation of this model did have mixed results, as discussed now. During the cross validation between the agent based model and equations based model, it was expected to have results that were similar, but the validation produced results that were surprisingly comparable,

[3]https://www.rte.ie/news/ireland/2021/1121/1262274-nphet-antigen-tests/

including very high correlations. Similarly, the sensitivity analysis produced excellent results with the agent based model showing it was responsive to changes to parameter values. Changing the values of the transmissibility and level 5 lockdown start date did affect how the outbreak developed in the simulation, which was as expected. One limitation found during the sensitivity analysis was discussed in Section 4.1.3, where the model performed better in terms of disease spread when PCR testing was not performed. This was due to the accuracy level of the PCR test and 100% compliance of infected agents remaining in isolation. This limitation did not affect the results of this study since testing was enabled for all scenarios, and a small amount of such leakage of false negatives back into the simulation environment may reflect behaviour in the real world. Finally, comparing the data generated by the simulation to the real world data for this specific outbreak did generate a high correlation which showed that the model was linearly related to the real world outbreak. However, the simulated data and the real world data were not from the same distribution when the simulated data, which was a sample, was scaled up to full size. This may be due to relying solely on disease attributes from the literature, instead of fine tuning their values against real world data using a grid search type function. When the simulated data was scaled up by a factor of 4, however, it did appear to be from the same distribution as the real world data. The real world data was also validated when tested whether it fell into the distribution of simulated outcomes by calculating the z-score of cumulative confirmed cases. While these results are mixed, they do show that the overall pattern of behaviour of the model was comparable to the real world outbreak in county Carlow and the experiment could progress with confidence.

The data generated by the experiment compared each random testing scenario to the base model and the results showed a statistically significant difference in means between the base scenario and two scenarios where 40% and 50% of testing capacity were reserved for random testing. Two additional scenarios of 20% and 30% were also very close to being statistically significant. These results showed that, in this simulation, random testing augmenting a close contact testing regime could help to

reduce the spread of COVID-19 such that the number of deaths could be reduced, compared to close contact tracing alone. The detection rate of infected agents across all random testing scenarios was approximately 1.6% (number of detected cases from random testing / total random tests performed). Even with such a small detection rate, this had a significant impact on the spread of the disease in the simulation. The agents selected for random testing were asymptomatic and had not been previously infected. This approach sought to take as targeted approach as possible to detect asymptomatic infected agents since symptomatic agents were already being tested through the standard PCR testing regime. This also demonstrates that both testing regimes work best when used together, and not with one replacing the other. A real world regime may dynamically adjust the percentage of random testing during an outbreak depending on demand, but this study assumed that the percentage of capacity reserved for random testing remained constant during each scenario.

## 5.4 Contributions and impact

The purpose of building epidemiology models, such as the one in this study, is to inform decisions on response policies and project their potential impact on the incidence of the disease. The results of the study focused on the outcome in the simulated scenarios, but they could also have implications for real world policy decisions. The study has shown that even a modest increase in the detection of asymptomatic infected people can have a significant impact on the spread of the disease, and although the implementation of a random testing regime may be difficult to coordinate in the real world, the benefits may outweigh the concerns. In addition, this epidemiology model has been built with multiple interventions available which allows additional complex scenarios to be constructed. The model can also be extended with additional societal and disease dynamics as the pandemic moves into new phases, for example the addition of vaccination.

This model has been built for a specific outbreak in a specific county in Ireland. However, the combination of the NetLogo model and Python scripts presented in this study offer a powerful tool set that can be used by researchers to build local level scenarios in other counties quickly[4]. The pre-processing Python script[5] has been generalised so that a county and sample size can be customised by changing two variables at the beginning of the script. The script outputs the relevant small areas, household, and agent data which are imported automatically by the NetLogo software during the initialisation process. Such customisation offers the ability to quickly respond to changing conditions in specific areas, or to implement a targeted response across a heterogeneous environment.

## 5.5 Future Work and recommendations

Some areas for future work have been discussed in this study, but are listed here for completeness. In addition, a number of other ideas for future work have also been included.

- Social network analysis could be used to build sub-communities within the simulation. People with larger social networks may be at a higher risk of becoming exposed to the disease. Expanding on this idea, it may be interesting to investigate whether personality type could affect the size of people's social networks, and therefore affect the risk of becoming exposed to the disease.

- Vaccination data has not been included in this model since the vaccination programmes were not prevalent during the time of the outbreak being modeled. This would be relatively easy to add to the model once the data were available. It would be interesting to simulate what level of vaccination would be required for herd immunity, taking the prevalence of new, more transmissible variants into account.

[4]https://github.com/johnmcdonne11/dissertation

[5]https://github.com/johnmcdonne11/dissertation/blob/main/2.%20Preprocessing.ipynb

- Since newer variants appear to be less severe, examining the risk of severe disease to an un-vaccinated cohort within a population of mostly vaccinated people could be used to measure the impact of COVID-19 on health system capacity (Abdullah et al., 2022).

- This study presented random testing as a non-pharmaceutical response to the pandemic, which was based on the PCR test. However, the use of antigen testing earlier in the pandemic may have offered similar or better results. This could also be an interesting topic for further research.

- Including geographically aware maps could be used to examine the role of movement on the spread of the pandemic. Mobile data is also available from Google[6] and Apple[7] which could be leveraged as part of this research.

[6]https://www.google.com/covid19/mobility/

[7]https://covid19.apple.com/mobility

# References

Abdullah, F., Myers, J., Basu, D., Tintinger, G., Ueckermann, V., Mathebula, M., ... Jassat, W. (2022). Decreased severity of disease during the first global omicron variant covid-19 outbreak in a large hospital in tshwane, south africa. *International Journal of Infectious Diseases*, *116*, 38-42. Retrieved from `https://www.sciencedirect.com/science/article/pii/S120197122101256X` doi: https://doi.org/10.1016/j.ijid.2021.12.357

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., ... Vespignani, A. (2010, June). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, *10*(1), 190.

Al-qaness, M. A. A., Ewees, A. A., Fan, H., & Abd El Aziz, M. (2020). Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *Journal of Clinical Medicine*, *9*(3). doi: 10.3390/jcm9030674

Barbarossa, M. V., Fuhrmann, J., Meinke, J. H., Krieg, S., Varma, H. V., Castelletti, N., & Lippert, T. (2020, 09). Modeling the spread of covid-19 in germany: Early assessment and possible scenarios. *PLOS ONE*, *15*(9), 1-22. Retrieved from `https://doi.org/10.1371/journal.pone.0238559` doi: 10.1371/journal.pone.0238559

Barek, M. A., Aziz, M. A., & Islam, M. S. (2020). Impact of age, sex, comorbidities and clinical symptoms on the severity of covid-19 cases: A meta-analysis with 55 studies and 10014 cases. *Heliyon*, *6*(12), e05684. Retrieved from `https://www`

.sciencedirect.com/science/art\icle/pii/S2405844020325275 doi: https://doi.org/10.1016/j.heliyon.2020.e05684

Bastos, L. S., Ranzani, O. T., Souza, T. M. L., Hamacher, S., & Bozza, F. A. (2021, August). COVID-19 hospital admissions: Brazil's first and second waves compared. *The Lancet Respiratory Medicine*, *9*(8), e82–e83. Retrieved 2022-01-08, from https://doi.org/10.1016/S2213-2600(21)00287-3 (Publisher: Elsevier) doi: 10.1016/S2213-2600(21)00287-3

Bootsma, M. C. J., & Ferguson, N. M. (2007). The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proceedings of the National Academy of Sciences*, *104*(18), 7588–7593. Retrieved from https://www.pnas.org/content/104/18/7588 (Publisher: National Academy of Sciences _eprint: https://www.pnas.org/content/104/18/7588.full.pdf) doi: 10.1073/pnas.0611071104

Brazeau, N., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., ... others (2020). Report 34: Covid-19 infection fatality ratio: estimates from seroprevalence.

Breban, R., Vardavas, R., & Blower, S. (2007, March). Theory versus data: how to calculate R0? *PloS one*, *2*(3), e282–e282. Retrieved from https://pubmed.ncbi.nlm.nih.gov/17356693 (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0000282

Carfì, A., Bernabei, R., Landi, F., & for the Gemelli Against COVID-19 Post-Acute Care Study Group. (2020, 08). Persistent Symptoms in Patients After Acute COVID-19. *JAMA*, *324*(6), 603-605. Retrieved from https://doi.org/10.1001/jama.2020.12603 doi: 10.1001/jama.2020.12603

Carpenter, C., & Sattenspiel, L. (2009, June). The design and use of an agent-based model to simulate the 1918 influenza epidemic at Norway House, Manitoba. *American journal of human biology : the official journal of the Human Biology Council*, *21*(3), 290–300. (Place: United States) doi: 10.1002/ajhb.20857

Chen, J., Qi, T., Liu, L., Ling, Y., Qian, Z., Li, T., ... Lu, H. (2020, May). Clinical progression of patients with COVID-19 in Shanghai, China. *The Journal of infection*, *80*(5), e1–e6. doi: 10.1016/j.jinf.2020.03.004

Choi, S., & Ki, M. (2020, March). Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiol Health*, *42*(0), e2020011–0. Retrieved from `https://doi.org/10.4178/epih.e2020011` (Publisher: Korean Society of Epidemiology) doi: 10.4178/epih.e2020011

Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schnemann, H. J., ... Schnemann, H. J. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The Lancet*, *395*(10242), 1973-1987. Retrieved from `https://www.sciencedirect.com/science/art\icle/pii/S0140673620311429` doi: https://doi.org/10.1016/S0140-6736(20)31142-9

Drew, D. A., Nguyen, L. H., Steves, C. J., Menni, C., Freydin, M., Varsavsky, T., ... Spector, T. (2020). Rapid implementation of mobile technology for real-time epidemiology of covid-19. *Science*, *368*(6497), 1362-1367. doi: 10.1126/science.abc0473

Du, Z., Xu, X., Wu, Y., Wang, L., Cowling, B. J., & Meyers, L. A. (2020). Serial interval of covid-19 among publicly reported confirmed cases. *Emerging infectious diseases*, *26*(6), 1341.

E. Frias-Martinez, G. Williamson, & V. Frias-Martinez. (2011, October). An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (pp. 57–64). (Journal Abbreviation: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing) doi: 10.1109/PASSAT/SocialCom.2011.142

El-Sayed, A. M., Scarborough, P., Seemann, L., & Galea, S. (2012, February). Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations*, *9*(1), 1. Retrieved from https://doi.org/10.1186/1742-5573-9-1 doi: 10.1186/1742-5573-9-1

Evans, S. J., & Jewell, N. P. (2021). Vaccine effectiveness studies in the field. *New England Journal of Medicine*, *385*(7), 650-651. Retrieved from https://doi.org/10.1056/NEJMe2110605 doi: 10.1056/NEJMe2110605

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., ... others (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. imperial college covid-19 response team. *Imperial College COVID-19 Response Team*, 20.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., ... Imperial College COVID-19 Response Team (2020, August). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. Retrieved from https://doi.org/10.1038/s41586-020-2405-7 doi: 10.1038/s41586-020-2405-7

Godio, A., Pace, F., & Vergnano, A. (2020). SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence. *International Journal of Environmental Research and Public Health*, *17*(10). doi: 10.3390/ijerph17103535

Greer, S. L., King, E. J., da Fonseca, E. M., & Peralta-Santos, A. (2020). The comparative politics of covid-19: The need to understand government responses. *Global Public Health*, *15*(9), 1413-1416. Retrieved from https://doi.org/10.1080/17441692.2020.1783340 (PMID: 32564670) doi: 10.1080/17441692.2020.1783340

Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Saptarshi Majumdar, ... Samuel Webster (2020). *Oxford COVID-19 Government Response Tracker.* Blavatnik School of Government. Retrieved 2020-11-01, from https://www.bsg.ox.ac.uk/research/researc\h-projects/coronavirus-government-respons\e-tracker

He, D., Dushoff, J., Day, T., Ma, J., & Earn, D. J. D. (2013, September). Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131345. Retrieved 2020-12-11, from `https://doi.org/10.1098/rspb.2013.1345` (Publisher: Royal Society) doi: 10.1098/rspb.2013.1345

He, S., Peng, Y., & Sun, K. (2020, August). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics*, *101*(3), 1667–1680. Retrieved from `https://doi.org/10.1007/s11071-020-05743-y` doi: 10.1007/s11071-020-05743-y

Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Rico, M. S., ... Leleu, H. (2020, September). A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine*, *26*(9), 1417–1421. Retrieved from `https://doi.org/10.1038/s41591-020-1001-6` doi: 10.1038/s41591-020-1001-6

Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., ... others (2020). The continuing 2019-ncov epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in wuhan, china. *International journal of infectious diseases*, *91*, 264–266.

Hunter, E., & Kelleher, J. D. (2020, October). A framework for validating and testing agent-based models: a case study from infectious disease modelling. Toulouse, France. Retrieved 2022-12-01, from `https://doi.org/10.21427/2xjb-cq79` doi: doi:10.21427/2xjb-cq79

Hunter, E., & Kelleher, J. D. (2021). Adapting an agent-based model of infectious disease spread in an irish county to covid-19. *Systems*, *9*(2). Retrieved from `https://www.mdpi.com/2079-8954/9/2/41` doi: 10.3390/systems9020041

Hunter, E., Mac Namee, B., & Kelleher, J. (2020). A hybrid agent-based and equation based model for the spread of infectious diseases. *Journal of Artificial Societies & Social Simulation*, *23*(4).

Ibarra-Vega, D. (2020, August). Lockdown, one, two, none, or smart. Modeling containing covid-19 infection. A conceptual model. *Science of The Total Environment*, *730*, 138917. Retrieved 2020-11-21, from http://www.sciencedirect.com/science/arti\cle/pii/S0048969720324347 doi: 10.1016/j.scitotenv.2020.138917

Inje, B., Kumar, S., & Nayyar, A. (2019). Swarm Intelligence and Evolutionary Algorithms in Disease Diagnosis - Introductory Aspects. In *Swarm Intelligence and Evolutionary Algorithms in Disease Diagnosis* (pp. 1–18). CRC Press.

Ivorra, B., Ferrández, M., Vela-Pérez, M., & Ramos, A. (2020, September). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, *88*, 105303. Retrieved from http://www.sciencedirect.com/science/arti\cle/pii/S1007570420301350 doi: 10.1016/j.cnsns.2020.105303

Kaushal, V., & Srivastava, S. (2021). Hospitality and tourism industry amid covid-19 pandemic: Perspectives on challenges and learnings from india. *International Journal of Hospitality Management*, *92*, 102707. Retrieved from https://www.sciencedirect.com/science/article/pii/S0278431920302590 doi: https://doi.org/10.1016/j.ijhm.2020.102707

Kennelly, B., O'Callaghan, M., Coughlan, D., Cullinan, J., Doherty, E., Glynn, L., ... Queally, M. (2020). The COVID-19 pandemic in Ireland: An overview of the health service and economic policy response. *Health Policy and Technology*, *9*(4), 419–429. Retrieved from https://www.sciencedirect.com/science/article/pii/S2211883720300952 doi: https://doi.org/10.1016/j.hlpt.2020.08.021

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., ... Klein, D. J. (2021, 07). Covasim: An agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, *17*(7), 1-32. Re-

trieved from https://doi.org/10.1371/journal.pcbi.1009149 doi: 10.1371/journal.pcbi.1009149

Khatami, F., Saatchi, M., Zadeh, S. S. T., Aghamir, Z. S., Shabestari, A. N., Reis, L. O., & Aghamir, S. M. K. (2020, December). A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis. *Scientific Reports*, *10*(1), 22402. Retrieved from https://doi.org/10.1038/s41598-020-80061-2 doi: 10.1038/s41598-020-80061-2

Kozyreva, A., Lorenz-Spreen, P., Lewandowsky, S., Garrett, P. M., Herzog, S. M., Pachur, T., & Hertwig, R. (2021, September). Psychological factors shaping public responses to COVID-19 digital contact tracing technologies in Germany. *Scientific Reports*, *11*(1), 18716. Retrieved from https://doi.org/10.1038/s41598-021-98249-5 doi: 10.1038/s41598-021-98249-5

Lam, S., Lombardi, A., & Ouanounou, A. (2020). Covid-19: A review of the proposed pharmacological treatments. *European Journal of Pharmacology*, *886*, 173451. Retrieved from https://www.sciencedirect.com/science/article/pii/S0014299920305434 doi: https://doi.org/10.1016/j.ejphar.2020.173451

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.

Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., ... Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine*, *9*(2), 538.

Liu, M., Ning, J., Du, Y., Cao, J., Zhang, D., Wang, J., & Chen, M. (2020). Modelling the evolution trajectory of covid-19 in wuhan, china: experience and suggestions. *Public health*, *183*, 76–80.

Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., ... Ramsay, M. (2021). Effectiveness of covid-19 vaccines against the b.1.617.2 (delta) variant. *New England Journal of Medicine*, *385*(7), 585-594. Retrieved from https://doi.org/10.1056/NEJMoa2108891 (PMID: 34289274) doi: 10.1056/NEJMoa2108891

Mardani, R., Ahmadi Vasmehjani, A., Zali, F., Gholami, A., Mousavi Nasab, S. D., Kaghazian, H., ... Ahmadi, N. (2020, April). Laboratory Parameters in Detection of COVID-19 Patients with Positive RT-PCR; a Diagnostic Accuracy Study. *Archives of academic emergency medicine*, *8*(1), e43–e43. Retrieved from https://pubmed.ncbi.nlm.nih.gov/32259132 (Publisher: Shahid Beheshti University of Medical Sciences)

Morawska, L., Tang, J. W., Bahnfleth, W., Bluyssen, P. M., Boerstra, A., Buonanno, G., ... Yao, M. (2020). How can airborne transmission of COVID-19 indoors be minimised? *Environment International*, *142*, 105832. Retrieved from https://www.sciencedirect.com/science/article/pii/S0160412020317876 doi: https://doi.org/10.1016/j.envint.2020.105832

Murphy, J., Vallières, F., Bentall, R. P., Shevlin, M., McBride, O., Hartman, T. K., ... Hyland, P. (2021, January). Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. *Nature Communications*, *12*(1), 29. Retrieved from https://doi.org/10.1038/s41467-020-20226-\9 doi: 10.1038/s41467-020-20226-9

Muthukaruppan, S., & Er, M. (2012, October). A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. *Expert Systems with Applications*, *39*(14), 11657–11665. Retrieved from http://www.sciencedirect.com/science/arti\cle/pii/S0957417412006331 doi: 10.1016/j.eswa.2012.04.036

Nagura-Ikeda, M., Imai, K., Tabata, S., Miyoshi, K., Murahara, N., Mizuno, T., ... Kato, Y. (2020, August). Clinical Evaluation of Self-Collected Saliva by

Quantitative Reverse Transcription-PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription-Loop-Mediated Isothermal Amplification, and a Rapid Antigen Test To Diagnose COVID-19. *Journal of clinical microbiology*, *58*(9). doi: 10.1128/JCM.01438-20

O'Driscoll, M., Dos Santos, G. R., Wang, L., Cummings, D. A., Azman, A. S., Paireau, J., ... Salje, H. (2021). Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, *590*(7844), 140–145.

Panovska-Griffiths, J., Kerr, C. C., Stuart, R. M., Mistry, D., Klein, D. J., Viner, R. M., & Bonell, C. (2020). Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second covid-19 epidemic wave in the uk: a modelling study. *The Lancet Child & Adolescent Health*, *4*(11), 817-827. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352464220302509` doi: https://doi.org/10.1016/S2352-4642(20)30250-9

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., ... Gruber, W. C. (2020). Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *New England Journal of Medicine*, *383*(27), 2603-2615. Retrieved from `https://doi.org/10.1056/NEJMoa2034577` (PMID: 33301246) doi: 10.1056/NEJMoa2034577

Pradhan, D., Biswasroy, P., Naik, P. K., Ghosh, G., & Rath, G. (2020). A review of current interventions for covid-19 prevention. *Archives of medical research*, *51*(5), 363–374.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., ... Klepac, P. (2020, January). The effect of control strategies that reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China. *medRxiv*, 2020.03.09.20033050. Retrieved from `http://medrxiv.org/content/early/2020/03/\12/2020.03.09.20033050.abstract` doi: 10.1101/2020.03.09.20033050

Railsback, S., Ayllon, D., Berger, U., Grimm, V., Lytinen, S., Sheppard, C., & Thiele, J. C. (2017). Improving execution speed of models implemented in netlogo. doi: 10.18564/jasss.3282

Redmond, P., McGuinness, S., et al. (2020). Essential employees during the covid-19 crisis. *ESRI Surv. Stat. Rep. Ser*.

Saravanan, M., Karthikeyan, P., Arathi, A., Kiruthika, M., & Suganya, S. (2013). Mobile Agent-Based Approach for Modeling the Epidemics of Communicable Diseases. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 16–20). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2492517.2492612` (event-place: Niagara, Ontario, Canada) doi: 10.1145/2492517.2492612

Shim, E., Tariq, A., Choi, W., Lee, Y., & Chowell, G. (2020). Transmission potential and severity of covid-19 in south korea. *International Journal of Infectious Diseases*, *93*, 339–344.

Shuvo, S. B., Molokwu, B. C., & Kobti, Z. (2020). Simulating the Impact of Hospital Capacity and Social Isolation to Minimize the Propagation of Infectious Diseases. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3451–3457). Virtual Event, CA, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3394486.3412859` (Type: 10.1145/3394486.3412859)

Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., & Silva, R. C. (2020, October). COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals*, *139*, 110088. Retrieved 2020-11-16, from `http://www.sciencedirect.com/science/arti\cle/pii/S0960077920304859` doi: 10.1016/j.chaos.2020.110088

Singh, M., Marathe, A., Marathe, M. V., & Swarup, S. (2018). Behavior Model Calibration for Epidemic Simulations. In *Proceedings of the 17th International Confer-*

*ence on Autonomous Agents and MultiAgent Systems* (pp. 1640–1648). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems. Retrieved from https://dl.acm.org/doi/abs/10.5555/323738\3.3237943 (event-place: Stockholm, Sweden)

Smirnova, A., deCamp, L., & Chowell, G. (2019, November). Forecasting Epidemics Through Nonparametric Estimation of Time-Dependent Transmission Rates Using the SEIR Model. *Bulletin of Mathematical Biology*, *81*(11), 4343–4365. Retrieved from https://doi.org/10.1007/s11538-017-0284-3 doi: 10.1007/s11538-017-0284-3

Tahamtan, A., & Ardebili, A. (2020). Real-time rt-pcr in covid-19 detection: issues affecting the results. *Expert review of molecular diagnostics*, *20*(5), 453–454.

Tisue, S., & Wilensky, U. (2004). Netlogo: A simple environment for modeling complexity. In *International conference on complex systems* (Vol. 21, pp. 16–21).

Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., & Marathe, M. (2018, March). Using data-driven agent-based models for forecasting emerging infectious diseases. *The RAPIDD Ebola Forecasting Challenge*, *22*, 43–49. Retrieved from http://www.sciencedirect.com/science/arti\cle/pii/S1755436517300221 doi: 10.1016/j.epidem.2017.02.010

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., ... others (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, *20*(6), 669–677.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., ... others (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. *Jama*, *323*(11), 1061–1069.

Whitelaw, S., Mamas, M. A., Topol, E., & Spall, H. G. C. V. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*, *2*(8), e435–e440. Retrieved from https://www.sciencedirect.com/

science/article/pii/S2589750020301424 doi: https://doi.org/10.1016/S2589-7500(20)30142-4

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., ... He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3). Retrieved from http://jtd.amegroups.com/article/view/363\85

Zeng, J.-H., Liu, Y.-X., Yuan, J., Wang, F.-X., Wu, W.-B., Li, J.-X., ... others (2020). First case of covid-19 complicated with fulminant myocarditis: a case report and insights. *Infection*, *48*(5), 773–777.

Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., ... others (2020). Changes in contact patterns shape the dynamics of the covid-19 outbreak in china. *Science*, *368*(6498), 1481–1486.

# Appendix A

# Code used in this study

The following code was used to perform exploratory data analysis, data pre-processing, model validation, and experiment results in Jupyter notebooks using Python, as well as building and executing the model in NetLogo. All code is stored in the following GitHub repository and is described in list below.

https://github.com/johnmcdonne11/dissertation

- 1. EDA.ipynb. This notebook was used to perform basic exploratory data analysis to understand the datasets used in this study.
- 2. Preprocessing.ipynb. This notebook was used to build a sample population of people, households, and small areas
- 3. Model Validation.ipynb. This notebook analysed the datasets produced by the NetLogo experiments for validating the model described in Section 4.1.
- 4. Results.ipynb. This notebook analysed the dataset produced by the main experiment as described in Section 4.2.
- covid-sim1.nlogo. This is the NetLogo file that contains all code and interface elements used for all experiments.

# Appendix B

# Screenshots of NetLogo Model

The following screenshots show the NetLogo model interface as well as some individual elements of the output. Figure B.1 shows a close up view of the simulated environment so that workplaces, households, hospitals, and person agents are visible. The colour of the people represents their epidemiology status. Figure B.2 shows the interface designed for researchers to run the model and receive real-time feedback, and Figure B.3 includes additional parameters for the disease dynamics. Finally, Figure B.4 shows the Behaviour Space module used to design experiments using the model.

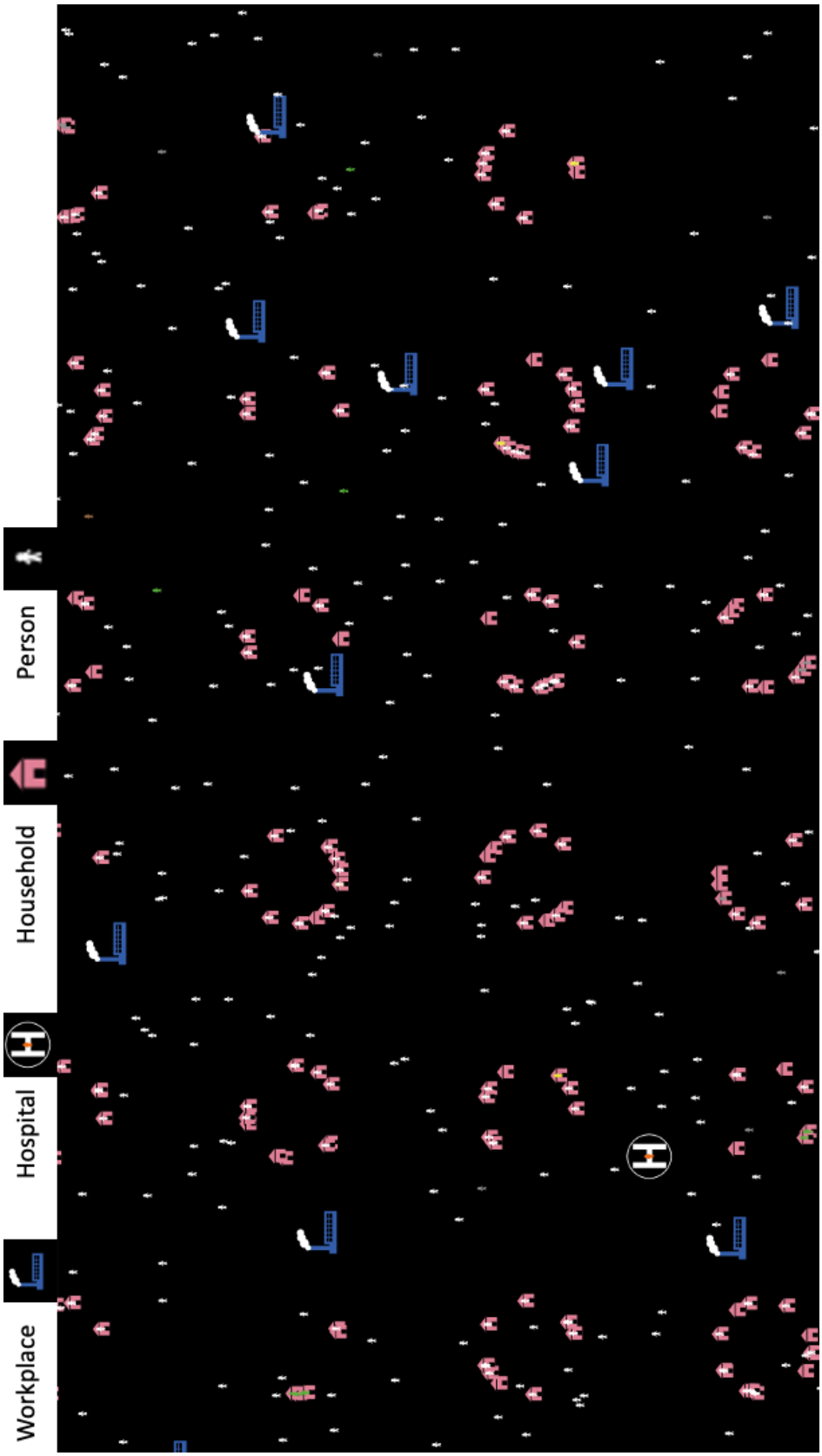



Figure B.1: Simulation Environment

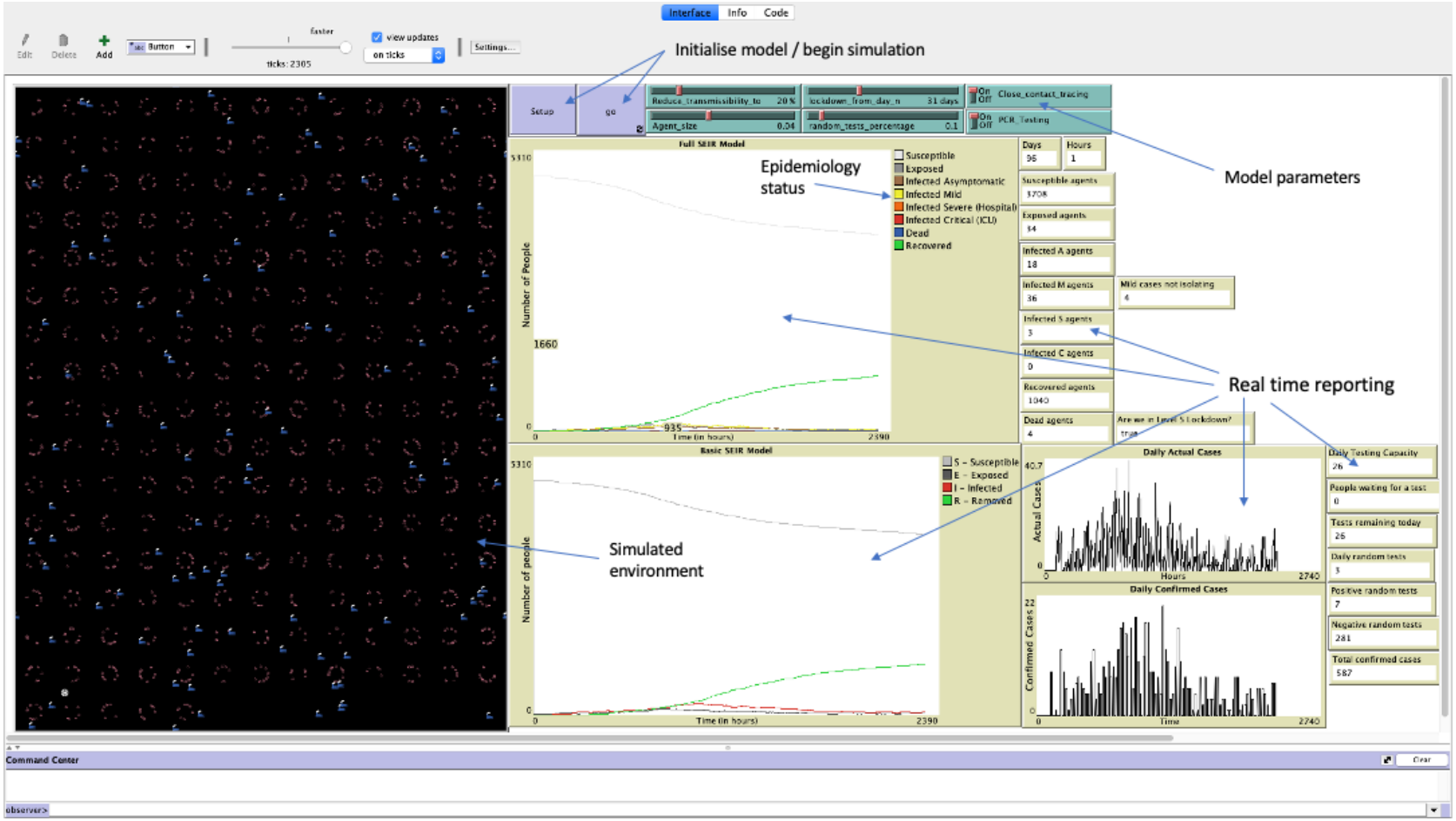



Figure B.2: Simulation Interface

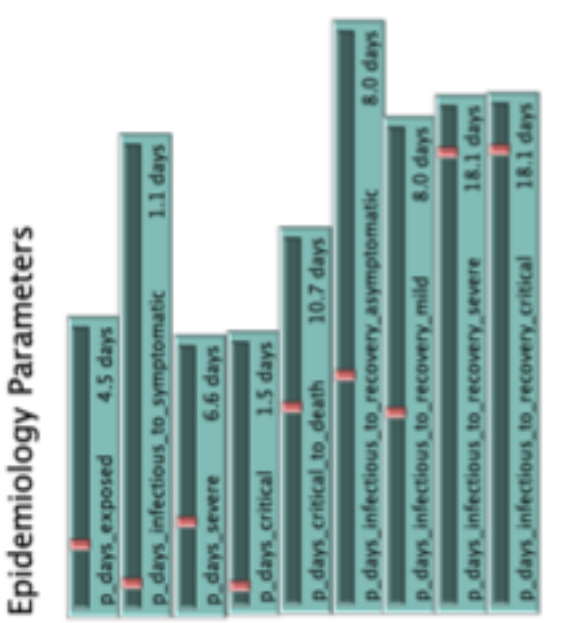



Figure B.3: Epidemiology parameters in the simulation

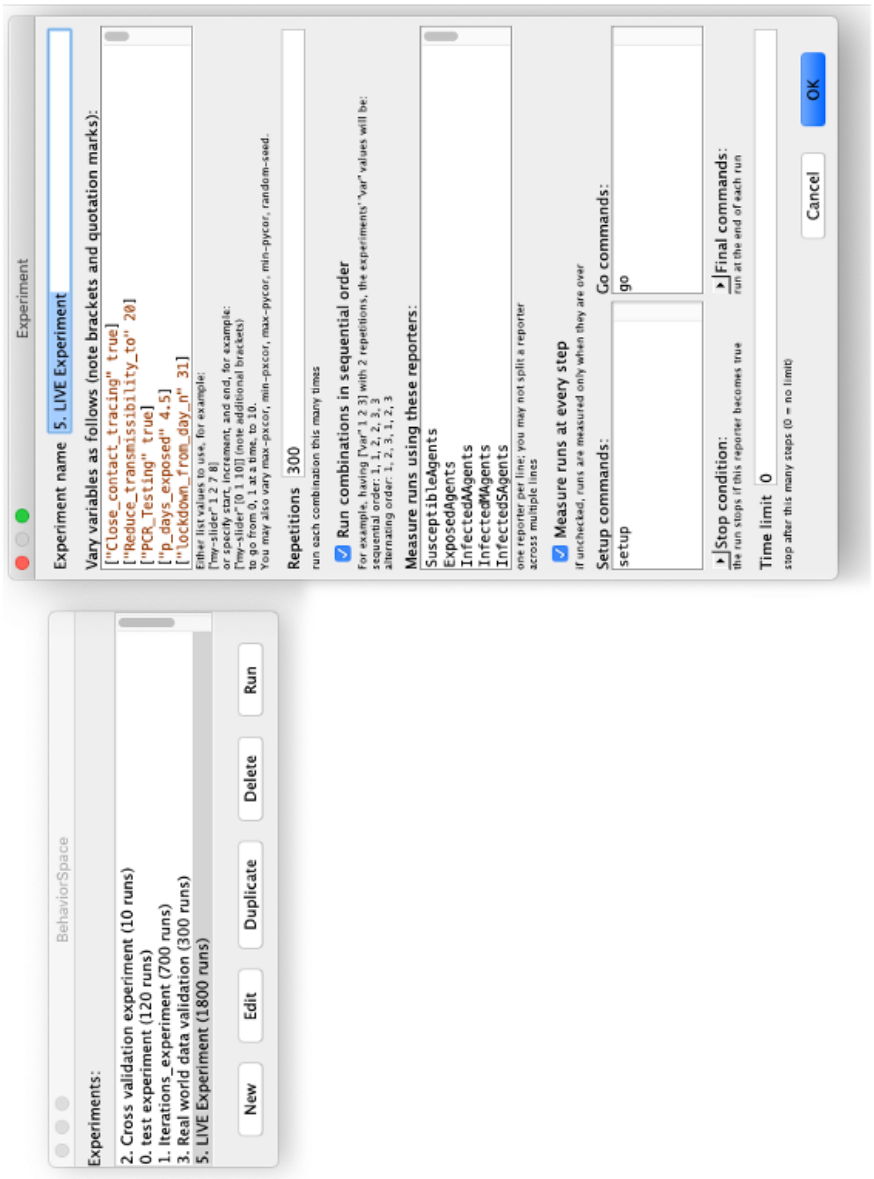



Figure B.4: Behaviour space used to generate experiments