A data-driven framework for managing and monitoring novel and endemic infectious diseases in Australia: Four essays

Chloe Morgan Rice

A thesis submitted for the degree of

Doctor of Philosophy

Crawford School of Public Policy

Australian National University

July 2024

© Copyright by Chloe Morgan Rice, 2024

All Rights Reserved

Declaration

This thesis has not previously been submitted for a degree at this, or any other university or institution. To the best of the author's knowledge, it contains no other material previously published or written by another person, except where due reference is made in the thesis.

19 July 2024

Acknowledgements

I would like to acknowledge the phenomenal support, patience and delicious cooking provided by my wonderful husband Chris Rice over the course of this very long, part-time PhD. I am very grateful to my family (Stewart, Sharyn, Elise and Kate Burns) for their enthusiastic encouragement, my faithful dogs (Milo and Duffy) for their companionship and emotional support, my mentor Dr Allyson Essex, for her guidance, and my thesis supervisors (Professor Quentin Grafton, Professor Long Chu and Professor Tom Kompas) for their valuable input and supervision. My sincere thanks also to the Department of Health and Aged Care for the provision of study leave enabling the completion of this thesis. Third party high level copy-editing of this thesis was provided by Maurice Neville, editor affiliated with the Canberra Society of Editors.

Disclaimer

This research has informed my paid work at the Australian Government Department of Health and Aged Care throughout the COVID-19 pandemic. The models described in this thesis were used to support the emergency response to COVID-19, covering the development of disease spread and vaccine allocation models to support decision making and contestability of advice. This research contributed to my receipt of a Public Service Medal awarded by the Governor General of Australia, awarded in my maiden name (Chloe Burns PSM) on January 26, 2023. Further details on how the contents of this thesis benefitted the Australian public are available at the award specification here:

https://honours.pmc.gov.au/honours/awards/2012795

This work continues to inform public policy outcomes, with the spine of the models presented in this thesis being used to develop the Health Outcomes Modelling and Evaluation (HOME) model at the Department of Health and Aged Care – a 25-year patient journey simulator for evidence-based policy intervention testing.

The views, opinions and conclusions contained in this thesis are my own and should not be interpreted as official Australian Government or Department of Health and Aged Care positioning or advice.

Abstract

Over the past decade there has been a substantial increase in the number of severe infectious disease outbreaks worldwide. Amalgamated with climate change, increased air travel and population growth, there is a strong need for foundational pandemic preparedness plans featuring data and evidence-based models at their core.

This research shifts the dial of what should be considered fundamental to pandemic preparedness and planning from a data and modelling perspective. Due to recent technological advances, the role and importance of data in decision making is still relatively novel. However, this does not mean that the establishment of data-driven functions should be delayed.

This thesis develops a framework for managing and monitoring novel and endemic infectious diseases in Australia, with a focus on the role of individual behaviours in influencing disease spread outcomes. The methodology has covered the development of an agent-based microsimulation model with individual behaviours, an early warning system to monitor infectious disease cases and potential waves, and an evidence-based scenario tool for monitoring disease spread, effectively embedding a “pandemic pulse” into preparedness capability.

This framework was established via the use of Agent Based Models and econometric techniques to generate a simple-to-use series of tools that can be embedded into business-as-usual pandemic planning and preparedness activities. This research found that incorporation of behavioural parameters greatly enhances the way disease spread trajectories can be modelled, which when coupled with econometric early warning systems can be used to generate a scenario tool for modelling the likely impact of interventions prior to the signalled wave.

This research allows a better understanding of emerging and existing infectious diseases and how they might affect the Australian population. By improving our understanding of likely disease outcomes, decision makers are empowered to inform public health measures that target the minimisation of individual harm and mortality associated with an infectious disease.

Table of Contents

1. Introduction 1

2. Literature review 9

2.1 Evolution of the COVID-19 pandemic in Australia 10

2.2 Methods and approaches to modelling pandemics 11

2.2.1 Importance of external factors in building disease spread models 16

2.3 The role of early warning systems for pandemic preparedness 21

2.4 Data-driven policy decision making 23

2.5 Conclusions 27

3. Essay one: A survey-driven microsimulation framework to measure the effect of behaviour propagation on the spread of COVID-19 29

3.1 Introduction 29

3.2 Methods 33

3.2.1 State variables and scales 34

3.2.1 Parameters 35

3.2.2 Process overview, restrictions and scheduling 38

3.2.3 Evaluation and sensitivity testing 47

3.3 Results 51

3.3.1 Behaviour trends 51

3.3.2 Second wave results 54

3.3.3 Evaluation of second wave results 56

3.3.4 Marginal utility of fearful and careful behaviours 61

3.3.5 Second wave sensitivity analysis 62

3.3.6 Preliminary third wave results 66

3.4 Discussion 66

4. Essay two: An input-driven microsimulation framework to measure the effect of behaviour propagation on endemic infectious disease spread 70
4.1 Introduction 70
4.2 Methods 73
4.2.1 State variables and scales 74
4.2.2 Parameters 76
4.2.3 Process overview, restrictions and schedules 78
4.2.4 Evaluation and sensitivity testing 82
4.3 Results 83
4.3.1 Disease trajectory accuracy 83
4.4 Discussion 87
5. Essay three: An econometric early warning system for signalling COVID-19 waves 90
5.1 Introduction 90
5.2 Methods 94
5.2.1 Modelling volatility 95
5.2.2 Specification of ARCH models 95
5.2.3 Forecast evaluation: variance series selection 97
5.2.4 Velocity of new case volatility 98
5.3 Results 99
5.3.1 Forecasting evaluation 99
5.3.2 Variance series 101
5.3.3 National level signals 101
5.3.4 State and Territory signals 106
5.3.5 Performance metrics 113
5.4 Discussion 115
6. Essay four: Embedding a pandemic pulse into preparedness capability: model outputs and policy implications 117
6.1 Introduction 117

6.2 Methods....119
6.2.1 State variables and scales....119
6.2.2 Parameters....123
6.2.3 Process overview, restrictions and scheduling....125
6.2.4 Evaluation and sensitivity testing....127
6.3 Results....128
6.3.1 Infection curves....130
6.3.2 Impact on cases due to signal intervention....131
6.4 Policy implications....132
6.5 Discussion....136
7. Conclusions....138
8. References....145
Appendix A....161

List of Figures

Figure 1.1 Framework for managing and monitoring novel and endemic infectious diseases in Australia 6
Figure 1.2 Timeline of analysis (essays) against Victorian daily new COVID-19 cases 7
Figure 2.1. CASES: Cumulative COVID-19 cases in Australia, by calendar year 11
Figure 2.2. DEATHS: Cumulative COVID-19 deaths in Australia, by calendar year 11
Figure 3.1. Behavioural states and disease spread outcomes 32
Figure 3.2. High level model overview 36
Figure 3.3. Proportion of the population by behavioural compartment (second wave date range) 40
Figure 3.4. Daily activity movement bias 41
Figure 3.5. Contact and infection space 44
Figure 3.6. Average and observed proportion of the population by behavioural compartment 52
Figure 3.7. Interactions between average behavioural compartments and average confirmed infections 53
Figure 3.8. COVID-19 infection case counts 54
Figure 3.9. Temporal comparison of COVID-19 confirmed and exposed cases 55
Figure 3.10. Transmission potential against the proportion of the population confirmed 55
Figure 3.11. Model comparisons against observed confirmed COVID-19 cases 59
Figure 3.12. Infection accuracy metrics and accompanying figures 60
Figure 3.13. Average maximum cumulative infections under varying eligibility criteria 63
Figure 3.14 Average maximum cumulative infections complacent transmission scenarios 64
Figure 3.15 Average maximum cumulative infections careful transmission scenarios 64
Figure 3.16 Average maximum cumulative infections fearful transmission scenarios 65
Figure 3.17 Average maximum cumulative infections careful movement reduction scenarios 65
Figure 3.18. Preliminary third wave results 66
Figure 4.1. High level model overview 75
Figure 4.2. Proportion of population infected – simulated and actual 86
Figure 5.1. National level analysis periods – new cases of COVID-19 94
Figure 5.2. New cases variance series (volatility) Australia 101
Figure 5.3. Period 1 new cases of COVID-19 and estimated signal (National) 102

Figure 5.4. Period 2 new cases of COVID-19 and estimated signal (National) 103
Figure 5.5. Period 3 new cases of COVID-19 and estimated signal (National) 104
Figure 5.6. Truncated period 3 new cases of COVID-19 and estimated signal (National) ... 105
Figure 5.7 Period 4 new cases of COVID-19 and estimated signal (National) 106
Figure 5.9. Periods 1 – 4 new cases of COVID-19 and estimated signal New South Wales 109
Figure 5.10. Periods 1 – 4 new cases of COVID-19 and estimated signal Victoria 109
Figure 5.11. Periods 1 – 4 new cases of COVID-19 and estimated signal Queensland 110
Figure 5.12. Periods 1 – 4 new cases of COVID-19 and estimated signal Western Australia 110
Figure 5.13. Periods 1 – 4 new cases of COVID-19 and estimated signal South Australia .. 111
Figure 5.14. Periods 1 – 4 new cases of COVID-19 and estimated signal Tasmania 111
Figure 5.15. Periods 1 – 4 new cases of COVID-19 and estimated signal Australian Capital Territory 112
Figure 5.16. Periods 1 – 4 new cases of COVID-19 and estimated signal Northern Territory 112
Figure 6.1. High level overview of model 122
Figure 6.2 Signal detection, action and simulation dates against new daily COVID-19 cases in Victoria 126
Figure 6.3 Average proportion of the population in careful and complacent compartments for scenario 2 (10% boost to careful population), scenario 3 (20% boost to careful population) and scenario 4 (30% boost to careful population) 129
Figure 6.4. Total infections by scenario 131

List of Tables

Table 2.1. Comparison of modelling approaches 16
Table 2.2. EWS methodologies in the literature 23
Table 2.3. Efficacy of policy recommendations and/or mandates on disease spread 25
Table 3.1. Model parameters 37
Table 3.2. Cumulative infections (proportion of the population) as at 31 January 2021 39
Table 3.3. Infection risk calculation 44
Table 3.4. Additive transmission rates 45
Table 3.5. Sensitivity tests 50
Table 3.6. Granger-causality results (p-values) 56
Table 3.7. Significant univariate least squares regression coefficients 56
Table 3.8. Eligible simulations based on model type 57
Table 3.9. Model accuracy metrics 60
Table 3.10. Probability of reduced transmission 61
Table 4.1. Model parameters 77
Table 4.2. Model estimate and actual seasonal outbreak results 85
Table 4.3. Model accuracy 87
Table 5.1. Descriptive statistics of new COVID-19 cases (daily) Nationally and by State and Territory 93
Table 5.2. National model evaluation 100
Table 5.3. State and Territory model evaluation results 101
Table 5.4. Signal performance metrics (National) 114
Table 5.5. Signal performance metrics – prior-to-peak (State and Territory) 114
Table 6.1. Model parameters 124
Table 6.2. Sensitivity analysis 127
Table 6.3. Cumulative infections by scenario 130
Table 6.4. Eligible simulations by scenario 130
Table 6.5. Impact due to signal intervention on cumulative case counts 132

Abbreviations and acronyms

ABM	Agent Based Model
ABS	Australian Bureau of Statistics
AIHW	Australian Institute of Health and Welfare
CSD	Critical Slowing Down
DALY	Disability Adjusted Life Year
EWS	Early Warning System
FRB	Fear Related Behaviour
ODE	Ordinary Differential Equation
PCR	Polymerase Chain Reaction
PPE	Personal Protective Equipment
RAT	Rapid Antigen Test
SEIR	Susceptible – Exposed – Infected – Recovered
SEIRS	Susceptible – Exposed – Infected – Recovered – Susceptible
SIR	Susceptible – Infected – Recovered
SIS	Susceptible – Infected – Susceptible

1. Introduction

Human behaviour is the ruling force of infectious disease spread. Individual movements, interactions and choices have been, and will continue to be, a primary driver of infectious disease related clinical and economic outcomes. This thesis builds on the already substantial literature developed in response to COVID-19 regarding mathematical disease spread modelling (e.g. Chang et al. 2020 Johnston and Pell, 2020; Kochanczyk et al. 2020; Pedro et al. 2020; Steinegger et al. 2020), early warning system modelling (e.g. Brett et al. 2020 and Miller et al. 2017) and methods for embedding a data-driven approach into policy making (e.g. Fearnley and Dixon, 2020).

There has been an exponential increase in the available research on infectious disease dynamics in recent years. Nevertheless, there is a scarcity of research focussing on the role of human behaviour and how such behaviour may change in response to novel infectious diseases such as COVID-19. Prior to the COVID-19 pandemic, few studies incorporated the scale and capricious nature of behavioural responses at the individual level, or modelled how volatility in individual behaviours may influence the overall spread of disease. In addition, the role of behaviour in implementing and embedding early warning systems has not been considered, meaning that critical points for evidence-based interventions to promote personal protective behaviours may have been missed. In this thesis, I argue that current and future monitoring and management of novel and endemic infectious diseases require greater consideration of human behaviour and underlying disease spread mechanisms to be able to appropriately plan and respond to health-related threats.

Several different types of modelling frameworks have been developed to assess the spread of infectious disease across various transmission modes. Standard compartmental models such as the Susceptible – Infected – Recovered (SIR), Susceptible – Infected – Susceptible (SIS) and Susceptible – Exposed – Infected – Recovered (SEIR) approaches are popular for modelling the spread of infectious disease. These models, however, lack the nuance and sophistication of other methodologies such as agent-based models (ABMs) and machine learning approaches. This thesis argues that ABMs are a valuable method for modelling the role of individual behaviours on the spread of novel and endemic infectious diseases. ABMs have been deployed over a broad range of applications, including cholera (Crooks and Hailegiorgic, 2014), avian influenza (Rao and Rao, 2009), ebola virus (Shultz et al. 2014), toxoplasmosis (Jiang, Sullivan and Zhao, 2012), tuberculosis (Segovia-Juarez, Ganguli and Kirschner, 2004), antibiotic-

resistant Staphylococcus aureus (Macal et al. 2014), influenza (Carpenter and Sattenspiel, 2009; Khalil et al. 2012; Jacintho et al. 2010), dengue fever (Wang et al. 2010), mosquito-borne disease (Mniszewski et al. 2014) and more recently, COVID-19 (Aleta et al. 2020; Chang et al. 2020; D'Orazio, Bernardini and Quagliarini, 2020; Morgunov et al. 2020; Nguyen & Prokopenko. 2022; Chang et al. 2023).

It is not new to model the impact of human behaviour on disease outcomes – infectious disease in particular. Differential SIR/SIS/SEIR equation models, such as those developed by Epstein et al. (2008) and Perra et al. (2011), are key examples of such work. Building on this work, the structure of ABMs makes them a good candidate for modelling human behaviour, by virtue of their ability to unpack the complex nature of human interactions with others and the environment (Macal and North, 2005; Barrett et al. 2008; El-Sayed et al. 2012). In this thesis, multiple ABMs are designed and deployed to evaluate the individual level impacts of three possible behavioural states, adapted from the definition of 'fear' detailed in Steimer (2002):

- Fearful: A motivational state aroused by specific stimuli that give rise to extremely defensive behaviour or escape.
- Careful: A motivational state aroused by specific stimuli that give rise to moderately defensive behaviour or escape.
- Complacent: A non-motivational state that gives rise to no change in behaviour.

While the above definitions have been developed from a biological perspective, they have been applied in this thesis as a method of measuring emotional response within a mathematical framework. Fearful behaviours have been previously described as a "highly contagious phenomenon" (Espinola et al. 2016; Shultz et al. 2016) often expressed in the form of fear-related behaviours (FRBs) that either work to help or hinder the spread of disease (Schultz et al. 2016). However, as evident throughout the COVID-19 pandemic, there are portions of the population that are not specifically fearful of COVID-19, rather careful enough to ensure they make rational decisions about their daily lives in light of their perceived opportunity costs. On the other end of the spectrum are complacent individuals, who instead place higher value on being able to continue ordinary daily living, regardless of the perceived or actual threat of disease.

There are a number of advantages to incorporating human behaviour in disease spread models. A conscious understanding of human behaviour impacts our ability to effectively develop

policies promoting engagement in preventative behaviours (Pakpour and Griffiths, 2020) including what policies would work best, and how readily the policies will be adopted by the population. For example, studies have found that fear of COVID-19 was the only significant predictor of increased social distancing and hand hygiene (Harper et al. 2020). By extension, allowing for differentiation between endogenous (voluntary – induced by fearfulness or carefulness) and exogenous (forced – via government or institutional restrictions or mandates) behaviours, it is possible to derive a proxy indicator for what is best referred to as a "pandemic pulse". In this thesis, econometric tools are deployed to estimate the pandemic pulse, representative of an Australian COVID-19 Early Warning System (EWS). This EWS, while deployed on COVID-19 data, could be readily adapted to other novel or endemic infectious diseases, should appropriate data be available.

Following previous research detailing the role of perception severity and susceptibility in motivation to act (Witte et al. 1998), this thesis details that individuals endogenously make behavioural adjustments to the threat of disease regardless of whether the threat is perceived (i.e. not infected yet cognisant of potential future infection) or realised (i.e. infected). The consequences of these behavioural adjustments are assumed to manifest themselves in the form of these FRBs or other behaviour-driven responses, flowing on to reduced activity and a slower spread of infection (Perra, et al. 2011), all else being equal.

Using publicly available data, this thesis demonstrates the importance of incorporating human behaviour into infectious disease modelling using an ABM framework. The significance of this modelling lies within the application of the EWS framework on top of the ABM, allowing government and policy makers to effectively design pandemic preparedness and mitigation strategies ahead of time. By targeting forward-looking policy response to population-level behavioural indicators, the ability for countries such as Australia to endure future novel infectious disease outbreaks is improved.

There are four primary categories of critical research in this field. First, this thesis explores the evolution of COVID-19 in Australia to date, highlighting the changing nature of infections over the past three years. Second, methods and approaches to modelling infectious disease outbreaks are discussed, underscoring the influential analysis of concurrent transmission of disease and fear, detailed in Epstein et al. (2008), involving a combination of agent-based Susceptible – Infected – Recovered models with differential equations in a hybrid format. Third, early warning system signals, as a tool for pandemic preparedness, demonstrate common

practice methods such as mean and variance signal analysis (Brett et al. 2020), and wastewater analysis (Bibby et al. 2021), as tools to identify the potential onset of future COVID-19 cases. Finally, this review turns to a focus on the role of government and institutional policy responses implemented in response to COVID-19, and the potential for these policies to be enhanced by data-driven early warning systems.

This thesis takes these four primary categories of critical research to develop a three-tiered framework for managing and monitoring novel and endemic infectious diseases (Figure 1.1):

Pillar one: Agent-based microsimulation models with individual behaviours as a key driver of disease transmission. Building on the importance of including behavioural factors in disease spread models, this thesis demonstrates a mathematical framework to measure the effect of human behaviour propagation on disease spread using ABM techniques under two different approaches. The Australian National University National Computing Infrastructure (Gadi) environment was used to develop these microsimulation models.

A) **Survey-driven approach (essay one, Chapter 3).** Given the wealth of data that has become available over the course of the COVID-19 pandemic, this approach leverages detailed survey data to proxy human behaviour. This framework is applied to observed COVID-19 case counts in Victoria, Australia, and demonstrates the way in which behavioural factors can track real-life observed COVID-19 case counts. Using publicly available data, this essay describes a stochastic ABM based on a Susceptible – Exposed – Infected – Recovered – Susceptible (SEIRS) framework, with time-varying transmission radii and movement patterns informed by YouGov COVID-19 Behaviour Tracker (Jones, 2020) proxies for fearful, careful and complacent population compartments. These compartments are probabilistically matched to the representative Australian population to inform likelihood of disease transmission and associated outcomes.

B) **Input-driven approach (essay two, Chapter 4).** Detailed survey data is usually not readily available or in an accessible format for inclusion in modelling. This approach is applied to observed clinical influenza notifications in Canberra, Australia, with the intention of demonstrating the way in which fear-related behaviours can be used to track observed case counts. Using publicly available data, this essay describes a stochastic ABM based on a Susceptible – Infected – Recovered (SIR) framework, with

input assumptions for the transmission of fearful behavioural status between agents. Due to data availabilities, scalars for environmental disease transmission or community-based fear transmission were introduced into this approach.

Pillar two: Data-driven EWS to monitor infectious disease cases and potential waves (essay three, Chapter 5). A simple econometric approach is used to signal future waves of COVID-19, developed at the National and State and Territory scales to support the case for early intervention that encourages personal protective behaviours such as social distancing and facemask wearing. This EWS, referred to as a pandemic pulse, forms a self-stabilising theory of COVID-19 whereby short-term patterns revolve around long term oscillations in the endemic prevalence of disease.

Pillar three: Embedding a pandemic pulse into preparedness capability (essay four, Chapter 6). This thesis deploys a number of scenarios within the aforementioned ABM framework to model COVID-19 case outcomes under various early intervention strategies informed by the pandemic pulse EWS.

Fundamentally, this thesis builds on existing literature by addressing the following research questions:

(a) *What are the main factors that influenced the spread of disease throughout the COVID-19 pandemic, and how can a better understanding of these factors in models be used to inform pandemic planning and preparedness?*
(b) *Novel and endemic infectious diseases continue to pose a threat to the population with ongoing waves of infection. How successful are simple econometric models of case velocity in establishing a fit-for-purpose early warning system to support policy decision making?*
(c) *How can an integrated model of human behavioural states and early wave detection be deployed in practice to reduce harm and mortality incurred from the spread of disease?*

The timeline of the research pertaining to COVID-19 is illustrated in Figure 1.2. New daily cases of COVID-19 in Victoria are represented by the black line, with the corresponding time periods of analysis for each essay highlighted in the coloured bars below the chart. Due to the longer spanning time horizons, the full capture of essay three (January 2020 to September 2022) is not illustrated in this figure. In addition, essay two (covering the 2017 and 2019 influenza seasons) has not been included in this visualisation.

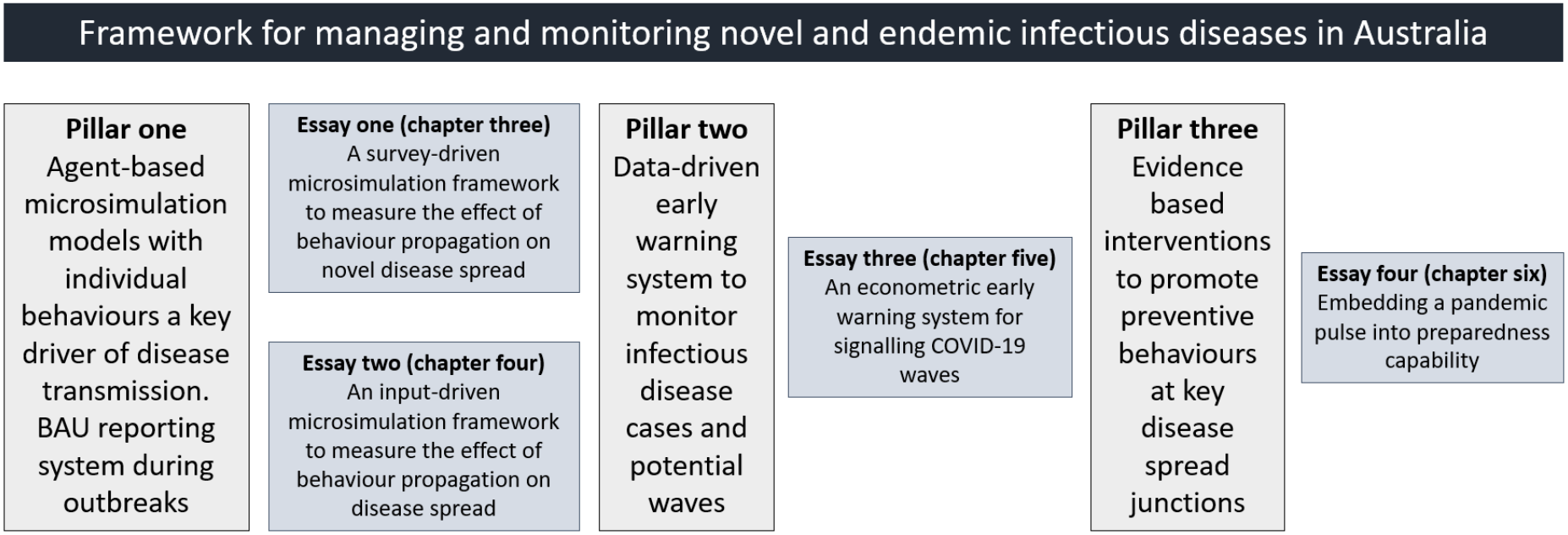


Figure 1.1 Framework for managing and monitoring novel and endemic infectious diseases in Australia

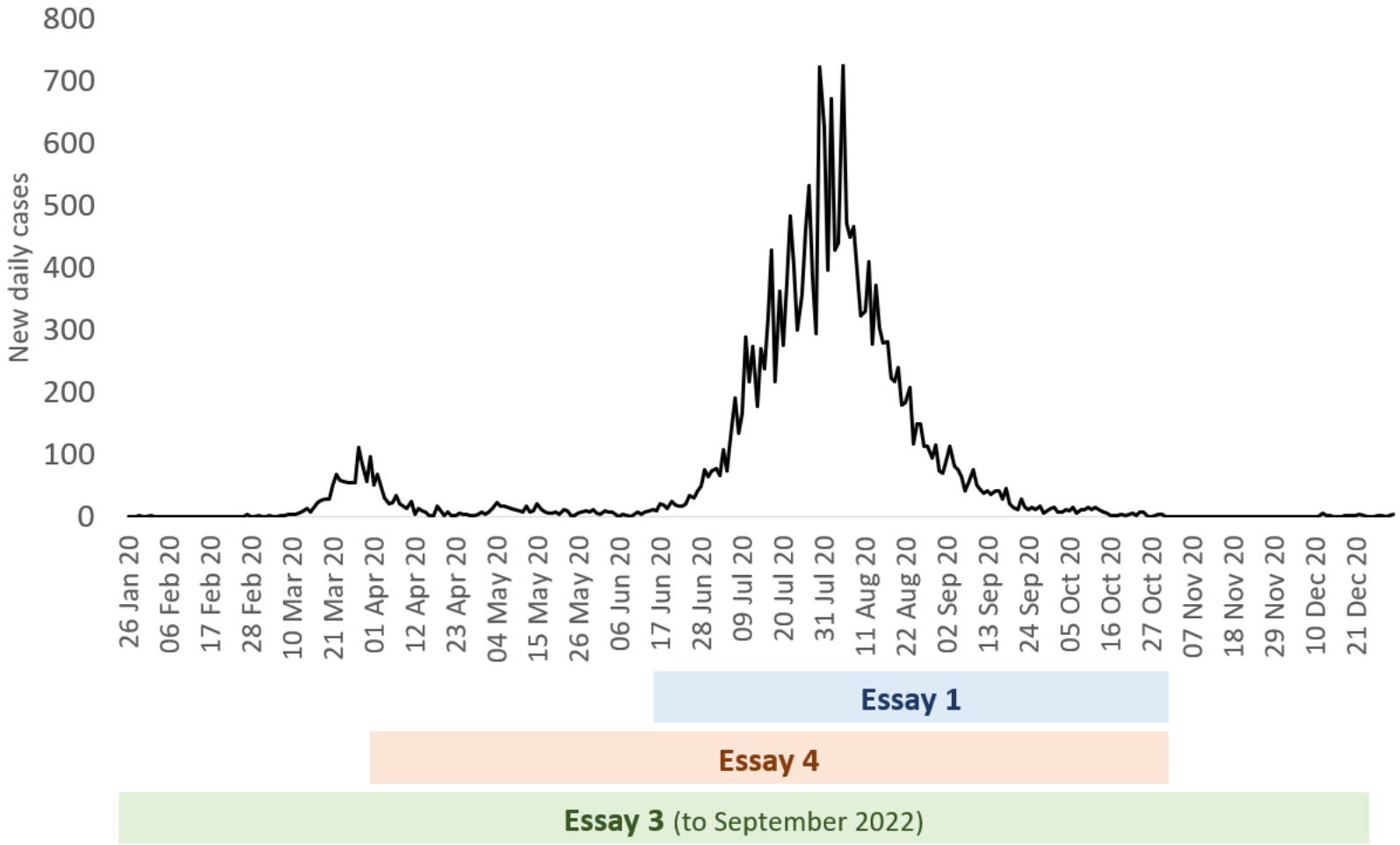


Figure 1.2 Timeline of analysis (essays) against Victorian daily new COVID-19 cases

There are a number of data limitations that pertain to this research. First, the introduction of Rapid Antigen Tests (RATs) from November 2021 indicates the potential for COVID-19 case counts to be under-reported, compared with true case counts. In addition, due to lack of reporting, explicit COVID-19 variant trends have not been included in this analysis. Finally, from September 2022 up to the point of writing, the States and Territories moved to a weekly reporting system, thus reducing the integrity of the time series data used to inform the EWS developed in this thesis.

The remainder of this thesis is structured as follows. Chapter 2 provides a brief literature review of infectious disease outbreaks in Australia, with a focus on COVID-19. This is followed by a discussion of the approaches to both modelling infectious diseases and the role of models in signalling future waves of infection to inform pandemic preparedness and response. Chapter 3 (essay one) introduces the survey-driven ABM for modelling COVID-19 disease spread in Victoria, Australia. Chapter 4 (essay two) introduces the input-driven ABM for modelling clinical influenza notifications in Canberra, Australia. Chapter 5 (essay three) details the simple framework for signalling future waves of COVID-19 in Australia using case volatility metrics to form an EWS representative of a pandemic pulse. Chapter 6 (essay four) brings together Chapters 3 and 5 to model a series of scenarios relating to COVID-19 case outcomes should early warning systems had been activated in Victoria, Australia. Finally, Chapter 7 provides a summary of the key findings and makes concluding remarks about the main themes covered. This thesis includes an appendix which reports additional empirical results.

2. Literature review

This chapter provides a targeted overview of aspects of contemporary infectious disease modelling in the Australian context. Given the wealth of publications in the infectious disease space, this review focuses on agent-based modelling of communicable diseases, specifically COVID-19 and influenza.

This review centres on COVID-19 trends from January 2020 to December 2022. At the time of writing, COVID-19 is in a stage of transition, where social attitudes to the disease are weakening despite consistent waves of infection. Despite the substantial number of mechanisms by which pandemics can be modelled using mathematical frameworks, agent-based models (ABMs) are highly suitable for understanding disease dynamics, despite their computational limitations. By accounting for individual level behaviours in agent-based models, policy interventions such as facemask recommendations can be modelled to assist in pandemic preparedness and planning. To assist with this preparedness and planning, there are a number of mechanisms by which an Early Warning System (EWS) may be deployed, which are discussed in this review. As mandated restrictions become increasingly unpalatable, both politically and socially, early warning systems built on available data become increasingly important to ensure that policy can be designed ahead of time to mitigate the future spread of disease and minimise mortality and morbidity.

In general, the severity of an infectious disease outbreak is closely linked to the behaviours and decisions made by individuals. Section 2.1 provides an overview of the evolution of the COVID-19 pandemic in Australia. The review highlights the role of human behaviour from peak crisis points through to endemic spread of the disease. Section 2.2 reviews the literature that illustrates practical methods for modelling infectious disease outbreaks, highlighting the importance and role for including behavioural parameters in mainstream models. Section 2.3 summarises the key literature on early warning systems deployed in a COVID-19 or similar context and introduces the concept of econometric modelling – commonly used in financial analysis – in developing an early warning system for pandemic preparedness based on volatility metrics. Finally, section 2.4 concludes the literature review by assessing the importance of such modelling and pandemic preparedness considerations on policy decision making.

2.1 Evolution of the COVID-19 pandemic in Australia

The novel coronavirus (COVID-19) is an infectious respiratory disease first identified in 2019 by health officials in China. Since identification of the disease, there have been hundreds of millions of infections and millions of deaths and morbidity globally (World Health Organization, 2023). From 2020-2022, COVID-19 spread rapidly in Australia, with over 10 million reported cases at the time of writing (CovidLive, 2023), approximately 40% of the Australian population.[1]

A key characteristic of COVID-19 is the high level of transmissibility and the ability to mutate into new variants that impact patient risk and fatality. Due to this, substantial efforts were put in place globally to attempt to curb the spread of the disease and work towards "zero COVID" – i.e. zero COVID-19 cases in country. Government-mandated policies primarily targeted social engagement and population mobility. In addition, a number of mandates were enforced with respect to vaccination and the use of personal protective equipment (PPE) such as face masks. The key dates for Federal-level government restrictions include:

- **18 March 2020 (7-day rolling average COVID cases: 79, deaths: 0)[2]**
    - A ban on non-essential indoor gatherings of 100 or more individuals;
    - A ban on outdoor gatherings of 500 or more individuals;
    - Essential travel only;
    - Limit of two visitors to aged care facilities each day;
    - 1.5 metres social distancing; and
    - Engagement of 20,000 student nurses to help respond to the pandemic.
- **20 March 2020 (7-day rolling average COVID cases: 79, deaths: 0)**
    - Closure of international borders to any travellers that are not an Australian resident, citizen or family member of a resident or citizen.
- **23 March 2020 (7-day rolling average COVID cases: 186, deaths: 1)**
    - Closure of:
        - Pubs and registered clubs
        - Gyms and other indoor sporting venues

[1] Noting that this proportion is likely to be lower given the lack of reporting on COVID-19 re-infection in the population and suspected under-reporting of positive Rapid Antigen Test (RAT) test results.
[2] Aforementioned CovidLive data used to inform these metrics.

- Cinemas, entertainment venues, casinos and nightclubs
- Restaurants and cafes (except for takeaway and delivery)

Restrictive decision making and announcements then transitioned to State and Territory specific management plans. This resulted in Western Australia being closed to the rest of Australia until 3 March 2022, and Victoria enduring six lockdowns.

Figures 2.1 and 2.2 illustrate the cumulative trend in COVID-19 cases and deaths in Australia (CovidLive, 2023) by calendar year. Looking at these charts, the clear-cut waves experienced can be identified. However, what these charts show in later years (see 2022 panel) is the changing nature of COVID-19 spread from short, sharp waves (see 2020 panel) to slower, more sustained waves akin to endemic disease.

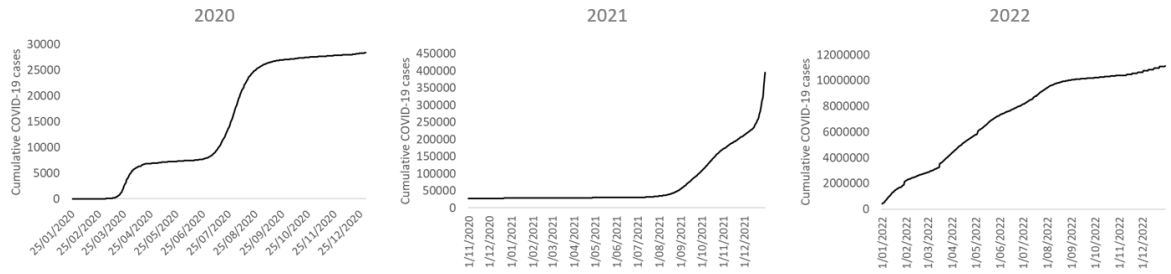


Figure 2.1. CASES: Cumulative COVID-19 cases in Australia, by calendar year

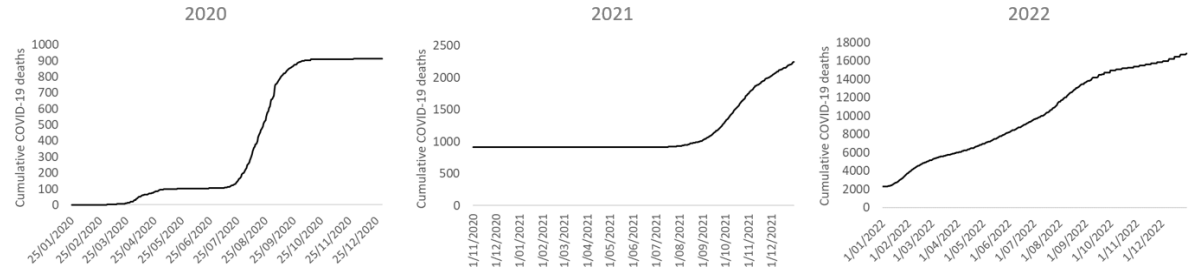


Figure 2.2. DEATHS: Cumulative COVID-19 deaths in Australia, by calendar year

2.2 Methods and approaches to modelling pandemics

Significant effort has been put in place to design mitigation strategies to curb the spread of infectious diseases. Behind these strategies often lies a mathematical model of varying complexity, used to model and understand disease spread dynamics under different scenarios.

Broadly, there are four main approaches for modelling the spread of infectious disease: ordinary differential equation (ODE) models; time series models; meta-population models; and Agent-Based Models (ABMs).

ODE models are widely used to model infectious diseases, due to their ability to model the compartmental nature of disease progression and recovery in a straightforward and computationally efficient manner. Used to track the movement of individuals through compartments, the most common ODE representations of infectious diseases include Susceptible – Infected – Recovered (SIR), Susceptible – Exposed – Infected – Recovered (SEIR) or some other adaptation of the above frameworks.

Many publications present ODEs as a practical modelling solution to the ongoing COVID-19 pandemic, including Bai (2020), Kohler-Rieper et al. (2020), and Ahmed et al. (2021). For those looking to undertake their own modelling, the common R package EpiModel (Jenness et al. 2018) provides a pre-built ODE framework for modelling the spread of infectious disease.

Despite ODEs being easy to use and efficient methods of modelling, a trade-off exists between output accuracy and the number of limiting assumptions applied to the models. For example, applying dynamic transmission rates to a set SEIR structure involves a substantial increase in the number of equations in the system, requiring some judicious assumption regarding when a transmission rate is changed and for what proportion of each compartment. A good example of behavioural adjustment in an ODE framework was developed by Del Valle (2013), where the population was split into two categories, determined by the prevalence of disease: those who do not adjust their behaviour, and those that do adjust their behaviour. In this case, the addition of this equation allowed for the effects of behaviours to be modelled within the overarching ODE framework. However, many ODEs do not include such equations, exposing model outputs to the potential bias incurred from the assumption of a uniform probability for parameters such as individual level interactions and social behaviours (Zacherson et al. 2022). In addition, because the time lag between the time at which individuals become sick is often 1 – 2 days before they are observed as confirmed cases, studies such as Menéndez (2020) have resorted to using delayed differential equations to account for this time lag.

Time series models of case counts have, likewise, become favoured in recent years. Leveraging recent advances in data science, models ranging from neural networks to state-space exponential models, and even basic regression methodology such as ARIMA, have been

deployed to predict future COVID-19 related cases, hospital and mortality outcomes (Bevenuto et al. 2020; Hernandez-Matamoros et al. 2020; Melin et al. 2020; Wang et al. 2020; Safi and Sanusi, 2021). These models are typically very quick to run, relying on historical data to infer statistical trends.

While useful for predicting short-term changes to disease trends, for example predicting the number of COVID-19 cases for the next month, these models rely on either (a) substantial historic information to gather trends (such as seasonal or time-dependent trends), or (b) threshold number of active cases to project future cases *from*. In the early days of the pandemic, satisfying (a) was impossible due to a lack of historic data necessary to inform the projection. Meanwhile, during pandemic "lulls" (b) was near impossible to satisfy for locked-down parts of the world with few cases of COVID-19, for attempting to determine the optimal next best step for pandemic preparedness and planning. These models are also not convenient for modelling on smaller geographic regions, meaning that their predominant use is in monitoring the spread of disease at larger geographic levels (i.e. National).

Meta-population models sit between ODEs and ABMs. Zacherson et al. (2022) define a meta-population model by constructing a compartmental model of disease transmission that is adjusted to account for the dynamic population structures provided in ABMs, while retaining some aggregated features. These meta-population models capture some detailed aspects of the population while grouping other aspects through aggregation and equation coupling. In this specific instance, the authors find that the meta-population model can produce reasonably close results to a similarly specified ABM when additional parameter tuning is conducted, however in the absence of this tuning the meta-population model consistently over-estimates peak prevalence of disease and under-estimates epidemic duration.

Further, a prior study conducted by Ajelli et al. (2010) compared model outputs from a stochastic meta-population and ABM in the context of a hypothetical scenario based in Italy. The authors likewise found that meta-population models over-estimate peak prevalence, however demonstrated overall alignment in model outputs with the ABM results. Despite this alignment, it was found that meta-population models yield significantly less information compared to ABMs.

By virtue of this design, meta-population models have a significantly lower computational burden. However, generating realistic outputs from these models remains difficult due to the

inaccuracies derived from a lack of individual-level mixing, which is further exaggerated when evaluating interactions across the spatial plane.

ABMs have been widely used to model the spread of various infectious and non-infectious diseases, however to a much lesser extent than ODEs and time series applications. Largely avoided due to the sheer size and computational power required to run the models, it must be noted that ABMs are a better candidate for modelling the spread of infectious disease due to their ability to unpack the complex nature of human interaction with others and the environment (Macal and North, 2005; Barrett et al. 2008; El-Sayed et al. 2012) in addition to being able to incorporate the natural history of infectious diseases (Chang et al. 2022).

A recent ABM published by Nguyen and Prokopenko (2022) incorporates behavioural factors in a different (yet relevant) manner to this thesis. The authors generate willingness to pay thresholds per Disability Adjusted Life Year (DALY) based on a simulated fraction of individuals who comply with social distancing over time. This allows for a robust assessment of social distancing intervention sustainability, balanced against economic costs and health benefits. Where this thesis differs is in the treatment of non-compliant (referred to in this thesis as "complacent") agents. In this thesis, complacent agents individually determine their movement distance or activity based on their assigned behavioural status each day, thus informing overall disease spread outcomes. Importantly, Nguyen and Prokopenko (2022) note that future research should include behavioural factors such as social media messaging and personal perceptions of risk. This is considered in detail in Chapter 4 (essay two) with the construction of spatial scalars for "fear hotspots".

Other dominant COVID-19 ABMs include the open-source Covasim model (Kerr et al. 2021) which covers the interactions between demographics, social networks, school, workplace and community on disease spread. COVID-ABS (Silva et al. 2020) is a similar model, predominantly used for modelling the impact of intervention on health and economic outcomes. ABMs have proven to be highly effective at modelling individual-level decisions in choosing to adhere (or not adhere) to social distancing or mask wearing policies, as demonstrated in papers such as in Chang et al. (2020) and Thompson et al. (2020).

A well-validated ABM developed by Nguyen et al. (2023) has recently been used to simulate the effects of various policies, variants and population demographics on disease spread outcomes. This study evaluated inequalities across the Statistical Area level 2 regions of Australia by constructing an infection incidence Lorenz curve. While confirming that

population growth does increase pandemic peaks, the burden of such peaks was found to be higher in a smaller fraction of regions as compared to the whole, with the exception of the Omicron variant which found all regions sitting on the line of equality due to the highly transmissible nature of disease, and inability of intervention policies to slow the spread of disease. Findings from this research conducted by Nguyen et al. (2023) demonstrates the importance of using ABMs to simulate the spread of disease as it is the only modelling framework which enables complex, nonlinear interactions to be evaluated in a meaningful way.

ABMs have also become more commonly used for simulating behaviour in other fields. Novel research embedding a Belief – Desire – Intention (BDI) framework enables modelling of individual goals within the context of their environment (Singh et al., 2016), including extensions to allow prior learning to inform choices (Singh et al., 2020). Applications of this BDI framework range from improving planning for bushfire (Singh and Padgham, 2015) and flood response (Padgham et al., 2014) in the Australian context.

Notwithstanding these benefits, there are two primary limitations associated with the use of ABMs in modelling infectious disease. First, calibration is a concern as a substantial number of assumptions need to be made in order to estimate parameters crucial to the model's development. In the absence of detailed unit record (person level) data, parameter identification becomes problematic, as highlighted in Bobashev et al. (2007), Farmer and Foley (2009), Venkatramanan et al. (2017) and Zacherson et al. (2022). In addition, these models are calibrated by running simulations that are validated against real-life outcomes and as such can be unreliable when an inadequate amount of 'fitting' data exist to tune the parameters. Secondly, a significant amount of computational power is required to run ABMs, even those which have invested in efficient parallel algorithms, thus imposing a costly computational and parametric burden (Bobashev et al. 2007). The use of ABMs requires a fine balance between the number of agents and parameters to ensure that model parsimony is appropriately offset by mechanistic rigour and accurate outputs (El-Sayed et al. 2012).

To conclude, Table 2.1 provides a high-level summary of the benefits and limitations of each mainstream modelling approach for infectious disease spread.

Table 2.1. Comparison of modelling approaches

ODE	Time Series	Meta-population	ABM
• Medium accuracy • Fast computations • Low ability to represent interactions in fine detail	• High accuracy • Fast computations • Low ability to represent interactions in fine detail	• Medium accuracy • Fast computations • Medium ability to represent interactions in fine detail	• High accuracy • Slow computations • High ability to represent interactions in fine detail

2.2.1 Importance of external factors in building disease spread models

To better understand and project future waves of infectious disease, it is important to understand the key factors driving increased cases. A review of the literature suggests that there are three primary drivers of increased spread or severity of disease: variant emergence, individual behaviours, and seasonality.

Variant emergence

At the time of writing there are five COVID-19 variants of concern, all of which are in active circulation (World Health Organization, 2023a). As new variants emerge, it has been observed that new combinations of transmissibility, lethality or severity materialise. For example, the Omicron variant (B.1.1.529) is reportedly less severe than the Delta variant with regard to hospitalisation and mortality, despite having substantially higher transmissibility (Nyberg et al. 2022). As COVID-19 continues to spread, it is anticipated that new variants of the disease will emerge, likely with differing characteristics and severity.

In addition, influenza-type disease features multiple different strains, which can change every year (Department of Health and Aged Care, 2022a). The World Health Organization (2023b) releases their recommendation for the best flu vaccine combination each year for the upcoming flu season, which in the Australian context must then be approved by the Therapeutic Goods Association.

Individual behaviours

With many infectious diseases, infection begins and can be spread prior to the onset of symptoms. In the absence of rigorous contact tracing, which is subject to lags and can only act on chains of identified transmission, this characteristic means that the only lines of defence between individuals infecting one another are by vaccination and personal protective behaviours, such as wearing a mask and social distancing.

Individual decisions to engage in ad-hoc or routine events have been analysed with regard to their impact on the spread of infectious disease. Research focusing on influenza has shown that mass gatherings (for example attending school, sporting events or music festivals) have the potential to create a new seed of infection in a region (Gutierrez et al. 2009) and potentially increase the risk of disease transmission, depending on the level of crowd density (Ishola and Phin, 2011).

Shi et al. (2010) developed a computational model to investigate varying levels of social mixing and contact patterns during mass gatherings and holiday travel, with the intention of determining whether such activities increase the spread of influenza. The authors evaluated changes to parameters such as the proportion of the population in attendance, the initial reproduction number and the length of the gathering or travel. The findings suggest positive effects on disease transmission for both mass gatherings and holiday travel, when such events occur close to the epidemic peak. This positive effect was reported to decline substantially when the time between the peak and the events was increased.

Further, there is some evidence to suggest that the amount of air ventilation at mass gatherings plays a role in the risk of infectious disease transmission. Based on a review undertaken by Ishola and Phin (2011), it was concluded that open-air sporting events are not related to increased spread of infectious disease. For example, based on the control measures implemented at the Asian Youth Games hosted in Singapore in 2009, no additional outbreaks were detected within the athlete village or any other venues, despite the event being held during an ongoing influenza pandemic (Lim et al. 2009).

Individual protective behaviours and vaccination must be viewed as two interdependent parts. While the benefits of vaccination are well known in the context of COVID-19 (for example Pfizer Phase II/III trial reporting 2-dose vaccine efficacy in reducing symptomatic flow on cases of COVID-19 is reported to be 95% (Department of Health and Aged Care, 2022)), these benefits must be weighed against the potential perception of security across the population

when higher levels of vaccination are achieved, particularly as vaccine efficacy wanes over longer time periods and fewer individuals seek out additional booster vaccinations (Catching et al. 2021).

One crucial component of individual protective behaviours is the use of facemasks. Several studies have sought to measure the efficacy of mask wearing, with studies such as Alihsan et al. (2022) estimating a 45% reduction in the probability of COVID-19 infection compared to no mask wearing, based on a sample of 13 studies. Facemask wearing is complex, as the act of putting the mask on oneself increases the act of facial touching, which has been found to theoretically increase the probability of infection (Stutt et al. 2020). Globally, mask shortages led to individuals crafting their own fabric masks, and further research has demonstrated conflicting outcomes as to whether these non-medical masks are indeed effective at limiting the spread of disease (Santarsiero et al. 2021). Given that 100% compliance with mask wearing across the population is unrealistic, model results suggest that the optimal mask wearing scenario is one by which the risk of an already infected individual is lowered, in lieu of reducing the risk of a susceptible individual from becoming infected (Rao et al. 2021). Across the majority of studies, the importance of coupled face mask wearing and social distancing measures was highlighted as best practice for effective individual protective behaviours.

A cross-sectional survey of mask wearing in five cities across Australia, the United Kingdom and the United States (MacIntyre et al. 2021) found that the predominant predictors of mask wearing included the existence of mandates, being younger than 50 years of age, educational attainment, and being aware of COVID-19 from either knowing someone infected or having symptoms. Confidence in government (in this instance, State/Territory government) was found to be a positive predictor of mask wearing (OR = 1.33 [CI 1.11 – 1.60]), again demonstrating the importance of trust in decision makers to enable positive behavioural change.

The role of confidence and trust in governments was echoed in Seale et al. (2020) which found this parameter to be a critical component of individual protective behaviour adoption. In this study, the authors found that 84.9% of respondents (n = 1,420) had performed at least one of the recommended hygiene-related behaviours over the last month, and 93.4% had performed at least one avoidance-related behaviour. The authors found a statistically significant relationship between trust in government and adoption of these avoidance behaviours.

For long-term protection from COVID-19, individual protective behaviours must be considered as important as vaccination. As discussed in Ferguson (2007), it is critical to model these individual behaviours and understand how they change over time. Specifically, Ferguson (2007) notes that individuals make decisions regarding the perceived level of risk, informed by current case counts and mortality statistics. By adjusting traditional disease spread models with behavioural patterns, the ability of governments and decision makers to predict and control epidemics is substantially improved. Following this, Catching et al. (2021) present an agent-based model measuring the efficacy of individual protective behaviour on disease spread outcomes, finding that face masks provide optimal protection even if adherence to social distancing is poor. However, as the world becomes more accepting of surging COVID-19 cases, there is concern that individual willingness to wear a mask in the absence of mandates will continue to decline, despite the evidence indicating that masks are an effective strategy to curb the spread of disease (Catching et al. 2020).

Meacci and Primicerio (2021) build on the concept of willingness to engage in individual protective behaviours by considering the effects of pandemic fatigue on the spread of COVID-19 in Italy in 2020. “Pandemic fatigue” is referred to as a state of de-motivation to follow recommended behaviours to prevent the spread of disease to themselves and others. Using an ODE framework, the model establishes a framework for measuring “lowering of the guard and negligence in observing precautionary practices”. Results from this model highlight that pandemic fatigue can be mitigated by fear of the disease, which the authors found to be driven by the proportion of deaths associated with the outbreak. The authors subsequently noted that the incorporation of environmental (spatial plane) effects on behavioural decision making would be a useful addition to understanding pandemic fatigue. In accordance with this, the models described in this thesis are spatially explicit to ensure that the effects of the surrounding environment on human behaviour are accounted for.

The importance of modelling a scale of human behaviour, specifically non-conformist behaviour, was highlighted in Javarone (2014). In this study, the author develops a behaviourally adjusted ABM for modelling opinion dynamics. At each modelled time step, individuals are assigned as either conformists or non-conformists with regard to their immediate social circle and the broader global population. The results from this analysis found that in simple networks, non-conformist behaviours play a central role in opinion dynamics, with conformists having a stabilising effect. The reverse was found for more complex social networks.

To measure the broad scale of human behaviour, this thesis covers the three behavioural states of fearfulness, carefulness and complacency. Fear has previously been defined as "a motivational state aroused by specific stimuli that give rise to defensive behaviour or escape" (Steimer, 2002) and is reported to be highly contagious between individuals (Espinola et al. 2016; Shultz et al. 2016). Harper et al. (2020) suggest that fear of COVID-19 was the only significant predictor of increased social distancing and hand hygiene. The logic is clear: individuals afraid of contracting disease are less likely to leave the house, to not wear a mask, and to engage with others at a close distance. Referred to as "fear related behaviours", these defensive reactions to the ongoing threat of disease either help or hinder the ongoing spread (Shultz et al. 2016). At the other end of the spectrum, complacency has been formally defined as "self-satisfaction, especially when accompanied by unawareness of actual dangers or deficiencies" (Merriam Webster, 2023).

While individuals express a degree of complacency as part of everyday life, the application to infectious disease spread is that individuals may choose to disengage from individual protective behaviours due to their own lack of knowledge, unawareness, or consumption of misinformation about the actual harms of their behaviour. For others, complacency captures individuals who have normalised the concept of infection and the implications of their potential spread of the disease on to others around them. Regardless of the reason, these individuals are hypothesised to be disengaged with protective behaviours and returning to normal activity of daily living. Between fearful and complacent lies careful behaviours. Defined as being "marked by wary, caution or prudence" (Merriam Webster 2023a) individuals encapsulated in this category strike a balance between being afraid of COVID-19 and going about their activities of daily living in a considered manner. These behaviours are discussed further in Chapter 3.

Analysis conducted by Del Valle et al. (2013) explicitly modelled the effect of fear-based isolation within an ABM structure, in addition to school closures and vaccine developments. In this model, individuals who choose to isolate due to fear do so for 7 to 14 days at a time. Once a threshold proportion of the population that is symptomatic is reached, the proportion of the population isolating increases linearly. Once the epidemic peak has been reached, the proportion of the population choosing to isolate then starts to decline. Highlighting the importance of informed, rational decision making by individuals, the authors found that when the self-isolating proportion of the population reaches 15%, there is a decrease in the death rate and clinical attack rate. When this proportion was increased to greater than 50%, it was found that the epidemic was split into two waves, resulting in an overall higher number of cumulative

infections compared to the baseline model. This highlights the importance of intervention timing (i.e. ensuring interventions are not relaxed before the number of cases have fallen below a specified threshold) and design (ensuring appropriate, targeted interventions are implemented) given the potential to trigger multiple waves of infection.

Seasonality

Emerging studies have established the potential for COVID-19 to be influenced by the seasons (Nichols et al. 2021), similar to many other respiratory viruses such as influenza, which likewise escalate in case numbers over the winter months (Brett et al. 2017).

Using a SEIR model incorporating seasonal factors, Liu et al. (2021) demonstrate that seasonality increases both infectivity and mortality of COVID-19 in the colder (winter) seasons. Using simulation methodology, Liu et al. (2021) suggest that between 40 to 60% of COVID-19 cases in the winter months are associated with the weather conditions, in addition to suggesting that both infectivity and mortality are bolstered by these seasonal effects.

2.3 The role of early warning systems for pandemic preparedness

Early warning systems refer to a process or processes that can be implemented should a predefined chain of information and sensors indicate the likelihood of a future event. In the context of COVID-19 endemicity, early warning systems become particularly pertinent if the relevant governing authority wishes to prevent large waves of infection, which could be attributable to previously discussed factors such as new variants, seasonality or individual level behaviours. Similar to influenza, when authorities are alerted to a potentially severe influenza season, protections can be put in place to protect those most at risk of morbidity or mortality (such as the elderly, or immunocompromised individuals).

EWS include signals for critical slowing down (CSD), which refers to the time or likelihood of a system to return to equilibrium after a perturbation, measured by variance and autocorrelation in the series (Dakos et al. 2008). For example, CSD is used to measure a range of tipping points, including large scale climate shifts (Dakos et al. 2008). A wide range of different approaches currently exist for EWS in an infectious disease context. As the COVID-19 pandemic has revealed, global capability to signal pandemic severity is critically underdeveloped (Fearnley and Dixon, 2020). Table 2.2 summarises some recent methodologies for signalling infectious disease outbreaks.

For studies such as Bari et al. (2021) and Krieg et al. (2020), it can be argued that the reliability of a media-based or a google-search based EWS would substantially diminish over time, as the concept of COVID-19 becomes more integrated with everyday living. Second, many of the bespoke data collections developed in response to COVID-19 – including some of the data collections referenced within this thesis – have now been discontinued, and as such the utility of these early warning systems are null unless such data was recollected or substituted with an alternate data source. Without such data to inform long term behavioural trends, signals must rely on information that is on hand to formulate an appropriate EWS.

Other studies, such as Brett et al. (2020), rely on metrics such as mean and variance to understand CSD prior to disease re-emergence. These metrics are considered the optimal form of early warning systems for obvious and tangible reasons, as they enable identification of when new COVID-19 cases trend away from the average number of cases, therefore signalling a movement away from the status quo (i.e. no wave vs wave). Chapter 4 (essay three) of this thesis discusses more sophisticated methods of statistical early warning systems using daily COVID-19 data.

Finally, studies such as Bibby et al. (2021) present a practical solution to signalling the onset of disease, given that the EWS would only provide a modest four to six-day lead in period for preventative protection behaviours to be reinforced across the population. Ideally, an EWS would have a longer lead in time allowing for considered and tailored policy implementation to curb the spread of disease.

Table 2.2. EWS methodologies in the literature

Reference	Description
Brett et al. (2017)	Dynamical systems theory approach which simulated a birth-death-immigration model to develop early warning systems via various statistical techniques, including mean, variance, coefficient of variation and autocorrelation.
Brett et al. (2020)	Another dynamical systems study comparing increasingly complex disease spread models to prove the existence of CSD prior to disease re-emergence, with the best performing EWS being mean and variance over the simulated data.
Miller et al. (2017)	Seasonal analysis to detect CSD prior to disease emergence in simulations via seasonal trend removal and amplitude. Seasonal trends were observed to be a limitation to infectious disease early warning systems, and the authors note that periodic detrending and differencing did not improve prediction and, in fact, introduced false patterns in the time series.
Bari et al. (2021)	Google search queries capturing trends in the number of individuals querying the browser for key COVID-related phrases, including estimates of isolation and mobility.
Krieg et al. (2020)	Multi-modal data collation activities including demographic data, news data, web trends, media bias data and mobility reports contrasted against the number of COVID-19 cases.
Bibby et al. (2021)	Wastewater monitoring as a mechanism for COVID-19 EWS from the onset of viral shedding prior to the presentation of disease.

2.4 Data-driven policy decision making

The conglomeration of the role of human behaviour and early warning systems is necessary in order to define appropriate long term pandemic mitigation strategies. While typically presented as a blunt statistical instrument (refer to section 2.3), the role of an EWS is ideally embedded within key decision making and communication structures.

As discussed in Fearnley and Dixon (2020):

> "Early warning systems are much more than simple systems that provide a siren or warning to move. For an EWS to be effective it must be embedded in an extensive system of observation and communication that integrates different expert and policy cohorts, thresholds or tipping points, communication mediums and iconographies, for the provision of timely warnings to people with the aim of minimising loss of life and reducing the social and economic impact of disasters."

The connection between the disease, the EWS and the outcome lies within policy intervention. In the context of COVID-19 and other infectious diseases, available policy levers include (where politically or socially palatable) lockdown, facemasks, social distancing, and vaccination recommendations and/or mandates.

In this thesis, these terms are defined as:

- **Lockdown:** A mandated restriction requiring individuals to remain at home unless they are an essential worker or for emergency or essential purposes.
- **Facemask:** A facial covering that limits the transmission of airborne particles from one individual to another.
- **Social distancing:** A set distance maintained between two or more individuals to limit the number of infectious airborne particles being passed from one individual to another.
- **Vaccination:** Specifically designed vaccines to limit the spread of disease and/or reduce the severity of infection.

The EWS policy mechanism follows that upon a positive signal, levers would be implemented to reduce the future spread of the disease, thus preventing a potential wave of infection. Literature to date suggests strong evidence for the effectiveness of each policy lever, as summarised in Table 2.3.

Table 2.3. Efficacy of policy recommendations and/or mandates on disease spread

Lever	Summary
Lockdowns	Based on a study of 27 countries from different continents, Meo et al. (2020) found that 15 days post-lockdown there was a declining trend in new COVID-19 cases and the disease growth factor, however no significant decline in overall prevalence and mortality. Counterfactual analysis undertaken by Chang et al. (2022) suggests that, in the context of the Delta variant, a high proportion of the population (over 70%) would need to adhere to strict stay-at-home restrictions for there to be a meaningful decrease in new daily cases over a 45-day intervention period. Considering the proportion of the population that are essential workers, this high ratio signals that alternative behaviours are required to help mitigate the spread of disease. A cross-country study of lockdowns was undertaken by Chu et al. (2023) which evaluated impacts on Australia, Canada and the United Kingdom. In this study, the authors deployed temporal before and after change analysis using publicly available data to demonstrate that mandatory lockdowns were effective at reducing adverse outcomes across three domains: COVID-19 cases, hospitalisations, and ICU cases. The results suggest that appropriately timed lockdowns are more effective than the cumulation of the individual protective behaviours together.
Facemasks	Facemasks have been traditionally used in medical settings to prevent the spread of respiratory viral infections. A rapid review of 58 studies demonstrated substantial benefits of wearing facemasks in reducing respiratory virus transmission (Abboah-Offei et al. 2021). In terms of overall quantification, a Canadian study (Karaivanov et al. 2021) estimated a 25% reduction in the weekly number of new COVID-19 cases as a result of mask mandates. A mathematical model developed by Tracht et al. (2010) demonstrated that a 20% reduction in influenza cases could be observed under a scenario whereby 10% of the population wear N95 masks measured at 20% effectiveness of reduced susceptibility and infectivity.

Social distancing	Based on a simulation study, social distancing was found to substantially reduce infection rates when social distancing interventions were adhered to, between the ranges of 80% to 100% (Daghriri and Ozmen, 2021). Referring to a more real-life study, it was found that the impacts of social distancing on reducing COVID-19 infections vary across the ten studied countries, with some countries taking longer than others to see the benefits of social distancing on reduced COVID-19 cases (Thu et al. 2020). The authors note that this is likely to be due to the difference in the actual implementation of the social distancing measures and the number of active cases in the population. A study by Chang et al. (2020) found that higher levels of compliance with social distancing (greater than 80%) are required to control the spread of disease, compared to lower levels of social distancing (less than 80%) which was reported to be unsuccessful at generating any marginal benefit for the intervention period of 13 weeks.
Vaccination	The efficacy of COVID-19 vaccinations has been well documented by the responsible pharmaceutical companies. A summary of the efficacy of the dominant vaccinations in Australia on preventing symptomatic COVID-19 infection following Agency of Clinical Innovation (2022) is provided below: • AstraZeneca: 55.7% • Pfizer (≥ 7 days after second dose): 92% • Moderna: 89.6% These figures must be considered allowing for waning vaccine effectiveness. In the absence of widespread, continued, booster uptake, the potential for previously administered vaccines is known to diminish. A systematic review and meta-analysis conducted by Menegale et al. (2023) found that vaccine effectiveness (both infection and symptomatic disease for the Omicron variant) was less than 20% at six months post the original vaccine cycle administration. The efficacy of influenza (H1N1) vaccines is less clear. A systematic review undertaken by Osterholm et al. (2012) noted that influenza vaccines are capable of providing moderate protection against disease, however this is not consistent across all seasons of the years.

In a report published by the United Nations Office for Disaster Risk Reduction (UNDRR, 2006) the required factors for implementation of early warning systems into policy include:

- Comprehension of the most likely risks, likelihood of disasters and their consequences should they arise;
- Establishment of priority systems for allocation of resources in the event of a signal;
- Clear responsibilities by stakeholders, in the Australian context, by the Federal and State and Territory Governments;
- Strong legal frameworks;
- Effective communication strategies;
- Mechanisms to obtain resources as required.

Referred to as the basic elements for integrating early warning systems into public policy, this thesis provides the foundation for monitoring future waves of infection in the context of COVID-19.

2.5 Conclusions

This targeted literature review is limited to an assessment of ABM, intervention effectiveness and early warning systems for policy decision support in the Australian context. The nature of COVID-19 spread in Australia is evolving into a fluctuating state of disease waves, akin to that of influenza. This review has highlighted the role and importance of including behavioural parameters into disease spread models given the considerable impact human behaviour has on sustained disease transmission. While studies such as Chu et al. (2023) have demonstrated the effectiveness of timely lockdowns, the range of intervention options for reducing disease transmission remain – at least without great resistance – within the bounds of individual protective behaviours.

Gaps in the current literature exist with relation to the integration of a multi-layered behavioural status into ABMs. Leading models developed by Epstein et al. (2008) and Perra et al. (2011) apply behavioural dynamics using ordinary differential equations to model fear dynamics and the resultant impacts on disease spread. Later studies, such as Nguyen and Prokopenko (2022) built an ABM to estimate the impact of varying levels of social distancing compliance on willingness to pay thresholds per DALY. However, in this model individuals were not able to make decisions relating to their compliance status.

To address these key gaps, this thesis develops an ABM encompassing several behaviour states, interactions and probabilistic decision points for modelling the spread of disease. In the context of influenza (Chapter 4, essay two), this thesis addresses key future research suggestions to incorporate the effects of social media messaging and personal perceptions of risk (Nguyen and Prokopenko, 2022).

In addition, this targeted review has demonstrated that data-driven pandemic preparedness by way of an EWS is not mature or incorporated as business-as-usual reporting in the Australian setting. Several EWS approaches spanning dynamic systems (Brett et al. 2020), time series (Miller et al. 2017), multi-modal (Krieg et al. 2020) and wastewater analyses (Bibby et al. 2021) have been developed to generate useful signals for disease emergence in the infectious disease context. This thesis argues that a simple, easy-to-use, time-series approach to signal generation is the most appropriate tool to support on the ground decision making and intervention. To ensure that policy interventions are targeting reduced mortality and morbidity from disease, these systems must be implemented in an integrated manner to ensure the ongoing monitoring and evaluation of likely population outcomes from disease spread.

3. Essay one: A survey-driven microsimulation framework to measure the effect of behaviour propagation on the spread of COVID-19

3.1 Introduction

The role and importance of individual behavioural decisions in infectious disease spread has never been as clear as it has been throughout the COVID-19 pandemic. Understanding the spread of fear and concern amongst population groups during a crisis is vital to designing policies that balance fear related protective behaviours against adverse health and economic outcomes.

It is not new to model the impact of human behaviour on disease (particularly infectious disease) outcomes. Differential equation models, such as those developed by Epstein et al. (2008) and Perra et al. (2011), are key examples of such work. More recently, Catching et al. (2021) presented an agent-based model to measure the efficacy of individual protective behaviours, noting concern that individual willingness to wear a mask in the absence of mandates will continue to decline, despite the evidence indicating that masks are an effective strategy to curb the spread of disease. Notwithstanding, there is still a lack of evidence incorporating specific behavioural states and the impact of these states on disease spread outcomes. In this essay I argue that a more nuanced view of behavioural response – characterised by time-varying contact and movement trends – is critical for modelling epidemic trends. By assigning a dynamic behavioural parameter to individuals, referred to in this essay as 'agents', long-term behavioural trends can be incorporated into disease spread models.

In Chapter 2, human behaviour was identified to be one of the most important factors for understanding the spread of infectious disease. For modelling purposes, it is posited that agents can be either fearful, careful or complacent in response to COVID-19, with these overarching categories illustrative of the individual protective behaviours contained within. Following the commentary in Chapter 2 (section 2.1), this essay constructs a spatial Agent Based Model (ABM) to model the complex interactions between agents and their environment (Macal and North, 2005; Barrett et al. 2008; El-Sayed et al. 2012).

The purpose of this model is to estimate the individual level impacts of three possible behavioural states, adapted from the definition of fear detailed in Steimer (2002):

- Fearful: A motivational state aroused by specific stimuli that give rise to extremely defensive behaviour or escape.
- Careful: A motivational state aroused by specific stimuli that give rise to moderately defensive behaviour or escape.
- Complacent: A non-motivational state that gives rise to no change in behaviour.

Inspired by these definitions, behavioural status is modelled using time series survey data from the YouGov COVID-19 Behaviour Tracker (Jones, 2020). The data repository, available through GitHub, provides behavioural data and analysis indicating how populations across the globe responded to the COVID-19 pandemic, including going outdoors, testing, personal protective behaviours and compliance with restrictions, for 30 countries, including Australia.

While this chapter provides the first attempt at classification of behavioural states into three definitions (fearful, careful, complacent), it must be noted that individual responses to pandemic threats are much more complex than these three behavioural states. This chapter therefore serves as a step towards a model of human behaviour by establishing a compliance-behaviour stratification within the modelled populations.

Under the general framework of this thesis, this essay answers the question:

What are the main factors that influenced the spread of disease throughout the COVID-19 pandemic, and how can a better understanding of these factors in models be used to inform pandemic planning and preparedness?

The data for Australia were broken down by date of response, age, gender and state/territory of residence. As this essay is focused on measuring the impact of behaviours on disease spread transmission to improve planning and preparedness, focus has been placed on question i_12_health_6: *"[Have you] avoided going out in general?"* to which individuals were able to respond with: *"Always", "Frequently", "Sometimes", "Rarely"* or *"Not at all".* Based on these response options, this dataset was used to proxy each behavioural compartment using the following assumptions:

- fearful agents report "*Always*" avoid going out in general;
- careful agents report either "*Sometimes*" or "*Frequently*" avoid going out in general, and;

- complacent agents report either "*Rarely*" or "*Not at all*" avoiding going out in general.

Data containing individual responses to these questions were collected from the public repository for the period of April 2020 to February 2021. Due to low or no responses on some days of the pandemic, data cleaning was undertaken as discussed in section 3.2. Data analysis and simulation modelling was undertaken using R programming software run in parallel within the Australian National University National Computing Infrastructure (Gadi) environment.

The general argument of this essay is illustrated in Figure 3.1. Agents who are fearful or careful regarding COVID-19 are likely to change their behaviours, resulting in altered disease spread dynamics. Complacent agents are assumed in this model to not change their daily behaviours at all. By including these individual-level, daily decisions in an ABM framework, the accuracy of disease spread models can be improved.

Using behaviour as a determinant for person-to-person interaction and transmission variability, this essay develops and deploys a behaviour adjusted disease spread model for COVID-19 in the State of Victoria, Australia. Victoria was selected due to the high number of new daily cases in the region, allowing for robust modelling of a simulated fraction of the population. Due to the granularity of the model developed with agent interactions, State/Territory level simulations were determined to be the highest level of population aggregation useful for this style of modelling. Simpler versions of this model (with fewer nuanced effects such as behaviourally adjusted transmission rates) have been deployed at the National level for disease spread modelling, with high levels of accuracy in relation to new cases of COVID-19. However, the framework presented here highlights the importance of including such nuanced effects in agent-based disease spread models. The model presented in this essay is intended to serve as a case study for incorporating behavioural dynamics in disease spread models as a standard metric for inclusion. Focusing on the first large wave of COVID-19 infection in Victoria, some further extensions into third wave dynamics under the same modelling framework are later discussed.

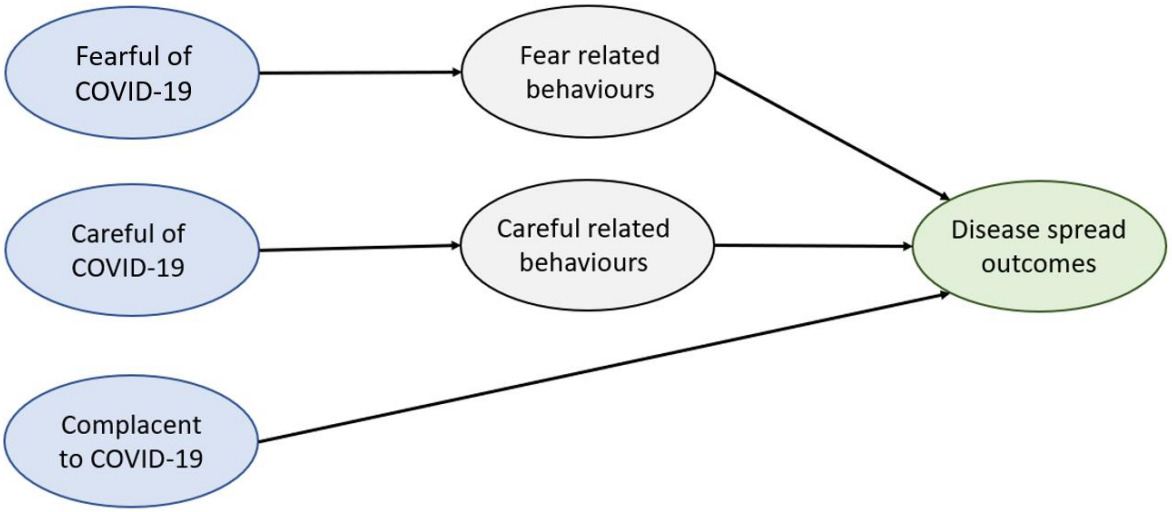


Figure 3.1. Behavioural states and disease spread outcomes

The results of this essay show that the incorporation of behaviourally adjusted transmission probabilities yields a substantially improved estimation of COVID-19 spread compared to publicly available disease spread models constructed under similar conditions. Compared to the observed number of cases, the model presented in this essay had the highest correlation (0.86) in active cases, with only a 0.002% difference in the cumulative proportion of the population infected. Evaluating the unique transmission probabilities in more detail, it was found that there is decreasing marginal utility over time for higher proportions of fearful and careful agents in the total populations. This implies that fearful and careful behaviours may not be sustainable over the long term and may be better suited to short term response rather than daily activity. This analysis also found Granger-causal relationships between new COVID-19 infections and increasing numbers of careful and fearful agents. This implies that as COVID-19 infections increase, the number of careful and fearful agents is posited to also increase, while the number of complacent agents was posited to decrease under the same Granger-causal hypothesis.

The findings of this essay build the first pillar of the framework contained within this thesis. Rapidly constructed in response to the pandemic, survey-based behavioural trackers (such as the YouGov tracker used in this essay) improve the ability to model disease spread dynamics and future trajectories. Coupled with statistical early warning systems (see Chapter 4 and 5), the ability of governments to plan and prepare for new or emerging infectious disease threats is substantially improved. This response function means that policies can be targeted towards

the most effective measures to reduce disease spread, with positive flow-on effects to disease associated morbidity and mortality.

This essay is organised as follows. Section 3.2 explains the methods behind the model and introduces the approach to derive behaviourally adjusted transmission and movement parameters. Section 3.3 presents the main results from running the ABM, followed by section 3.4 containing important implications and conclusions for health modelling and policy during pandemics.

3.2 Methods

The model developed in this essay is a spatially explicit Agent Based Model (ABM) designed to simulate the spread of COVID-19 under a SEIRS (Susceptible – Exposed – Infected – Recovered – Susceptible) framework on a population of 6,649 agents (a 0.1% sample) in the State of Victoria, Australia. As discussed in Chapter 2, due to the sheer computational power required to run large ABMs, smaller fractions of the population must be simulated upon to generate results in a timely manner. As a result of this computational burden, a 0.1% sample of the population was selected for modelling purposes. Larger population fractions (greater than 10% sample) can be simulated with this model framework, time permitting. The model contains the option for parallel processing which can be constructed across simulations, however due to the time-dependent nature of each timestep (in other words, the outcomes of today are dependent on the interactions of previous days) the efficiency of the algorithm is still contained to the length of one full simulation run.

This ABM framework allows for a number of key infectious disease spread dynamics:

- Ability to transmit disease during the asymptomatic period.
- Ability to be re-infected with the disease.
- Varying daily movement characteristics based on behavioural status.
- Varying infection probabilities based on individual behavioural characteristics.

The SEIRS model developed in this essay is overlaid with estimates of the population reported to be fearful, careful or complacent in response to COVID-19, which is probabilistically linked to the representative synthetic population. Results have been validated per observational data from July 2020 to December 2020.

3.2.1 State variables and scales

A high-level summary of the modules and processes for the model described in this essay is provided in Figure 3.2. Designed to simulate daily time steps, this model tracks the daily movements and infection likelihoods of agents within their physical environment, as a function of their behavioural status (fearful, careful, complacent). The synthetic population is constructed of three age brackets: *young* (0 – 19 years), *adult* (20 – 64 years) and *old* (65+ years). Due to the lack of available evidence at the time of model build and deployment, gender disaggregated estimates have not been presented here.

Once established, this synthetic population undertakes fixed daily movement activities – to work, school or leisure – which are statistically biased (in the model) towards city centres (business areas). Agents who are careful or fearful of COVID-19 have behaviourally adjusted movement patterns and distances, discussed later in this section. Throughout engagement within an individual's home location and daily activity location, they contact other agents within the bounds of their contact space through which disease may be transmitted. The daily activity location is truncated for careful agents who may still go outside (to get groceries, medicines, etc.) but avoid any other usual daily activities observed prior to the outbreak. For fearful agents, this daily movement is reduced to zero.

Contact between two agents occurs when they inhabit the same 3-metre radius, defined as the radius of a circle with the respective individual at the epicentre. Referred to as the contact space to which disease may be transmitted, the probability of disease transmission varies with the behavioural status of the individual, discussed later in this section. When contact is made between an infectious and non-infectious individual, probability of infection is determined by the behaviourally adjusted transmission rate β.

This essay provides validated results for the first and second major waves of COVID-19 in Victoria. Due to the status of COVID-19 in Australia at this time, agents are assumed to be highly *unlikely* to be re-infected with COVID-19, and as such exits from the infected state are comprised of either recovery ($\gamma_{disease}$) and return to the population or death via the clinical mortality rate *m*. Introduction of new positive cases from international travel was excluded due to the closure of the Australian borders during the period of this analysis. Long term patient outcomes, such as long COVID, have not been included in this model, however future research should seek to include such effects for model completeness. Out of interest, some insights are provided for the third (more severe) wave of COVID-19 whereby agents are able to be re-

infected with COVID-19 under the SEIRS framework. Due to the balanced nature of the variables and the relatively short time period of analysis, natural mortality and reproduction (births) have not been included in this analysis.

3.2.1 Parameters

The full list of input parameters used to inform the model described in this essay is provided in Table 3.1. Parameters included in this model are all publicly available, found from Australian statistical collection agencies and academic institutions for the region of Victoria, Australia. Note that due to the time at which this model was built, some parameters may appear to be outdated. These parameters would be updated prior to use in the current setting, however inputs current at the time of writing this essay have been included below.

INPUTS

Spatial module

- Set spatial boundary in line with area of target location (x, y)
- Assume everyone moves each day
- Set daily movement distance by age group (average, standard deviation)
- Cluster agents around city centre

Demographic module

- Set number of agents
- Assign age (young = 0 – 19 years, adult = 20 – 64 years, old = 65+ years)
    - Assign mortality rates for clinical infection

Interaction module

- Assign agent interaction radii and space in terms of a 3D cylinder of interaction space
- Assume probability of contact with overlapping radii

Disease module

- Set initial prevalence of disease
- Set contact probability
- Set infection probability
- Set recovery rate
- Set maximum days of infection

Behaviour module

- Set behavioural status (fearful, careful, complacent)

Monitoring

- Measuring days infected, days fearful, days careful, days complacent, outbreak days, home location and daily activity location (e.g. occupation)

CALCULATIONS

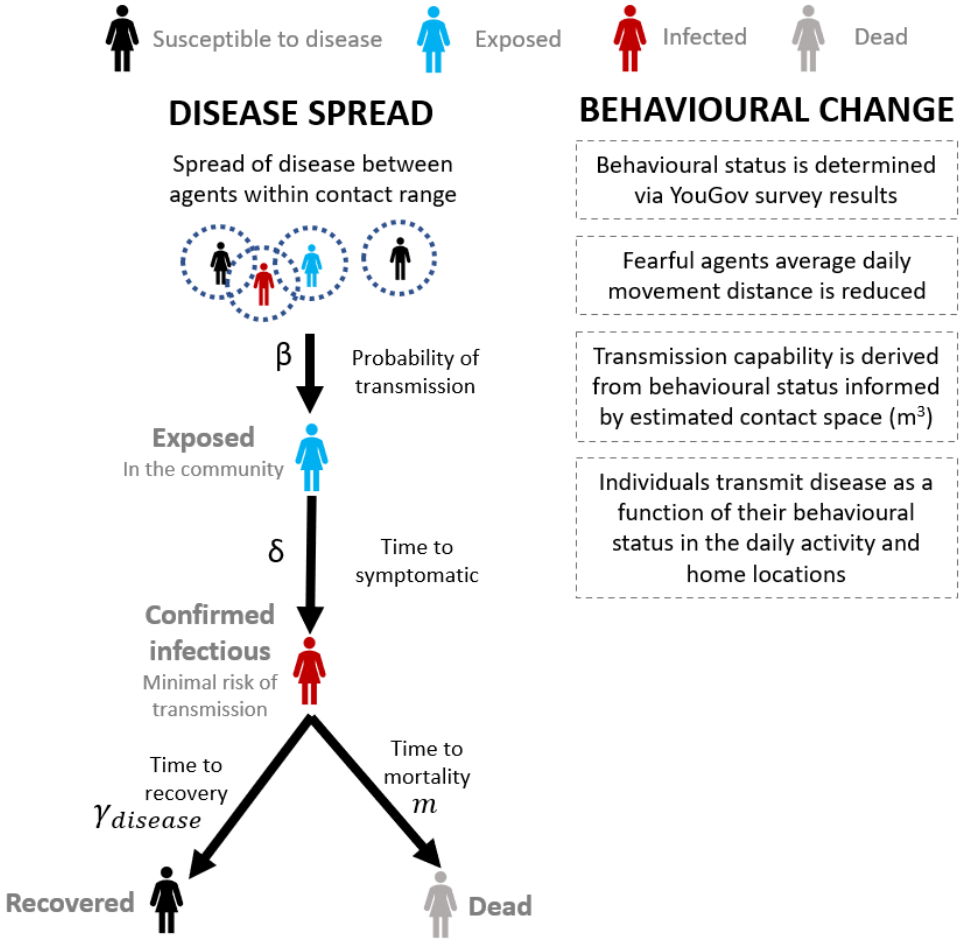


Figure 3.2. High level model overview

Table 3.1. Model parameters

Parameter	Value	Source
Model cycle		
Cycle length	1 day	Input
Cycle duration	150 days	Input
Population parameters		
Population size	6,649 people (0.1% sample)	ABS (2021)
Land size	227.82 km^2 (axes = 15.09 km)	Geoscience Australia (2001)
Proportion of population in each age category	Young (0 – 19) = 24.0% Adult (20 – 64) = 60.5% Old (65+) = 15.4%	ABS (2021)
Movement parameters		
Contact radius	Baseline: 3m radius	Input
Mean distance moved by age class (adjusted to reflect map size and population density)	Young (0 – 19) = 0.01 km Adult (20 – 64) = 0.02 km Old (65+) = 0.01 km	ABS (2018)
Standard deviation distance moved by age class	Young (0 – 19) = 0.02 km Adult (20 – 64) = 0.02 km Old (65+) = 0.02 km	ABS (2018)
Disease parameters		
Initial infected	2	Input
Probability of infection if a contact is made (β)	Fearful = 0.023 Careful = 0.031 Complacent = 0.338	Calculation – see *scheduling 3*
Exposure – duration (days) to develop symptoms (δ)	$1/7$: midpoint of estimated 0 to 14 days to develop symptoms	HealthDirect (2020)
Probability of recovery (γ)	$1/14$: most people will recover between 2 to 3 weeks from onset of symptoms	HealthDirect (2020)
Maximum life of infection	8 weeks	Constraint*
Clinical case fatality rate (m)	Young (0 – 19) = 0.0% Adult (20 – 64) = 0.1% Old (65+) = 17.0%	ABS (2020)
Behavioural parameters		
Initial population proportion	Fearful = 36% Careful = 58% Complacent = 6%	Jones (2020) – see Figure 3.3.

*While important, the effects of long COVID have not been included in this analysis.

3.2.2 Process overview, restrictions and scheduling

This section provides an overview of the processes and schedules deployed in the microsimulation model presented in this essay.

Overview and parameter restrictions

The model was run for 1,000 simulations of a 150-day outbreak. Following Kohler et al. (2000) the model took the average of all simulations where the following conditions were satisfied:

- Second wave impacts: cumulative minimum of 0.15% of the population infected, and cumulative maximum of 5% of the population infected.
- Third wave impacts: cumulative minimum of 0.15% of the population infected, and no restriction on the cumulative maximum of the population infected.

These restrictions were important to modelling the early waves of the pandemic for a number of reasons. First, in the second wave experienced by Victoria, a less infectious variant of COVID-19 was in circulation, resulting in a dampened number of cases in comparison to more infectious variants such as the later observed Omicron variant. In order to ensure the cumulative maximum restriction was appropriate, a comparison of the cumulative number of COVID-19 cases globally (as at 31 January 2021) was collated, to represent the transmissibility of the variant in circulation at that point in time. Table 3.2 provides a summary of the cumulative number of infections as a proportion of the total population in similar Western and European countries. Based on this evidence, the average cumulative proportion of the population infected with COVID-19 at this point in time was found to be 4.6%, which for the purposes of this thesis was rounded up to 5% for application to the Australian population.

As previously mentioned, the advent of more infectious strains of COVID-19, such as the Omicron variant, meant that restrictions on cumulative infections was no longer necessary, and as such any third wave impacts have not been subject to cumulative maximum restrictions, in line with both domestic and global trends. Regarding the cumulative minimum restrictions, due to the small population size in the model (0.1% population) a cumulative minimum restriction was required to ensure that the simulation was successful, i.e. additional cases were generated outside of the initial starting infected cases. Excluding these simulation results ensured that model runs where zero new COVID cases were being recorded were not included in the final results.

Table 3.2. Cumulative infections (proportion of the population) as at 31 January 2021

Country	Cumulative proportion of population infected %
United States	7.9%
United Kingdom	5.6%
France	4.8%
Germany	2.6%
Canada	2.1%
Average	**4.6%**

Source: Our World in Data (2021)

Scheduling 1: Assign behavioural status to synthetic population

Once the population age group and movement parameters have been set, the next stage in generating the synthetic population is to assign the temporal behavioural status parameter. Following the description provided in section 3.1 of this Chapter, behavioural states are informed by the question i_12_health_6: *"[Have you] avoided going out in general?"* to which individuals were able to respond with: *"Always", "Frequently", "Sometimes", "Rarely"* or *"Not at all".* Based on these response options, this data was used to proxy each behavioural compartment using the following assumptions:

- fearful agents report "*Always*" avoid going out in general;
- careful agents report either "*Sometimes*" or "*Frequently*" avoid going out in general, and;
- complacent agents report either "*Rarely*" or "*Not at all*" avoid going out in general.

Figure 3.3 illustrates the proportion of the total Victorian population in each behavioural compartment, based on these definitions for the period of analysis in which the second wave of infections was occurring. The numerator represents the number of responses by behavioural compartment, and the denominator represents the total number of respondents across all behavioural compartments. For days (time steps) where there were low or no responses, the previous day's results were carried over. Following this calculation, agents are assigned a behavioural compartment (fearful, careful or complacent) for each day of the outbreak. Given that the YouGov survey data includes age indicators, the compartment is probabilistically

matched to the synthetic population (split out by age group) in the simulation model. Note that as of 29 March 2022 data collection for the survey ceased.

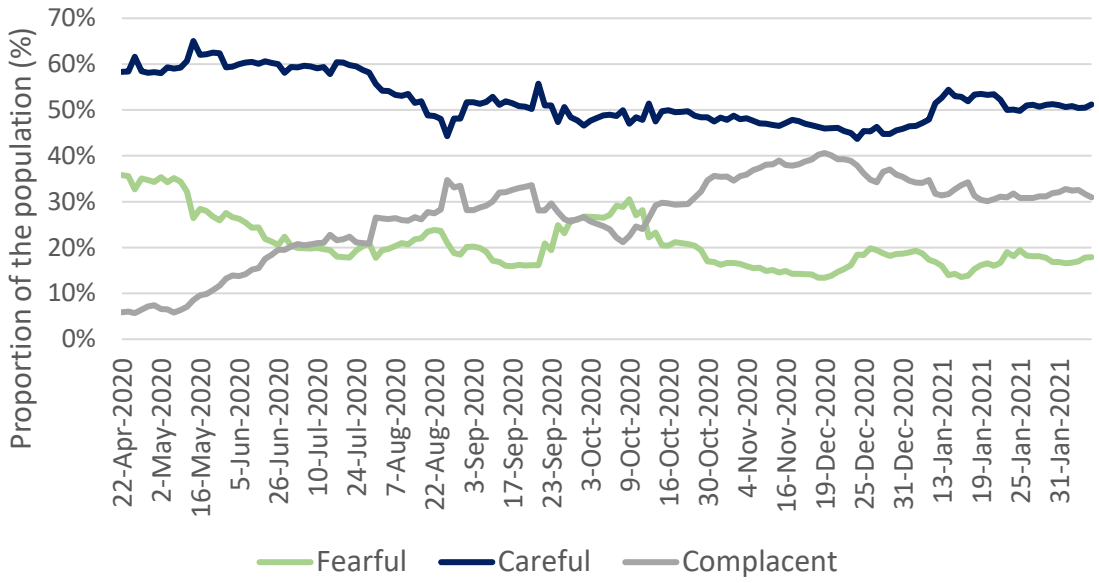


Figure 3.3. Proportion of the population by behavioural compartment (second wave date range)

Scheduling 2: Daily movement

Once the synthetic population (including behavioural status) has been established, the model then starts the simulation by initiating a daily movement for each individual agent relating to their daily activities, which may include: employment, schooling, leisure, shopping or so on. For simplicity, agents move to a fixed *activity* (x_2, y_2) location in each time step prior to moving back to their fixed *home* (x_1, y_1) location at the end of the time step (i.e. end of each day). Distance (D) travelled during this activity movement is determined by a Gaussian random normal distribution defined as:

$$D = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad [1]$$

where: μ = mean of the series, σ = standard deviation of the series, σ^2 = variance of the series and x = a normal random variable.

The mean and standard deviation parameters used to define the distance travelled are detailed in Table 3.1. Based on the assignment of the *home* and *activity* locations, it is implicit that when

one agent's contact area (3m radius) overlaps with another's, they share the same residence or activity location, such as a school classroom or an office.

Due to the city-central nature of Victoria, daily movement directions are biased towards the city centres, where it is assumed the majority of occupations and leisure activities are concentrated. Movement direction is calculated as the radian arctangent between the x-axis and the vector from the origin (city centre) to the home (x_1, y_1) coordinates. The distance travelled is applied to the cos(x) and sin(y) of the calculated theta radians. Figure 3.4 provides a schematic representation of this process for a hypothetical individual in the simulation.

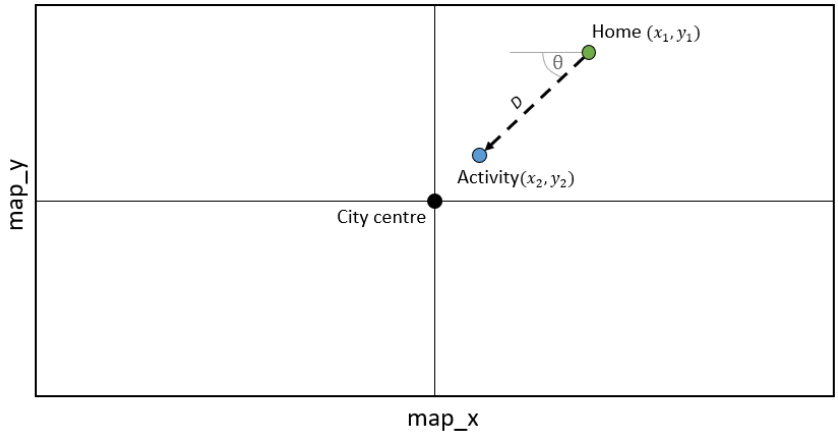


Figure 3.4. Daily activity movement bias

Daily movement is one of the key model interactions influenced by the assignment of behavioural status. When behavioural status is applied to the synthetic population, the following adjustments are made to daily movement (refer to Table 3.1 for further details):

- *Fearful* agents do not move to their daily *activity* location at all. They remain fixed at the *home* location until such time they are no longer fearful.
- *Careful* agents reduce their daily *activity* distances by a factor of two, halving the distance otherwise travelled. This adjustment is intended to capture agents who are cautious of COVID-19, undertaking necessary daily activity within their immediate locale.
- *Complacent* agents do not adjust their daily *activity* locations or distances at all. They continue with daily movements and usual patterns.

In this model, fearful agents may still transmit infection within their household, however at a substantially reduced rate of transmission (see *Scheduling 3*). Careful and complacent agents are able to transmit or become exposed to infection in either the *home* or *activity* locations.

During the period of this analysis, two major lockdowns were enforced in Victoria. Given that the underlying behavioural survey data (Jones, 2020) was collected each day throughout the lockdowns, the movement patterns in the simulation model are considered to be reflective of real-life outcomes. There is the option in the simulation model for an enforced lockdown to be applied to the full population, however the concept of the entire population remaining in their home and not leaving for any activity (such as going to the supermarket or for an approved walk outdoors) is considered to be unrealistic for modelling purposes.

Finally, confirmed infected cases (see Scheduling 3) simulate the effect of isolation restrictions and do not move to their daily *activity* location, remaining fixed at their *home* location until such time they recover from COVID-19.

Scheduling 3: Contact area and disease transmission

Once the synthetic population has been established and set with movement distances and locations (subject to behavioural adjustment) the model then assigns a contact area for each individual agent. This contact area represents a three-metre radius stemming from the agent themselves, forming a circle around the agent representing the potential for airborne transmission of COVID-19 from one agent to another. The design is that when agent radii overlap, this transmission probability is set based on disease qualities and behavioural status.

Figure 3.5 illustrates this concept. Building on the construct of an individual having a 3-metre radius around them forming their potential infection space, Figure 3.5 demonstrates the three-dimensional properties of human interaction, forming a cylinder around the individual. For simplicity, it is assumed that an agent cannot be greater than 2 metres tall in real life. Therefore, to calculate the behaviourally adjusted transmission rate, the area of the base of the cylinder must be estimated.

The World Health Organization suggests that COVID-19 infection primarily occurs between "*people who are in close contact with each other, for example at a conversational distance*" (World Health Organization, 2021), however it can occur at distances greater than 1.8 metres (6 feet) (United States Environmental Protection Agency, 2023). In order to capture these effects, the agent radius of three metres has been split up into three parts. The first sub-radius, *r1*, is the 1.5 metre radius stemming directly from the individual. This is the area, based on the available evidence, in which disease transmission is most likely to occur. The second sub-radius, *r2*, is a 0.5 metre radius, making the total radius of *r1* and *r2* equal to two metres, a

close approximation of the WHO estimation of 1.8 metres. Naturally, *r3* then represents the remaining one metre radius to make up the full three metre radius.

The way in which these radii spaces are interacted with in the model are adjusted by behavioural status:

- *Complacent* agents (that are not complying to social distancing restrictions) allow other agents to interact within their 1.5-metre (*r1*) boundary. This allows other agents to enter the closest possible level of personal space thus allowing for maximum potential disease transmission (*r1 + r2 + r3*).
- *Careful* agents (that are mindful of social distancing restrictions) maintain the minimum 1.5-metre social distancing requirement, meaning that the space of potential disease transmission for careful agents is reduced to radii *r2* and *r3* (*r2 + r3*).
- *Fearful* agents observe a minimum of 2-metre radius from others, further reducing the potential for disease transmission to *r3* only.

Following these adjustments, the calculation of risk for an individual to contract COVID-19 from another infected or exposed agent is defined as:

$$\beta_{complacent} = \left[[V_{r1} \cdot \alpha_1] \cdot 100\right] + \left[[V_{r2-r1} \cdot \alpha_2] \cdot 100\right] + \left[[V_{r3-r2-r1} \cdot \alpha_3] \cdot 100\right] \quad [2]$$

$$\beta_{careful} = \left[[V_{r2-r1} \cdot \alpha_2] \cdot 100\right] + \left[[V_{r3-r2-r1} \cdot \alpha_3] \cdot 100\right] \quad [3]$$

$$\beta_{fearful} = \left[[V_{r3-r2-r1} \cdot \alpha_3] \cdot 100\right] \quad [4]$$

where V represents the volume of the cylinder for each subset and α represents the concentration of COVID-19 particles (copies/units^3). Given the behavioural status of each agent, the probability of viral copies inhaled and deposited in the airways is determined by multiplying the volume of the cylinder the interacting agent crosses into with the associated airborne viral concentration, α. The transmission mechanism between agents is discussed later in this section.

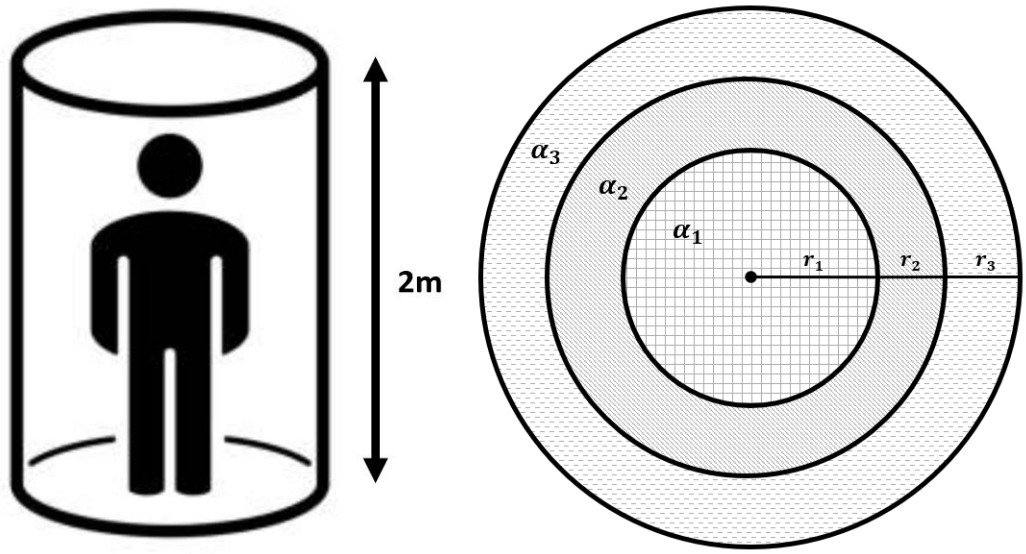


Figure 3.5. Contact and infection space

Note - Base of cylinder features. r1 = 1.5m (0.023 units in model terms), r2 = 0.5m (0.008 units), r3 = 1m (0.015 units). α_1 = low air concentration = 2.1 copies/units^3, α_2 = medium air concentration = 2.3 copies/units^3, α_3 = high air concentration = 63.2 copies/units^3. Not pictured, height of cylinder = 2m (0.03018 units). Diagram is not to scale.

Table 3.3 details the calculation steps to determine the infection risk per contact, based on equations [2 – 4]. Estimates of COVID-19 particle concentration were sourced from Riediker and Tsai (2020).

Table 3.3. Infection risk calculation

Circle radii	Volume (cm^3)	Concentration of COVID-19 particles (copies/ cm^3)	Probability of transmission within space (copies/ units^3)	Total transmission probability reference
r1 = 1.5m	0.21	1.46	0.31	(c)
r2 = 0.5m	0.17	0.052	0.01	(b)
r3 = 1m	0.47	0.048	0.02	(a)

In this model, the estimated transmission rates (probability of transmission within space column in Table 3.3) are applied additively as a function of the assigned behavioural radius. Table 3.4 summarises this effect.

Table 3.4. Additive transmission rates

Behavioural status	Total transmission probability	Justification
Fearful (a)	0.023	Fearful agents assumed to not let other agents within 2m boundary
Careful (a) + (b)	0.031	Careful agents assumed to keep 1.5m boundary from other agents
Complacent (a) + (b) + (c)	0.338	Complacent agents assumed to keep no boundary and do not adhere to any social distancing recommendations

The application of the behaviourally adjusted transmission rate is informed by degrees of social distancing and other personal preventive behaviours such as mask wearing. In this calculation, fearful agents are found to have only a minor decrease in transmission probability compared to careful agents (0.008 difference in probability). Section 3.3.4 discusses the implications of the fearful and careful transmissions probabilities on marginal utility of reduced infection.

The mechanism for disease transmission based on these total transmission probabilities follows that when an infectious agent (representative agent *i*) and non-infected agent (representative agent *j*) cross boundaries, the pairwise interaction follows Keeling and Rohani (2011):

$$\varphi_j = \vartheta\beta \sum_{i\,\epsilon\, infectious} K_T(d_{ij}) \qquad [5]$$

Where d_{ij} is the distance between agents *j* and *i* and K_T measures how transmission decreases with distance. The probability of contact ϑ signals the likelihood of contact being made once *j* enters *i's* contact area. The probability of infection (β) determines whether or not agent *j* becomes infected following the interaction.

To reduce computation time, disease transmission is limited to the static *home* and *activity* locations. Future iterations of this model would benefit from building in disease transmission during the movement step (i.e. allowing disease to be spread while commuting from a home location to the activity location). Depending on the goals and priorities of the disease spread modelling, this functionality can be enabled as required.

It must be noted that during the time of analysis, the level of vaccination (first and second doses) was very low. As such, vaccination has not been explicitly accounted for in this model.

When assessing future waves (beyond the scope of this Chapter's analysis) a dampening factor on transmission probabilities and mortality outcomes can be applied in line with the evidence generated in clinical trials.

Scheduling 4: Exposure and case confirmation

Once infected by another infected agent (*scheduling 3*), agents enter an asymptomatic exposure period defined as the average duration of time to develop symptoms $\delta = \frac{1}{D}$ where it is assumed to take seven days to develop symptoms and become a confirmed case, calculated as the midpoint of the estimated 0 to 14 days to develop symptoms (see Table 3.1). For the purposes of this model, this assumption is viable as at this stage of the pandemic, the only official way to become a confirmed case was via polymerase chain reaction (PCR) test. While exposed, agents are able to spread COVID-19 to other susceptible agents.

Scheduling 5: Recovery and reinfection

Recovery from COVID-19 is defined as a function of the average duration of infection $\gamma = \frac{1}{D}$ where it is assumed that most individual's symptoms will resolve within 14 days from the onset of symptoms (see Table 3.1). This estimate reflects current research suggesting that most people recover from COVID-19 within two to three weeks from the onset of symptoms (HealthDirect, 2020). Once recovered, agents are no longer able to transmit COVID-19 to other susceptible agents.

Due to the low rates of COVID-19 infection and re-infection during the first two waves in Victoria, the re-infection parameter was turned off for the purposes of the results presented in this Chapter. For future projections (such as the third wave impacts briefly discussed in this essay), the potential for reinfection eight weeks after recovery was enabled. The probability of reinfection after this eight week period was fixed to the behaviourally adjusted transmission rate described in *scheduling 3*.

Scheduling 6: Clinical mortality from COVID-19

Agents infected with COVID-19 are at risk of clinical mortality due to their illness. The probability of dying from COVID-19 varies by age group (see Table 3.1). The model can be adjusted for the impact of vaccination on mortality outcomes, however as previously stated, the low levels of vaccination in Australia throughout this time period of analysis precludes any

such effects. As previously mentioned, long COVID and other disease-related morbidity is not currently considered as a simulated patient outcome, however future research should look to consider this health state for model completeness.

3.2.3 Evaluation and sensitivity testing

Further details regarding the evaluation and sensitivity testing of the model presented in this essay are provided below.

Evaluation

Five key metrics were used to evaluate the outputs of the model described in this Chapter. Evaluation was undertaken by looking at cumulative infections, the Pearson correlation test, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Mean Squared Errors (MSE)

MSE is one of the more important metrics for this evaluation due to the tendency to penalise large forecast errors more than other accuracy measures. In the context of COVID-19 cases, it is important to select a model that avoids large errors. MSE is represented as:

$$MSE = \sum_t^n \frac{(y_t - \hat{y}_t)^2}{n} \quad [6]$$

where y_t represents the actual observed value and $\hat{y}_t$ represents the fitted value, estimated by the variance series, both at time t.

Root Mean Squared Errors (RMSE)

Taken as the square root of the MSE, RMSE measures results in the same units as the target variable, and is represented as:

$$RMSE = \sqrt{\sum_t^n \frac{(y_t - \hat{y}_t)^2}{n}} \quad [7]$$

Mean Absolute Percentage Error (MAPE)

Another metric for measuring model accuracy is MAPE, which expresses accuracy as a function of the absolute percentage error $\left|\left(\frac{y_t-\hat{y}_t}{y_t}\right)\right| \cdot 100$ on the fitted values:

$$MAPE = \sum_{t=1}^{n} \frac{\left|\left(\frac{y_t-\hat{y}_t}{y_t}\right)\right| \cdot 100}{n} \quad [8]$$

In addition, Granger causality testing was deployed to investigate any potential Granger-causal relationships between the outputs. Granger causality testing was used to evaluate the relationship between total infections (including both confirmed and exposed agents) and each behavioural category. For estimation purposes, a bivariate Granger-causality model of the following form is specified:

$$\pi_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i \pi_{t-i} + \sum_{i=1}^{n} \gamma_i h_{t-i} + \varepsilon_t \quad [9]$$

where π_t represents the day-on-day total COVID-19 infections and h_t represents the day-on-day behavioural state ($\pi_t = fearful, careful, complacent$). In equation [9], the lagged term π_{t-1} appears as an explanatory variable, indicating the cause-and-effect relationship between these series. In the specification, the subscript i can take an infinite value, however in this instance it is assumed to be restricted to the course of observation. The limited sense definition of Granger-causality (precedence), following Leamer's (1985) terminology, holds that for the sum of the coefficients of COVID-19 infections must not be statistically zero following the regression of behavioural status on lagged COVID-19 infections and lagged behavioural status such that $\sum \gamma_i$ $(i = 1, 2, \dots)$ is not statistically zero. Due to the design of the model presented in this essay, where behavioural interactions are built-in to the changing number of COVID-19 infections, a one-way relationship is explored.

Given the vast number of disease spread models currently available for use, this essay sought to evaluate the model outputs against two well-known epidemiological models available in the R software, SimInf and EpiModel.

The SimInf package is defined by a series of continuous-time Markov chains using the Gillespie stochastic simulation algorithm (Widgren et al. 2019). This package includes births, deaths and movements as scheduled events at predefined stages in time. This package was deployed as a simple SEIR model, run over 150 days and 1000 simulations. Recovery and exposure durations were specified as per the parameters detailed in Table 3.1. A specific

limitation of this model is that it only allows a singular transmission rate, meaning that the multiple transmission rates developed in this thesis are required to be simplified into a singular transmission rate for comparison. To accommodate this, the sum of the transmission rates defined in Table 3.4 were weighted by the proportion of the population in each behavioural compartment as at day one of the simulated process, equating to a transmission rate roughly equal to 0.046.

The EpiModel package includes a range of options for modelling infectious disease outbreaks, however for the purposes of comparison a stochastic individual SIR contact model again using the same parameters (where possible) as those listed in Table 3.1 was constructed using this package. Due to similar limitations in this package in using multiple transmission rates, the same weighted transmission rate used in the SimInf (0.046) package was deployed. The EpiModel package requires the input of an "act rate" as in the number of transmissible acts per partnership per unit of time. As this essay focuses on transmission in the home and the activity locations as two separate "acts", an act rate of two was specified. Note that both comparison models were subject to the same inclusion and exclusion criteria.

Finally, estimates of the marginal utility of fearful and careful behaviours, compared to complacent behaviours, are estimated. Following the basic principles of the law of diminishing marginal returns, the utility function for fearful and careful agents (using complacent agents as the baseline) follows equation [10]:

$$U = F^{w_1} C^{w_2} \qquad [10]$$

where U = utility, F = Fearful and C = Careful. Weights (w_1 and w_2) are calculated as the difference between transmission probabilities for complacent agents, careful and fearful agents respectively (see Table 3.10 in *results*).

Sensitivity testing

Due to the computationally complex nature of the model developed in this Chapter, sensitivity testing was limited to running a smaller number of simulations to generate sensitivity results for discussion. This section has focused on varying the exclusion criteria applied to the cumulative maximum and minimum proportion of the population infected, the transmission rate and the movement distance reduction factor applied to agents in the careful compartment. Variation in the transmission rate was based on incremental decreases and increases beyond the set parameters. The full list of sensitivity tests is listed in Table 3.5.

Table 3.5. Sensitivity tests

Description	Scenarios	Source
Variation in exclusion criteria based on total proportion of the population infected (Min, Max)	1. (0.05%, 2%) 2. (0.1%, 2%) 3. (0.15%, 2%) 4. (0.2%, 2%) 5. (0.25%, 2%) 6. (0.3%, 2%) 7. (0.05%, 3%) **8. (0.1%, 3%) – modelled output** 9. (0.15%, 3%) 10. (0.2%, 3%) 11. (0.25%, 3%) 12. (0.3%, 3%) 13. (0.05%, 4%) 14. (0.1%, 4%) 15. (0.15%, 4%) 16. (0.2%, 4%) 17. (0.25%, 4%) 18. (0.3%, 4%)	Original model output
Variation in transmission rate (Fearful, Careful, Complacent)	1. (0.02, 0.03, 0.1) 2. (0.02, 0.03, 0.2) 3. **(0.02, 0.03, 0.3) – modelled output** 4. (0.02, 0.03, 0.4) 5. (0.02, 0.03, 0.5) 6. (0.02, 0.01, 0.3) 7. (0.02, 0.02, 0.3) 8. (0.02, 0.04, 0.3) 9. (0.02, 0.05, 0.3) 10. (0, 0.03, 0.3) 11. (0.01, 0.03, 0.3)	100 runs of each scenario subject to exclusion/inclusion criteria

	12. (0.03, 0.03, 0.3) 13. (0.04, 0.03, 0.3)	
Variation in individual movement distance by careful agents	**1. Fearful do not move, careful reduce by a factor of two, complacent continue regular movement – modelled output.** 2. Fearful do not move, careful reduce by a factor of 1.5, complacent continue regular movement. 3. Fearful do not move, careful continue regular movement, complacent continue regular movement.	100 runs of each scenario subject to exclusion/inclusion criteria

3.3 Results

Results from the model described in this essay are presented in this section. The findings primarily focus on the second wave of infection in Victoria, Australia, with some brief analysis on the third wave to support findings presented later on in the overarching thesis.

3.3.1 Behaviour trends

As discussed in section 3.2, the behavioural parameters introduced in this model play a significant role in how disease is spread. This section of the results discusses the underlying trends in each behavioural compartment (*fearful*, *careful*, *complacent*). Figure 3.6 replicates the proportion of the population by behavioural compartment, per Figure 3.3, with the addition of the horizontal dotted lines to represent the average proportion of the population for each behavioural compartment for the entire analysis period (July 2020 to December 2020).

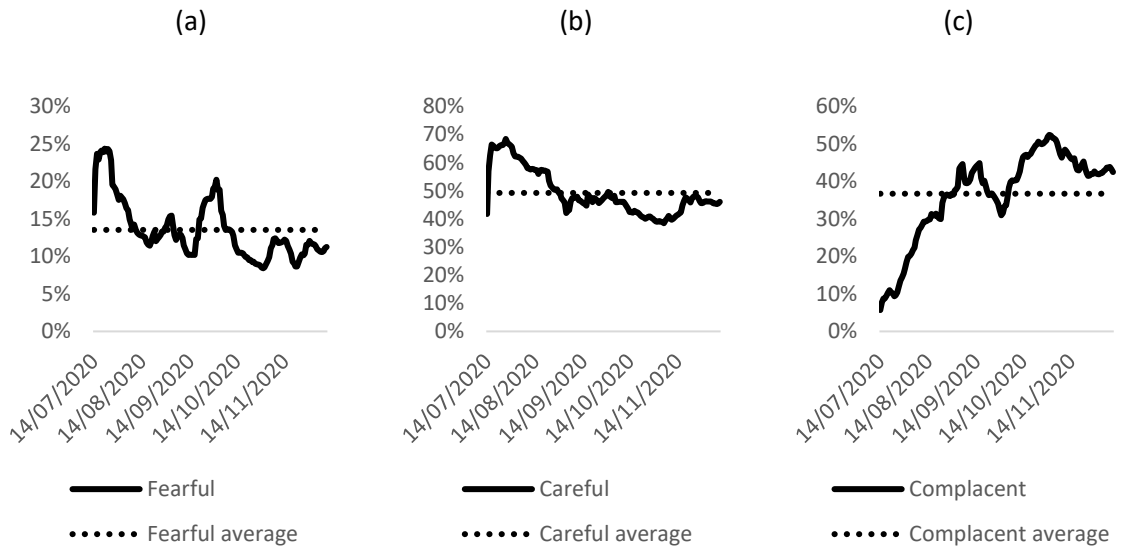


Figure 3.6. Average and observed proportion of the population by behavioural compartment

There are three primary drivers of the second wave in Victoria that can be observed visually, including:

1. High initial proportion of agents in the careful category (Figure 3.6 pane B) at the time of sampling. The proportion of agents in this category remained above average for the first 49 time steps (days) of the observation period, meaning that the peak of the second wave had already passed by the time the proportion of agents in the careful category had returned to average levels.
2. Higher than average number of agents in the fearful category over the first 28 time steps in the model (Figure 3.6 pane A).
3. An initially low proportion of agents in the complacent category (Figure 3.6 pane C), which despite proportional increases (as actual COVID-19 case numbers began to drop) did not become significant in predicting a third wave. In this instance, the number of modelled cases fell below the minimum threshold of infection probability (represented as the smallest proportion of the population that could be infected – 1 person) prior to any substantial increase. The introduction of an interstate case, international arrival that is quarantine exempt, international quarantine leak or new variant would override this modelling limitation.

These drivers, overlaid with confirmed infections, are explored further in Figure 3.7.

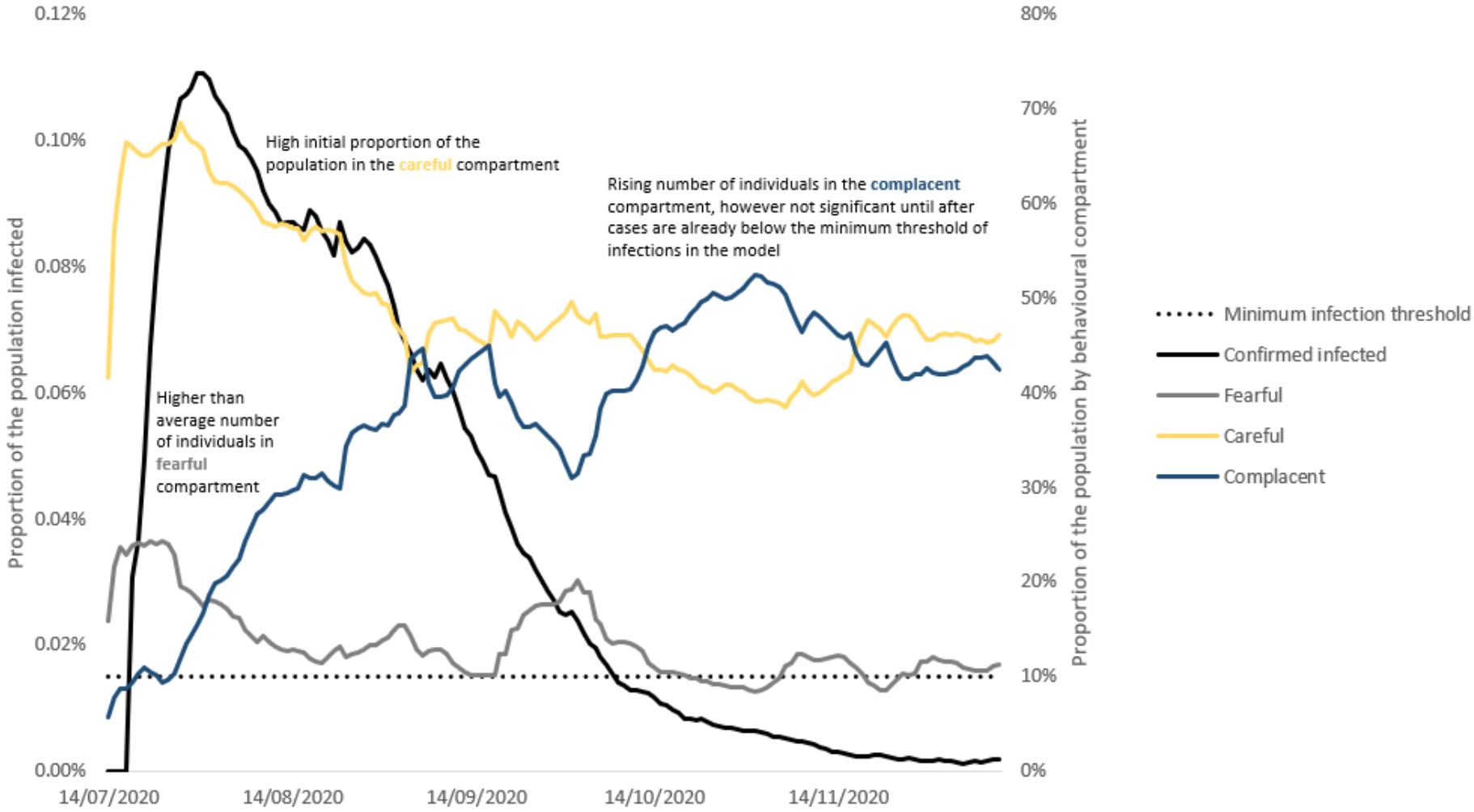


Figure 3.7. Interactions between average behavioural compartments and average confirmed infections

3.3.2 Second wave results

The primary reason for evaluating the second wave (or in other words, the first substantial wave) of COVID-19 in the Australian context is the high emotional tensions and behavioural reactions in response to the advent of an unknown and unanticipated disease. The argument here is that there is a high likelihood that the population would respond in a similar manner to another new or novel infectious disease, and as such there is merit in evaluating these preliminary days of the COVID-19 pandemic to generate insights that can be leveraged in future outbreaks of novel infectious disease. Overall, 376 simulations met the eligibility criteria for analysis (37.6% success rate). The results contained herein refer to these eligible simulations.

The average proportion of the population in the susceptible, exposed, confirmed infected and recovered classes is presented in Figure 3.8. In this figure, confirmed infections follow exposed agents with a lag.

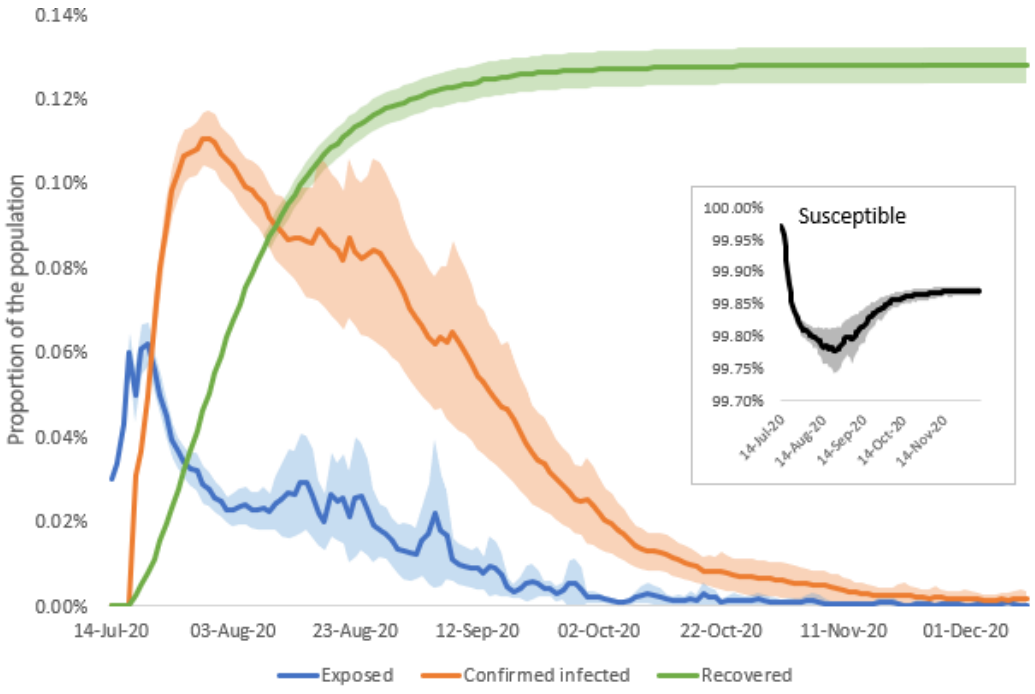


Figure 3.8. COVID-19 infection case counts

These findings are explored further in Figures 3.9 and 3.10. In Figure 3.9, it becomes clear that not all exposed cases become confirmed infected cases. There are three potential explanations for this. First, there has long been speculation of what the "hidden" COVID impact is, particularly with regard to people not coming forward for PCR testing, or in more recent times, not reporting positive RAT tests. Second, it could be deduced that some infected individuals

may not take a COVID-19 (RAT or PCR) test at all. Third, it is possible that confirmed cases in this outbreak had on average, fewer than one close contact.

Further, Figure 3.10 details the temporal transmission potential of COVID-19, calculated by adding the exposed (asymptomatic, able to spread disease) and confirmed infections (confirmed cases) together, then weighted by the respective behavioural transmission rate (see Table 3.4 in section 3.2.3). In this figure, the number of confirmed cases declines with a lag against the transmission potential. This reiterates that as the number of confirmed cases in the population declines, the transmission potential likewise declines with a lag representative of the average duration of infection. In this instance the transmission potential (x-axis) was not significant enough to support continuing increases in confirmed cases.

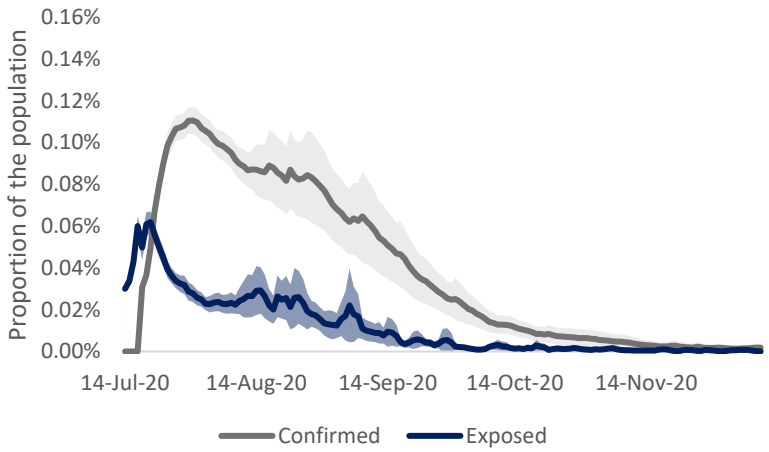


Figure 3.9. Temporal comparison of COVID-19 confirmed and exposed cases

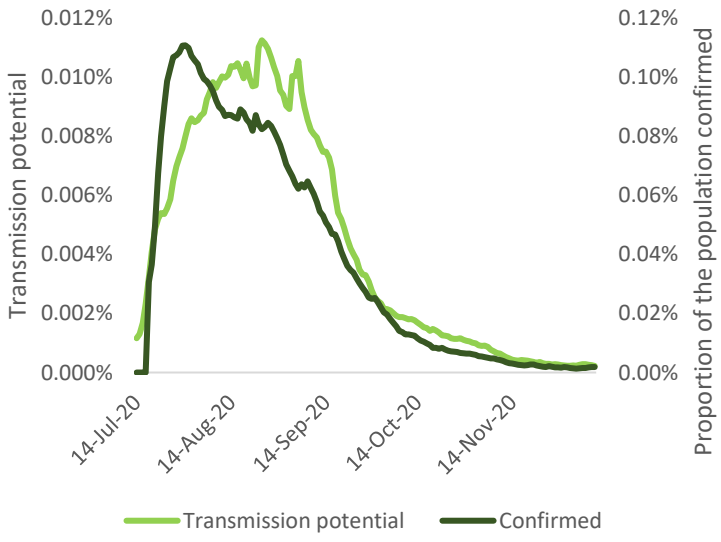


Figure 3.10. Transmission potential against the proportion of the population confirmed

3.3.3 Evaluation of second wave results

As discussed in Section 3.2, the outputs from this model are evaluated with regard to three key metrics, in addition to Granger-causality testing to test for any potential Granger-causal relationships between the model outputs.

As a means of defining the lagged term as an explanatory variable indicating the cause-and-effect relationship between the series, regression and Granger-causality testing was applied to empirically test the possible relationships that were not manually defined in the model. Granger causality testing was used to evaluate COVID-19 infections (the sum of both confirmed and exposed cases) and behavioural status across the entire study period. Lags were selected for up to 7 lag periods (i.e. days, inferring a 1-week lag period). Significant p-values from the Granger-causality tests are listed in Table 3.6.

Table 3.6. Granger-causality results (p-values)

Test	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
$\pi \rightarrow h_{fearful}$	.	.	.	0.0124	.	.	.
$\pi \rightarrow h_{careful}$	8e-07	.	.	.	.	.	.
$\pi \rightarrow h_{complacent}$	0.0412	0.0139	0.0450	.	.	.	0.0313

. = not significant, π = total COVID-19 infections (exposed and confirmed), h_{fear} = proportion of population in fearful compartment, $h_{careful}$ = proportion of population in careful compartment, $h_{complacent}$ = proportion of the population in complacent compartment.

To determine the direction of the Granger-causal relationship, a univariate least square regression was used to define each relationship that yielded a significant Granger-causal relationship. The direction of the relationships is described in Table 3.7.

Table 3.7. Significant univariate least squares regression coefficients

Relationship	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
$\pi \rightarrow h_{fearful}$				+			
$\pi \rightarrow h_{careful}$	+						
$\pi \rightarrow h_{complacent}$	-	-	-				-

+ = positive relationship, - = negative relationship. Note: only tests that yielded significant Granger-causal relationships were subject to least squares regression.

Based on these findings, the following conclusions can be drawn:

1. New COVID-19 infections (exposed agents) Granger-cause an increase in fearful agents at a lag-length of 4 days
2. New COVID-19 infections Granger-cause an increase in careful agents at a lag-length of 1 day
3. New COVID-19 infections Granger-cause a decrease in complacent agents at a lag length of 1, 2, 3 and 7 days.

Based on this evidence it can be concluded that new COVID-19 infections improve the forecasting performance of behavioural status at these specified lag lengths.

Accuracy and comparison to other publicly available epidemiological models

The model presented in this Chapter, alongside the SimInf and EpiModel packages, have been validated against the observed data for active COVID-19 infections in Victoria, Australia, from 19 March 2020 to 23 November 2020 (CovidLive, 2023). Table 3.8 compares the total number of simulations meeting the eligibility criteria for each model, including the model developed in this essay, referred to as "model". Of note is the particularly low success rate of the SimInf package, whereby the majority of simulation outputs fell below the minimum eligibility criteria of 0.15%, defined as the threshold for the minimum proportion of the population infected on the modelled sample size.

Table 3.8. Eligible simulations based on model type

Eligibility criteria	Model	SimInf	EpiModel
Total simulations meeting eligibility criteria (%)	37.6%	1.6%	47.4%

Figure 3.11 illustrates the visual accuracy of each model with respect to total confirmed infections of COVID-19. Given that the observed data recorded by the Victorian Government reporting entity has no way of determining the number of infected agents who have been exposed yet not confirmed as having the virus, exposed agents from the microsimulation model outputs have been excluded from this analysis.

From Figure 3.11 it can be concluded that while the SimInf maintains a low and relatively steady epidemic peak, the EpiModel outputs assume a constant upwards trajectory of active

COVID-19 cases. Note here that the outputs presented in this figure represent total daily cases, not cumulative cases as suggested by the shape of the curve generated by the EpiModel outputs. It is highly possible that the dilution of the behaviourally adjusted transmission rate used in the model derived in this Chapter into a singular transmission rate, that can be used in EpiModel and SimInf, is driving these results, however despite visual differences the SimInf results perform surprisingly well in the accuracy metrics described in Figure 3.12 and Table 3.9.

On the other hand, looking to the number of confirmed cases in the model derived in this Chapter, visually there is some difference in the timing of the epidemic peak, which could be partially attributable to the initial starting conditions (two confirmed cases) on a relatively small sample size (0.1% of the population). The accuracy of the epidemic curve would likely be improved should additional simulation runs be deployed.

Despite the clearly incorrect visual shape of the curve, the SimInf R package produces strong results with respect to cumulative average infections. There is an argument here, that if the primary objective is to predict cumulative active infections or cumulative proportion of the population infected, then the SimInf package would be a timesaving, less computationally intensive solution. However, this comes at the expense of a more accurate disease trajectory (see Figure 3.11) which is critical for identifying peak active cases required to inform pandemic planning and preparedness.

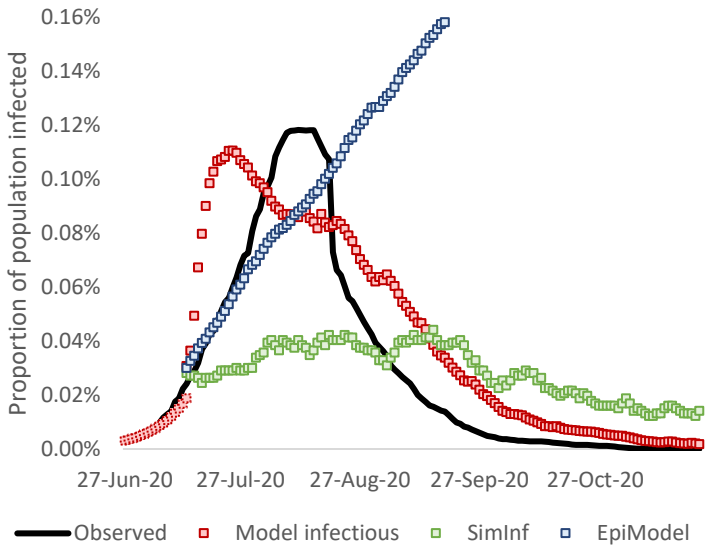
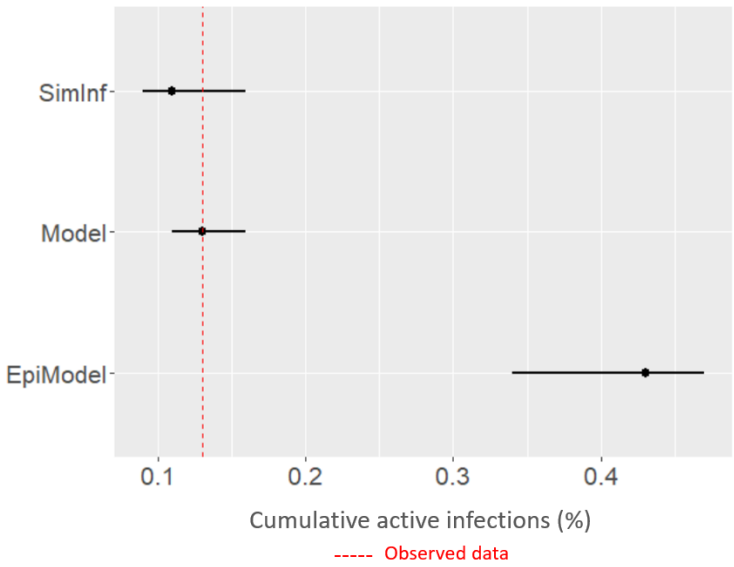


Figure 3.11. Model comparisons against observed confirmed COVID-19 cases

Note: (1) All three models were subject to a smoothing process between zero and two cases due to the initial conditions (two infectious agents) in each model. Given the sample size differences, the smoothing process is required to then align model outputs to observational data. (2) Due to the small sample size of the model (0.1% sample) there is some difference in the shape of the curve in the earlier time steps.

Figure 3.12 and Table 3.9 detail both infection and accuracy metrics for each model type. In terms of cumulative active infections, the model developed in this Chapter outperforms the SimInf and EpiModel applications with respect to cumulative active infections, correlation, MSE, RMSE and MAPE. As discussed previously, despite visual differences the SimInf package performs surprisingly well compared to the much more considered model developed in this Chapter. Benefits and limitations of these two models under different circumstances are discussed further in section 3.4.



Metric (proportion of the population)	Observed	Model	SimInf	EpiModel
Cumulative active infections (%)	0.13	0.13 [0.11 – 0.16]	0.11 [0.09 – 0.16]	0.43 [0.34 – 0.47]

Figure 3.12. Infection accuracy metrics and accompanying figures

Table 3.9. Model accuracy metrics

Scenario	Difference in cumulative proportion of the population infected	Correlation (active cases)	MSE	RMSE	MAPE (%)
Observed outbreak compared to Model	0.002%	0.86	$5.2e^{-8}$	$2.2e^{-4}$	980
Observed outbreak compared to SimInf	0.02%	0.63	$1.1e^{-7}$	$3.4e^{-4}$	5,362
Observed outbreak compared to EpiModel	-0.30%	-0.73	$3.7e^{-6}$	$1.9e^{-3}$	130,365

3.3.4 Marginal utility of fearful and careful behaviours

The real-life application of the behaviourally adjusted transmission rate is informed by degrees of social distancing and other personal preventive behaviours such as mask wearing. In this calculation, fearful agents are found to have only a minor decrease in transmission probability compared to careful agents (0.008 difference in probability). This is expected, especially in *home* locations where there is a point by which you can only be so socially distanced or protected from infection.

Table 3.10 details the weights calculated as the difference between transmission probabilities for complacent agents, careful and fearful agents respectively.

Table 3.10. Probability of reduced transmission

Behavioural status	Total transmission probability	Probability of reduced transmission	Ratio
Complacent	0.338		
Careful	0.031	0.307	1/3
Fearful	0.023	0.315	1/3

Therefore, the marginal utility of fearful agents indicates that as more agents become fearful, utility (measured in terms of reduced transmission of COVID-19) increases (see equation [11]).

$$MU_F = \frac{\partial U}{\partial F} = \frac{F^{\frac{1}{3}-1}C^{\frac{1}{3}}}{3} = \frac{F^{-\frac{2}{3}}C^{\frac{1}{3}}}{3} = \frac{C^{\frac{1}{3}}}{3F^{\frac{2}{3}}} > 0 \qquad [11]$$

where MU_F = marginal utility for fearful agents. Taking a step further to look at the rate of the marginal utility, the derivative indicates that while utility increases as more people become fearful, it increases at a decreasing rate (see equation [12]).

$$\frac{dMU_F}{dF} = \frac{-2F^{-\frac{2}{3}-1}C^{\frac{1}{3}}}{3(3)} = \frac{-2F^{-\frac{5}{3}}C^{\frac{1}{3}}}{9} = \frac{-2C^{\frac{1}{3}}}{9F^{\frac{5}{3}}} < 0 \qquad [12]$$

Due to a very close ratio for careful agents, this proof has not been replicated for this compartment. However, it can be concluded that for both fearful and careful agents there is a decreasing marginal utility over time. This suggests that once a certain proportion of the population is fearful and/or careful, the returns to reduced COVID-19 disease transmission diminish. This is particularly important as COVID-19 becomes endemic in the population, and individuals must make personal decisions about how best to protect themselves against severe

illness or harm. It may not be in the best interest for the entire population to become fearful of COVID-19, particularly in light of the severe economic consequences, given how the returns on reduced infection begin to diminish.

3.3.5 Second wave sensitivity analysis

Sensitivity analysis was conducted to evaluate three key metrics: the exclusion criteria, the transmission rate combinations and the distance reduction factor applied to travel for careful agents. The outcome of interest is the average maximum cumulative infections in the population as a metric for variance in the number of total infections over the analysis period.

Eligibility criteria

Figure 3.13 illustrates the average maximum cumulative infections under varying eligibility criteria. For reference, the combination of eligibility criteria used in the results presented in section 3.3.2 and 3.3.3 is represented by the magenta diamond.[3] This figure suggests that changing the eligibility criteria substantially impacts the maximum cumulative infections reported from the successful simulations. While the justification for the exclusion criteria are valid (see section 3.2.3), the impact of these restrictions on reducing the overall cumulative infections is not to be ignored. For further waves of infection (i.e. third and fourth waves) these restrictions are removed from the modelling outputs.

[3] Note that scaling for this figure is 6,649 persons.

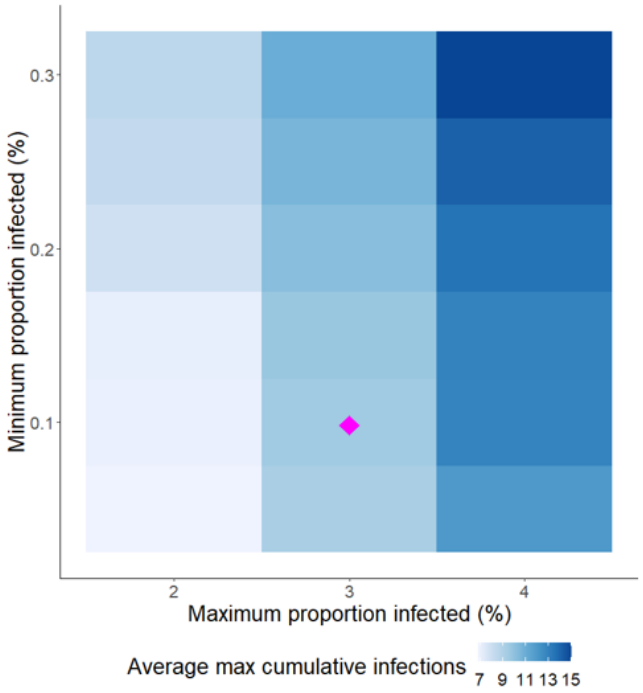


Figure 3.13. Average maximum cumulative infections under varying eligibility criteria

Note: The magenta diamond (◆) represents model outputs presented in this Chapter.

Transmission rate

Another key variable that was subject to sensitivity testing was the transmission rate. Transmission rate sensitivity tests were run for 100 simulations of each scenario. Given that the primary model presented in this essay was run for 1,000 simulations, the results will not be directly comparable due to the dampening effects of more simulations on model outputs. However, model responses to the varying transmission rates can be compared across the scenarios presented here noting that the changing transmission probabilities will be offset by the proportion of the population in each behavioural compartment (see Figure 3.7). As a result, it is unlikely that an increase in the transmission rate will result in a commensurate increase in average maximum cumulative infections for all scenarios.

Figure 3.14 illustrates the average maximum cumulative infections for varying levels of transmission for agents in the complacent compartment. In this analysis, increasing the transmission probability for the complacent compartment (all else equal) resulted in an overall trend towards increased average maximum cumulative infections. The average number of eligible simulations for this compartment was 45% of the 100 simulations run per scenario.

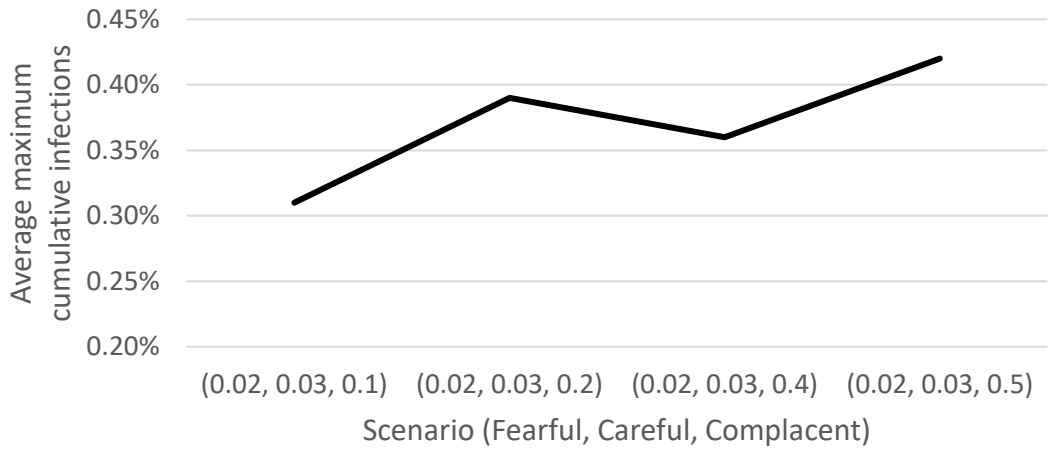


Figure 3.14 Average maximum cumulative infections complacent transmission scenarios

Figure 3.15 demonstrates that as the transmission potential for the careful compartment increases (all else equal), there is a point at the 0.05 probability that the average maximum cumulative infections is reduced. The average number of eligible simulations for this compartment was 47% of the 100 simulations run per scenario.

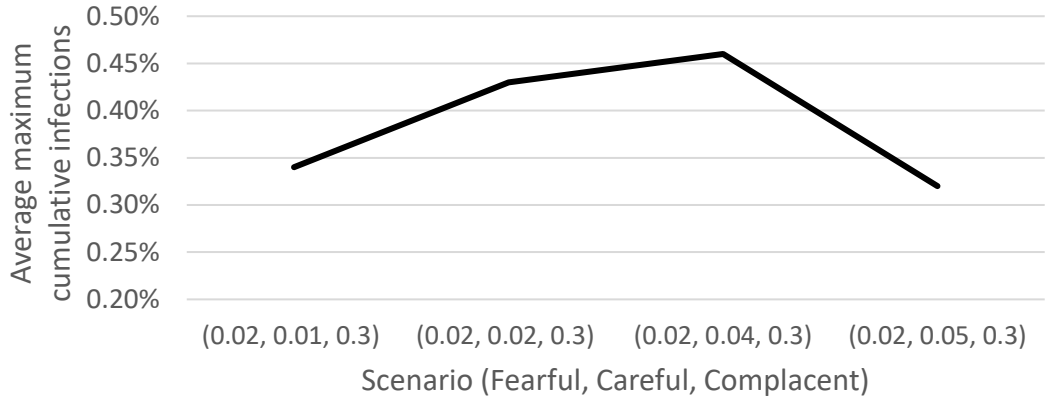


Figure 3.15 Average maximum cumulative infections careful transmission scenarios

Finally, varying the transmission rate for the fearful compartment (all else equal, Figure 3.16) illustrates a general increasing trend in average maximum cumulative infections, which begins to reduce towards the higher end of fearful transmission. The average number of eligible simulations for this compartment was 43% of the 100 simulations run per scenario.

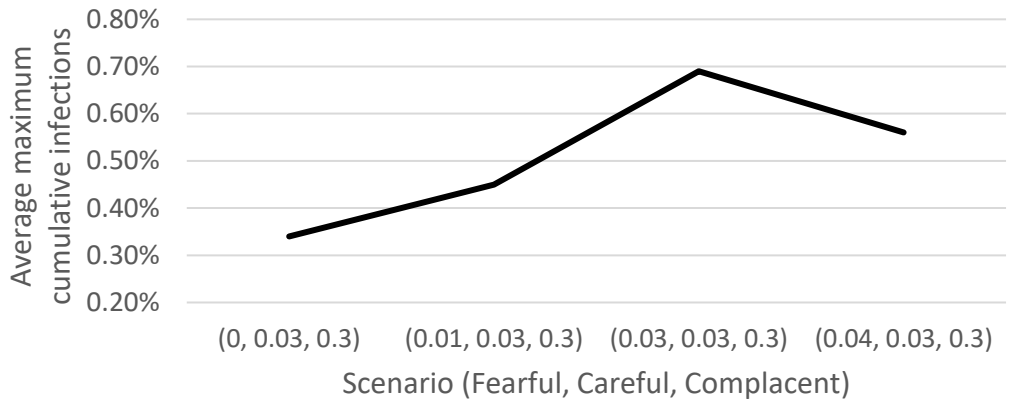


Figure 3.16 Average maximum cumulative infections fearful transmission scenarios

Careful compartment distance reduction factor

The model presented in this essay makes the assumption that agents in the careful compartment reduce their daily *activity* distances by a factor of two, halving the distance otherwise travelled.

Figure 3.17 illustrates average maximum cumulative infections under the selected scenarios of applying no factor reduction, a 1.5x factor reduction to movement and a 3x factor reduction to careful agent movement. The average number of eligible simulations for this compartment was 40% of the 100 simulations run per scenario. From this chart, it is clear that reducing the distance travelled by agents in the careful compartment reduces the average maximum cumulative infections.

Movement assumptions for fearful (do not move) and complacent (continue regular movement) were not subject to sensitivity testing.

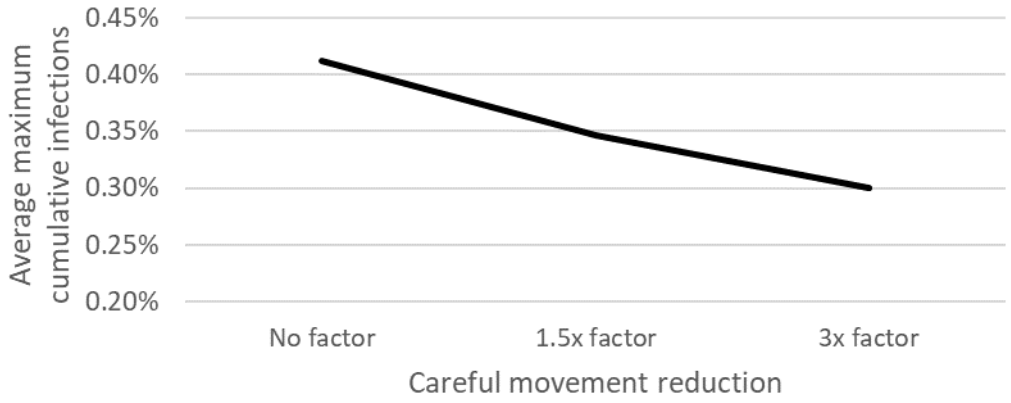


Figure 3.17 Average maximum cumulative infections careful movement reduction scenarios

3.3.6 Preliminary third wave results

Preliminary analysis was undertaken to evaluate third wave effects via the removal of simulation eligibility (cumulative infections) criteria. Figure 3.18 illustrates these findings, with high visual performance against the third, more severe wave observed in Victoria. Visually, it can be noted that the model missed the second wave of infection, which can be attributable to the differing calibration effects specified for this model. Slight error bounds can be observed around the lower proportions of active infections, which is expected given the low sample size simulated in the model.

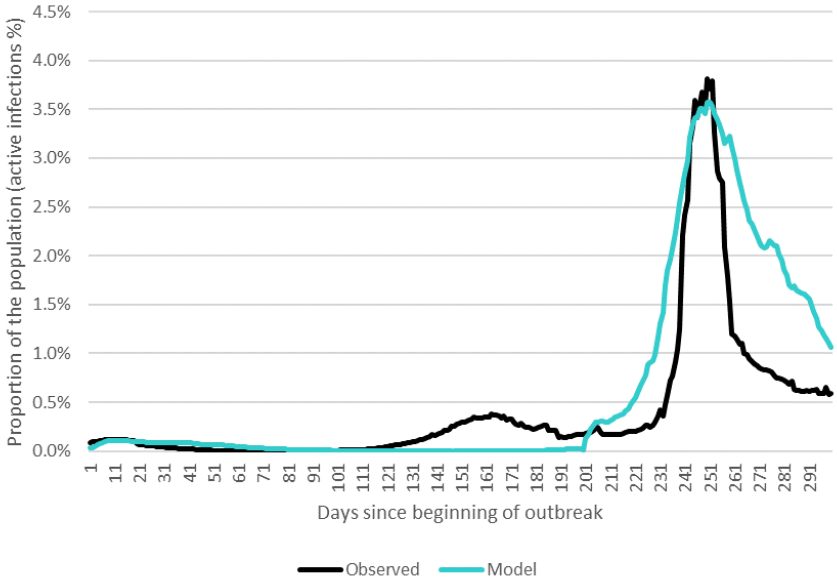


Figure 3.18. Preliminary third wave results

3.4 Discussion

This essay has sought to answer the question of what factors influence the spread of infectious disease, and how a better understanding of these factors can be used to inform pandemic planning and preparedness. To do so, this essay has established the foundational model for this thesis, namely a behaviourally influenced Susceptible – Exposed – Infected – Recovered – Susceptible (SEIRS) model of COVID-19 disease transmission. The application of this model has been in the State of Victoria, Australia, for the second and third waves of COVID-19 infection whereby individual agents were temporally classed as either *fearful*, *careful* or *complacent* in regards to the ongoing COVID-19 pandemic. The primary reason for evaluating the second wave of COVID-19 in the Australian context is the high emotional tensions and

behavioural reactions in response to the outbreak of an unknown and unanticipated disease. As such, this modelling capability is intended to assist with pandemic preparedness and planning for similar novel infectious disease outbreaks in the Australian context.

The aim of this essay has been to highlight the importance of incorporating behavioural parameters into infectious disease spread modelling methodologies. This essay has developed a mechanism by which behavioural factors can be accounted for via use of spatially explicit Agent Based Model (ABM) modelling infrastructure, allowing for nuanced effects on movement and disease transmission to be modelled. Based on these more fine-grain details, it was found that this model outperforms other popular disease spread models including SimInf and EpiModel using the same or similar parameter fits. This improved accuracy was based on visual accuracies (peak and shape of the curve) and statistical tests such as correlation coefficients and accuracy scores.

The analysis also highlighted a number of key issues in the field of disease spread modelling and emergency preparedness. Simulation outputs provided advice for the number of "hidden" or unreported COVID-19 cases in the population, represented as the number of agents that are "exposed" to infection however never become a confirmed case. In addition, based on the unique transmission probabilities modelled, it was found that compared to the probability of disease transmission for complacent agents, there is decreasing marginal utility over time for growing proportions of the population being fearful or careful of COVID-19. Using the Granger-causality approach, the existence of any linkage between new cases of COVID-19 and the number of agents in each behavioural compartment was investigated. The empirical results suggest Granger-causal relationships resulting in an increased number of fearful and careful agents in response to new COVID-19 infections. On the other hand, a Granger-causal relationship indicating fewer complacent agents in response to increasing new infections was observed.

Sensitivity analysis found the model to be sensitive to changes in the eligibility criteria, suggesting that there are potentially improved ways in which slower spreading, less severe variants can be modelled instead of flat restrictions based on the experiences of similar countries. This is a limitation of the research and will be investigated in future research. Sensitivity analysis on the transmission rate combination highlighted complex model responses to varying transmission rates, likely to reflect the proportion of the population in each behavioural compartment at a given point in time. Analysis on the distance reduction factor for

agents in the careful compartment demonstrated a reduction in the average cumulative infections for increasing reduction factors.

Compared to targeting mortality as an outcome, COVID-19 infections were specifically chosen as the outcome of interest due to the ability to then calculate the flow-on effects to hospitalisation and death. Future models should seek to include additional calibration to COVID-19 related deaths specifically, with reference to the calculation of years of life lost due to infection.

The implications of the findings presented in this Chapter are twofold. First, this Chapter demonstrates the role and importance of incorporating behavioural parameters into agent-based disease spread models. The individual actions of people have the potential, on aggregate, to shape the trajectory of outbreaks. As a consequence, these individual actions can shape the degree of morbidity or mortality experience for the population more broadly. This Chapter has argued that a more nuanced view of behavioural response – characterised by time-varying contact and movement trends – is not only possible within an ABM framework, but critical for modelling epidemic outcomes. Second, by incorporating these behavioural parameters it is possible to plan the incremental effect of new policies, such as recommended mask wearing use and social distancing, on disease trajectories, such as peak timing and height. By leveraging survey-based behavioural trackers, such as the one used in this Chapter, the ability to model individual level outcomes from the bottom up is greatly enhanced. Chapter 5 will discuss the role of overlaying models such as the one described in this Chapter for improving pandemic preparedness using early warning system activation measures.

Future research in this area should aim to address some of the key limitations discussed throughout this essay, including the appropriateness of applying a cumulative maximum infections restriction on the number of cases for waves characterised by variants that are less transmissible than others, such as the Omicron variant. Explicit fitting of parameters (such as delay distributions) to observed data was not undertaken, however is recommended as a further extension of the work to increase the level of realism in the simulated outputs. An example could include fitting the model to the current reproduction number as per Aleta et al. (2021). Additional advances in parallel computing in R or transfer of model infrastructure to a more efficient platform would be beneficial for running larger population proportions, thus minimising the number of repeated simulations that must be run in order to generate robust estimates. An alternative solution would be to investigate the robustness of an ABM

constructed at the household level, acknowledging the limitations associated with standardising daily movement patterns for varying age and demographic characteristics contained within the household composition. Finally, this analysis would be improved by providing gender disaggregation, in addition to any other demographic disaggregation that may improve the fidelity of the interactions each agent is having with their surrounding environment.

4. Essay two: An input-driven microsimulation framework to measure the effect of behaviour propagation on endemic infectious disease spread

4.1 Introduction

Influenza is an airborne communicable disease that affects the upper airway and lungs of both humans and animals, typically during the colder months of the year. Over a hundred years after the 1918 influenza pandemic, the influenza virus (in all its subtypes) continues to be recognised as one of the most complicated public health problems faced by health authorities. In the same way that individual behavioural decisions impact the spread of novel infectious diseases, more well-known infectious diseases – such as influenza – are subject to the same risks. Novel infectious diseases in the current digital landscape are highly monitored and heavily researched. This allows for a wealth of data to be input into mathematical models estimating the trajectory of disease spread. In this essay, I argue that robust disease spread models can still be established in the absence of such data. Using influenza as a case study, an input-driven microsimulation framework has been established to measure the effect of behaviour propagation on disease spread.

Disease and fear transmission have been most notably modelled by Epstein et al. (2008), involving the combination of an agent-based Susceptible-Infected-Recovered (SIR) differential equation model with fear dynamics. In Epstein's model, individuals can contract fear through either disease-infected, fearful, or disease-infected and fearful individuals. Those fearful individuals were then removed from the population. This essay develops upon this research by building in a removal function for fearful individuals, positing that even though an agent may become fearful, they can still return to their activities of daily living once they are no longer fearful. In reality, individuals may choose to self-isolate (or, in model terms, be removed from participating in daily activities) either for fear of their own infection, family or other vulnerable members of their community. Incorporation of movement parameters is well established, particularly in the influenza context. For example, Balcan et al. (2009) successfully integrated mobility and transportation data into modelling to estimate novel influenza A(H1N1) transmission potential for 220 different countries.

Other research has noted the importance of behavioural change on the epidemiology of an outbreak, including that:

- Depending on the prevalence, disease spread slows considerably if people enter a state of low activity (for example, reduced travel) (Del Valle et al. 2005), and;
- The dissemination of health information can reduce the prevalence of infection, if it results in individuals seeking earlier treatment or engaging in preventative action (Kiss et al. 2010).

In Chapter 3 (essay one), individuals (referred to in the model as 'agents') were capable of three behavioural states in response to COVID-19: fearful, careful or complacent. These behavioural states were informed by readily available, global survey data provided by the YouGov COVID-19 Behaviour Tracker (Jones, 2020) and were input into a spatial agent-based model (ABM) to simulate disease spread. As data sources such as this become more infrequent or are discontinued, similar problems will be faced by COVID-19 as less novel infectious diseases such as influenza.

The model presented in this essay addresses these data limitations by switching to an input-driven microsimulation framework for modelling the trajectory of disease spread during severe influenza seasons. Using the existing definition of fear, the following changes have been applied in this essay, compared to essay one:

- Behavioural states limited to "fearful" and "not fearful".
- Introduction of "hotspot" scalars:
    - Environmental scalars allowing agents to become infected by touching surfaces or objects in their contact radii that have been previously touched by infected agents within a set time period.
    - Fear scalars allowing social media and rumours to influence individual views and perceptions of the severity of an outbreak based on a pre-specified retention rate.

In this essay, the term "scalar" refers to a graduated series of decreasing steps, for example the infectiousness of a previously touched surface (such as a doorknob) decreasing at rate x over time. The incorporation of these scalars is made possible by more detailed research being available in the influenza setting, compared to what was available (at the time of writing) for the COVID-19 setting. Including a fear scalar is an experimental addition, following notes by

Kasperson et al. (1988) detailing how risk events that pose even a minor physical consequence to the individual have the ability to bring about “extraordinarily severe” social impacts and public concern. The importance of including such scalars was highlighted in Nguyen and Prokopenko (2022) who noted that future research should seek to model the effects of behavioural factors such as social media messaging and personal perceptions of risk. Specifically, this parameter was constructed in light of other social learning frameworks, such as that described in Busetta et al. (2001). The authors of this research established a novel agent breakdown into suggesting agents (external observers) and service agents (agents doing the work) where agents can engage in communication and overhearing to pass on information and update work preferences. The model contained in this essay builds on the concept of overhearing by establishing a spatial environmental scalar for fear, represented as a fear hotspot. Data analysis and simulation modelling was undertaken using R programming software run in parallel within the Australian National University National Computing Infrastructure (Raijin, now known as Gadi) environment.

Following Chapter 3 (essay one), this essay builds on the first research question, within the context of a lack of granular behavioural survey data:

What are the main factors that influenced the spread of disease, and how can a better understanding of these factors in models be used to inform pandemic planning and preparedness?

Considering these new model additions and removals, this essay develops and deploys a behaviour-driven disease spread model for influenza in the region of Canberra, Australia. The model presented in this essay is intended to serve as a case study for incorporating behavioural dynamics in disease spread models, despite known data limitations. Model outcomes across multiple influenza seasons both with and without behavioural adjustment are presented.

The results of this essay show that incorporation of fearful behaviour in disease spread modelling demonstrates the greatest estimation power in severe seasonal influenza outbreaks (such as the 2017 season), as compared to less severe influenza seasons (such as the 2016 season). The working hypothesis is that when so few agents are being infected (in less severe seasons) the uptake of fear and associated fear related behaviours does not yield any change to the infection trajectory. On the other hand, in severe influenza seasons, the incorporation of the behavioural parameter improves the disease estimation process. The mean absolute percentage error (MAPE) for the behaviourally adjusted scenarios with fear was significantly lower,

reported at 24.8% and 160.9% compared to the non-adjusted scenarios without fear reported at 50.8% and 234.9% for the 2016 and 2017 seasonal outbreaks respectively. This implies that the behaviourally adjusted scenarios with fear provide a more accurate prediction for both seasonal outbreaks. Granger-causality testing found that the actual mechanism of the fear related behaviours is protective of infection with influenza.

These findings support continued research into the incorporation of behavioural parameters into disease spread models and pandemic preparedness activities, regardless of the operating availability of data. Whether a survey-driven or input-driven approach is applied, these models are readily coupled with statistical early warning systems (see Chapters 5 and 6) to assist with crisis response and planning functions.

This essay is organised as follows. Section 4.2 explains the methods behind the model and adjustments made to accommodate an input-driven microsimulation approach. Section 4.3 presents the primary results from running the ABM, followed by section 4.4 providing important implications and conclusions for health modelling and policy during pandemics.

4.2 Methods

The model developed in this essay is a spatially explicit Agent Based Model (ABM) designed to simulate the spread of influenza under a SIR (Susceptible – Infected – Recovered) framework on a population of 4,257 agents (a 1% sample) in the city of Canberra, Australia. The rationale for choosing Canberra is that it is a relatively dense, co-located region where impacts in one suburb have the potential to greatly influence the behaviours of others living in the same or nearby suburbs. As discussed in Chapter 2 and Chapter 3, due to the sheer computational power required to run large ABMs, smaller fractions of the population must be simulated upon in order to generate results in a timely manner. Larger population fractions (greater than 10% sample) can be simulated with this model framework, time permitting. The model contains the option for parallel processing which can be constructed across simulations, however due to the time-dependent nature of each timestep (in other words, the outcomes of today are dependent on the interactions of previous days) the efficiency of the algorithm is still contained to the length of one full simulation run.

This model simulates the following dynamics:

- Transmission of disease from interaction with other infected agents or surfaces.

- Transmission of fearful status from interaction with other fearful agents or hotspot presence.
- Varying daily movement characteristics based on behavioural status.

The SIR model developed in this essay is overlaid with the starting population of fearful agents (two people) which is generated at the beginning of the simulation. Results have been validated per observational data from influenza seasons in 2016 and 2017.

4.2.1 State variables and scales

A high-level summary of the modules and processes for the model described in this essay is provided in Figure 4.1. Following a similar pattern to Chapter 3 (essay one), this model tracks the daily movements and infection likelihoods of agents within their physical environment as a function of their behavioural status (fearful or not fearful). Due to increased data availability, the synthetic population is constructed of five age brackets: *young* (0 – 5 years), *school* (6 – 17 years), *work* (18-64 years), *retired* (65-84 years) and *older* (85 years and above). Due to the lack of available evidence at the time of modelling for gender-disaggregated impacts of communicable disease infection (i.e. workplace imbalances in certain professions, such as nursing and childcare), gender disaggregated insights have not been featured in this model.

This synthetic population undertakes fixed daily movement activities which are statistically biased towards the city centre. Throughout movement to and from an individual's home location and daily activity location, they make contact with other agents within the bounds of their contact space to which disease and/or fear may be transmitted. Disease and/or fear may also be transmitted while within the fixed home location or fixed daily activity location.

This model provides validated results for 2016 and 2017 influenza seasons in Canberra, Australia.

INPUTS

Spatial module
- Set spatial boundary in line with area of target location (x, y)
- Assume everyone moves each day
- Set daily movement distance by age group (average, standard deviation)
- Cluster agents around city centre

Demographic module
- Set number of agents
- Assign age (young = 0 – 5 years, school = 6 – 17 years, work = 18-64 years, retired = 65-84 years, older = 85 years and above)
    - Assign mortality rates for clinical infection

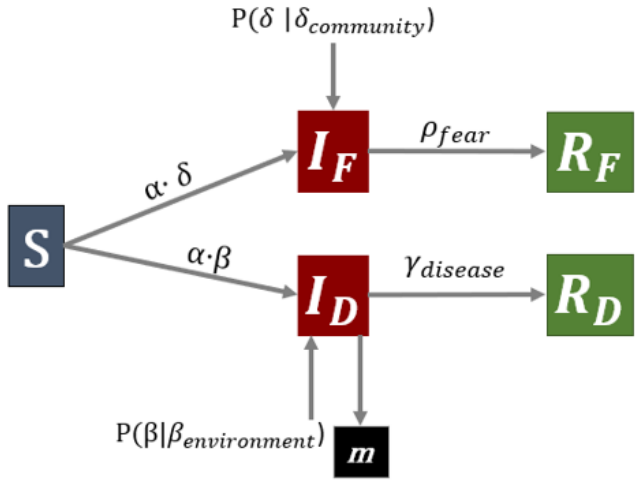
Interaction module
- Assign agent interaction radii and space in terms of a 3D cylinder of interaction space.
- Assume probability of contact with overlapping radii

Disease module
- Set initial prevalence of disease
- Set contact probability
- Set infection probability
- Set recovery rate
- Set maximum days of infection
- Set environment infectivity retention rate and scalar

Behaviour module
- Set behavioural status (fearful, not fearful)
- Set community fearfulness retention rate and scalar

Monitoring
- Measuring disease status (susceptible, infected, recovered), fearful status (susceptible, positive or recovered), days infected, days fearful, outbreak days, home location and daily activity location (e.g. occupation)

CALCULATIONS



Where:
S = susceptible (to infection or fear)
α = contact rate
β = transmission rate of infection. This is a function of the infection status of the interacting agent and the infectivity of the surrounding environment ($\beta_{environment}$)
δ = transmission rate of fear. This is a function of the fear status of the interacting agent and the level of fear in the surrounding community ($\delta_{community}$)
I_D= Infected disease
I_F= Positive fearful
m = clinical mortality rate
$\gamma_{disease}$= recovery rate from infection
ρ_{fear} = recovery rate from fear
R_F = recovered from fear
R_D = recovered from infection

Figure 4.1. High level model overview

4.2.2 Parameters

The input parameters used to inform the model described in this essay is provided in Table 4.1. Parameters included in this model are all publicly available, sourced from Australian statistical collection agencies and the Australian Government Department of Health and Aged Care Pandemic Influenza Planning Assumptions. Note that due to the time at which this model was built, some parameters may appear to be outdated. These parameters would be updated prior to use in the current setting, however inputs current at the time of writing this essay have been included below.

To maintain an accurate population density for Canberra given a 1% population sample, the size of the modelled surface area has been reduced to 24.76 km^2. Agents not infected at the beginning of the simulation ($t = 0$) are assumed to be susceptible to contracting the novel strain of influenza. In other words, natural immunity is not factored into the modelling assumptions.

Table 4.1. Model parameters

Parameter	Value		Source
Population density Canberra	179 people per square km		(ABS, 2019)
Population size	4,257 people (1% sample)		(ABS, 2019a)
Proportion of population in each age and gender category*	*Female* Young = 3.9% School = 6.8% Working = 32.9% Retired = 6.0% Older = 0.9%	*Male* Young = 4.1% School = 7.3% Working = 32.1% Retired = 5.3% Older = 0.6%	(ABS, 2019a)
Contact space (α)	1-metre radius around each agent		(Perez & Dragicevic, 2009)
Proportion of agents moving by age and gender category	*Female* Young = 47.0% School = 87.4% Working = 64.3% Retired = 50.0% Older = 20.0%	*Male* Young = 47.0% School = 88.4% Working = 71.5% Retired = 50.0% Older = 20.0%	Young: (ABS, 2020a) and (National Centre for Education Statistics, 2019)[1] School: (ABS, 2020b) Working: (ABS, 2019b) Retired: assumption Older: assumption
Mean distance moved by age class (adjusted to reflect map size and population density)	Young = 0.2 km School = 0.27 km Working = 0.33 km Retired = 0.2 km Older = 0.1 km		Young: assumption School: (ABS, 2018a) Working: (ABS, 2018a) Retired: assumption Older: assumption
Standard deviation distance moved by age class	Young = 0.3 km School = 0.38 km Working = 0.40 km Retired = 0.3 km Older = 0.15 km		Young: assumption School: (ABS, 2018a) Working: (ABS, 2018a) Retired: assumption Older: assumption
Initial infected	2		Initial condition

Probability of infection if a contact is made (β)	Scenario one (2016): 4% Scenario two (2017): 5%	Initial condition
Probability of recovery (γ)[t]	¼: one week less three days to start recovering	(Department of Health, 2014), (Rakowski et al. 2010)
Maximum life of infection	30 days	(Department of Health, 2014)
Clinical case fatality rate (m)	Young = 2.5% School = 1.0% Working = 1.5% Retired = 2.0% Older = 2.5%	(Department of Health, 2014)
Environment infectivity retention rate (rr)	0.007	Assumption
Environment infectivity scalar	0.001	Assumption
Initial infected with fear	2	Initial condition
Probability of becoming fearful if a contact is made (δ)	10%	(Epstein et al. 2008)
Probability of recovery from fear (ρ)[t]	$1/12$: two weeks less two days to start recovering	Assumption
Maximum life of being fearful	500 days	Assumption
Community infectivity retention rate (rf)	0.99	Assumption
Community infectivity scalar	0.01	Assumption

[*] Young = 0–5 years, School = 6–17 years, Working = 18–64 years, Retired = 65–84 years, Older = 85+.

[1] Young movement characteristics are based on US data, averaged with ABS (2020a), as some agents in the Young category will be attending school.

4.2.3 Process overview, restrictions and schedules

This section provides an overview of the processes and schedules deployed in the microsimulation model. The model built is based on a stochastic model, where parameters are

held constant although each run is different. In this model, deviations are represented as random perturbations away from, and transient-like return towards, the deterministic attractor (Keeling & Rohani, 2011). Sustained cycles are balanced on the probability that epidemic oscillations around normal endemic prevalence are activated by the stochastic nature of the model.

Overview and parameter restrictions

The model was run for 200 simulations of a 200-day outbreak, across two scenarios (400 total runs). These scenarios track two separate seasonal influenza outbreaks experienced in the Canberra region in 2016 (a relatively moderate seasonal influenza outbreak) and 2017 (a relatively severe seasonal influenza outbreak). The results presented in this essay follow Kohler et al. (2000) by taking the average proportion of the population infected for each scenario.

Scenario restrictions were applied to the cumulative number of infected agents. If the cumulative number of infected agents did not reach 0.1% of the population the simulation was excluded from the results. The reason for this restriction is to ensure that only simulations resulting in an outbreak (i.e. a successful run) are reported on.

Scheduling 1: Assign behavioural status and initial infected to synthetic population

Once the population age group and movement parameters have been set, the next stage in generating the synthetic population is to randomly assign the number of fearful and infected agents. In this model, a starting stock of two agents is selected to be infected and fearful respectively.

Scheduling 2: Daily movement

Over the course of one simulated day, moving agents travel from their *home* location to their *activity* location (school, work, shopping mall, etc.), before returning to their *activity* location at the end of each day. *Activity* locations are biased towards the city centre. Agents may have different interactions or contacts with random agents each day, excluding those living in their residence, which are fixed. An agent that is assigned the fearful status is one that will actively seek protective behaviour, whereby once they return to their *home* location (or in the case of non-movers, remain in their *home* location), they cease to continue moving to their *activity* location and elect to remain at *home*.

When an individual becomes fearful, they cease any movement to their *activity* or place of education and other daily movement activities, instead remaining at *home*. This acts as a self-isolation measure to limit exposure to disease while the agent is fearful. Agents then move from the infected (disease) / fearful (behaviour) to the recovered (disease) / removed (behaviour) class at probabilities $\gamma_{disease}$ and ρ_{fear}. As this process is occurring, agents infected with influenza (disease) exit the population via the clinical mortality rate *m*. Due to the short time span of analysis, reproduction and natural mortality have not been included in this model.

As per prior research in the field (Crooks and Hailegiorgis, 2014), this essay classes emergence as the overarching culmination of individual interactions between agents over time to form something larger (Kelly et al. 2013). This is particularly true in concentrated distributions around the centre of a city. This essay addresses concerns raised by Bobashev (2007) regarding potentially spurious results gained from lack of disaggregation, by grouping the model population into 5 separate age categories.

Note here that the 'Canberra, Australia' region is a unique Territory by virtue of being relatively small, dense around the city centre, and isolated from nearby towns. For this reason, Canberra is easier to model, noting that the general pathway of an individual on their daily commute is for the most part predictable. Additional complexity (primarily in the form of computational power) would be introduced into a model attempting to calculate the spread of disease in a city that is more dispersed or complicated.

Scheduling 3: Contact area and disease transmission

Once the synthetic population has been established and set with movement distances and locations (subject to behavioural adjustment), the model then assigns a contact area for each individual agent. Due to the simplified nature of this model (from a behavioural viewpoint), a 1-metre radius[4] stemming from the agent themselves is used to form the circle space representing the potential for airborne transmission of influenza from one agent to another.

In this model, fear is a status that is transmitted from one agent to another, due to the lack of available survey data indicating behavioural status at various points in the influenza season. As a result, this essay does not calculate a behaviourally adjusted transmission rate given that

[4] Note that this model was developed in 2019, pre COVID-19 pandemic. As such, assumptions such as this do not reflect the most up-to-date social distancing and disease transmission advice.

fearful agents return to their *home* location and are removed of their ability to transmit infection to other agents.

The mechanism for disease and fear transmission based on these total transmission probabilities follows that when an infectious (representative agent *i*) and non-infected agent (representative agent *j*) cross boundaries, the pairwise interaction follows Keeling and Rohani (2011).

$$\varphi_j = \vartheta\beta \sum_{i\,\epsilon\, infectious} K_T(d_{ij}) \qquad [13]$$

where d_{ij} is the distance between agents *j* and *i* and K_T measures how transmission decreases with distance. The probability of contact ϑ signals the likelihood of contact being made once *j* enters *i's* contact area. The probability of infection (β) (interchanged with fear, δ) determines whether or not agent *j* becomes infected following the interaction.

Scheduling 4: Environmental transmission

Agents may also become infected by touching surfaces or objects in their contact distance that have been previously touched by infected agents. These surfaces or objects may continue to be infectious even after the infected agent has left. Here, the probability of becoming infected (β) is a function of whether the environment ($\beta_{environment}$) is infectious or not P(β|$\beta_{environment}$). Measured on a scale from 0 to 1, the ability of the environment to retain infectious material decays exponentially such that, for a given infectivity of the environment at time *t*, at time *t+1* the level of infectious material in the environment is $\cdot\, rr$, where *rr* is equal to the pre-specified retention rate.

Social media and rumours play an important role in forming individual views and perceptions of the severity of an outbreak. If there is significant concern from reputable sources, it is likely that this will contribute to individual levels of fear in the general population. The probability of becoming fearful (δ) is a function of whether the surrounding community ($\delta_{community}$) is a 'fear hotspot' or not P($\delta|\delta_{community}$). The community does not remain a fear hotspot for the entire life of the outbreak, however. Measured on a scale from 0 to 1, the ability of the community to retain significant fear messaging decays exponentially such that, for a given level of fear in the community at time *t*, at time *t+1* the level of fear in the community is $\cdot\, rf$, where *rf* is equal to the pre-specified retention rate.

Scheduling 5: Recovery, removal and reinfection

The recovery and removal rates for influenza and fear are functions of the average duration of infection (*D*). In other words, the recovery rate for disease (γ) is calculated as:

$$\gamma = \frac{1}{D} \quad [14]$$

where it is assumed that, for influenza, most symptoms will resolve within one week of onset ($\gamma = 0.14$) (Department of Health, 2014). Each agent in the model is therefore subject to three days of illness before they start recovering at a rate of 0.25 ($\gamma = 0.25$) from day four onwards. This is consistent with the approach taken by a similar study focusing on the spread of influenza in Poland (Rakowski et al. 2010). For fear, it has been assumed that agents will no longer be fearful within two weeks of onset ($\rho = 0.07$) with a two-day lead-in time in the model ($\rho = 0.08$).

Scheduling 6: Clinical mortality

Agents infected with influenza are at risk of clinical mortality due to their illness. The probability of dying from influenza varies by age group (see Table 4.1).

4.2.4 Evaluation and sensitivity testing

Model outputs are contrasted with actual observed outbreaks using the total proportion of the population infected in terms of maximum, minimum and standard deviation of those infected, dead and fearful. Actual outbreaks are measured as the total number of infected persons in the Canberra region by year, as reported by the Australian Government Department of Health (Department of Health, 2019) divided by the estimated resident population by year, as reported by the Australian Bureau of Statistics (ABS, 2019).

The model validation process follows testing the model's ability to predict historic outcomes in order to ascertain its relevance for future applications. Five key metrics were used to evaluate the outputs of the model described in this Chapter. Evaluation was undertaken by looking at cumulative infections, the Pearson correlation test, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). For definitions of each evaluation metric, refer to equations [6] – [8] in section 3.2.3 of Chapter 3.

The p value <0.05 was considered statistically significant. Sensitivity is measured in this essay by varying the transmission rate of disease spread. Future research would benefit from an increased understanding of temporal fear spread parameters.

Causality testing

To further identify the significance of incorporating fear into the model, this essay then uses a Granger-causality test to examine the relationship between infection and fear. A bivariate Granger-causality model of the following form is specified:

$$m_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i m_{t-i} + \sum_{i=1}^{n} \gamma_i f_{t-i} + \varepsilon_t \quad [15]$$

$$f_t = \beta_0 + \sum_{i=1}^{n} \beta_i f_{t-i} + \sum_{i=1}^{n} \delta_i m_{t-i} + \vartheta_t \quad [16]$$

where m_t represents the per-day proportion of the population infected with influenza, and f_t represents the per-day proportion of the population that is fearful. The cause-and-effect relationship between both series is indicated using the lagged terms m_{t-i} and f_{t-i} as explanatory variables. Following Leamer's (1985) definition of Granger-causality, infection can granger-cause fear provided that the sum of the coefficients of infection must not be statistically zero following the regression of fear on lagged fear and lagged infection, thus implying that $\sum \gamma_i$ (i = 1, 2, …) is not statistically zero in equation [15]. The second potential finding is that fear can granger-cause infection, provided that $\sum \delta_i$ (i = 1, 2, …) is not statistically zero in equation [16]. The third potential finding would be a two-way causation, whereby both conditions hold simultaneously.

4.3 Results

Results from the model described in this essay are presented below. The results focus on the 2016 and 2017 influenza seasons in Canberra, Australia.

4.3.1 Disease trajectory accuracy

Both scenarios tested using this model generate an accurate representation of the 2016 and 2017 seasonal influenza outbreaks in Canberra, Australia with respect to the proportion of the population infected.

Table 4.2 presents the results of both the microsimulation model proposed in this essay, and the actual outcomes of the corresponding seasonal influenza outbreak in the region of interest.

For each simulation model, not all simulations resulted in greater than 5 cases of influenza (0.1% of the population), meaning that only a proportion (62% and 59% for the 2016 and 2017 outbreaks respectively) of results were included for analysis.

The simulated seasonal outbreak and the reported historic data were similar. For scenario 1a (2016 model estimate with fear), the peak proportion of the population infected[5] differed by 0.003% compared to the reported historic data for 2016. The same difference was found between the reported and estimated standard deviation. Scenario 2a (2017 model estimate with fear) differed by 0.025% in terms of the peak proportion of the population infected, and 0.0004% in terms of the series' standard deviation.

Given the small population size of Canberra, influenza-related mortality is not available for each historic seasonal influenza outbreak.

[5] Peak proportion of the population infected refers to the maximum number of daily infections for the modelled wave, averaged across all eligible simulations.

Table 4.2. Model estimate and actual seasonal outbreak results

Scenario	% simulations resulting in an outbreak	Infected (avg %)			Mortality (%)			Fearful (avg %)		
		Max	Min	SD	Max	Min	SD	Max	Min	SD
Scenario 1a 2016 model estimate with fear	62	0.19	0	0.06	2.42	0	0.29	72.23	0	15.58
Scenario 1b 2016 model estimate without fear	67	0.19	0	0.07	1.22	0	0.22	-	-	-
Scenario 1c Actual (2016 outbreak)	-	0.19	0	0.06	-	-	-	-	-	-
Scenario 2a 2017 model estimate with fear	59	0.33	0	0.11	1.90	0	0.36	72.89	0.01	15.56
Scenario 2b 2017 model estimate without fear	60	0.65	0	0.20	2.21	0	0.52	-	-	-
Scenario 2c Actual (2017 outbreak)	-	0.30	0	0.11	-	-	-	-	-	-

Illustrations depicting the cross-over between the simulated and actual proportion of the population infected are provided in Figure 4.2. The solid lines detail the proportion of the population infected at each timestep (average for the model, total for actual) with corresponding 95% confidence intervals. The blue line represents laboratory confirmed influenza cases as a proportion of the population, compared to the model detailed in this essay, both with behavioural adjustment (transmission of fear, red line) and without behavioural

adjustment (no transmission of fear, green line). Overall, the model provides an accurate representation of what occurred in previous outbreaks.

In both scenarios, Figure 4.2 illustrates how the model estimates a greater proportion of the population being infected in the earlier days of the outbreak. The results in both panes of this figure demonstrate a more accurate estimation of disease spread in the model incorporating behavioural adjustment (red line). In the case of the 2017 seasonal influenza outbreak, the substantial difference between the model outputs without behavioural adjustment and the observed infections may be explained by the simulated individual protective behaviours. In other words, the incorporation of the fearful effect in this model are appropriate in trending the model outputs back towards the observed data.

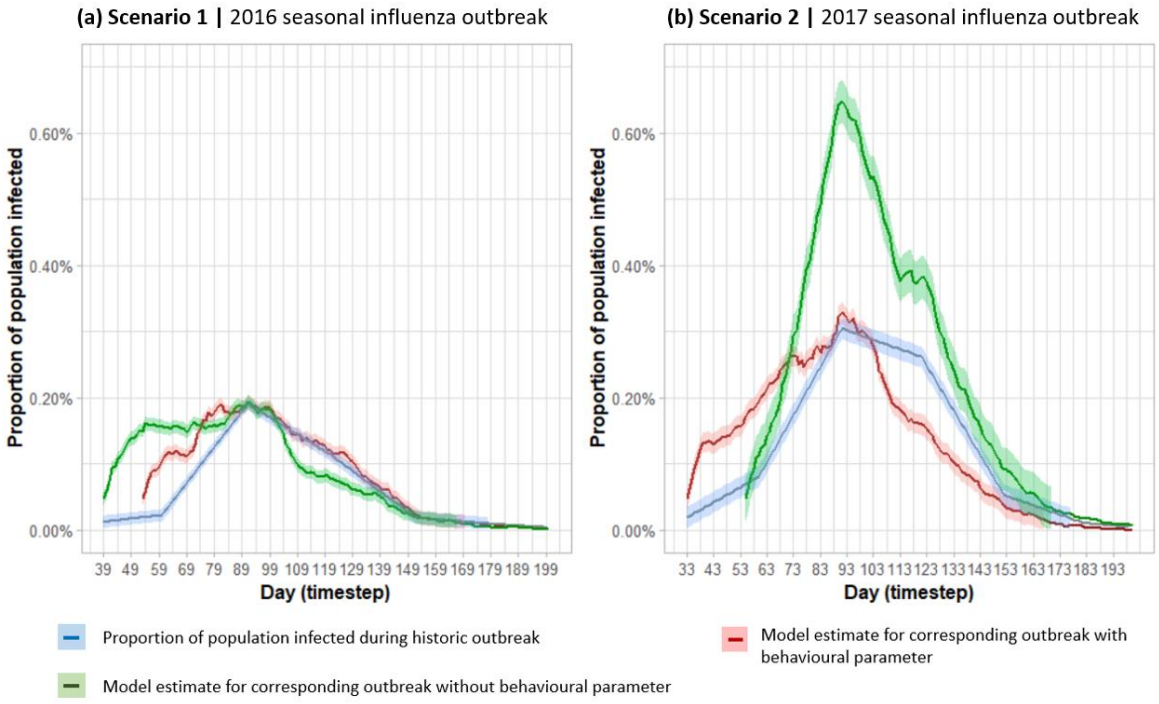


Figure 4.2. Proportion of population infected – simulated and actual

Table 4.3 shows that the model outputs are highly correlated with the historic seasonal outbreak data for the region at 86% and 81% for the 2016 and 2017 years respectively. In addition, the series with behavioural adjustment for fear performed better than the unadjusted scenarios across all other performance metrics (MSE, RMSE and MAPE).

Table 4.3. Model accuracy

Scenario	Difference in peak proportion of the population infected	MSE	RMSE	MAPE (%)
Scenario 1a (with fear) and scenario 1c (2016 observed cases)	0.003%	$6.1e^{-8}$	0.0002	24.8
Scenario 1b (without fear) and scenario 1c (2016 observed cases)	0.002%	$1.8e^{-7}$	0.0004	50.8
Scenario 2a (with fear) and scenario 2c (2017 observed cases)	0.025%	$1.1e^{-6}$	0.001	160.9
Scenario 2b (without fear) and scenario 2c (2017 observed cases)	0.34%	$3.6e^{-6}$	0.002	234.9

Given that the models incorporating behavioural parameters provide a more accurate representation of prior outbreaks, granger-causality testing was used to determine whether any causal relation exists between infection and fear, and vice versa. Note that both series are non-stationary.

Using a singular lag period, infection was found to granger-cause fear ($p < 0.001$), and fear was found to *not* granger-cause infection ($p > 0.1$). These results suggest that a fearful response is protective of further infection – this is likely due to the self-isolation parameter as detailed in the model structure, whereby agents do not attend work, school and other locations due to their fear of contracting influenza.

4.4 Discussion

This essay sought to further highlight the importance of individual behaviours as the primary factors of infectious disease spread, including how a better understanding of these factors can be used to inform pandemic planning and preparedness.

To do so, this essay built upon the foundational model for this thesis to establish a framework which can be used in the absence of behavioural indicators. The model described in this Chapter is a Susceptible – Infected – Recovered (SIR) model of influenza transmission in the Canberra region of Australia. By evaluating a more severe seasonal influenza season (2017) against a less severe seasonal influenza season (2016), this Chapter has demonstrated the value of ABMs in monitoring seasonal outbreaks to support planning and preparedness. In this model, agents were able to contract "fear" through engagement with others or through a community diffusion process enabled by a spatial scalar. This thesis argues that incorporation of this behavioural parameter – regardless of the quality of data available – should be considered in modelling for policy decision making.

This essay highlights the importance of incorporating behavioural parameters into infectious disease spread modelling methodologies. The essay developed a mechanism by which behavioural factors can be accounted for via use of spatially explicit Agent Based Model (ABM) modelling infrastructure, allowing for nuanced effects on movement and disease transmission to be modelled. Based on these more fine-grain details, it was found that the model with the behavioural parameter for fearful behaviour built in performed substantially better than the models that did not incorporate fear. This improved performance was based on visual performance, where the model with behavioural adjustment was much more closely aligned with the overall disease trajectory of the observed actual data, in addition to metrics such as the peak proportion of the population infected, correlation, MSE, RMSE and MAPE.

The mean absolute percentage error (MAPE) for the behaviourally adjusted scenarios with fear was significantly lower, reported at 24.8% and 160.9% compared to the non-adjusted scenarios without fear reported at 50.8% and 234.9% for the 2016 and 2017 seasonal outbreaks respectively. Granger-causality testing implies that the actual mechanism of the fear related behaviours is protective of infection with influenza.

In addition, this essay provides the foundation for defining an influenza-specific transmission rate for the Canberra, Australia region covering the 2016 and 2017 influenza seasons. An estimated transmission rate of 0.04 and 0.05 was calculated for the 2016 and 2017 seasons respectively, noting that the rate of contact was determined via a distance measure around each agent. By modelling the interaction between individuals and their environment, a more holistic understanding of disease transmission and social learning (Funk et al. 2010) is established.

This essay reinforces the findings of Chapter 3 (essay one) with regard to the importance of understanding the true dynamics of an outbreak. Per the review conducted by Funk et al. (2010), there has been significant attention placed on the impact of self-initiated actions on the spread of infectious disease. Individual responses, or fear related behaviours, have the potential to change the fate of an outbreak, and therefore need to be considered in tandem with disease dynamics. This essay has argued that a more nuanced view of behavioural response – characterised by time-varying contact and movement trends – is not only possible within an ABM framework, but critical for modelling epidemic outcomes. Secondly, by incorporating these behavioural parameters it is possible to plan the incremental effect of new policies. For example, recommended mask wearing use and social distancing, on disease trajectories, on peak timing and height.

In order to better understand and forecast the implications of previous or future outbreaks, it is critical that policy makers are equipped with the tools necessary to understand the underlying dynamics of infectious disease transmission. Once this is established, it would then be possible to better understand the flow-on socio-economic consequences of outbreaks to inform optimal intervention strategies. Future research in this area should seek to further refine the model via parameter fitting to observed data, and detailed representation of behavioural parameters within agent-based modelling and seek to better understand the underlying mechanism of fear and fear related behaviours.

5. Essay three: An econometric early warning system for signalling COVID-19 waves

5.1 Introduction

Modelling the estimated trajectory of disease spread is a critical component of emergency preparedness, necessary for estimating future likely outcomes in a complex environment. Chapters 2 and 3 (essays one and two) of this thesis discussed behaviour-focussed approaches to refining disease spread estimation. In this essay, I argue that the necessary step *prior to* implementation of disease spread models is to deploy robust early warning systems that enable evidence-based decisions ahead of time. Early warning systems have great potential to allow a lead-in time to enact policy change or influence behaviour, thus ultimately reducing the overall burden of a potential future wave on health outcomes.

Developing signal models for any unexpected or future event is difficult, and is even more complicated when faced with noisy, seasonally affected data, as observed in COVID-19 case data across the globe. Carefully designed early warning systems are most useful when determining the likelihood of a new wave of COVID-19 (or other novel communicable diseases) that may be attributable to a new variant. Often the full characterisation of new variants or diseases (including infectious periods, transmissibility and severity) takes months to compose, however early warning systems can detect changes in case volatility (as discussed in this essay) preceding a potential future wave.

Many of the bespoke data collections that captured individual level behaviours, including the Australian Bureau of Statistics COVID-19 surveys and the YouGov Imperial College of London surveys referenced in Chapter 3 are already, or are in the process of, being discontinued. Without such data to inform long term behavioural trends, signals must rely on information that is available for analysis to formulate an early warning system (EWS). This is why this essay examines the underlying properties of the data *in addition to* individual behavioural characteristics for endemic waves of COVID-19.

Under the framework of this overarching thesis, this essay answers the following question:

Novel and endemic infectious diseases continue to pose a threat to the population with ongoing waves of infection. How successful are simple econometric models of case velocity in establishing a fit-for-purpose early warning system to support policy decision making?

To answer this question, this essay develops and compares a range of GARCH variance series that are used to depict new COVID-19 case volatility, which in turn is transformed into an estimate of velocity to derive signals for emerging waves of COVID-19 ahead of time. The methods are then used to reinforce the concept of a self-stabilising theory of COVID-19 having both autoregressive and moving average components. This essay evaluates the GARCH, T-GARCH and E-GARCH models to capture velocity in new COVID-19 cases at the National and State and Territory level in Australia. For each model, the error measures of mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) are compared to derive an overall model rank for highest performance.

Introduced by Engle (1982), the Autoregressive Conditional Heteroskedasticity (ARCH) models were initially used to capture volatility, however these models were soon improved upon in the form of the generalised ARCH (GARCH) developed by Bollerslev (1987). Further extensions to develop the threshold GARCH (T-GARCH) (Glosten et al. 1993) and exponential GARCH (E-GARCH) (Nelson, 1991) have been instrumental in improving econometric understandings of volatility. While typically applied to the stock market, this essay seeks to apply similar foundations such that high performing volatility models are used to derive an estimation of velocity that can detect future spikes in COVID-19 cases.

The early warning systems developed in the literature vary wildly from very simple approaches, such as the standard deviation and Z score approach used in Reintjes et al. (2000), to the machine learning based least absolute shrinkage and selection operator models used in Shi et al. (2016), and the ensemble models used in Gao et al. (2020). Despite the complexity of the approach, one thing rings true for all EWS models: sensitivity to changes in disease reporting. This has continued to be the case with COVID-19 reporting, with the switch to weekly case counts as of September 2022. While this makes it difficult to produce early warning systems with the publicly available data, State and Territory Governments, who would still be receiving daily case count data, are able to deploy methods such as the one described in this essay to support pandemic preparedness and planning.

Other approaches to early warning systems that take a different data frame include papers on social media indicators (Google searches, tweets), including Zhang et al. (2022) who deployed natural language processing and linear regression to enhance COVID-19 surveillance in the United States. While social media approaches can be useful for understanding “hidden” effects, as in people who may not be getting tested (i.e. not showing up in the data) or are indeed

infectious and returning negative test results, the ability for individuals to publish factually incorrect information makes such indicators unreliable. Reintjes et al. (2000) noted three fundamental pillars for a successful EWS indicators: evidence-based, continuously collected, and are representative of the population. Particularly as COVID-19 becomes more normalised in society and culture, early warning systems based on social media are becoming increasingly less capable of predicting future waves of infection.

Wastewater early warning systems are another approach to the detection of COVID-19 or other infectious diseases in the population. In Australia, the various States and Territories undertake their own sewerage water analysis to detect viral shedding, providing a number of days to a few weeks warning of a potential outbreak of disease. A review conducted by Bibby et al. (2021) noted how much of the literature on wastewater early warning systems used retrospective analysis to inform their findings, meaning that critical elements of shedding time and clinical testing time have potentially been missed. The authors found that the maximum lead time a wastewater EWS could feasibly provide is four days' notice from disease onset. Further research by Kumar et al. (2021), investigating the correlation between COVID-19 cases and wastewater COVID-19 concentrations, found that this lead time can be expanded to up to 1 – 2 weeks. While this lead time does not provide sufficient time to implement an effective intervention strategy, wastewater analysis is useful during an outbreak to better understand disease under-reporting (Medema et al. 2020) or as a second or third level investigatory option should earlier early warning systems generate inconclusive results.

In this essay, an EWS is built upon new cases of COVID-19 in Australia, both Nationally and by individual State and Territory. It builds upon the disease spread model for Victoria described in Chapter 3 (essay one) by demonstrating the power of a signal that could ultimately influence positive behaviours, such as carefulness in the population ahead of peak disease spread periods, without having to resort to stringent policies or closing down parts of the economy. This is discussed further in Chapter 6 (essay four).

Table 5.1 details the descriptive statistics of new COVID-19 cases across Australia for the period 25 January 2020 to 9 September 2022 (959 observations). More recent data are not available due to the decision to publish weekly case estimates from September 2022 onwards. Without the publicly available data to inform the models, much of the integrity of the signals is lost. The summary statistics provided in this table illustrate the vastly different scenarios each State and Territory experienced over the observation period. Understanding the nature of

the curve of an infection wave, in addition to the more frequent waves experienced towards the end of the observation period, shows that all series are right skewed. Uniform leptokurtic distributions were observed, however for New South Wales and Victoria there was excess positive kurtosis with scores of 30.90 and 21.58 respectively.

Table 5.1. Descriptive statistics of new COVID-19 cases (daily) Nationally and by State and Territory

Statistic	National	NSW	VIC	QLD	WA	SA	NT	TAS	ACT
Mean	10,627	3,642	2,774	1,683	1,209	745	101	261	216
Median	147	18	41	2	1	1	0	0	0
Maximum	150,702	92,264	51,356	23,630	17,105	6,091	1,383	2,478	1,860
Skewness	2.28	3.96	3.47	2.61	2.71	1.73	2.61	2.01	1.64
Kurtosis	9.51	30.90	21.58	11.28	10.67	4.73	10.42	6.33	4.39

Note: NSW = New South Wales, VIC = Victoria, QLD = Queensland, WA = Western Australia, SA = South Australia, NT = Northern Territory, TAS = Tasmania, and ACT = Australian Capital Territory.

To evaluate the EWS developed in this essay, the ongoing pandemic has been split up into four time periods representing the unique waves of COVID-19 up until September 2022. Figure 5.1 illustrates the number of new cases across each time period at the National level:

- Period 1: 25 January 2020 to 30 April 2020
- Period 2: 1 May to 11 November 2020
- Period 3: 12 November 2020 to 14 March 2022
- Period 4: 15 March 2022 to 9 September 2022

State and Territory illustrations have not been included here for brevity.

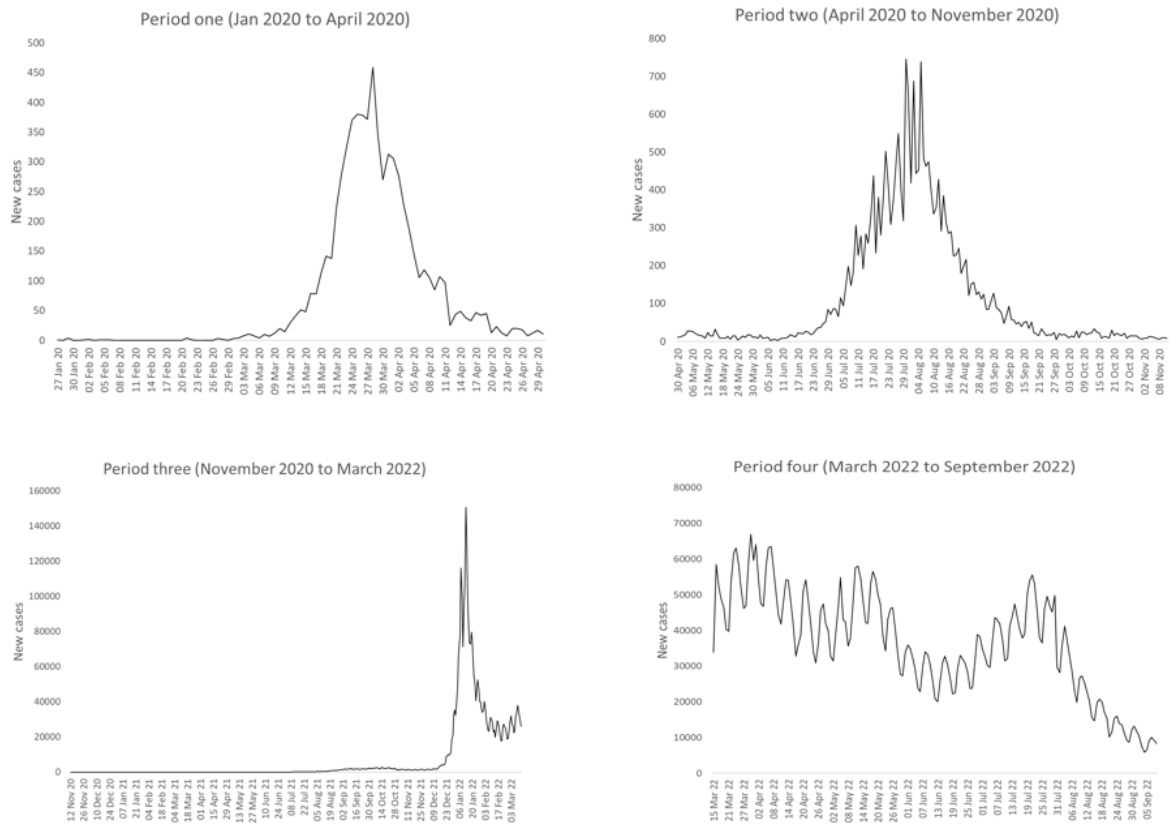


Figure 5.1. National level analysis periods – new cases of COVID-19

Note differing scales on y axis.

This essay is organised as follows. Section 5.2 explains the methods for the EWS. Section 5.3 summarises the forecasting performance of the selected signals prior to presenting the signals themselves. Section 5.4 provides a brief review of the results and discusses implications and conclusions for emergency preparedness and planning.

5.2 Methods

The EWS developed in this essay is a GARCH-variant variance series transformed into a signal indicator via a simple calculation of the variance series slope. This section describes the selected specifications for developing the signal, which are compared against each other using various forecasting evaluation techniques. This is followed by the method to calculate velocity of new cases. For reporting, the total number of cases for the entire Australian population are presented first, followed by State and Territory breakdowns.

5.2.1 Modelling volatility

Change in COVID-19 cases (including escalation or de-escalation of cases) is, in the simplest representation, a metric of volatility. Volatility is typically computed as a measure of variance in daily new cases (σ^2) and in this essay is represented as:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(c_i - \mu) \quad [17]$$

where μ and c are the mean number of new COVID-19 cases and the number of COVID-19 cases respectively. Number of observations is represented by *n*, in this case the number of days new cases were reported n= 959. Higher volatility in cases can therefore be indicated by a larger σ^2.

5.2.2 Specification of ARCH models

Three variants of the generalised autoregressive conditional heteroskedasticity (GARCH) model were compared in this essay to derive the EWS: GARCH, E-GARCH and T-GARCH. Further analysis to produce state-space quasi maximum likelihood estimates were also undertaken and can be provided upon request.

Generalised ARCH (GARCH)

Following Enders' (2014) account of the work undertaken by Bollerslev (1986), the GARCH model allows conditional variance to be an autoregressive moving average (ARMA) process. In accordance, the error term of the new daily COVID-19 cases series (ε_t) is referred to as:

$$\varepsilon_t = v_t\sqrt{h_t} \quad [18]$$

where v_t represents a white noise process and h_t represents the conditional variance of the error term of new daily COVID-19 cases ε_t. Letting the standard deviation of v_t equal to one ($\sigma_v^2 = 1$), and setting the conditional variance h_t as:

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \quad [19]$$

From this, it can be deduced that the conditional variance of ε_t is an ARMA process by first taking the expected value of the error term of new COVID-19 cases ε_t: $E\varepsilon_t = Ev_t(h_t)^{1/2} = 0$ and rearranging to arrive at the final representation of ε_t:

$$h_t = E\varepsilon_{t-1}\varepsilon_t^2 \qquad [20]$$

By allowing the conditional variance to be an ARMA process, the long-lagged effects of a shock with fewer parameters can therefore be evaluated (Hill, Griffiths and Lim. 2011). Here, the conditional variances of the disturbances in the $\{y_t\}$ series, in this case, new COVID-19 cases, acts like an ARMA process and therefore captures key features of COVID-19 transmission not covered by standard mean and variance analysis. GARCH models are unique as they allow for both autoregressive (past values used to predict future values) and the moving average components in the heteroskedastic variance (Enders, 2014).

Threshold GARCH (T-GARCH)

Building on the standard GARCH model, the threshold GARCH (T-GARCH) model allows for the effects both above and below a predefined threshold to have a differing impact on volatility. This is commonly used to account for the impact of "good news" and "bad news" on stock returns, so has limited influence in the context of COVID-19 cases, with case generation being non-sentient. Following Enders (2014), the T-GARCH is defined as:

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \lambda_1 d_{t-1}\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad [21]$$

where the term ε_{t-1} indicates the value of the dummy variable d_{t-1}. When ε_{t-1} is less than zero the dummy variable is equal to one, and when greater than or equal to zero the dummy variable is equal to zero. This allows for the threshold effect to be captured, and in this instance, allows negative shocks to have a stronger impact on volatility.

Exponential GARCH (E-GARCH)

Perhaps more suited to modelling new cases of COVID-19 is the exponential GARCH (E-GARCH) model. Proposed by Nelson (1991), the E-GARCH model is defined as:

$$\ln(h_t) = \alpha_0 + \alpha_1\left(\frac{\varepsilon_{t-1}}{h_{t-1}^{0.5}}\right) + \lambda_1\left|\frac{\varepsilon_{t-1}}{h_{t-1}^{0.5}}\right| + \beta_1 \ln(h_{t-1}) \qquad [22]$$

Given the log-linear form of the equation, coefficients are able to be negative as h_t is taken in log form. Furthermore, the size and persistence of shocks appears to be easier to interpret as ε_{t-1} is standardised to a unit free measure. Leverage effects can also be built in given the sign of the coefficient λ_1.

Note that the construction of the E-GARCH y value is different to that of the T-GARCH and GARCH specifications. This is a limitation of the analysis by which performance metrics (MSE, RMSE, MAPE) are not directly comparable across the specifications. For the purposes of this analysis, the E-GARCH model has been compared to ensure that a range of model options are available for comparison. This is discussed further in section 5.3.

Strengths and weaknesses of different GARCH models

From the above brief review of GARCH, T-GARCH and E-GARCH, this essay hypothesised that E-GARCH would be the best fit for modelling volatility in new COVID-19 cases. There are a number of weaknesses associated with the standard GARCH model, notably:

1. Excess kurtosis in forecasting volatility (assumption of a Gaussian distribution).
2. Requirement for symmetry in estimates resulting in biased predictions.
3. Requirement for all estimated coefficients to be positive.

The asymmetric specifications of T-GARCH and E-GARCH address point (2) by introducing the threshold indicator to allow for negative shocks (T-GARCH), and logarithmic transformations enabling the parameter restrictions on GARCH models to be relaxed (E-GARCH). E-GARCH also allows for coefficients to be negative, meaning that exponential effects are captured more effectively.

5.2.3 Forecast evaluation: variance series selection

In order to select the optimal GARCH variance series to generate the signal, this essay compares each model's accuracy in predicting case counts, standardised to a simple slope to enable comparison. This essay deploys the most widely used measures including, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Given that one model rarely performs best across all performance metrics, a ranking system has been developed to determine the overall best model for measuring volatility on COVID-19

cases. Forecast evaluation was only conducted at both the National and State and Territory level.

Mean Squared Errors (MSE)

MSE is one of the more important metrics for this study due to the tendency to penalise large forecast errors more than other accuracy measures. In the context of COVID-19 cases, it is important to select a model that avoids large errors. MSE is represented as:

$$MSE = \sum_t^n \frac{(y_t - \hat{y}_t)^2}{n} \quad [23]$$

where y_t represents the actual observed value and $\hat{y}_t$ represents the fitted value, estimated by the variance series, both at time t.

Root Mean Squared Errors (RMSE)

Taken as the square root of the MSE, RMSE measures results in the same units as the target variable, and is represented as:

$$RMSE = \sqrt{\sum_t^n \frac{(y_t - \hat{y}_t)^2}{n}} \quad [24]$$

Mean Absolute Percentage Error (MAPE)

Another metric for measuring prediction accuracy is MAPE, which expresses accuracy as a function of the absolute percentage error $\left|\left(\frac{y_t - \hat{y}_t}{y_t}\right)\right| \cdot 100$ on the fitted values:

$$MAPE = \sum_{t=1}^n \frac{\left|\left(\frac{y_t - \hat{y}_t}{y_t}\right)\right| \cdot 100}{n} \quad [25]$$

5.2.4 Velocity of new case volatility

Velocity, commonly measured as the speed of something in a particular direction, is particularly important for developing a COVID-19 EWS.

Taking the selected GARCH variance series for new case volatility, velocity (V) is measured as:

$$V = \frac{(\sigma_t^2 - \sigma_{t-1}^2) - (\sigma_{t-1}^2 - \sigma_{t-2}^2)}{|(\sigma_{t-1}^2 - \sigma_{t-2}^2)|} \quad [26]$$

where (σ^2) measures new case volatility at time *t*. In terms of simple rise/run metrics, here the rise is being measured in terms of growth (velocity) over time, as by using daily data that are consistent over the entire time period, the run is equal to one. This estimated daily change is what is referred to as the velocity of new case volatility.

This volatility metric is intended to measure underlying human behaviours that are implicit in the change in new COVID-19 cases. When there is a large degree of behavioural variance, this essay hypothesises that a change to the volatility (GARCH variance series) of new COVID-19 cases would become evident.

Signals are compared against observed COVID-19 cases over the defined windows of analysis. In some instances, "appropriate" signals are generated at an inappropriate time. For example, a signal may be generated over six months in advance of the actual peak of the wave. These "fake" signals are referred to in this essay as ineligible signals, as they are too unreliable to be considered seriously for policy implementation over such a sustained period of time, for a wave that is potentially no threat to the target population. In such circumstances, it is recommended that policies for second or third-order investigation (such as wastewater analysis or deployment of an alternative EWS) be implemented to determine the most logical course of action. All analysis was conducted in Eviews software. Data and estimation outputs can be made available upon request.

5.3 Results

The results presented in this section cover the selection of the variance series, leading to National and State and Territory splits of volatility estimation and signal generation.

5.3.1 Forecasting evaluation

Table 5.2 details the MSE, RMSE and MAPE for each model and time period for National level data. A ranking process has been used as one model rarely performs the best across all

three metrics. The "final rank" is the sum of all other measurement ranks. A lower score implies the best performance.

Table 5.2. National model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	**GARCH**	0.1306	1	0.3615	1	32.4497	2	4
	T-GARCH	0.1306	2	0.3615	2	32.4497	1	5
	E-GARCH	0.2644	3	0.5142	3	32.9406	3	8
2	GARCH	0.1561	2	0.3951	2	54.2805	2	6
	T-GARCH	0.1561	3	0.3951	3	54.2805	3	9
	E-GARCH	0.0872	1	0.2954	1	35.8286	1	3
3	GARCH	0.3218	2	0.5673	2	158.6813	2	6
	T-GARCH	0.3218	3	0.5673	3	158.6813	3	9
	E-GARCH	0.1955	1	0.4422	1	103.7668	1	3
4	GARCH	0.0132	2	0.1148	2	82.0434	2	6
	T-GARCH	0.0132	3	0.1148	3	82.0434	3	9
	E-GARCH	0.0085	1	0.0919	1	46.2719	1	3

As expected, E-GARCH outperforms T-GARCH and GARCH, with the exception of period 1, which as illustrated previously in Figure 5.1 has substantially fewer cases on the curve, resulting in a much less "exponential-like" nature to the wave. Period 1 was also characterised by importation of cases from international sources prior to Australian border closures, which would account for the shape of this wave at the National level. Note again that due to the y-value of E-GARCH being different to T-GARCH and GARCH, these models are not directly comparable across the selected performance metrics. Nevertheless, the T-GARCH and GARCH specifications performed in a highly similar fashion, again referencing earlier comments that the threshold effect would not hold much power in a disease spread context. As such, it is important to include the E-GARCH model in this analysis to ensure the best estimate of new case volatility. The difference between the point estimates is so precise that it cannot be shown within the demonstrated four decimal points. As such, the GARCH model has been selected for developing the variance series in period 1, where the E-GARCH model was selected for the remaining time periods at the National level. Appendix A contains evaluation data for each State and Territory. A summary of the selected parameters is listed in Table 5.3 below.

Table 5.3. State and Territory model evaluation results

Period	NSW	VIC	QLD	WA	SA	NT	TAS	ACT
1	T-GARCH	T-GARCH	GARCH	T-GARCH	E-GARCH	E-GARCH	E-GARCH	E-GARCH
2	E-GARCH	E-GARCH	GARCH	T-GARCH	T-GARCH	E-GARCH	E-GARCH	T-GARCH
3	E-GARCH	T-GARCH	E-GARCH	T-GARCH	E-GARCH	GARCH	E-GARCH	T-GARCH
4	T-GARCH	E-GARCH	E-GARCH	T-GARCH	E-GARCH	GARCH	GARCH	T-GARCH

5.3.2 Variance series

The selected variance series (as a representation of volatility) are illustrated in Figure 5.2.

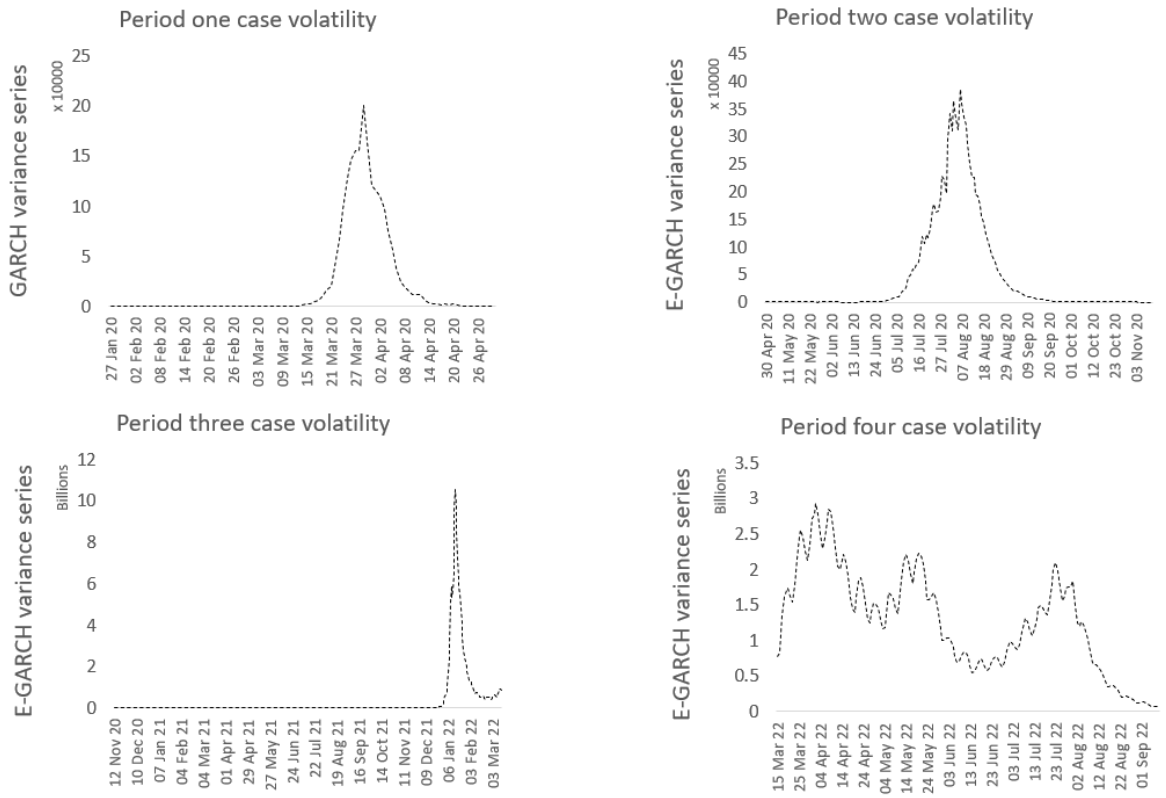


Figure 5.2. New cases variance series (volatility) Australia

5.3.3 National level signals

Applying the methodology described in section 5.2, Figure 5.3 illustrates the estimated signal (velocity of new cases measured as per the calculation provided in equation [26] and illustrated as the red series) for the first period of analysis, representative of the first wave of COVID-19 experienced in Australia. An exceptionally strong signal is observed approximately 36 days prior to the actual peak of the wave, or approximately 22 days prior to daily new cases

exceeding 50 cases per day. This strong signal can be attributed to the GARCH variance series rapidly changing from very small levels of variance to substantially larger levels of variance in the space of a week (see Figure 5.2).

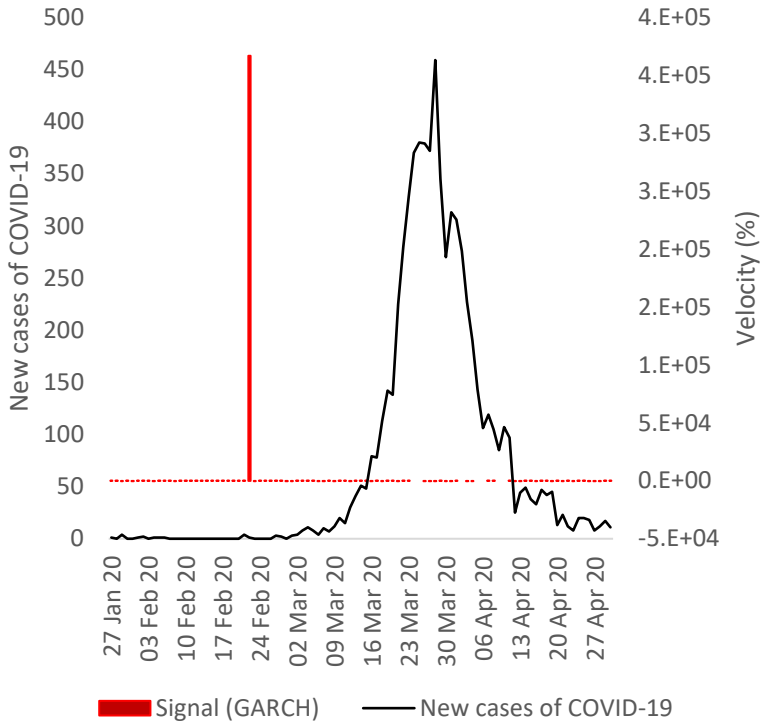


Figure 5.3. Period 1 new cases of COVID-19 and estimated signal (National)

Period 2 (Figure 5.4), now leveraging the E-GARCH signal, saw a larger wave of new cases of COVID-19 than period 1, and instead of a strong, singular signal, saw a series of four signals ahead of the peak of the wave. The first of these four signals occurred approximately 48 days prior to the peak of the wave, or 16 days prior to daily cases exceeding 50. In this figure there is also a clear negative signal correlating with the wave's gradual conclusion. One of the primary benefits of the E-GARCH model is the ability to generate negative coefficients by taking the logarithm of h_t. While useful in signalling that the number of new cases per day are indeed going down, this "negative signal" occurs after cases are already declining.

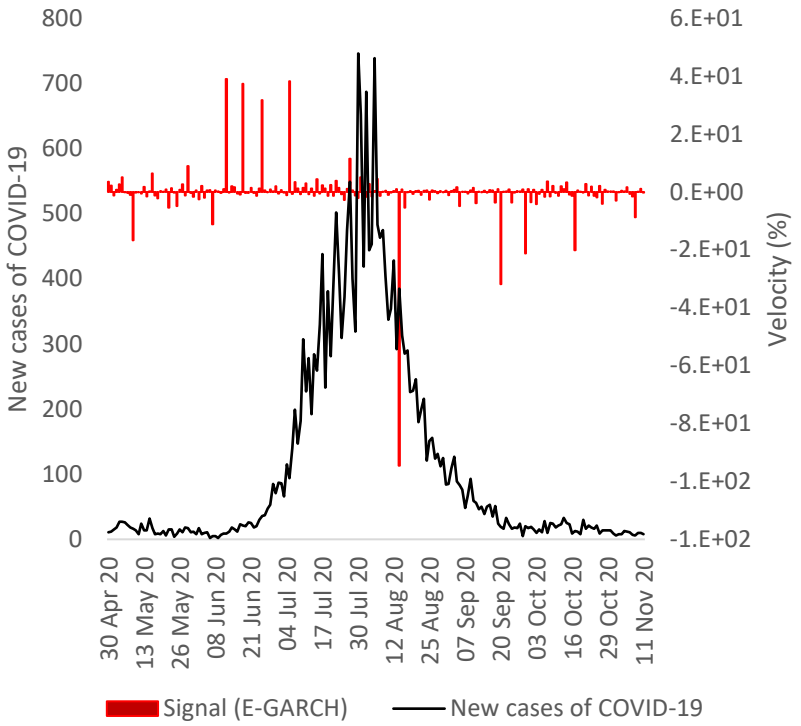


Figure 5.4. Period 2 new cases of COVID-19 and estimated signal (National)

Period 3 (Figure 5.5) saw the biggest wave of COVID-19 observed in Australia over the defined period. The signal in Figure 5.5 is hard to ignore – occurring up to 6 months ahead of the true peak of that wave. This was driven by the Omicron variant of COVID-19, where the preceding peak in September to October 2021 was attributable to the pre-existing Delta variant. Importantly, this wave of Delta infections had been developing prior to the detected signal. Coupled with a partial lockdown announced in Sydney at the end of June 2021, this signal does not qualify as a true signal of the upcoming wave, given that it was simply trigged by an increase in incidence from a baseline of zero cases to mid-June 2021.

Zooming in further on this time period, Figure 5.6 illustrates a strong signal 57 days prior to the peak of the wave, when cases were already at approximately 1,800 new cases per day nationally. We can split these signals up into two labels – signal A for the strong signal in July 2021 (Figure 5.5) and signal B for the strong signal in November 2021 (Figure 5.6). Signal A is reflective of a strong uptick in projected cases, signalling an upcoming wave when only 116 new cases per day were being reported at the time. By then, when signal B is flagged, this suggests that the wave is going to continue to worsen, with the lead time of 57 days to dampen the curve. Both signals are again confounded by several factors, including the Delta wave, increasing levels of population vaccination, and the nation-wide relaxation of restrictions in

December 2021. Given the changing environment at the time of this analysis, weighted consideration should be placed on the validity of these signals. This figure demonstrates the interchange from the Delta to the Omicron wave, and particularly for signal B (Figure 5.6), demonstrates how the uptick in cases that was measured was unrelated (from a variant perspective) to the oncoming Omicron wave.

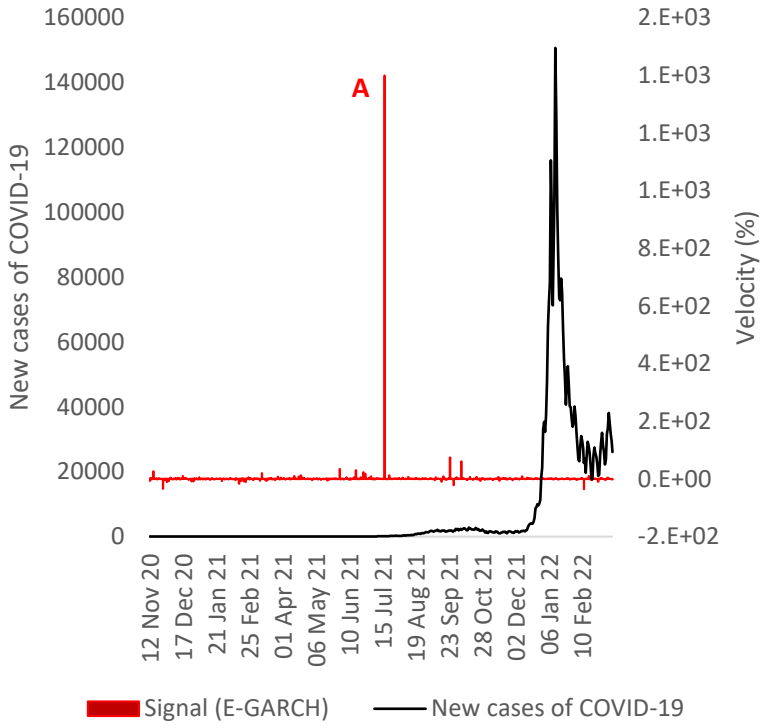


Figure 5.5. Period 3 new cases of COVID-19 and estimated signal (National)

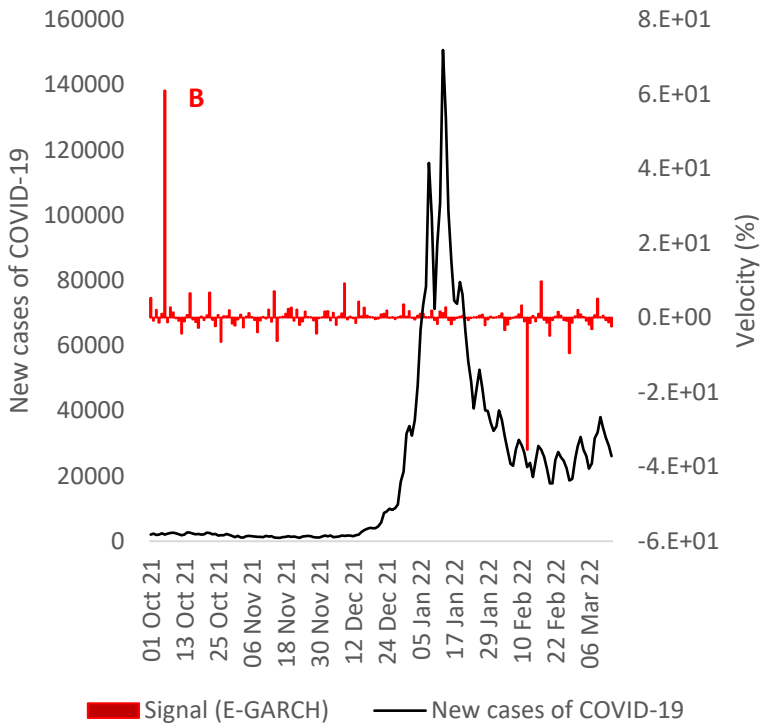


Figure 5.6. Truncated period 3 new cases of COVID-19 and estimated signal (National)

The final period (period 4, Figure 5.7) was specifically chosen to determine how the signal works in a less defined (i.e. no specific wave to evaluate) environment. Figure 5.7 reinforces that the signals are not ideal for predicting downturns, often signalling a downturn when the reduction in cases is already occurring. Small lead times to "pulsating waves" (i.e. waves that have no zero beginning or end point) provides a 35-day lead-in effect to the observed peak of the wave in July 2022. Overall, it can be concluded that the signals are less useful outside of a defined wave period.

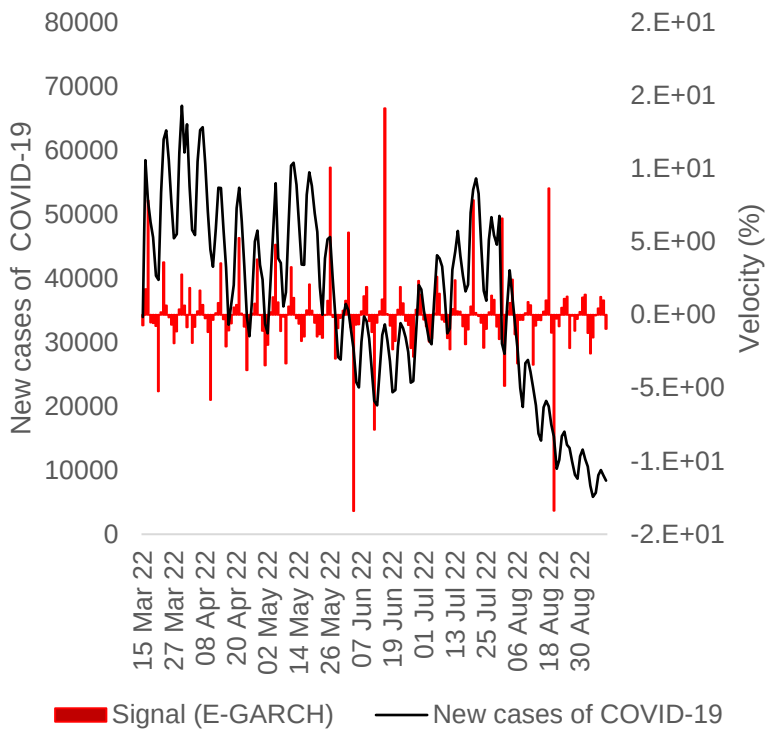


Figure 5.7 Period 4 new cases of COVID-19 and estimated signal (National)

5.3.4 State and Territory signals

The above analysis was repeated for each individual State and Territory to determine the applicability of the signal generation process at smaller geographic levels. For each State and Territory, findings suggest:

- **New South Wales (Figure 5.8):** Strong signal ahead of waves in both period 1 and period 3. Period 3 signals performed much more accurately than at the National level, given the weaker signal as the wave was commencing, followed by a much stronger signal as the wave continued to escalate. Noting again, this signal was likely confounded by the interchange from the Delta to the Omicron wave. The case data in period 2 was much noisier than the National aggregate figures, however a signal can still be identified. For period 4 the signal appears to be non-functional.
- **Victoria (Figure 5.9):** Strong signal ahead of waves in both period 1 and period 3. The signal in period 3 was several months ahead of the actual wave, so for the purposes of policy implementation is considered non-functional, particularly given that this signal occurred during a period where the only new cases of COVID-19 were imported from

New South Wales in December 2021 and January 2022. This is therefore a "fake" signal as it is unreasonable to expect behaviour change for such a sustained period of time. In this instance, secondary methods of investigation (such as wastewater analysis) would be recommended. For period 2, there were a number of signals in the lead-up to the peak of the wave, one of which signalled the origination of new cases in hotel quarantine, which had seeded into the broader community by May 2020. This signal will be evaluated in further detail in essay four. Finally, signals in period 4 performed better than the National analyses, indicating the downturn over July and August 2022 ahead of time.

- **Queensland (Figure 5.10):** Strong signal ahead of the wave in period 1. Noisy data was observed in period 2, accompanied by low case counts. Period 3 was characterised by a negative contraindicating signal, followed by weaker positive signals. Period 4 demonstrated overall reactionary signals to pre-existing up or down trends.
- **Western Australia (Figure 5.11)**: Strong signal ahead of the waves in period 1 and three. There was a very short time period between the signal and the wave in period 2, noting that very low case counts were being observed in Western Australia at that time. Period 4 signalled the upcoming downtrend in cases ahead of the peak of the wave.
- **South Australia (Figure 5.12)**: Multiple signals ahead of wave in period 1, despite relatively low case counts. Extremely low case counts in period 2. Strong signal ahead of the larger wave in period 3, several months in advance hence making this another non-functional, "fake" signal as it was followed by months of near-zero transmission of COVID-19 in the population and was likely to have been triggered by the opening of borders in the state over that period of time. Negative signal in period 4 ahead of the gradual decline in cases over the second half of 2022.
- **Tasmania (Figure 5.13)**: Very low case counts until period 3, where a positive signal is observed ahead of the wave. This signal is also likely to be reflective of border openings during this time window. In period 4, a negative signal was observed ahead of the general downturn in cases. Despite low case counts, a negative signal was observed in period 1.
- **Australian Capital Territory (Figure 5.14)**: Positive signal ahead of wave in period 1. Low case counts in period 3 where a positive signal ahead of the wave was detected, again several months in advance so ineligible for policy action. Given the low case counts in the Territory, it is plausible that this signal was the culmination of two

consecutive days of notified positive cases in November 2021. A signal was also detected ahead of the wave pulse in period 4.

- **Northern Territory (Figure 5.15)**: Low case counts until period 3. Several signals ahead of the wave in period 3 over the course of several months which may be attributed to the cluster of cases in Katherine in November 2021. In period 4, negative signals contraindicated the uptick in cases observed over July and August 2022.

Accompanying figures listed on the following pages.

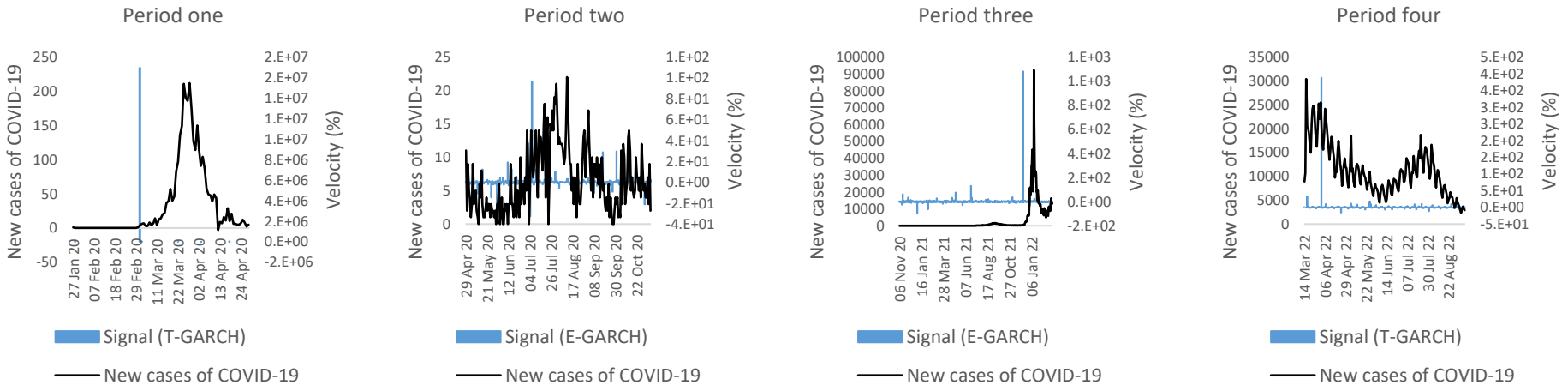


Figure 5.8. Periods 1 – 4 new cases of COVID-19 and estimated signal New South Wales

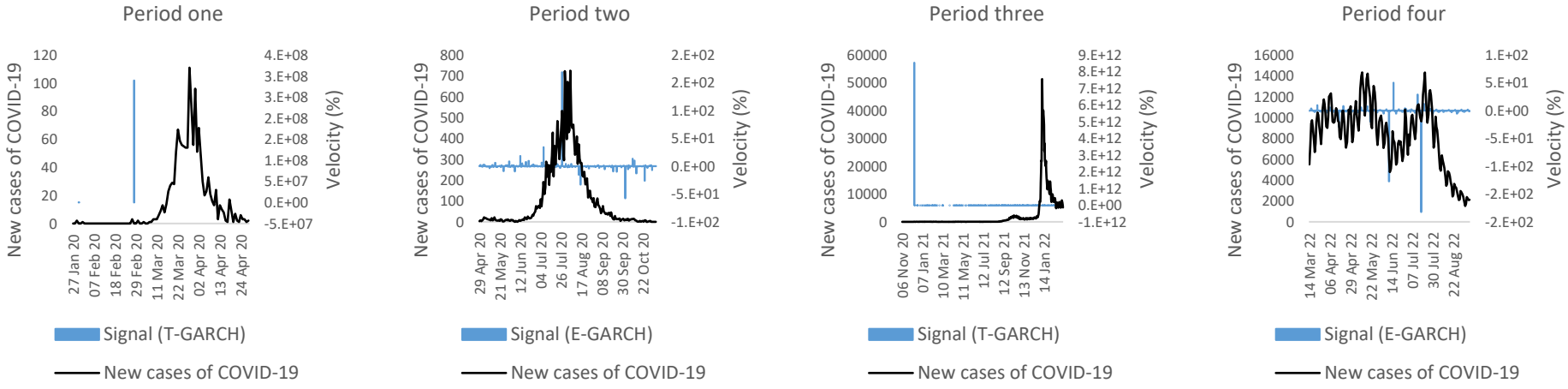


Figure 5.9. Periods 1 – 4 new cases of COVID-19 and estimated signal Victoria

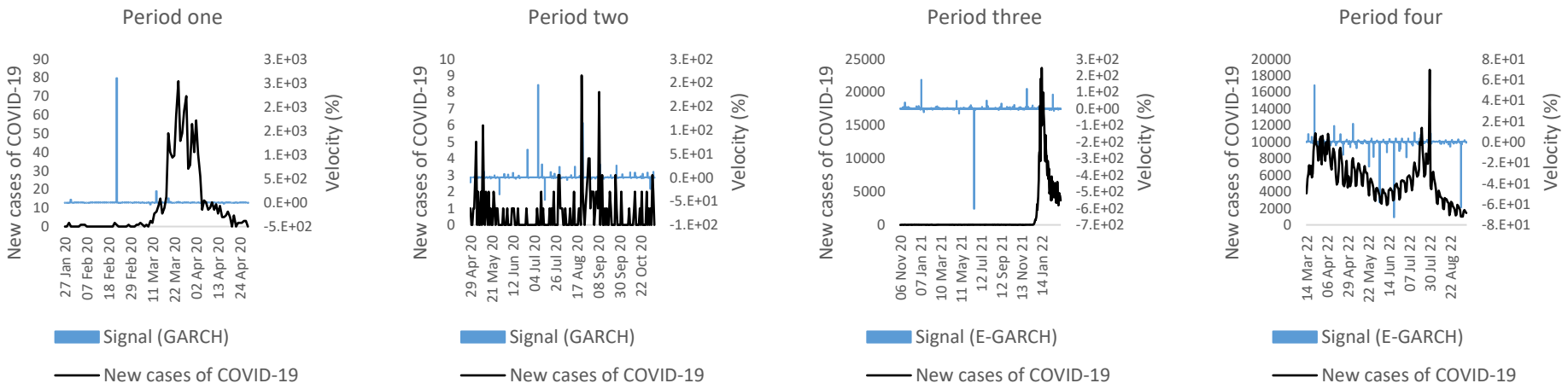


Figure 5.10. Periods 1 – 4 new cases of COVID-19 and estimated signal Queensland

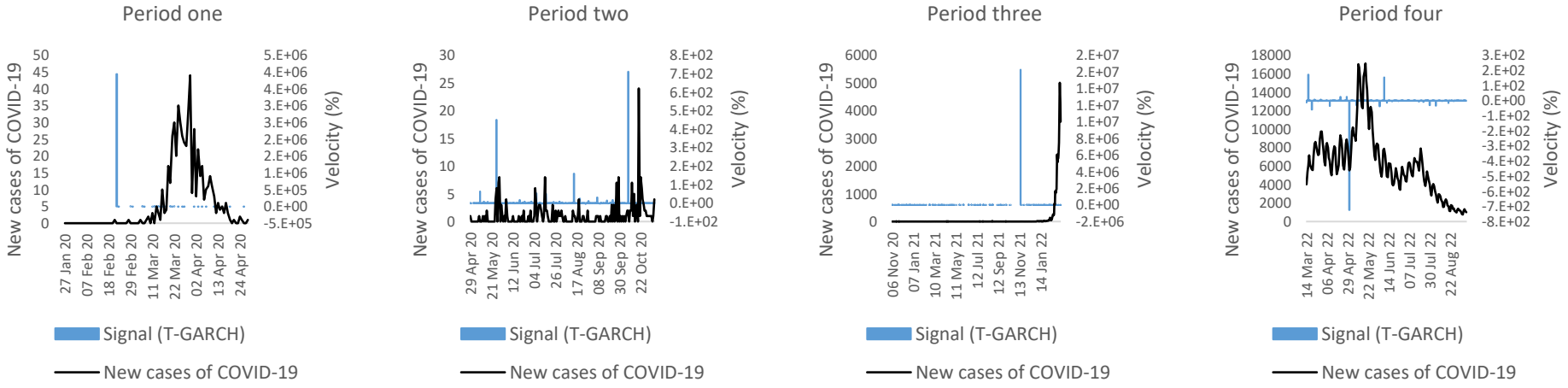


Figure 5.11. Periods 1 – 4 new cases of COVID-19 and estimated signal Western Australia

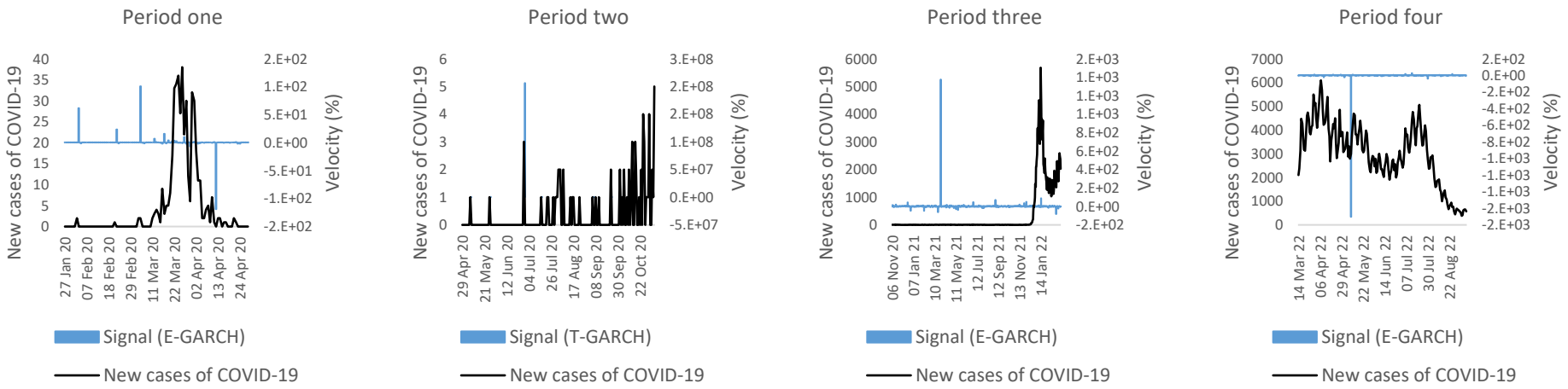


Figure 5.12. Periods 1 – 4 new cases of COVID-19 and estimated signal South Australia

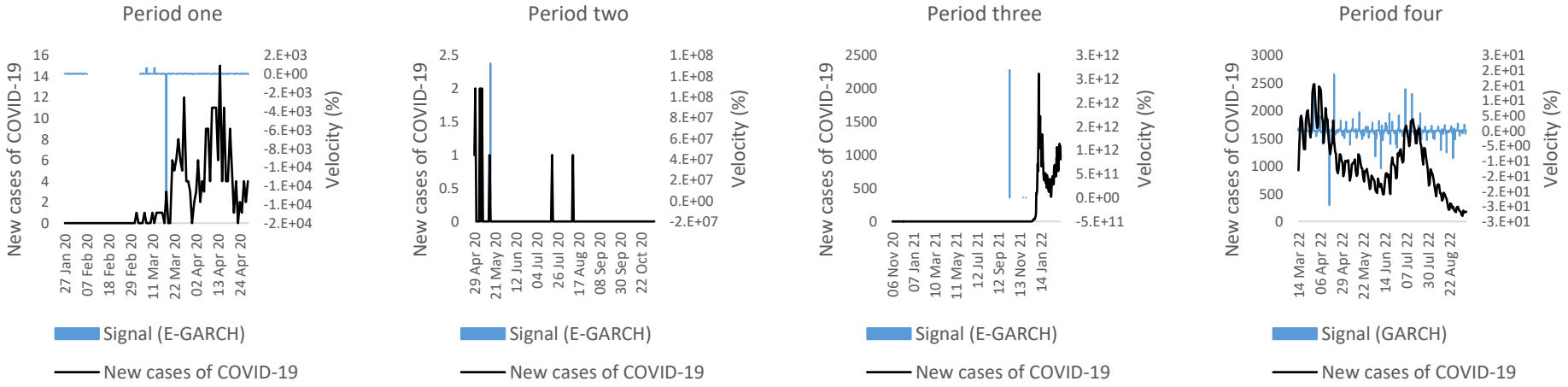


Figure 5.13. Periods 1 – 4 new cases of COVID-19 and estimated signal Tasmania

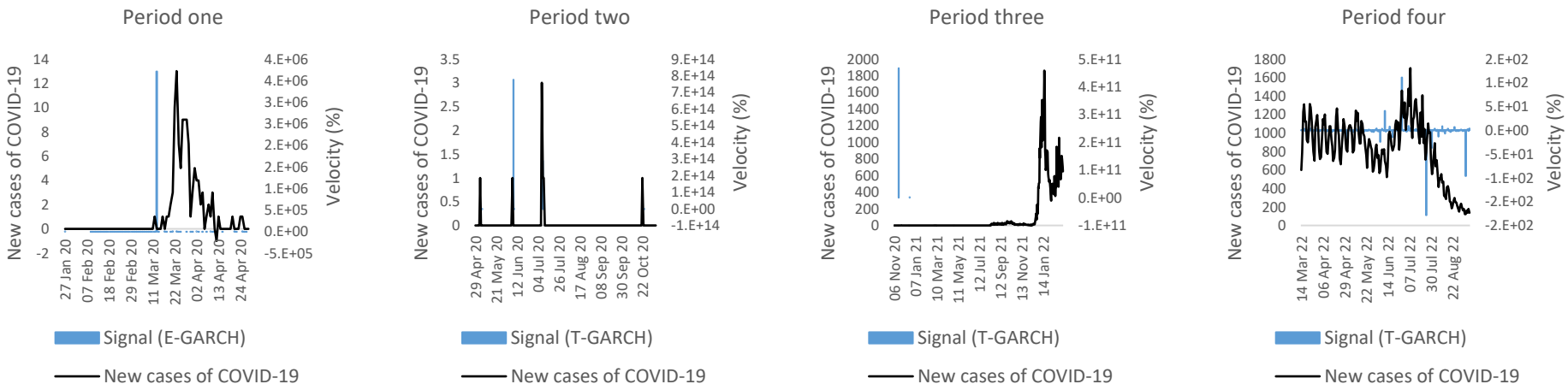


Figure 5.14. Periods 1 – 4 new cases of COVID-19 and estimated signal Australian Capital Territory

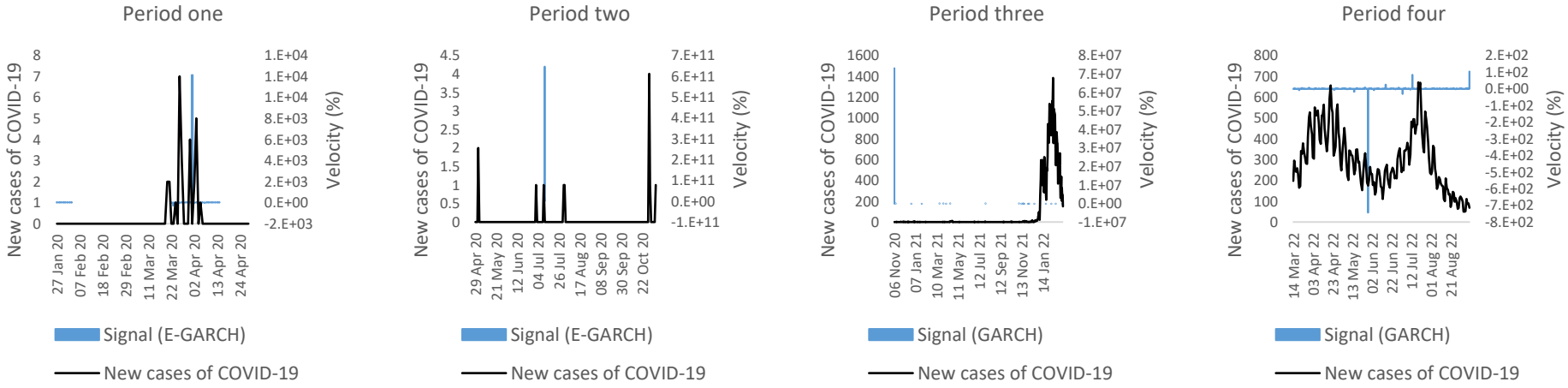


Figure 5.15. Periods 1 – 4 new cases of COVID-19 and estimated signal Northern Territory

5.3.5 Performance metrics

Tables 5.4, 5.5 and 5.6 provide National and State and Territory breakdowns of the accuracy of the selected signal in predicting future waves of infection. Metrics for evaluation include:

- Whether the signal occurred prior to the peak of the wave:
    - Time between signal and wave peak (days).
    - The signal was accurate and timely, or a "fake" signal that is ineligible for policy intervention.
    - Signal occurs greater than 100 days or less than 20 days ahead of the peak of the wave.
- Whether the signal occurred prior to >= 50 new cases per day:
    - Time between signal and ~ 50 new cases per day.

Due to the amount of noise in period 4, this time period has been excluded from performance analysis.

Table 5.4 illustrates that the EWS was successful in signalling future waves in a timely manner for periods 1 and 2 of this analysis. For period 3, the first signal (signal A) was an ineligible signal, being so early from the peak of the third period wave. Nevertheless, the follow-up signal (signal B) was both timely and accurate at the National level, noting that it was likely confounded by the interchange from the Delta to the Omicron wave. This suggests that caution and common sense must be applied when evaluating signals during periods of variant interchange or low disease transmission, as per the sustained period of low cases featured in the period 3 observation window. Such caution should also be applied in complex environmental circumstances, for example changes in the dominant circulating variant of disease, vaccination rates and mandated interventions. As previously discussed, in such contexts it would be prudent to have second and third level investigatory options, such as wastewater analysis, for complex circumstances or when case counts are low.

Table 5.4. Signal performance metrics (National)

Period	Signal occurred prior to peak?	Time between signal and peak (days)	Signal occurred prior to >= 50 cases/day?	Time between signal and threshold (days)
1	Yes	36	Yes	22
2	Yes	48	Yes	16
3	Ineligible – signal A	180	Yes – signal B	57

Table 5.5 details the prior-to-peak performance metrics by State and Territory. Given the additional level of geographic granularity, some States and Territories were unable to be considered for a signal given the low number of cases during the relevant period of analysis. Out of a total 18 signals generated, 9 of these signals (50%) were considered to be timely and accurate (*eligible* scenarios) given the selection criteria detailed above.

Table 5.5. Signal performance metrics – prior-to-peak (State and Territory)

Period	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
1	Y (23 days) *Eligible*	Y (29 days) *Eligible*	Y (31 days) *Eligible*	Y (37 days) *Eligible*	Y (21 days) *Eligible*	N* *Ineligible*	Y (10 days) *Ineligible*	LCC
2	Y (37 days) *Eligible*	Y (48 days) *Eligible*	LCC	LCC	LCC	N* *Ineligible*	LCC	LCC
3	Y (34 days) *Eligible*	N* *Ineligible LCT*	Y (46 days) *Eligible*	N* *Ineligible*	N* *Ineligible LCT*	N *Ineligible LCT*	N* *Ineligible*	N* *Ineligible*

LCC = Low case count. Signal is unable to be reliably generated when new cases are low.

LCT = Low Community Transmission. These signals are likely to be the result of imported cases or border re-openings, as they have occurred during periods of low to no community transmission.

N* = Greater than 100 days between estimated signal and wave. It is unrealistic to expect policy or protective behaviours to be sustained for longer than 100 days.

In Table 5.6 these results are further examined regarding which signals were generated prior to 50 new cases of COVID-19 per day. In this scenario, there were six signals across the three time periods that successfully indicated a future wave of infection prior to 50 new cases per day. This is an important metric, however in an environment of ongoing low-level COVID-19 transmission across Australia (for example, at the time of writing the number of new cases per

week nationally are in the 1,000 range) this metric is less important than what is detailed in Table 5.5.

Table 5.6. Signal performance metrics – prior to >=50 cases/day (State and Territory)

Period	NSW	VIC	QLD	WA	SA	TAS	ACT	NT
1	Y (16 days)	Y (22 days)	Y (26 days)	NC	NC	-	NC	-
2	NC	Y (18 days)	-	-	-	-	-	-
3	NC	-	Y (16 days)	-	-	-	-	-

NC = New cases do not reach >= 50

5.4 Discussion

This essay has developed and deployed an econometric early warning system (EWS) in the form of a case velocity signal to support policy decision making. Analysis followed comparison of a number of GARCH indicators, frequently used in the financial literature to model volatility and change over time. The E-GARCH model performed best across most time periods of analysis and produced insightful signals for pre-warning upcoming COVID-19 waves of infection. The signals were found to perform best at the National level, where more data was available to generate the signals. This was found to be a significant limitation of the EWS, particularly for States/Territories that had very low case counts or low community transmission, such as the Australian Capital Territory, Northern Territory and Tasmania. Another finding was that while the signal is successful in predicting upcoming waves, it is not suitable for predicting downturns in cases, often signalling the downwards trend after cases had already begun to fall.

GARCH-based early warning systems are a suitable and practical tool to deploy as an EWS in the COVID context. Particularly when used at the National level, such signals can greatly benefit government and individuals to employ preventive policies or behaviours that can ultimately dampen the curve and minimise loss and harm from the spread of disease. As seen in this essay, the EWS can be used at lower geographic levels (State or Territory), however is most useful when the region of focus is experiencing many cases. Many (if not most) models have difficulty predicting outcomes when small case numbers are involved, or in complex environmental circumstances including changes to circulating variants, vaccination rates and

mandated interventions. Early warning systems, such as the one described in this essay, represent low effort high reward outcomes for policymakers.

In a post-COVID restrictions world, the Australian population is unlikely to respond well to mandated restrictions on movement or activities in future. As such, the critical lead-in times developed by the EWS in this essay have the potential to influence individual behaviour ahead of a potential future wave. In practice, should a signal emerge that COVID cases are beginning to increase rapidly, National or State governments could then actively promote protective behaviours such as mask wearing, hand washing and social distancing, to limit the spread of disease. This would be particularly pertinent should a new, more lethal or transmissible variant of COVID-19 be introduced to Australia. It also applies to any new diseases that have equal or greater potential to harm the population.

Building upon Reintjes et al. (2000) and their three pillars of a successful EWS, this model provides a systematic, evidence-based indicator that can be readily deployed to the daily COVID-19 new case data continuously collected by Government for the entirety of the Australian population. As COVID-19 becomes more normalised, it is anticipated that hidden (undetected, not reported) cases may distort early warning systems such as the one described in this essay. To overcome this problem, there needs to be continuous investment in up-to-date simulation models, such as the one described in essay one, to ensure not only that estimates of disease trajectory are continuously updated, but also to provide additional evidence to EWS models should any observational data be lacking.

Finally, in light of reduced testing (both PCR and RAT) for COVID-19, it is recommended that analysis to develop an EWS using weekly hospital data (from 9 September 2022 to 10 November 2023) is undertaken. The intention of this analysis would be to determine whether the aforementioned EWS modelling framework would be suitable for signalling future waves based on the number of hospitalised cases. Thus, future research should seek to tailor these models to hospital admissions to have a broader spectrum of tools available. Future research in this field should look to further enhance the models developed in this essay and seek to better understand the drivers for ineligible signals that are generated either too early or too late for effective policy intervention, in addition to establishing a precedent for activation of second and third tier investigatory measures (such as wastewater analysis). Further applications should also look to apply this framework to other infectious diseases such as influenza, tuberculosis and malaria.

6. Essay four: Embedding a pandemic pulse into preparedness capability: model outputs and policy implications

6.1 Introduction

Pandemic preparedness is a multifaceted concept. It covers adequate vaccination stock (Newland et al. 2021), treatment availability (Jennings et al. 2008), workforce protection and support (Kidd, 2020), and planning. This essay focuses on the importance of planning to be able to implement pre-defined, evidence-based policies to mitigate the risk of harm or mortality from disease.

In September 2020 the Independent Panel for Pandemic Preparedness and Response was formed. Their report, *COVID-19: make it the last pandemic* was submitted to the World Health Assembly in May 2021. This report contains recommendations for action that will reduce the likelihood of another pandemic, like COVID-19, occurring again. Within this report, the authors call for:

> "surveillance and alert systems at national, regional and global levels to be redesigned, bringing together their detection functions – picking up signals of potential outbreaks – and their relay functions – ensuring that signals are verified and acted upon. Both must be able to function at near-instantaneous speed… it will also require a commitment to open data principles as the foundation of a system that can adapt and correct itself."

As evidenced by the COVID-19 pandemic, there is still work to be done with respect to automated surveillance systems for infectious disease (not restricted to COVID-19). In this essay I demonstrate the potential impact of early intervention on COVID-19 disease spread in Victoria, Australia, should an EWS, such as the one described in Chapter 5 (essay three), have been deployed to nudge individual behaviour. It is important to note that this essay looks at the role of community engagement (Lal et al. 2020) in promoting careful behaviours that ultimately dampen case counts and improve health outcomes. In other words, the dissemination of health information can reduce the prevalence of infection if it results in individuals seeking earlier treatment or engaging in preventative action (Kiss et al. 2010). Reduced cases are modelled using the microsimulation model defined in Chapter 3 (essay one), with varied input parameters to adjust for the increased community engagement.

The policies implemented throughout the COVID-19 pandemic are well documented. As discussed in Chapter 2, policies include mask mandates, social distancing, venue capacity limits and lockdowns. The effectiveness of these policies has also been well documented – an

Australian study found that facemasks demonstrated protective efficacy in excess of 80% for clinical influenza (MacIntyre et al. 2008). In addition, Adjodah et al. (2021) demonstrate a statistically significant decrease in new cases, deaths and hospital admissions associated with mask mandates up to 40 days after the introduction of the mandated policy. The authors note, through analysis of a novel survey, that community attitudes towards masks are associated with reduced COVID-19 cases and deaths. This is important as while the global population further "normalises" the concept of COVID-19, these community attitudes separate from government messaging will be critical to managing the spread of disease.

Timing and adherence to behavioural interventions play a critical role in the success of interventions on disease spread outcomes. A study by Chang et al. (2020) found that a 3-day delay in the implementation of intervention measures resulted in an average of 23.56 additional days required to maintain the intervention to achieve the desired state of disease suppression. A further study by Chang et al. (2023) noted that it is difficult to assess the likelihood of a population to adopt personal protective behaviours, such as social distancing, particularly when pandemic fatigue is considered in the individual opportunity cost calculation. In this study, the authors developed a dynamic model of social distancing adherence, which further highlighted the propensity of the population to vary their adherence levels over time. As such, this essay considers various proportions of the population likely to enter the "careful" behavioural compartment, representative of varying levels of communication campaigns targeting improved individual protective behaviours ahead of a future wave of infection.

Under the overarching framework of this thesis, this essay answers the question of:

How can an integrated model of human behavioural states and early wave detection be deployed in practice to reduce harm and mortality incurred from the spread of disease?

To do so, this thesis applies a novel application of an EWS signal coupled with a multi-layered behaviour focused ABM. This supports the existing literature by quantifying the effectiveness of proactive policies, such as mask wearing and social distancing recommendations, which could easily be scaled up or down during future waves of COVID-19, or similar infectious disease waves. In other words, this essay explores how data-driven strategies can be dovetailed with effective public policy to empower individuals with timely, accurate knowledge to proactively reduce the spread of disease.

This essay is organised as follows. Section 6.2 explains the method behind the model and the adjustments made vis à vis the model presented in Chapter 3 (essay one). Section 6.3 presents

the main results from running the microsimulation model, followed by section 6.4 discussing the implications of deploying early warning systems in future pandemics, and the consequent ability of governments and individuals to respond to infectious disease threats.

6.2 Methods

The model deployed in this essay is the spatially explicit Agent Based Model (ABM) defined in Chapter 3 (essay one) of this thesis. This model is a Susceptible – Exposed – Infected – Recovered – Susceptible (SEIRS) model representative of the Victorian population in Australia run over 6,649 agents (a 0.1% sample). This model is overlaid with estimates of the population reported to be fearful, careful or complacent in response to COVID-19, which is then probabilistically matched to the representative synthetic population. Per previous essays, individuals are referred to as 'agents' in the model framework. As discussed in Chapter 2 (literature review) and Chapter 3 (essay one), due to the sheer computational power required to run large ABMs, smaller fractions of the population must be simulated upon to generate results in a timely manner. Larger population fractions (greater than 10% sample) can be simulated with this model framework, time permitting. The model contains the option for parallel processing which can be constructed across simulations, however due to the time-dependent nature of each timestep (in other words, the outcomes of today are dependent on the interactions of previous days) the efficiency of the algorithm is still contained to the length of one full simulation run.

In this essay, the ABM generates scenario outputs with adjustments made ahead of time using the early warning signal developed for Victoria in Chapter 5 (essay 3).

6.2.1 State variables and scales

A high-level summary of the modules and processes for this model is provided in Figure 6.1. The model operates in a very similar fashion to what is described in Chapter 3 (essay one), however contains the addition of the wave signal developed in Chapter 5 (essay three). As disease spreads through person-to-person contact in a spatial environment, the model detects when a signal has been activated for an upcoming increase in cases indicative of a large wave. From here, the government can choose to take action or no action in response to the signal, based on the available data and evidence. If action is taken, then the model increases the proportion of the population that is "careful" in response to the threat of disease.

To capture an upper and lower bound of the impact of this behaviour adjustment, four scenarios are investigated:

- Scenario one: No behaviour change
- For the duration of the signal, boosting the proportion of the population that is careful by:
    - Scenario two: 10%
    - Scenario three: 20%
    - Scenario four: 30%

Referring back to Chapter 3 (essay one), the transmission probability for careful agents (0.031) is substantially lower than for complacent agents (0.338), and as such the overall transmission potential is hypothesised to be reduced under the scenario conditions. Given the proportion of the population that is complacent is often less than the scenario four change (30% boost) to the careful population, this scenario represents the impact of having zero complacent individuals on overall disease spread. Fearful agents have not been adjusted in the scenario analysis as this type of behavioural response is often highly irrational or driven by unalterable personal conditions (such as being immunocompromised or high risk). In other words, these agents are unlikely to change their position following government intervention. In addition, as diseases such as COVID-19 become more normalised, the likelihood of an intervention to *induce* fearful behaviour is substantially reduced.

When the proportion of the population that is careful is increased, this is taken directly from the proportion of the population that is complacent, until such point that complacent agents bottom out to zero, which occurs at various points throughout the time series of each scenario. Given the unique emotional state often associated with fear, the proportion of the population that is fearful of COVID-19 does not change. In other words, the action taken increases the number of careful agents and decreases the number of complacent agents. The process to increase the number of careful agents could take a number of forms under the remit of community engagement, such as encouragement to wear masks or socially distance, and be adjusted based on the severity of the signal's prediction of cases. This style of approach is sustainable and empowers the population to make the best possible decision with the available information.

The provision of greater information to the general public reinforces the distinction between endogenous (voluntary) and exogenous (forced) individual protective behaviours. Leveraging

existing research highlights the role of perception severity and susceptibility in motivation to act (Witte et al. 1998), so increasing the amount of information individuals process has the potential to impact behavioural adjustments, regardless of whether the threat is perceived (i.e. the individual is not infected yet fearful of becoming infected) or realised (i.e. the individual is infected). The consequence of these behaviours flows on to reduced activity and, thus, a slower spread of infection, as demonstrated in previous research (Perra et al. 2011).

Following this behaviour adjustment, agents then go on to transmit disease as per the original model described in essay one, with the differential being the increased number of careful agents reducing the overall transmission of disease by virtue of their lowered transmission rate.

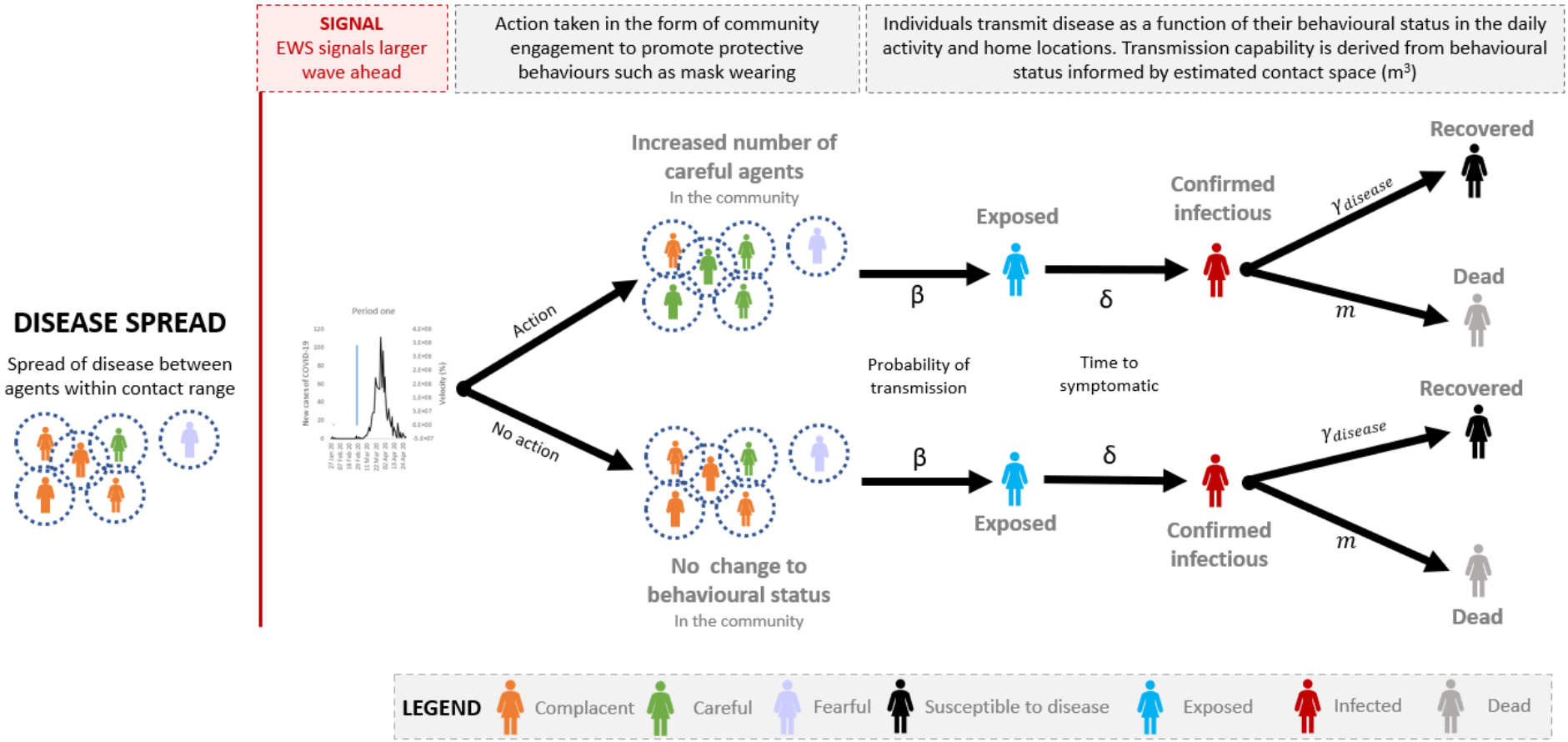


Figure 6.1. High level overview of model

6.2.2 Parameters

The full list of input parameters used to inform the model described in this essay is provided in Table 6.1. Parameters included in this model are all publicly available, sourced from Australian statistical collection agencies and academic institutions for the region of Victoria, Australia. Note that due to the time at which this model was built, some parameters may appear to be outdated. These parameters would be updated prior to use in the current setting, however inputs current at the time of writing this essay have been included in this table.

Table 6.1. Model parameters

Parameter	Value	Source
Model cycle		
Cycle length	1 day	Input
Cycle duration	75 days	Input
Signal duration	51 days	Essay 2
Population parameters		
Population size	6,649 people (0.1% sample)	ABS (2021)
Land size	227.82 km^2 (axes = 15.09 km)	Geoscience Australia (2001)
Proportion of population in each age category*	Young (0 – 19) = 24.0% Adult (20 – 65) = 60.5% Old (65+) = 15.4%	ABS (2021)
Movement parameters		
Contact radius	Baseline: 3m radius	Input
Mean distance moved by age class (adjusted to reflect map size and population density)	Young (0 – 19) = 0.01 km Adult (20 – 65) = 0.02 km Old (65+) = 0.01 km	ABS (2018)
Standard deviation distance moved by age class	Young (0 – 19) = 0.02 km Adult (20 – 65) = 0.02 km Old (65+) = 0.02 km	ABS (2018)
Disease parameters		
Initial infected	2	Input
Probability of infection if a contact is made (β)	Fearful = 0.023 Careful = 0.031 Complacent = 0.338	Calculation – see *scheduling 3*
Exposure – duration (days) to develop symptoms (δ)	$1/7$: midpoint of estimated 0 to 14 days to develop symptoms	HealthDirect (2020)
Probability of recovery (γ)	$1/14$: most people will recover between 2 to 3 weeks from onset of symptoms	HealthDirect (2020)
Maximum life of infection	8 weeks	Constraint*
Clinical case fatality rate (m)	Young (0 – 19) = 0.0% Adult (20 – 64) = 0.1% Old (65+) = 17.0%	ABS (2020)
Behavioural parameters		
Initial population proportion – dynamically adjusted based on EWS signals	Fearful = 36% Careful = 58% Complacent = 6%	Jones (2020)

6.2.3 Process overview, restrictions and scheduling

This section provides an overview of the processes, restrictions and schedules deployed in the microsimulation model presented in this essay.

Overview and parameter restrictions

The model was run for 650 simulations of a 201-day period with 75-day intervention period. This time period was designed to capture the signal period for Victoria as defined in essay three. The signal for the selected wave (period 2) runs from 25 April 2020 to 30 July 2020. The date 25 April was selected as it was exactly 20 days post the signal detected on 5 April 2020. This allows the government time, under a pre-existing pandemic management plan, to implement policy measures and communicate these measures to the public. The proportion of careful agents (10%, 20% and 30% boost scenarios) are increased over this signal period to determine the overall impact of the changed behavioural composition on disease spread outcomes.

A visual breakdown of these dates, overlaid with actual new daily COVID-19 cases in Victoria, is presented in Figure 6.2. As discussed in Chapter 5 (essay three), this signal was generated from a series of cases in hotel quarantine, which later seeded into the community in May 2020. Importantly, this signal alerts governments and decision makers of the change in case velocity such that interventions could be established and made ready for deployment in May, once the disease had progressed out to the broader community.

Following Kohler et al. (2000), the model took the average of all simulations where the cumulative minimum of 0.15% of the population infected, and no restriction on the cumulative maximum of the population infected.

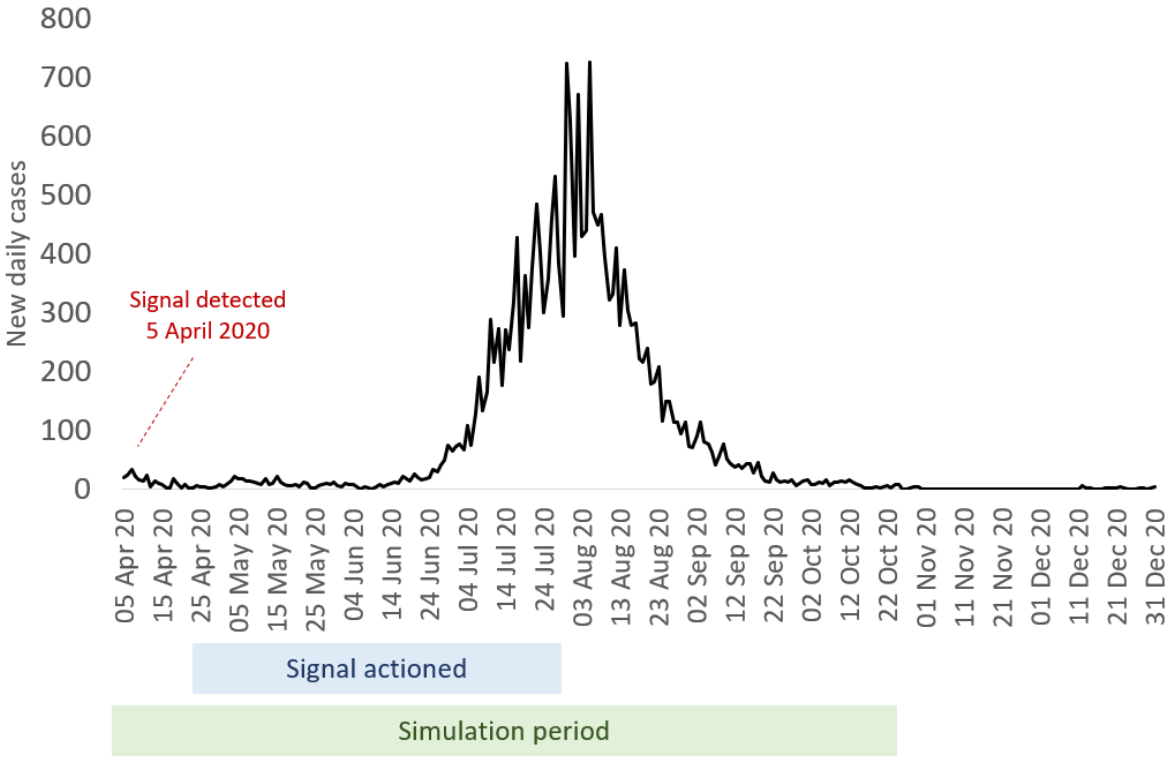


Figure 6.2 Signal detection, action and simulation dates against new daily COVID-19 cases in Victoria

Scheduling

The scheduling process for this ABM follows that as described in Chapter 3 (essay one) in section 3.2.3. For reference, the scheduling steps include:

1. Assign behavioural status to synthetic population
2. Daily movement
3. Contact area and disease transmission
4. Exposure and case confirmation
5. Recovery and reinfection
6. Clinical mortality from COVID-19

Data analysis and simulation modelling was undertaken using R programming software run in parallel within the Australian National University National Computing Infrastructure (Gadi) environment.

6.2.4 Evaluation and sensitivity testing

Evaluation

Each scenario was compared to the base model (no behavioural adjustment), which is reflective of the model presented in Chapter 3 (essay one), however over a shorter period. Differences were measured in terms of total infections across the duration of the simulation period. Finally, the impact of the signal on case reduction is quantified for each scenario using an average estimate of the change in cumulative cases. Reflective of the "power" of the scenario on reducing cumulative infections compared to the baseline, this calculation follows the ensuing simple methodology:

$$Impact = 1 - \frac{x-y}{x} \quad [27]$$

where *x* represents the total cumulative cases in the baseline scenario and *y* represents the total cumulative cases in the modelled scenario.

Sensitivity testing

The scenarios included in this essay serve as a sensitivity test for the model against varying uptake in careful behaviours. Table 6.2 summarises these scenarios.

Table 6.2. Sensitivity analysis

Description	Scenarios	Source
Increased proportion of the population that is careful, reduced proportion of the population that is complacent	1. Boost careful proportion of the population by up to 10%, simultaneously reducing complacent proportion of the population by up to 10% 2. Boost careful proportion of the population by up to 20%, simultaneously reducing the complacent proportion of the population by up to 20% 3. Boost careful proportion of the population by up to 30%, simultaneously reducing the complacent proportion of the population by up to 30%	650 runs of the microsimulation model

6.3 Results

Results from the model described in this essay are presented below. The results focus on the difference in reported cases for each scenario compared to the baseline model presented in Chapter 3 (essay one), albeit for a shorter period.

Figure 6.3 illustrates the average proportion of the population in the careful and complacent compartments across all simulations for each scenario. As prefaced in section 6.2.1, the proportion of the population in the complacent compartment trends towards a "bottoming out" of complacent agents in scenario four which is due to the magnitude of the booster applied to the careful population. In other words, the results from scenario four (30% boost to the careful compartment) would be interpreted as the "reduce complacency to zero" scenario by policy makers. Alternatively, scenario two and three present oscillating complacency, with scenario two being the most realistic representation of an early reduction in complacent agents, following a successful policy intervention, followed by low levels of complacency prior to the end of the intervention period.

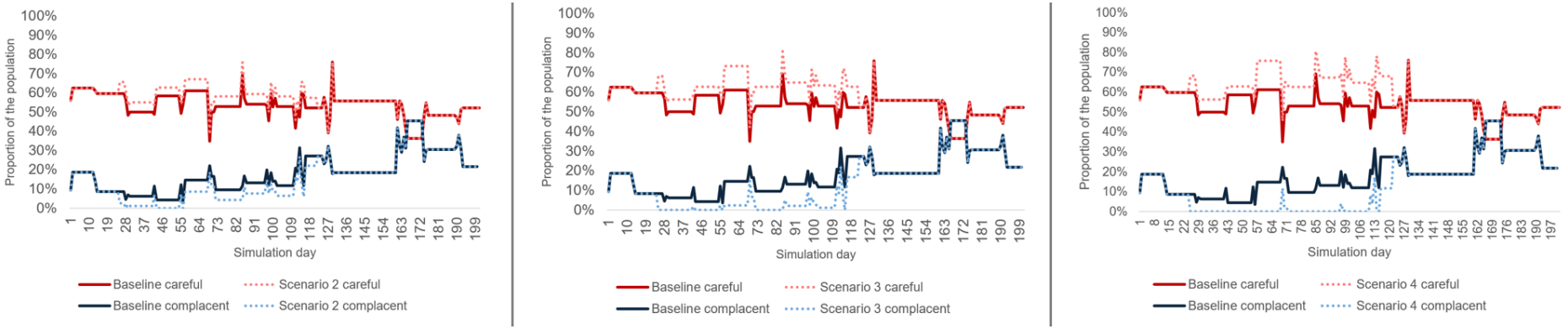


Figure 6.3 Average proportion of the population in careful and complacent compartments for scenario 2 (10% boost to careful population), scenario 3 (20% boost to careful population) and scenario 4 (30% boost to careful population)

Table 6.3 summarises the cumulative infections and peak number of cases per day by scenario. Due to the short time period of analysis (looking 51 days ahead), some scenarios did not result in concave distributions. These figures should be considered in light of the proportion of total simulations that met the eligibility criteria of minimum 0.15% population infected, implemented to ensure that non-starter simulations do not bias the model outputs (see Table 6.4).

Table 6.3. Cumulative infections by scenario

Scenario	Cumulative infections	Number of infections at peak of curve
Scenario one (baseline)	269	88
Scenario two (10% careful boost)	182	57
Scenario three (20% careful boost)	192	45
Scenario four (30% careful boost)	138	33

Table 6.4. Eligible simulations by scenario

Eligibility criteria	Scenario one	Scenario two	Scenario three	Scenario four
Total simulations meeting eligibility criteria (%) out of total 650 runs	70.5	71.7	68.8	70.2

6.3.1 Infection curves

Figure 6.4 illustrates total (exposed and confirmed) infections per day under each scenario. In this figure, the EWS signal (beginning on day 25 of the simulation and ending on day 121) is highlighted to demonstrate the time horizon of the increased proportion of the population in the careful behavioural compartment. In scenario three and four (more so in scenario four) the shape of the disease curve becomes increasingly flat, implying a stronger degree of control over the rate of disease spread. While considered to be unrealistic, scenario four represents the impact on caseloads should the number of complacent agents be reduced to zero for the

duration of the intervention period. This result is important for decision making as it suggests the lower bound to which cases may be reduced to under varying levels of policy intervention success.

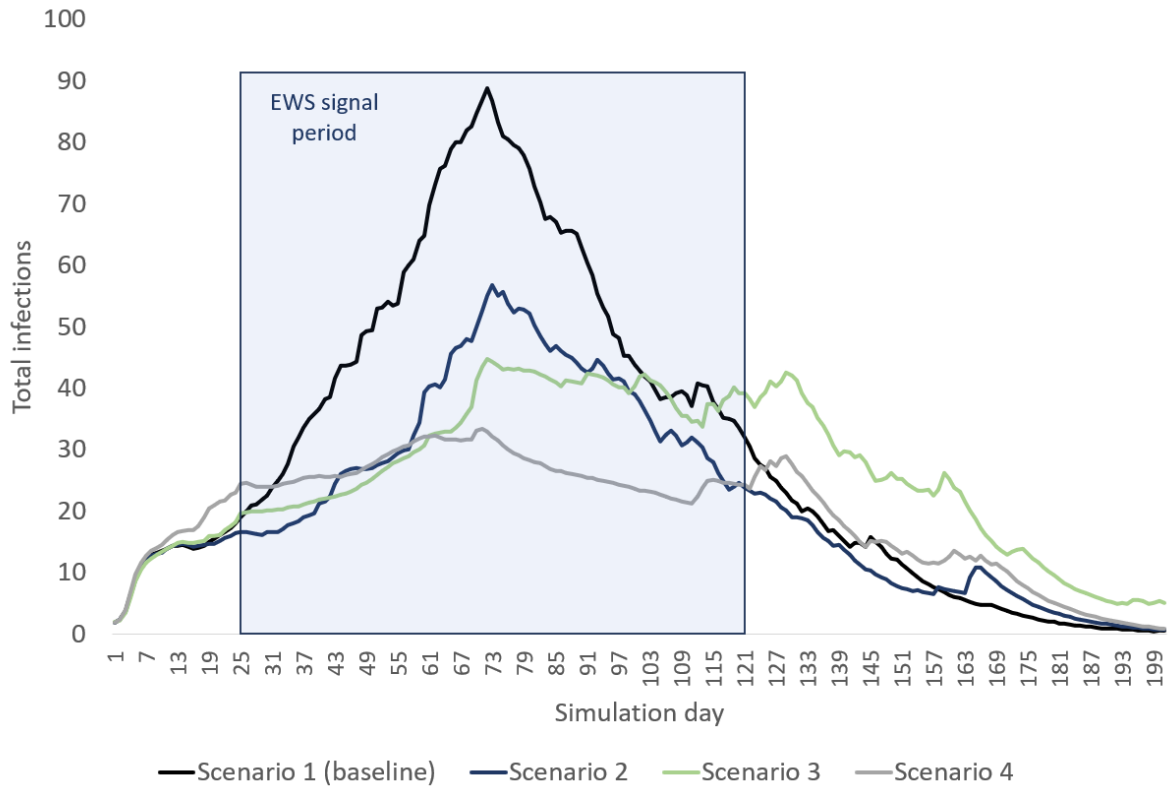


Figure 6.4. Total infections by scenario

6.3.2 Impact on cases due to signal intervention

Table 6.5 lists the impact calculation for each scenario, representing the overall dampening effect on modelled cumulative case counts. For example, the scenario two ('boost careful' by 10% scenario) has an impact calculation of 0.68, representing an averaged 68% of the baseline cumulative infections under this scenario. In addition to raw case counts, metrics such as this are useful tools to support decision making.

Table 6.5. Impact due to signal intervention on cumulative case counts

Scenario	Impact calculation
Scenario one (baseline)	N/A
Scenario two (10% careful boost)	0.68
Scenario three (20% careful boost)	0.71
Scenario four (30% careful boost)	0.51

The results in Table 6.5 suggest that all boosts to the proportion of the population that is careful yield a reduction in cumulative infections, as expected. Importantly, the marginal benefit of increasing the proportion of the population that is careful is negative when comparing scenario two and three. The impact of boosting the careful population by 30% (scenario four) has the strongest influence on reducing cumulative cases, where cases are expected to be 51% of baseline cumulative infections. These results must be viewed with consideration of the realities of being able to develop and successfully implement a public communications campaign for reducing the spread of disease. As previously discussed, scenario two is the most realistic scenario, which based on this analysis would yield significant benefits to reducing the spread of disease.

6.4 Policy implications

A data-driven approach to monitoring and managing infectious diseases is something that is not only possible in the current environment, but also necessary to ensure that countries across the globe are adequately prepared for the next crisis.

Over past decades, responses to severe influenza seasons have been largely reactive, and could have been managed more efficiently had a simple early warning system been put in place. Then, for diseases or disease variants of concern, the ability to monitor and predict the future spread of disease using data-driven foundations – as described in this essay – can be used to further tailor and nuance responses that limit the potential morbidity and mortality associated with disease related outcomes.

This essay is strongly aligned with the objectives of the World Health Organization's *Strategic Preparedness and Response Plan: April 2023 – April 2025* (World Health Organization, 2023c). This Strategic Preparedness and Response Plan (the plan) specifically states that:

"there is an opportunity to strengthen the public health foundation for future epidemic and pandemic response efforts",

and that

"…we cannot become complacent and fall victim to the cycle of panic and neglect."

With specific reference to the plan's goal of calibrating the response to COVID-19, this essay has presented fit-for-purpose solutions to embedding surveillance, via an early warning system, into business-as-usual management and monitoring operations. This essay has also discussed methods by which data-driven solutions can be used to adjust public health and social measures that support long term planning and preparedness of novel infectious disease monitoring.

Such data-driven approaches will change the way in which government decisions are made during periods of crisis or intense public scrutiny. It is the role of government to operate in a manner that serves the public interest, however, to date those decisions have not had an embedded data-driven foundation from which evidence must be generated. Embedding foundational practices such as those described in this essay will require new processes by which public servants and elected officials translate information into outcomes. The beginnings of this influence were seen throughout the COVID-19 pandemic. Nevertheless, due to the immaturity of the current systems for disease monitoring and management, including linkages between National and State and Territory data assets, Australia is now in the position to embed these lessons learnt and establish a foundation for disease management that will best serve the public interest for the next outbreak.

This essay has focused on the role of individual behaviours as a key driver of disease outcomes. By leveraging the Imperial College of London YouGov COVID-19 behaviour tracker (Jones, 2020), this essay was able to construct a behaviour-driven model of disease transmission. In future, the national statistics agency – the Australian Bureau of Statistics – should look to embed behaviour surveys into their regular rotation, on a timely basis, such that trends in fearfulness, carefulness and complacency can be baselined and then calibrated against changing perceptions and behaviours during periods of crisis. An alternative to conducting the

highly detailed survey conducted by Jones (2020) would be to emulate the survey established by Seale et al. (2020) which has already been successfully executed in the Australian context. This application would extend beyond pandemics, to help government target decision making and interventions during other events such as natural disasters or global events.

It is important to not dismiss the importance of on-the-ground knowledge, experience and policy wisdom when establishing a framework such as the one described in this essay. No realistic scenario would see data-driven tools automate decisions relating to disease preparedness and planning. Such tools can, however, be used to *support* decision making and expand the amount of information available to design appropriate and effective interventions.

Lockdown (restricted movement) measures were perhaps the most contentious intervention throughout the COVID-19 pandemic. Studies have found that lockdowns are most effective when timed appropriately (Chu et al. 2023), however it is unlikely that future justification of such stringent measures will be fairly tolerated by the public unless there is a strong threat to human health. This is particularly pertinent considering research undertaken by Chang et al. (2022) which found that, in the context of the Delta variant, over 70% of the population would need to adhere to the lockdown for there to be a meaningful decrease in new COVID-19 cases over a full 45-day intervention period. On social distancing interventions, high levels of adherence have also been noted as a requirement for effective disease control. Chang et al. (2020) developed an ABM that compared strategies across air travel, lockdowns, social distancing, and school closures. For a 13-week intervention period, it was found that greater than 80% of the population would be required to comply with social distancing to achieve true disease control. These studies highlight the importance of targeted and informed interventions that cover a spectrum of individual protective behaviours.

Nonetheless, the impact of interventions should not have the singular view of caseloads. Advancements in data linkage have the potential to monitor and track the impact of interventions (or disease itself) on the long-term health outcomes of individuals. For example, linkage of Admitted Patient Care, Emergency Department, Medicare Benefits Schedule and Pharmaceutical Benefits Scheme data to COVID-19 case data opens up the potential for models such as those described in this essay to evaluate the long-term health outcomes associated with disease, including (but not limited to): chronic health, mental health and mortality outcomes. Data linkage also opens the possibility of building in socio-economic factors that can help determine whether differing income, education and employment (for example) levels impact

upon disease-related (from infection from disease) or intervention-related (stemming from the interventions e.g. lockdowns themselves) health outcomes. By leveraging person-level, deidentified linked data, greater insights can be added to agent-based models such that post pandemic effects can be better evaluated.

Interventions and their flow-on effects should also consider impacts on population inequality with respect to geographic, occupational and other socio-economic outcomes. A novel application of the Lorenz Curve was developed by Nguyen et al. (2023) highlighted that, regarding COVID-19 infections, equality was only achieved for highly transmissible disease variants such as the Omicron variant. This means that for less transmissible variants there is potential for unequal outcomes across the Australian population, not considering the additional impacts of inequality with regard to income, housing and so on. Inequality is also influenced by the model assumptions used to construct disease spread models such as the ABM described in this essay. Continuous efforts must be made to ensure that model inputs are accurate and represent the population (gender, culturally and linguistic diverse, First Nations populations) to the best extent possible. This is to mitigate the potential risks associated with decision making using a biased model and thus making uninformed or biased decisions. The EWS developed in essay three and used in this essay has intentionally taken a data-driven approach to mitigate any potential bias when constructing a signal for future waves of disease spread.

Finally, it is important to consider the significance of model communication during a crisis. If data-driven outputs are being used to inform decision making in the public interest, it is important that the assumptions, methods, and underlying logic of such models are clearly communicated in a manner that allows contestability of outputs. The model described in this essay allows for differentiation between endogenous (voluntary – induced by fearfulness or carefulness) and exogenous (forced – via government or institutional restrictions or mandates) behaviours. In other words, individuals can only make adjustments to their behaviours should they be motivated to act following exposure to specific stimuli that give rise to such behaviours (Steimer, 2002). Research has shown that pandemic fatigue is more prevalent in younger populations, and older populations are less likely to engage in individual protective behaviours such as mask wearing (MacIntyre et al. 2021). In releasing communications to the public, governments must balance the motivational forces that promote personal protective behaviour against the potential to artificially increase the level of fear (therefore encouraging pandemic fatigue) in the population without the supporting evidence or justification.

When tracking cyclones, experts typically agree that an event will occur if multiple different models, based on different assumptions and methods, predict similar results. As such, no individual model is the single source of truth. Outside experts need to be able to understand the assumptions of models being used to support decision making in order for the evidence base to be contestable. The models presented in this essay have been selected as the foundational underpinnings of a disease spread framework because they address the key issues of (a) integrating human behaviour into models, and (b) effectively using the information available to generate robust insights that enable the timely implementation of interventions. That is not to say that these models are the final destination for data-driven decision making. As technology becomes more advanced the granularity and detail of the models described in this essay will greatly improve. In the same way, improvements to time series econometrics, data science, or other fields such as wastewater disease detection, will present improved opportunities for government to enhance decision making, and for individuals to make informed decisions regarding their personal protective behaviours.

6.5 Discussion

This essay has answered the question of how an integrated, fit-for-purpose model of human behavioural states and early wave detection can be deployed to reduce harm and mortality incurred from the spread of infectious disease.

To do so, this essay has brought together the findings of Chapter 3 and 4 (essay one and two) to deploy a fit-for-purpose scenario tool. In this essay, I have quantified the impact of implementing an early warning system method of communication or intervention on infectious disease case counts, with the intention of modelling the reduced cases associated with appropriate and timely action in response to the signal. As expected, the cumulative number of infections and resulting peaks of cases were reduced under scenarios whereby the proportion of the population that is careful is increased, commensurate with a decrease in the proportion of the population that is complacent. While these results are expected by virtue of their construction, the outputs of these scenario models can be used to determine the efficacy of different strengths of campaigns to educate and inform the public about the risk of infection. This can give rise to increased transparency in decision making, given the data-driven foundations of this model. Increased transparency is critical for ensuring public trust and

accountability for their own personal protective behaviours when a signal is raised for increasing cases of novel infectious diseases.

The implications of this research are twofold. First, this essay has demonstrated the ability of agent-based microsimulation models to provide rapid evaluation of the potential benefits of interventions to improve personal protective behaviours, such as social distancing, mask wearing or hand washing. The approach described in this essay represents a data-driven estimation of the impact of policies, meaning that a pre-emptive, considered approach can be taken towards current and future novel infectious disease outbreaks. The current approach to mitigating the spread of disease is still – largely – unplanned across the globe, meaning that there is still work to do to reach long term stability in case counts while protecting those most vulnerable.

Second, this essay articulates the practical way by which tools developed in the academic context can be deployed by governments and policy decision makers. Earlier versions of the base model developed in Chapter 3 (essay one) were used to ensure contestability in modelling outputs produced for the Australian COVID-19 outbreak, with a particular focus on modelling the scenario of personal protective behaviours on disease spread outcomes. Over the course of the pandemic the model has been continually refined, based on the factors which are important to government and individual level decision making. Behavioural adjustments lie at the heart of what is most important to disease outcomes, as evidenced by the COVID-19 pandemic to date.

Future research should look to enhancing the base model (essay one) and EWS (essay three), as previously discussed. As reporting of cases becomes increasingly unreliable (due to significant scaling back of testing availability and adherence), models such as this will become critical to estimating the number of "hidden" cases of COVID-19 or other novel infectious diseases. Finally, governments may have a longer or shorter time to implement interventions than what was modelled in this essay. Consideration of the impact of such time sensitivities should be evaluated in the context of the true impact an intervention may have on individual outcomes.

7. Conclusions

This thesis serves as a framework for decision makers to leverage data-driven, behaviour-focused evidence into business-as-usual disease monitoring, planning and preparedness activities. It also extends the range of applicability for Agent Based Models (ABMs) in the context of a three dimensional, behaviourally adjusted infectious disease transmission rate. This thesis has developed an ABM of infectious disease spread under two different data availability and disease settings, in addition to an early warning system framework. These models culminate into an easy-to-use scenario tool to predict future outcomes based on varying levels of public compliance. While COVID-19 is the primary focus of this thesis, consideration has also been given to influenza, which remains a high-burden infectious disease in the Australian setting. In this vein, the overarching framework developed in this thesis can be adapted to both new and old infectious diseases for monitoring, planning and preparedness activities.

Fundamentally, this thesis has built on the existing literature by addressing the following research questions:

(a) *What are the main factors that influenced the spread of disease throughout the COVID-19 pandemic, and how can a better understanding of these factors in models be used to inform pandemic planning and preparedness?* ***(essay one and essay two)***

(b) *Novel and endemic infectious diseases continue to pose a threat to the population with ongoing waves of infection. How successful are simple econometric models of case velocity in establishing a fit-for-purpose early warning system to support policy decision making?* ***(essay three)***

(c) *How can an integrated model of human behavioural states and early wave detection be deployed in practice to reduce harm and mortality incurred from the spread of disease?* ***(essay four)***

The first pillar of this thesis demonstrated that models to monitor and track the spread of disease can be built in a sophisticated yet simple manner, using publicly available data and open-source technology. These models were found to track disease trajectories well and were improved with the inclusion of behavioural parameters representing the human reaction to the spread of

disease. In Chapter 3 and 6 (essay one and four), agents were categorised into three behavioural compartments: fearful, careful and complacent. These compartments were based on real-time survey data collected throughout the COVID-19 pandemic. The results from this thesis overwhelmingly found that incorporation of behavioural parameters – whether survey-driven or input-driven – improved the overall accuracy of the ABM results.

The second pillar of this thesis established a data-driven early warning system to monitor infectious disease cases and potential waves. Chapter 5 (essay three) examined the role of a simple econometric early warning system (EWS) that could be used as a signal for potential future outbreaks. That essay argued that successful use and interpretation of an EWS would open up opportunities for governments and individuals to pre-emptively protect themselves against future waves of infection in the community. Given the known incubation times for COVID-19, early messaging to encourage personal protective behaviours (such as mask wearing or social distancing) has the potential to have profound impacts on disease transmission, in comparison to the current approach of reporting on cases once they have already begun to increase. Implications for hospitalisation and mortality rates follow. The EWS developed in essay three selected the optimal signal estimate for each Australian State and Territory based on results generated from GARCH, T-GARCH and E-GARCH specifications.

Combining the behaviourally adjusted simulations with a data-driven EWS, Chapter 6 (essay four) quantitatively applied the selected EWS from Chapter 5 (essay three) to projected COVID-19 case count outcomes using the model described in Chapter 3 (essay one). Following the signal, scenarios were applied to look at the impact of a 10%, 20% and 30% increase in the proportion of the population that is careful, in response to evidence-based interventions that promote preventive behaviours at key disease spread junctions. Simulating these actions on the core ABM simulation framework in a hypothetical timestep, the model results suggested that a strong intervention (i.e. zero complacency) demonstrated clear improvements to overall COVID-19 case counts. However, interventions resulting in a smaller boost to careful agents (10% and 20% boost), while still reducing overall case counts from no intervention, did not differ substantially from one another. This implies that interventions around that 10% to 20% careful booster mark have a similar overall effect. Consequently, scenario models such as this can be used to test different intervention options and their likely success rates, at varying levels of geographic granularity.

This thesis has predominantly focused on COVID-19, with a separate application of disease spread modelling for influenza (Chapter 4, essay two) to demonstrate the applicability of such modelling in the absence of detailed survey data. In practicality, the tools and techniques described in this thesis should be extended beyond to other novel and endemic infectious diseases in Australia. For example, by leveraging the data already collected by the National Notifiable Disease Surveillance System (NNDSS), simple time series models such as the EWS described in Chapter 5 (essay three) can be used to generate signals of future waves of infection, requiring little attention once the algorithm is established.

There are a number of limitations associated with the models presented in this thesis, most of which are largely reflected in the name itself – "models". Models are an estimation of reality or future events and are in no way oracles of the future. Models should be used as complementary tools, paired with on-the-ground knowledge, policy and environmental factors. When used appropriately, models can greatly enhance confidence in decision making by providing contestability for policy options using a data-driven evidence base. The need for a mature, embedded evidence base for monitoring and managing infectious disease emergencies was demonstrated across the globe throughout the COVID-19 pandemic. Implementing tools such as those described in this thesis would ensure that actionable, accurate and timely models could inform pandemic planning and preparedness in future.

The remaining limitations associated with this thesis can be captured under three themes: computational power, input accuracy and observational data limitations. Due to computational restrictions, the models presented in this thesis were restricted to relatively small percentage proportions of the Australian population. While unnecessary to simulate the entire Australian population, the application of ABMs continue to pose a costly computational burden (Bobashev et al. 2007) to models. In recent years there have been substantial advances in the development of graphics processing units (GPUs) and scalable cloud-based solutions, which will greatly enhance the ability of ABMs to be run over larger population sizes.

The models developed in this thesis were written in R and deployed on the Australian National University National Computing Infrastructure (Gadi) supercomputer. Despite the creation of efficient parallel algorithms to speed up processing time, the models presented in this thesis would benefit from being re-written in a more efficient programming language such as C++. In this thesis, the use of ABMs has required a carefully considered balance between the number of agents and parameters to ensure that model parsimony is appropriately offset by mechanistic

rigor and accurate outputs (El-Sayed et al. 2012). This echoes findings from other behaviour-adjusted ABMs such as Bulumulla et al. (2017) which found more accurate results using a simpler model, highlighting the need for detailed consideration when adding in model variables that increase computational complexity.

Input accuracy is another factor that will remain a limitation for these types of models into the future. With any new infectious disease, it is difficult to be able to adequately specify the unique characteristics of said disease in the early phase of an epidemic. Thus, modellers must make assumptions around transmission rates, mortality rates and so on, until such information is made available by peer reviewed research. In the same manner that simulation models are subject to bias from changing or unknown drivers of disease transmission, they are also subject to over-fitting from the incorporation of too many, potentially correlated, variables. Therefore, this thesis has sought to navigate the fine line between under-fitting and over-fitting in a changing environment.

Two specific developments hampered the calibration time for the simulation models presented in this thesis: the development of new variants of COVID-19, and the strong behavioural response of individuals and governments to the spread of disease. These developments mean that inputs in the models developed in this thesis had to be tested multiple times, thus slowing the time for the model to be calibrated. This is an ongoing limitation which can only be addressed insofar that models are updated once this new information is obtained, ensuring that results are generated with the most accurate inputs.

Even for known infectious diseases, such as influenza, some data are not collected. Chapter 4 (essay two) highlighted that new data collected throughout the COVID-19 pandemic – such as survey data detailing behavioural responses to the COVID-19 pandemic – simply do not exist for influenza. Following seminal work conducted by Epstein et al. (2008), this thesis detailed a novel solution to modelling fear by introducing it as a separate compartment that allows individuals to contract and recover from the behavioural state. Again, returning to COVID-19, this disease was different given that rapid antigen test (RAT) and polymerase chain reaction (PCR) tests provided a much more granular insight to the spread of disease compared to standard practice. Models built for estimating influenza trajectory – such as that described in Chapter 4 (essay 2) – are limited to clinically notified cases, reported by an Australian medical practitioner. This means that models produced for these disease types are naturally

biased towards modelling more severe cases of disease, implying that total disease counts will not reflect the true burden of disease in the population.

Observational data limitations have likewise influenced the fidelity of the findings presented in this thesis. While there has been an exponential increase in the amount of data made publicly available since the beginning of the COVID-19 pandemic, the collection and reporting of these data has been sporadic. To begin with, public reactions to COVID-19 have become normalised and, as such, from a certain point the reported RAT and PCR case data became unreliable. The models developed in this thesis were also left exposed when government decisions were made to change the reporting cadence of cases from daily to weekly. As demonstrated in Chapter 5 (essay three), more granular daily data is required to estimate the highest quality signals for future waves of infection. Given this data would still be available within State and Territory decision making authorities, these models remain highly applicable for continued use.

The amount of observational data available during the early stages of an outbreak also pose limitations regarding the tuning and calibration of models. The EWS described in this thesis relies on a history of time series data to be built. Earlier applications of the model developed (not included in this thesis) included an ETS (Error, Trend, Seasonal) estimate of case projections, in addition to the simulation structure from essays one and two, to estimate signals in the absence of detailed historic data. Despite this limitation, hospital data have become more freely available since the COVID-19 pandemic. Hospital data, while indicative of more severe cases of disease, can be used for both disease spread and EWS modelling. Future research should seek to tailor the EWS to hospital data to allow a broader spectrum of available signal tools.

The implications of this research are clear. From 1960 to 2020, there has been a tenfold increase in the number of natural disasters worldwide (Institute for Economics & Peace, 2020). There has also been an increase in the number of severe infectious disease outbreaks, including COVID-19, severe respiratory syndrome coronavirus, swine flu, Middle East respiratory syndrome, Ebola virus and Zika virus. Recent history shows a substantial increase in global air-travel, population density and population growth (Baker et al. 2022). As such, there is a strong need for governments worldwide to have standard pandemic preparedness plans, which should feature data and evidence-based models at their core.

This thesis has focused on what should be considered fundamental to pandemic preparedness and planning from a data and modelling perspective. Due to recent technological advances, the role and importance of data in decision making is still relatively novel. Nevertheless, this does not mean that establishment of data-driven functions should be delayed. This thesis has set the foundation for what processes need to be embedded into business-as-usual functions to support decision making. The models described in this thesis have consequently been designed to support crisis management at aggregate or granular geographic levels.

The practical application of the framework for monitoring and managing novel and endemic infectious diseases requires a dedicated allocation of resources by government to support these objectives. It requires some up-front investment in technical capability, IT infrastructure, data linkage and new data pipelines, likely already underway with the development of the new Australian interim Centre for Disease Control (CDC), established in January 2024. The models developed in this thesis provide the underpinning for effective decision making and can be scaled to suit the purposes of the issue at hand. This was the case during 2020 and 2021 where less sophisticated versions of the models described in Chapter 3 and 6 (essay one and four) were used to postulate internal contestability on disease spread outcomes within the Commonwealth Department of Health and Aged Care. The skeleton of this model was then re-deployed to build the first COVID-19 vaccine allocation model, used to provide internal advice estimating the time for full 2-dose vaccination across the Australian population for priority population groups. As a result, the models presented in this thesis are built around the policy problems facing Australia, allowing enhanced support for the monitoring and management of novel and infectious disease outcomes in the years to come.

In addition, this thesis has raised the need for regular, standardised surveys of human perception that can be used to support baseline models such as those described in this thesis. As discussed in Chapter 6 (essay four), this is essential to ensure that models can take a wider view than simple caseloads, while simultaneously seeking to build in metrics for understanding inequality across the Australian population. Building such models in greater detail will require a careful and parsimonious variable selection, that avoids any systematic bias to the greatest extent possible. Advancements in de-identified, whole of population linked data, such as the Person Level Integrated Data Asset (PLIDA) hosted by the Australian Bureau of Statistics or the National Integrated Health Services Information (NIHSI) asset hosted by the Australian

Institute of Health and Welfare, are examples of the types of datasets that could be used to enhance disease spread models.

The research undertaken in this thesis will be of increasing importance to governments and individuals so long as infectious disease outbreaks continue to place pressure on health and economic outcomes. Consideration of the tools highlighted in this thesis will ensure that the population is well-placed to manage and monitor future health-related threats worldwide.

8. References

Abboah-Offei, M. Salifu, Y. Adewale, B. Bayuo, J. Ofosu-Poku, R. & Opare-Lokko, E. B. A. (2021). A rapid review of the use of face mask in preventing the spread of COVID-19. *International journal of nursing studies advances*, *3*, 100013.

ABS. (2018). 2071.0.55.001 - Census of Population and Housing: Commuting to Work - More Stories from the Census, 2016. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2071.0.55.001~2016~Main%20Features~Journey%20to%20Work%20to%20Place%20of%20Work~60\

ABS. (2018a). 2071.0.55.001 - Census of Population and Housing: Commuting to Work - More Stories from the Census, 2016. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2071.0.55.001

ABS. (2019). 3101.0 - Australian Demographic Statistics, Jun 2019. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3101.0Main+Features1Jun%202019#:~:text=3101.0%20%2D%20Australian%20Demographic%20Statistics%2C%20Jun%202019&text=The%20preliminary%20estimated%20resident%20population,people%20since%2031%20March%202019.

ABS. (2019a). 3218.0 - Regional Population Growth, Australia, 2017-18. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02017-18

ABS. (2019b). 6202.0 - Labour Force, Australia, Jul 2019. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6202.0Main+Features1Jul%202019?OpenDocument=#:~:text=6202.0%20%2D%20Labour%20Force%2C%20Australia%2C%20Jul%202019&text=Employment%20increased%2024%2C600%20to%2012%2C915%2C200,increased%206%2C600%20to%20715%2

ABS. (2020). COVID-19 Mortality in Australia, Deaths registered to 31 August 2020. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/articles/covid-19-mortality-australia-deaths-registered-31-august-2020

ABS. (2020a). 4221.0 - Schools, Australia, 2019. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/AUSSTATS/abs@.nsf/7d12b0f6763c78caca257061001cc588/4058adac898288ebca25783b001a9da1!OpenDocument

ABS. (2020b). 4221.0 - Schools, Australia, 2018. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/ausstats/abs@.nsf/exnote/4221.0/

ABS. (2021). National, state and territory population. Retrieved from Australian Bureau of Statistics: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/dec-2020

Adjodah, D. Dinakar, K. Chinazzi, M. Fraiberger, S. P. Pentland, A. Bates, S. Staller, K. Vespignani, A. & Bhatt, D. L. (2021). Association between COVID-19 outcomes and mask mandates, adherence, and attitudes. *PloS one*, *16*(6), e0252315.

Agency for Clinical Innovation. (2022). Critical Intelligence Unit - Living Evidence - COVID-19 vaccines. Retrieved from NSW Government: https://aci.health.nsw.gov.au/covid-19/critical-intelligence-unit/covid-19-vaccines

Ahmed, D. A. Ansari, A. R. Imran, M. Dingle, K. & Bonsall, M. B. (2021). Mechanistic modelling of COVID-19 and the impact of lockdowns on a short-time scale. PLoS One, 16(10), e0258084.

Ajelli, M. Gonçalves, B. Balcan, D. Colizza, V. Hu, H. Ramasco, J. & Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, *10*, 1-13.

Aleta, A. Martin-Corral, D. Piontti, A. Ajelli, M. Litvinova, M. & Chinazzi, M. (2021). Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the covid-19 pandemic.(2020). *Publisher Full Text*.

Alihsan, B. Mohammed, A. Bisen, Y. Lester, J. Nouryan, C. & Cervia, J. (2022). The efficacy of facemasks in the prevention of COVID-19: a systematic review. *medRxiv*, 2022-07.

Bai, S. (2020). Simulations of COVID-19 spread by spatial agent-based model and ordinary differential equations. *International Journal of Simulation and Process Modelling*, *15*(3), 268-277.

Baker, R. E. Mahmud, A. S. Miller, I. F. Rajeev, M. Rasambainarivo, F. Rice, B. L. Takahashi, S. Tatem, A. J. Wagner, C. E. Wang, L. Wesolowski, A. & Metcalf, C. J. E. (2022). Infectious disease in an era of global change. *Nature Reviews Microbiology*, *20*(4), 193-205.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., Paolotti, D. Perra, N. Tizzoni, M. Broeck, W. Colizza, V. & Vespignani, A. (2009). Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine, 7*, 1-12.

Bari, A. Khubchandani, A. Wang, J. Heymann, M. & Coffee, M. (2021). COVID-19 early-alert signals using human behavior alternative data. *Social Network Analysis and Mining*, *11*, 1-17.

Barrett, C. L. Bisset, K. R. Eubank, S. G. Feng, X. & Marathe, M. V. (2008, November). Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing* (pp. 1-12). IEEE.

Benvenuto, D. Giovanetti, M. Vassallo, L. Angeletti, S. & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, *29*, 105340.

Bibby, K. Bivins, A. Wu, Z. & North, D. (2021). Making waves: plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research*, *202*, 117438.

Bobashev, G. V. Goedecke, D. M. Yu, F. & Epstein, J. M. (2007, December). A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches. In *2007 Winter Simulation Conference* (pp. 1532-1537). IEEE.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542-547.

Brett, T. Ajelli, M. Liu, Q. H. Krauland, M. G. Grefenstette, J. J. van Panhuis, W. G. Vespignani, A. Drake, J. M. & Rohani, P. (2020). Detecting critical slowing down in high-dimensional epidemiological systems. *PLoS Computational Biology*, *16*(3), e1007679.

Brett, T. S. Drake, J. M. & Rohani, P. (2017). Anticipating the emergence of infectious diseases. *Journal of The Royal Society Interface*, *14*(132), 20170115.

Bulumulla, C. Padgham, L. Singh, D. & Chan, J. (2017, May). The importance of modelling realistic human behaviour when planning evacuation schedules. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems* (pp. 446-454).

Busetta, P. Serafini, L. Singh, D. & Zini, F. (2001). Extending multi-agent cooperation by overhearing. In *Cooperative Information Systems: 9th International Conference, CoopIS 2001 Trento, Italy, September 5–7, 2001 Proceedings 9* (pp. 40-52). Springer Berlin Heidelberg.

Carpenter, C. & Sattenspiel, L. (2009). 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 Association*, *21*(3), 290-300.

Carter Hill, R. Griffiths, W. E. & Lim, G. C. (2011). Principles of econometrics. In *USA Library of Congress Cataloging*.

Catching, A. Capponi, S. Yeh, M. T. Bianco, S. & Andino, R. (2021). Examining the interplay between face mask usage, asymptomatic transmission, and social distancing on the spread of COVID-19. *Scientific Reports*, *11*(1), 15998.

Catching, A. Capponi, S. Yeh, M. T. Bianco, S. & Andino, R. (2020). Examining face-mask usage as an effective strategy to control COVID-19 spread. *MedRxiv*, 2020-08.

Chang, S. L. Cliff, O. M. Zachreson, C. & Prokopenko, M. (2022). Simulating transmission scenarios of the Delta variant of SARS-CoV-2 in Australia. *Frontiers in Public Health*, 10, 823043.

Chang, S. L. Harding, N. Zachreson, C. Cliff, O. M. & Prokopenko, M. (2020). Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature communications*, *11*(1), 5710.

Chang, S. L. Nguyen, Q. D. Martiniuk, A. Sintchenko, V. Sorrell, T. C. & Prokopenko, M. (2023). Persistence of the Omicron variant of SARS-CoV-2 in Australia: The impact of fluctuating social distancing. *PLOS Global Public Health*, 3(4), e0001427.

Chu, L. Grafton, R. Q. Kompas, T. & McLaws, M. L. (2023). *Effects of Closures and Openings on Public Health in the Time of COVID-19: A Cross-Country and Temporal Trend Analysis*. SAGE Open, 13(4), 21582440231207472.

CovidLive. (2023, January). CovidLive Dashboard. Retrieved from CovidLive: https://covidlive.com.au/

Crooks, A. T. & Hailegiorgis, A. B. (2014). An agent-based modeling approach applied to the spread of cholera. *Environmental Modelling & Software*, *62*, 164-177.

Daghriri, T. & Ozmen, O. (2021). Quantifying the Effects of Social Distancing on the Spread of COVID-19. *International Journal of Environmental Research and Public Health*, *18*(11), 5566.

Dakos, V. Scheffer, M. Van Nes, E. H. Brovkin, V. Petoukhov, V. & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, *105*(38), 14308-14312.

Del Valle, S. Mniszewski, S. & Hyman, J. (2013). Modeling the impact of behavior changes on the spread of pandemic influenza. *Modeling the interplay between human behavior and the spread of infectious diseases*, 59-77.

Del Valle, S. Hethcote, H. Hyman, J. M. & Castillo-Chavez, C. (2005). Effects of behavioral changes in a smallpox attack model. *Mathematical Biosciences*, *195*(2), 228-251

Department of Health. (2014). Pandemic Planning Assumptions: Evidence Summary. Retrieved from Department of Health: [No longer available]

Department of Health. (2019). National Notifiable Diseases Surveillance System. Retrieved from: http://www9.health.gov.au/cda/source/cda-index.cfm

Department of Health and Aged Care. (2022). Influenza (flu). Retrieved from Department of Health and Aged Care: https://www.health.gov.au/diseases/influenza-flu

Department of Health and Aged Care. (2022a). COVID-19 vaccines advice for providers [website no longer active]. Retrieved from Department of Health and Aged Care: https://www.health.gov.au/initiatives-and-programs/covid-19-vaccines/advice-for-providers/clinical-guidance/product-information

D'Orazio, M., Bernardini, G., & Quagliarini, E. (2020). *How to restart? An agent-based simulation model towards the definition of strategies for COVID-19" second phase" in public buildings.*

El-Sayed, A. M. Scarborough, P. Seemann, L. & Galea, S. (2012). Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations*, *9*(1), 1-9.

Enders, W. (2014). Applied econometric time series fourth edition. *New York (US): University of Alabama*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Epstein, J. M. Parker, J. Cummings, D. & Hammond, R. A. (2008). Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PloS One*, *3*(12), e3955.

Espinola, M. Shultz, J. M. Espinel, Z. Althouse, B. M. Cooper, J. L. Baingana, F. and Rechkemmer, A. (2016). Fear-related behaviors in situations of mass threat. *Disaster Health*, *3*(4), 102-111.

Farmer, J. D. & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*(7256), 685-686.

Fearnley, C. J. & Dixon, D. (2020). Early warning systems for pandemics: Lessons learned from natural hazards. *International Journal of Disaster Risk Reduction*, *49*, 101674.

Ferguson, N. (2007). Capturing human behaviour. *Nature*, *446*(7137), 733-733.

Funk, S. Salathé, M. & Jansen, V. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, *7*(50), 1247-1256.

Gao, Y. Cai, G. Y. Fang, W. Li, H. Y. Wang, S. Y. Chen, L. Yu, Y. Liu, D. Xu, S. Cui, P. Zeng,S. Feng, X. Yu, R. Wang, Y. Yuan, Y. Jiao, X. Chi, J. Liu, J. Li, R. Zheng, X. Song, C. Jin, N. Gong, W. J. Liu, X. & Gao, Q. L. (2020). Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nature Communications*, *11*(1), 5033.

Geoscience Australia. (2001). Land areas of States and Territories. Retrieved from Geoscience Australia: https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/area-of-australia-states-and-territories

Glosten, L. R. Jagannathan, R. & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779-1801.

Gutiérrez, I., Litzroth, A., Hammadi, S., Van Oyen, H., Gerard, C., Robesyn, E., ... & Wuillaume, F. (2009). Community transmission of influenza A (H1N1) v virus at a rock festival in Belgium, 2-5 July 2009. Eurosurveillance, 14(31), 19294.

Harper, C. A. Satchell, L. P. Fido, D. & Latzman, R. D. (2021). Functional fear predicts public health compliance in the COVID-19 pandemic. *International Journal of Mental Health and Addiction*, *19*, 1875-1888.

HealthDirect. (2020). COVID-19 Symptoms. Retrieved from HealthDirect: https://www.healthdirect.gov.au/covid-19/symptoms-and-medical-advice

Hernandez-Matamoros, A. Fujita, H. Hayashi, T. & Perez-Meana, H. (2020). Forecasting of COVID19 per regions using ARIMA models and polynomial functions. *Applied Soft Computing*, *96*, 106610.

Hill, R. C. Griffiths, W. E. & Lim, G. C. (2011). *Principles of Econometrics*. John Wiley & Sons.

Institute of Economics & Peace. (2020, September). *Ecological Threat Register 2020: Understanding Ecological Threats, Resilience and Peace*. Retrieved from Institute of Economics & Peace: http://visionofhumanity.org/reports

Ishola, D. A., & Phin, N. (2011). Could influenza transmission be reduced by restricting mass gatherings? Towards an evidence-based policy framework. Journal of epidemiology and global health, 1(1), 33-60.

Jacintho, L. F. Batista, A. F. Ruas, T. L. Marietto, M. G. & Silva, F. A. (2010, April). An agent-based model for the spread of the Dengue fever: a swarm platform simulation approach. In *Proceedings of the 2010 Spring Simulation Multiconference* (pp. 1-8).

Javarone, M. (2014). Social influences in opinion dynamics: the role of conformity. *Physica A: Statistical Mechanics and its Applications*, *414*, 19-30.

Jenness, S. M. Goodreau, S. M. & Morris, M. (2018). EpiModel: an R package for mathematical modeling of infectious disease over networks. *Journal of Statistical Software*, *84*.

Jennings, L. C. Monto, A. S. Chan, P. K. Szucs, T. D. & Nicholson, K. G. (2008). Stockpiling prepandemic influenza vaccines: a new cornerstone of pandemic preparedness plans. *The Lancet Infectious Diseases*, *8*(10), 650-658.

Jiang, W. Sullivan, A. M. Su, C. & Zhao, X. (2012). An agent-based model for the transmission dynamics of Toxoplasma gondii. *Journal of Theoretical Biology*, *293*, 15-26.

Johnston, M. D., & Pell, B. (2020). A dynamical framework for modeling fear of infection and frustration with social distancing in COVID-19 spread. *Mathematical Biosciences and Engineering: MBE*, *17(6)*, 7892-7915.

Jones, S. P. (2020). Imperial college london yougov covid data hub, v1. 0. *Imperial College London Big Data Analytical Unit and YouGov Plc*, *2*(5), 6.

Karaivanov, A. Lu, S. E. Shigeoka, H. Chen, C. & Pamplona, S. (2021). Face masks, public policies and slowing the spread of COVID-19: Evidence from Canada. *Journal of Health Economics*, *78*, 102475.

Kasperson, R. E. Renn, O. Slovic, P. Brown, H. S. Emel, J. Goble, R. ... & Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*, *8*(2), 177-187

Keeling MJ, Rohani P. *Modeling infectious diseases in humans and animals*. Princeton University Press; 2011 Sep 19.

Kelly, R. A. Jakeman, A. J. Barreteau, O. Borsuk, M. E. ElSawah, S. Hamilton, S. H. Henriksen, H. J. Kuikka, S. Maier, H. R. Rizzoli, A. E. Delden, H. & Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software*, *47*, 159-181.

Kerr, C. C. Stuart, R. M. Mistry, D. Abeysuriya, R. G. Rosenfeld, K. Hart, G. R. Núñez, R. C. Cohen, J. A. Selvaraj, P. Hagedorn, B. George, L. Jastrzebski, M. Izzo, A. S. Fowler, G. Palmer, A. Delport, D. Scott, N. Kelly, S. L. Bennette, C. S. Wagner, B. G, Chang, S. T. Oron, A. P. Wenger, E. A. Panovska-Griffiths, J. Famulare, M. & Klein, D. J. (2021). Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, *17*(7), e1009149.

Khalil, K. M. Abdel-Aziz, M. Nazmy, T. T. & Salem, A. B. M. (2012). An agent-based modeling for pandemic influenza in Egypt. *Handbook on Decision Making: Vol 2: Risk Management in Decision Making*, 205-218.

Kidd, M. R. (2020). Five principles for pandemic preparedness: lessons from the Australian COVID-19 primary care response. *British Journal of General Practice*, *70*(696), 316-317.

Kiss, I. Z. Cassell, J. Recker, M. & Simon, P. L. (2010). The impact of information transmission on epidemic outbreaks. *Mathematical Biosciences*, *225*(1), 1-10.

Kochańczyk, M. Grabowski, F. & Lipniacki, T. (2020). Dynamics of COVID-19 pandemic at constant and time-dependent contact rates. *Mathematical Modelling of Natural Phenomena*, *15*, 28.

Kohler, T. A. & Gumerman, G. G. (Eds.). (2000). *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*. Oxford University Press.

Köhler-Rieper, F. Röhl, C. H. & De Micheli, E. (2020). A novel deterministic forecast model for the Covid-19 epidemic based on a single ordinary integro-differential equation. *The European Physical Journal Plus*, *135*(7), 599.

Krieg, S. J. Schnur, J. J. Marshall, J. D. Schoenbauer, M. M. & Chawla, N. V. (2020). Pandemic pulse: unraveling and modeling social signals during the COVID-19 pandemic. *Digital Government: Research and Practice*, *2*(2), 1-9.

Kumar, M. Joshi, M. Patel, A. K. & Joshi, C. G. (2021). Unravelling the early warning capability of wastewater surveillance for COVID-19: a temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environmental Research*, 196, 110946.

Lal, A. Ashworth, H. C. Dada, S. Hoemeke, L. & Tambo, E. (2022). Optimizing pandemic preparedness and response through health information systems: lessons learned from Ebola to COVID-19. *Disaster Medicine and Public Health Preparedness*, *16*(1), 333-340.

Leamer, E. E. (1985, January). Vector autoregressions for causal inference?. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 22, pp. 255-304). North-Holland.

Lim, H. C., Cutter, J., Lim, W. K., Ee, A., Wong, Y. C., & Tay, B. K. (2010). The influenza A (H1N1-2009) experience at the inaugural Asian Youth Games Singapore 2009: mass gathering during a developing pandemic. British Journal of Sports Medicine, 44(7), 528-532.

Liu, X. Huang, J. Li, C. Zhao, Y. Wang, D. Huang, Z. & Yang, K. (2021). The role of seasonality in the spread of COVID-19 pandemic. *Environmental Research*, *195*, 110874.

Macal, C. M. North, M. J. Collier, N. Dukic, V. M. Wegener, D. T. David, M. Z. Daum, R. S. Schumm, P. Evans, J. A. Wilder, J. Miller, L. G. Eells, S. & Lauderdale, D. S. (2014). Modeling the transmission of community-associated methicillin-resistant Staphylococcus aureus: a dynamic agent-based simulation. *Journal of Translational Medicine*, *12*(1), 1-12.

Macal, C. M. & North, M. J. (2005, December). Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.* (pp. 14-pp). IEEE.

MacIntyre, C. R., Nguyen, P. Y., Chughtai, A. A., Trent, M., Gerber, B., Steinhofel, K., & Seale, H. (2021). Mask use, risk-mitigation behaviours and pandemic fatigue during the COVID-19 pandemic in five cities in Australia, the UK and USA: A cross-sectional survey. International Journal of Infectious Diseases, 106, 199-207.

MacIntyre, C. R. Dwyer, D. Seale, H. Fasher, M. Booy, R. Cheung, P. Ovdin, N. & Browne, G. (2008). The first randomized, controlled clinical trial of mask use in households to prevent respiratory virus transmission. *International Journal of Infectious Diseases*, *12*, e328.

Maupin, K. A., Swiler, L. P., & Porter, N. W. (2018). Validation metrics for deterministic and probabilistic data. Journal of Verification, Validation and Uncertainty Quantification, 3(3), 031002.

Meacci, L. & Primicerio, M. (2021). Pandemic fatigue impact on COVID-19 spread: A mathematical modelling answer to the Italian scenario. *Results in Physics*, *31*, 104895.

Medema, G. Heijnen, L. Elsinga, G. Italiaander, R. & Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. *Environmental Science & Technology Letters*, *7*(7), 511-516.

Melin, P. Monica, J. C. Sanchez, D. & Castillo, O. (2020, June). Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: the case of Mexico. In *Healthcare* (Vol. 8, No. 2, p. 181). MDPI.

Menegale, F. Manica, M. Zardini, A. Guzzetta, G. Marziano, V. d'Andrea, V. Trentini, F. Ajelli, M. Poletti, P. & Merler, S. (2023). Evaluation of Waning of SARS-CoV-2 Vaccine–Induced Immunity: A Systematic Review and Meta-analysis. *JAMA Network Open*, 6(5), e2310650-e2310650.

Menéndez, J. (2020). Elementary time-delay dynamics of COVID-19 disease. *medRxiv*, 2020-03.

Meo, S. A. Abukhalaf, A. A. Alomar, A. A. AlMutairi, F. J. Usmani, A. M. & Klonoff, D. C. (2020). Impact of lockdown on COVID-19 prevalence and mortality during 2020 pandemic: observational analysis of 27 countries. *European Journal of Medical Research*, *25*(1), 1-7.

Merriam-Webster. (2023, January). Careful (adjective). Retrieved from Merriam-Webster: https://www.merriam-webster.com/dictionary/careful

Merriam-Webster. (2023, January). Complacency (noun). Retrieved from Merriam-Webster: https://www.merriam-webster.com/dictionary/complacency

Miller, P. B. O'Dea, E. B. Rohani, P. & Drake, J. M. (2017). Forecasting infectious disease emergence subject to seasonal forcing. *Theoretical Biology and Medical Modelling*, *14*(1), 1-14.

Mniszewski, S. M. Manore, C. A. Bryan, C. Del Valle, S. Y. & Roberts, D. (2014, July). Towards a hybrid agent-based model for mosquito borne disease. In *Summer Computer Simulation Conference:(SCSC 2014): 2014 summer simulation multi-conference: Monterey, California, USA, 6-10 july 2014. Summer Computer Simulation Conference (2014: Monterey, California)* (Vol. 2014). NIH Public Access.

Goldstein, G. P., Morgunov, A., Nangalia, V., & Rotkirch, A. (2020). *Universal Masking is Urgent in the COVID-19 Pandemic: SEIR and Agent Based Models, Empirical Validation, Policy Recommendations.*

National Centre for Education Statistics. (2019). *Preschool and kindergarten enrolment.* Retrieved from National Centre for Education Statistics: https://nces.ed.gov/programs/coe/indicator_cfa.asp

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Newland, M. Durham, D. Asher, J. Treanor, J. J. Seals, J. Donis, R. O. & Johnson, R. A. (2021). Improving pandemic preparedness through better, faster influenza vaccines. *Expert Review of Vaccines*, *20*(3), 235-242.

Nichols, G. L. Gillingham, E. L. Macintyre, H. L. Vardoulakis, S. Hajat, S. Sarran, C. E. Amankwaah, D. & Phalkey, R. (2021). Coronavirus seasonality, respiratory infections and weather. *BMC Infectious Diseases*, *21*(1), 1-15.

Nguyen, Q. D., Chang, S. L., Jamerlan, C. M., & Prokopenko, M. (2023). Measuring unequal distribution of pandemic severity across census years, variants of concern and interventions. Population Health Metrics, 21(1), 17.

Nguyen, Q. D. & Prokopenko, M. (2022). A general framework for optimising cost-effectiveness of pandemic response under partial intervention measures. *Scientific Reports*, 12(1), 19482.

Nyberg, T. Ferguson, N. M. Nash, S. G. Webster, H. H. Flaxman, S. Andrews, N. Hinsley, W. Bernal, J. L. Kall, M. Bhatt, S. Blomquist, P. Zaidi, A. Volz, E. Abdul Aziz, N. Harman, K. Funk, S. Abbott, S. Hope, R. Charlett, A. Chand, M. Ghani, A. C. Seaman, S. R. Dabrera, G. De Angelis, D. Presanis, A. & Thelwall, S. (2022). Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B. 1.1. 529) and delta (B. 1.617. 2) variants in England: a cohort study. *The Lancet*, *399*(10332), 1303-1312.

Our World in Data. (2021). Coronavirus (COVID-19) cases. Retrieved from Our World in Data: *https://ourworldindata.org/covid-cases?country=IND~USA~GBR~CAN~DEU~FRA~AUS*

Padgham, L., Horne, R., Singh, D., & Moore, T. (2014). Planning for sandbagging as a response to flooding: A tool and case study. *Australian Journal of Emergency Management, The*, *29*(4), 26-31.

Pakpour, A. H. & Griffiths, M. D. (2020). The fear of COVID-19 and its role in preventive behaviors. *Journal of Concurrent Disorders*, *2*(1), 58-63.

Pedro, S. A. Ndjomatchoua, F. T. Jentsch, P. Tchuenche, J. M. Anand, M. & Bauch, C. T. (2020). Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. *Frontiers in Physics*, *8*, 574514.

Perez, L. & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, *8*(1), 1-17.

Perra, N. Balcan, D. Gonçalves, B. & Vespignani, A. (2011). Towards a characterization of behavior-disease models. *PloS One*, *6*(8), e23084.

Rakowski, F. Gruziel, M. Bieniasz-Krzywiec, Ł. & Radomski, J. P. (2010). Influenza epidemic spread simulation for Poland—a large scale, individual based model study. *Physica A: Statistical Mechanics and its Applications*, *389*(16), 3149-3165.

Rao, D. M. Chernyakhovsky, A. & Rao, V. (2009). Modeling and analysis of global epidemiology of avian influenza. *Environmental Modelling & Software*, *24*(1), 124-134.

Rao, I. J. Vallon, J. J. & Brandeau, M. L. (2021). Effectiveness of face masks in reducing the spread of COVID-19: a model-based analysis. Medical Decision Making, 41(8), 988-1003.

Reintjes, R. Baumeister, H. G. & Coulombier, D. (2001). Infectious disease surveillance in North Rhine-Westphalia: First steps in the development of an early warning system. *International journal of hygiene and environmental health*, *203*(3), 195-199.

Riediker, M. & Tsai, D. H. (2020). Estimation of viral aerosol emissions from simulated individuals with asymptomatic to moderate coronavirus disease 2019. *JAMA Network Open*, *3*(7), e2013807-e2013807.

Safi, S. K. & Sanusi, O. I. (2021). A hybrid of artificial neural network, exponential smoothing, and ARIMA models for COVID-19 time series forecasting. *Model Assisted Statistics and Applications*, *16*(1), 25-35.

Santarsiero, A. Giustini, M. Quadrini, F. D'Alessandro, D. & Fara, G. M. (2021). Effectiveness of face masks for the population. Annali di Igiene, Medicina Preventiva e di Comunita, 33(4).

Schoenharl, T. W. & Madey, G. (2008). Evaluation of measurement techniques for the validation of agent-based simulations against streaming data. In Computational Science–ICCS 2008: 8th International Conference, Kraków, Poland, June 23-25, 2008, Proceedings, Part III 8 (pp. 6-15). Springer Berlin Heidelberg.

Seale, H. Heywood, A. E. Leask, J. Sheel, M. Thomas, S. Durrheim, D. N. Bolsewicz, K. & Kaur, R. (2020). COVID-19 is rapidly changing: Examining public perceptions and behaviors in response to this evolving pandemic. PloS one, 15(6), e0235112.

Segaran, T. (2007). *Programming collective intelligence*. " O'Reilly Media, Inc.".

Segovia-Juarez, J. L. Ganguli, S. & Kirschner, D. (2004). Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model. *Journal of Theoretical Biology*, *231*(3), 357-376.

Shi, P., Keskinocak, P., Swann, J. L., & Lee, B. Y. (2010). The impact of mass gatherings and holiday traveling on the course of an influenza pandemic: a computational model. BMC public health, 10, 1-12.

Shi, Y. Liu, X. Kok, S. Y. Rajarethinam, J. Liang, S. Yap, G. Chong, C. Lee, K. Tan, S. Kuan Yew Chin, C. Lo, A. Kong, W. Ng, L. C.& Cook, A. R. (2016). Three-month real-time dengue

forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. *Environmental health perspectives*, *124*(9), 1369-1375.

Shultz, J. M. Cooper, J. L. Baingana, F. Oquendo, M. A. Espinel, Z. Althouse, B. M. Marcelin, L. H. Towers, S. Espinola, M. McCoy, C. Mazurik, L. Wainberg, M. L. Neria, Y. & Rechkemmer, A. (2016). The role of fear-related behaviors in the 2013–2016 West Africa Ebola virus disease outbreak. *Current Psychiatry Reports*, *18*, 1-14.

Silva, P. C. Batista, P. V. Lima, H. S. Alves, M. A. Guimarães, F. G. & Silva, R. C. (2020). 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.

Singh, D., Sardina, S., & Padgham, L. (2010). Extending BDI plan selection to incorporate learning from experience. *Robotics and Autonomous Systems*, *58*(9), 1067-1075.

Singh, D., & Padgham, L. (2015, May). Community evacuation planning for bushfires using agent-based simulation. In *AAMAS* (pp. 1903-1904).

Singh, D., Padgham, L., & Logan, B. (2016). Integrating BDI agents with agent-based simulation platforms. *Autonomous Agents and Multi-Agent Systems, 30*, 1050-1071.

Steimer, T. (2002). The biology of fear-and anxiety-related behaviors. Dialogues in clinical neuroscience, 4(3), 231-249.

Steinegger, B. Arenas, A. Gómez-Gardeñes, J. & Granell, C. (2020). Pulsating campaigns of human prophylaxis driven by risk perception palliate oscillations of direct contact transmitted diseases. *Physical Review Research, 2*(2), 023181.

Stutt, R. O. Retkute, R. Bradley, M. Gilligan, C. A. & Colvin, J. (2020). A modelling framework to assess the likely effectiveness of facemasks in combination with ‘lock-down’ in managing the COVID-19 pandemic. *Proceedings of the Royal Society A*, 476(2238), 20200376.

Thompson, R. N. Hollingsworth, T. D. Isham, V. Arribas-Bel, D. Ashby, B. Britton, T. Challenor, P. Chappell, L. H. K. Clapham, H. Cunniffe, N. J. Dawid, A. P. Donnelly, C. A. Eggo, R. M. Gilbert, N. Glendinning, P. Gog, J. R. Hart, W. S. Heesterbeek, H. House, T. Keeling, M. Kiss, I. Z. Kretzschmar, M. E. Lloyd, A. L. McBryde, E. S. McCaw, J. M. McKinley, T. J. Miller, J. C. Morris, M. O’Neill, P. D. Parag, K. V. Pearson, C. A. B. Pellis, L. Pulliam, J. R. C. Ross, J. V. Scalia Tomba, G. Silverman, B. W. Struchiner, C. J. Tildesley, M. J. Trapman,

P. Webb, C. R. Mollison, D. & Restif, O. (2020). Key questions for modelling COVID-19 exit strategies. *Proceedings of the Royal Society B, 287*(1932), 20201405.

Thu, T. P. B. Ngoc, P. N. H. & Hai, N. M. (2020). Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. *Science of the Total Environment, 742*, 140430.

Tracht, S. M. Del Valle, S. Y. & Hyman, J. M. (2010). Mathematical modeling of the effectiveness of facemasks in reducing the spread of novel influenza A (H1N1). *PloS One*, 5(2), e9018.

UNDRR. (2006). Integrating Early Warning into Relevant Policies. Retrieved from Second International Conference of Early Warning: https://www.unisdr.org/2006/ppew/info-resources/ewc2/upload/downloads/Policy_brief.pdf

United States Environmental Protection Agency. (2023). Indoor Air and Coronavirus (COVID-19) . Retrieved from United States Environmental Protection Agency: https://www.epa.gov/coronavirus/indoor-air-and-coronavirus-covid-19

Venkatramanan, S. Lewis, B. Chen, J. Higdon, D. Vullikanti, A. & Marathe, M. (2018). Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics, 22*, 43-49.

Wang, J. Xiong, J. Yang, K. Peng, S. & Xu, Q. (2010, June). Use of GIS and agent-based modeling to simulate the spread of influenza. In *2010 18th International Conference on Geoinformatics* (pp. 1-6). IEEE.

Wang, Y. Xu, C. Yao, S. Zhao, Y. Li, Y. Wang, L. & Zhao, X. (2020). Estimating the prevalence and mortality of coronavirus disease 2019 (COVID-19) in the USA, the UK, Russia, and India. *Infection and Drug Resistance*, 3335-3350.

Widgren, S., Bauer, P., Eriksson, R., & Engblom, S. (2019). SimInf: An R Package for Data-Driven Stochastic Disease Spread Simulations. *Journal of Statistical Software, 91*, 1-42.

Witte, K. Berkowitz, J. M. Cameron, K. A. & McKeon, J. K. (1998). Preventing the spread of genital warts: Using fear appeals to promote self-protective behaviors. *Health Education & Behavior*, *25*(5), 571-585.

World Health Organization. (2021). Coronavirus disease (COVID-19): How is it transmitted? . Retrieved from World Health Organization: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted

World Health Organization. (2023). Tracking SARS-CoV-2 variants. Retrieved from World Health Organization: https://www.who.int/activities/tracking-SARS-CoV-2-variants

World Health Organization. (2023a). WHO Coronavirus (COVID-19) Dashboard. Retrieved from World Health Organization: https://covid19.who.int/

World Health Organization. (2023b). Recommendations for influenza vaccine composition. Retrieved from World Health Organization: https://www.who.int/teams/global-influenza-programme/vaccines/who-recommendations

World Health Organization. (2023c). From emergency response to long-term COVID-19 disease management: sustaining gains made during the COVID-19 pandemic. May 2023.

Zachreson, C. Chang, S. Harding, N. & Prokopenko, M. (2022). The effects of local homogeneity assumptions in metapopulation models of infectious disease. *Royal Society Open Science, 9*(7), 211919.

Zhang, Y. Chen, K. Weng, Y. Chen, Z. Zhang, J. & Hubbard, R. (2022). An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. *Expert Systems with Applications, 198*, 116882.

Appendix A

The tables contained in this appendix provide detailed State and Territory GARCH model evaluation referred to in Essay 3: An econometric early warning system for signalling COVID-19 waves.

Table A.1. New South Wales model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.3817	2	0.6178	2	18.4874	1	5
	E-GARCH	250.5308	3	15.8282	3	30.0711	3	9
	T-GARCH	0.3774	1	0.6143	1	18.8435	2	**4**
2	GARCH	0.7608	3	0.8722	3	34.5047	3	9
	E-GARCH	0.5314	1	0.7290	1	26.3216	1	**3**
	T-GARCH	0.7436	2	0.8623	2	33.9922	2	6
3	GARCH	0.6460	3	0.8037	3	192.9339	2	8
	E-GARCH	0.5428	1	0.7368	1	207.6277	3	**5**
	T-GARCH	0.6328	2	0.7955	2	190.8449	1	5
4	GARCH	0.0257	2	0.1603	2	36.2189	2	6
	E-GARCH	0.0617	3	0.2483	3	61.0875	3	9
	T-GARCH	0.0252	1	0.1586	1	35.5117	1	**3**

Table A.2. Victoria model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.6068	2	0.7790	2	12.7446	3	7
	E-GARCH	0.6840	3	0.8271	3	12.4180	1	7
	T-GARCH	0.6068	1	0.7790	1	12.7446	2	**4**
2	GARCH	0.1990	3	0.4461	3	41.5698	3	9
	E-GARCH	0.1144	1	0.3382	1	27.6842	1	**3**
	T-GARCH	0.1990	2	0.4461	2	41.5698	2	6
3	GARCH	2.2074	2	1.4857	2	259.2704	3	7
	E-GARCH	7.0000	3	2.6458	3	190.3830	1	7
	T-GARCH	2.2074	1	1.4857	1	259.2704	2	**4**
4	GARCH	0.0100	3	0.1001	3	57.6909	3	9
	E-GARCH	0.0075	1	0.0864	1	42.3720	1	**3**
	T-GARCH	0.0100	2	0.1001	2	57.6909	2	6

Table A.3. Queensland model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	**GARCH**	0.1022	1	0.3196	1	14.5165	1	**3**
	E-GARCH	0.8374	3	0.9151	3	17.4327	3	9
	T-GARCH	0.1022	2	0.3196	2	14.5165	2	6
2	**GARCH**	0.9101	1	0.9540	1	14.0611	1	**3**
	E-GARCH	4.4994	3	2.1212	3	19.2252	3	9
	T-GARCH	0.9101	2	0.9540	2	14.0611	2	6
3	GARCH	2.1418	2	1.4635	2	73.8870	2	6
	E-GARCH	1.1395	1	1.0675	1	59.1631	1	**3**
	T-GARCH	2.1418	3	1.4635	3	73.8870	3	9
4	GARCH	0.0275	3	0.1659	3	131.2777	3	9
	E-GARCH	0.0246	1	0.1567	1	103.1355	1	**3**
	T-GARCH	0.0275	2	0.1659	2	131.2777	2	6

Table A.4. Western Australia model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.2221	2	0.4712	2	10.6704	2	6
	E-GARCH	2.5937	3	1.6105	3	31.6139	3	9
	T-GARCH	0.2221	1	0.4712	1	10.6704	1	**3**
2	GARCH	2.6033	2	1.6135	2	20.7818	2	6
	E-GARCH	7.41E+08	3	27333	3	92456	3	9
	T-GARCH	2.6033	1	1.6135	1	20.7818	1	**3**
3	GARCH	14.2171	2	3.7706	2	65.8988	2	6
	E-GARCH	1.61E+14	3	1.27e+7	3	41051412	3	9
	T-GARCH	14.2171	1	3.7706	1	65.8988	1	**3**
4	GARCH	0.0106	2	0.1029	2	55.8134	2	6
	E-GARCH	4.37E+6	3	2091	3	2.47E+5	3	9
	T-GARCH	0.0106	1	0.1029	1	55.8134	1	**3**

Table A.5. South Australia model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.2838	3	0.5327	3	17.8261	3	9
	E-GARCH	0.2804	1	0.5296	1	13.7539	1	**3**
	T-GARCH	0.2838	2	0.5327	2	17.8261	2	6
2	GARCH	0.4258	2	0.6525	2	10.9571	2	6
	E-GARCH	77.6118	3	8.8098	3	35.2881	3	9
	T-GARCH	0.4258	1	0.6525	1	10.9571	1	**3**
3	GARCH	3.7165	3	1.9278	3	94.6563	3	9
	E-GARCH	1.3274	1	1.1521	1	68.7120	1	**3**
	T-GARCH	3.7165	2	1.9278	2	94.6563	2	6
4	GARCH	0.0088	3	0.0936	3	47.3528	2	8
	E-GARCH	0.0088	1	0.0935	1	49.2701	3	**5**
	T-GARCH	0.0088	2	0.0936	2	47.3528	1	5

Table A.6. Tasmania model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.6537	2	0.8085	2	15.4040	2	6
	E-GARCH	0.0534	1	0.2311	1	4.0251	1	**3**
	T-GARCH	0.6537	3	0.8085	3	15.4040	3	9
2	GARCH	0.7159	2	0.8461	2	5.0892	2	6
	E-GARCH	0.0063	1	0.0796	1	0.6300	1	**3**
	T-GARCH	0.7159	3	0.8461	3	5.0892	3	9
3	GARCH	25.1730	2	5.0173	2	91.3576	1	5
	E-GARCH	0.0900	1	0.3001	1	118.7230	3	**5**
	T-GARCH	25.1730	3	5.0173	3	91.3576	2	8
4	**GARCH**	0.0163	1	0.1276	1	114.6875	1	**3**
	E-GARCH	45.8424	3	6.7707	3	2103.7009	3	9
	T-GARCH	0.0163	2	0.1276	2	114.6875	2	6

Table A.7. Australian Capital Territory model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	1.0329	3	1.0163	3	11.4066	2	8
	E-GARCH	0.5793	1	0.7611	1	8.0128	1	**3**
	T-GARCH	1.0307	2	1.0152	2	11.4561	3	7
2	GARCH	26.7855	2	5.1755	2	30.3869	2	6
	E-GARCH	78.3203	3	8.8499	3	44.8113	3	9
	T-GARCH	26.6004	1	5.1576	1	30.2824	1	**3**
3	GARCH	8.7285	2	2.9544	2	131.9799	2	6
	E-GARCH	47.1003	3	6.8630	3	542.3879	3	9
	T-GARCH	8.6578	1	2.9424	1	131.4358	1	**3**
4	GARCH	0.0217	2	0.1474	2	117.8096	2	6
	E-GARCH	2.2204	3	1.4901	3	468.5032	3	9
	T-GARCH	0.0216	1	0.1471	1	117.2068	1	**3**

Table A.8 Northern Territory model evaluation

Period	Model	MSE	Rank	RMSE	Rank	MAPE (%)	Rank	Final rank
1	GARCH	0.4244	2	0.6515	2	4.2746	2	6
	E-GARCH	0.0096	1	0.0982	1	0.8602	1	**3**
	T-GARCH	0.4244	3	0.6515	3	4.2746	3	9
2	GARCH	1.1680	2	1.0807	2	6.4188	2	6
	E-GARCH	0.0081	1	0.0898	1	0.8466	1	**3**
	T-GARCH	1.1680	3	1.0807	3	6.4188	3	9
3	**GARCH**	5.1532	1	2.2701	1	61.7233	1	**3**
	E-GARCH	10.1586	3	3.1873	3	548.5635	3	9
	T-GARCH	5.1532	2	2.2701	2	61.7233	2	6
4	**GARCH**	0.0220	1	0.1482	1	45.6197	1	**3**
	E-GARCH	7.1170	3	2.6678	3	442.3793	3	9
	T-GARCH	0.0220	2	0.1482	2	45.6197	2	6