Te Pūnaha Matatini
Data ▪ Knowledge ▪ Insight

This paper has not yet undergone formal peer review

Network-based simulations of re-emergence and spread of COVID-19 in Aotearoa New Zealand

last updated October 19, 2020, author list alphabetical

Emily Harvey[1,2], Oliver Maclaren[3], Dion O'Neale[1,4], Adrian Ortiz-Cervantes[1,4], Frankie Patten-Elliott[3], Steven Turnbull[1,6], Demival Vasques Filho[5,1], and David Wu[3]

[1]Te Pūnaha Matatini, Univeristy of Auckland, NZ
[2]M.E Research, Takapuna, Auckland, NZ
[3]Engineering Science, University of Auckland, NZ
[4]Department of Physics, University of Auckland, NZ
[5]Leibniz Institute for European History, Germany
[6]Faculty of Education and Social Work, University of Auckland, NZ

EXECUTIVE SUMMARY

We simulate the late July/early August re-emergence and spread of COVID-19 in Aotearoa New Zealand. We use a stochastic, individual-based network model of all ≈5 million individuals in Aotearoa, and run simulations for a period of 30 days. Based on these simulations, we calculate: the expected time to detection of the first case after initial seed cases; the number of cases at the time of detection; the time until detection of a first case outside of Auckland; and how the overall number of cases increases without intervention. Our model includes interaction pathways, referred to as 'contexts' in the network, broken down into network 'layers' representing home, work, school, and community structure. Each simulation starts from initial (seed) cases corresponding to the first detected re-emergence cases in August 2020. We run 50 realisations of each simulation for 30 days — each simulation scenario corresponding to one of three different levels of transmission rate. To model the behaviour of individuals in the weeks prior to the August 11th re-emergence, we assume a moderate rate of people getting tested if mildly symptomatic. No contact tracing or intervention is present in this scenario, other than cases that test positive being isolated to their dwelling.

Key points:

- Without intervention, and with a typical transmission rate and a moderate assumed level of symptomatic cases testing, the **first case gets detected by testing** a mean of **7 days** after initial seeding cases (median 7, inter-quartile range of [4,10]). Given a known first detection date of 11 August, and a generation time of around 5 days, this indicates that these 9–13 seed cases were infected by an earlier case or cases that had been **infected in late July/early August 2020**.
- Without intervention, and with a typical transmission rate and a moderate assumed level of symptomatic testing, the **first infected person outside the Auckland region is detected by testing** a mean of **17 days** after initial seeding cases (median 17, IQ range of [13,21]). Given the initial seeding used, and the subsequently reported cases outside of Auckland on 14 August, this suggests that the increased rates of testing and contact tracing brought forward the detection of cases outside the Auckland region by around one week.
- Without intervention, and with a typical transmission rate and a moderate assumed level of symptomatic cases testing, at **7 days** after the seeding cases our simulations have a mean of **29 total cases** (median 27, IQ range of [24, 32]). **17 days** after the seeding cases, our simulations have a mean of **141 total cases** (median 142, IQ range of [93,178]). This latter scenario corresponds to **approximately 21 August 2020**.
- Without intervention, and with a typical transmission rate and a moderate assumed level of symptomatic cases testing, the **first infected person outside the Auckland region** occurs after a mean of **5 days** and a median of 5 days after the seeding cases. **7 days** after seeding cases, there is transmission to outside of Auckland in **76% of model runs** and, **17 days** after seeding cases, in **100% of model runs**.

Introduction

Aotearoa's government has been praised for its science-informed approach to COVID-19, with much of this response informed by computer simulation and modelling. Following the first reported case of COVID-19 in Aotearoa on the 28th of February, 2020, early simulation models[1,2] informed an elimination strategy which saw the country enter a nationwide lock-down and international borders closed to everyone except citizens and residents. On the 21st of March, an Alert Level system[3] was introduced, with the immediate move to Level 2, all entry ports and borders closed to non-residents, and all returning citizens and residents required to self-isolate. Due to escalating cases this increased to a Level 3 lock-down by the 23rd of March, and a Level 4 lock-down by the 25th of March. During Alert Level 4, people were only allowed to leave home for essential work, essential (e.g. supermarket and pharmacy) shopping and exercise. Any social interactions outside of a direct household (bubble) were not permitted. Following these measures, case numbers dropped and Aotearoa gradually lifted lock-down restrictions; down to Level 3 on the 27th of April, Level 2 on the 13th of May, and Level 1 on the 8th June. Level 1 is described as 'Prepare', and is a state where the virus has been controlled in NZ, but where the continued uncontrolled spread of COVID-19 overseas means that there is the potential for imported cases.

There are a number of components to the government's Level 1 plan (preparing for a re-emergence of COVID-19) that were in place in July and early August:

- There was a compulsory 14 day managed isolation for any arrivals into the country, with very few exemptions granted;
- A national centre and system for rapid contact tracing had been established to coordinate and support regional contact tracing teams;
- COVID-19 testing was available at a number of GP clinics and community-based assessment centres (CBACs);
- Government issued QR codes that work with the NZ COVID Tracer app were available for businesses to generate and display for people to use to maintain records of the places they'd been;
- Anyone with cold or flu (COVID-like) symptoms was requested to stay home and call a Ministry of Health call centre (Healthline) to report symptoms and ask whether they should be tested for COVID-19

At the time of writing (the week of August 15th), a new outbreak had been detected. The first case was detected on the 10th of August in Auckland, after the person sought medical care due to symptoms, and subsequently went to get tested for COVID-19 at a CBAC. Initial contact tracing investigations found no direct link to anyone who had entered the country recently, or anyone who worked in a border or managed isolation facility. Contact tracing (backwards and forwards) was started immediately, with a number of people in the person's household testing positive by the time the government announced the case late on the 11th of August. Due to the number of cases already identified, with no clear link to the border, initial estimates were that there had been at least 2 or 3 generations of spread, and that there would most likely be a number of cases in the community across Auckland. The action was taken to immediately move the Auckland region to Level 3, with the rest of the country moving to Level 2, and travel restrictions across the regional boundary to prevent any new spread from Auckland. As with the initial outbreak of COVID-19 in Aotearoa, disease modelling is again playing a role in informing government response, and the current report is dedicated to this effort.

The highly responsive and informative nature of previous simulation models has been vital for rapid response, but it is important that they are improved upon in order to better represent the heterogeneous nature of people, communities, and disease spread. Many of the models used so far for rapid decision making have made convenient but often unrealistic assumptions about the homogeneity of both attributes and interactions of Aotearoa's population.

The current study

Traditional models of disease spread assume a well-mixed, homogeneous population. This approach is mathematically tractable but unrealistic and makes it difficult or impossible to track fine-grained behaviour of epidemic spread. Instead, modern models of disease spread use an underlying network of contacts or potential disease transmission routes[4]. Following these modern model approaches, we have built a preliminary version of a detailed interaction network of ≈5 million individuals for Aotearoa New Zealand. Each individual has the demographic characteristics of age, sex, ethnicity, and geographic location (Statistical Area (SA2)) of usual residence. Individuals are placed in dwellings, with other individuals, in the same geographic location (SA2). Besides dwellings, individuals have places of work and/or education, and participate in *community events* (socialising, shopping, etc). The latter include long-range travel within Aotearoa New Zealand.

The interaction network is a static representation of a typical week, with interaction patterns fixed over time. This means that there is no explicit separation weekdays and weekends, and community events do not turn on and off. This network provides the structure on which we implement a stochastic, dynamic contagion model. The contagion model allows us to include effects such as exposure to multiple infected individuals and different scenarios for social distancing, contact tracing, and/or quarantine policies. Despite the static interaction network, transmission probabilities within different contexts, as well as

behavioural parameters (e.g. testing rates), can be updated dynamically to reflect changes such as a move to a different Alert Level. The ability to modify transmission probabilities within specific contexts is a major advantage of the network approach over some of the early models. This gives us the ability to simulate the effects of different Alert Levels and non-pharmaceutical interventions (NPIs), and different transmission and disease progression by individual demographic factors and context type.

Methods

Interaction network

The interaction network models a population of ≈5 million individuals who have demographics (age, sex, and ethnicity) and location of usual residence at Statistical Area 2 (SA2) level from the StatisticsNZ 2018 Census[5]. The network includes an explicit layer for each of the three main contexts for interaction that people have: *home*, *work*, and *school*. These layers are built primarily using the linked data within the StatisticsNZ Integrated Data Infrastructure (IDI)[6], in conjunction with other official data sources including School Roll information from Ministry of Education[7]. Other than those three contexts/layers, we include a single additional layer — the *Community layer* — which is a proxy for all the other contact *events* people might have. The community layer is currently a very rough approximation of all other contacts, including interactions with known contacts (e.g. meeting up with a friend or family member who you don't live with, or attending events such as church or sports groups) as well as casual contacts with strangers (e.g. interactions while shopping or on public transport). The community layer is also used to include the effect of long-range travel — individuals can attend an event outside of their dwelling region. Individuals have a direct potential transmission route to others if they live in the same dwelling, are co-workers, go to the same school, or attend the same 'event'.

Individuals

The individuals in this model are the population of individuals from the Statistics NZ 2018 Census[5]. We extract counts of the population with usual residence in each SA2, by the demographic characteristics age (with age bands [0-14,15-29,30-64,65+]), sex (Male or Female), and ethnicity (Māori, Pacific peoples, or Other). This gives a base population of ≈4.7million individuals.

Some limitations of this approach are that, even with imputation from administrative data, the Census 2018 totals were lower than the Estimated Resident Population (ERP) at that time which was 4.9 million, and the population of Aotearoa has increased since then – estimated to pass 5 million in March 2020. In future we could increase the population of individuals in our network using population projections, but as we would be missing any linked data from the IDI[6] for them, we would simply be scaling up the structures we already have (e.g. dwelling sizes). This scaling without changing network structure is not expected to change the dynamics of the contagion spread.

Dwelling layer

This layer captures where people live and who they live with. We choose to use the unit of a 'Dwelling', rather than a 'Household', as we believe it better captures the number of people who an infected individual could be interacting with and potentially transmitting COVID-19 to. The definition of a dwelling used here matches that of the StatisticsNZ Census: "A dwelling is any building or structure — or its parts — that is used, or intended to be used, for human habitation"[5]. This covers all buildings with regular occupancy, including large dwellings like prisons, defence force bases, aged care facilities, and hostels. There can be more than one dwelling within a building. For example, an apartment building is made up of multiple individual dwellings. In the current implementation of the model, we do not categorise dwellings into different types, and have the same transmission rates within all dwellings. This means we will not be capturing any higher spread, e.g. in prisons or aged care facilities, in this model.

Our goal is to create dwellings with the right number of people living together, along with their ages, to capture inter-generational living effects. We construct this layer using aggregated confidentialised distributions and count-type information extracted from the IDI[6], using the 'dwelling (usual residence)' information from StatisticsNZ 2018 Census[5]. We start by building a list of all the dwelling sizes in each SA2 and what we call *age structure types* — the presence or absence of people within different age bands (roughly children, young adults, adults, and older). Each individual in NZ is linked to a single dwelling within their SA2. Note that we do not reconstruct the actual dwelling data; instead, we create an instance of a representative network, from an ensemble of possible networks, ensuring that features like the distribution of dwelling sizes and inter-generational living are accurately reproduced. Approximately 10% of people in the 2018 Census had no dwelling information, only an SA2 location. For people of Māori and Pacific peoples, this figure is closer to 20% who are missing dwelling information in the 2018 Census. To mitigate the effects of this missing data we accommodate these individuals by assigning them to additional representative dwellings that we cloned from other dwellings within the corresponding SA2, respecting the dwelling sizes and age structures in the local area.

The use of the 'usual residence' rather than 'census night' populations means that we do not capture the transient occupancy of hotels, for example, but instead capture the main place where people would be living. Although we believe that 'usual

residence' counts better reflect the place people would be living, especially during higher Alert Levels, there are some issues with data. Firstly, 'usual residence' is not consistently defined, e.g. high school students living in hostels are instructed to use their 'home' address rather than the hostel address, whereas university students living in hostels are instructed to use the hostel address. And secondly, the definitions are not consistently interpreted by people filling out the forms. We found this to be particularly problematic when the place people think of as 'home' and the place they are living is different. This showed up in the data for large dwellings such as prisons and defence force bases, and we believe there are some data quality issues for these large dwellings in particular that we aim to address through further investigation in future iterations of the network.

School layer

This layer captures where people go to school. We use Ministry of Education data[7] which give us the total roll (with prioritised ethnicity counts), co-educational status, year levels, and location at SA2 level for each school. We link individuals in the population to schools by matching their age, sex, and ethnicity. This process uses the distance between SA2s (based on adjacent SA2s with at least one land connection) to preferentially assign students to schools closer to them. This layer includes schooling from Early Childhood Education to Secondary School. We do not include tertiary education in this version of the network.

Workplace layer

This layer captures where people go to work. As with the home layer, we can only extract confidentialised distribution and count-type data from the IDI[6]. We consider workplaces by industry sector (ANZSIC06 level 1 code) within each Territorial Authority (TA), and ensure that there are the right number of firms in each firm size band [1–5,6–9,10–19,20–49,50–99,100+]. The exact firm size within a band we sample from the full list of firm sizes (not broken down by sector) within each TA. We then assign individuals to workplaces (workers) matching the demographics of the employees (age, sex, and ethnicity) at the sector level proportions for that TA. Similar to the school layer, we have a distance measure between adjacent TAs to assign workers to workplaces with probabilities based on the distance between home and workplace. This is important as many workers commute to workplaces in a different TA from where they live. Finally, we note that ≈5% of workers have more than one job, we randomly assign some workers to multiple jobs matching the numbers in each TA. The data we used for building this layer was the latest data in the IDI for tax (PAYE) payments, which was the month of January 2019.

Community layer

This layer captures all remaining interactions that individuals may have, and which are not already considered in the three previous layers. It includes, for instance, grocery shopping, playing sports, attending religious services, or participating in any other social activity. In the absence of real-world data on community interactions, this layer currently provides a very rough approximation of community interactions during a typical week. there is extensive literature on group structure and social networks, includeing the degree distributions found in them see, for example[8,9] and the references withing them. Typical social networks follow a heavy-tailed (power-law type) distribution for the number of interactions per individual. We have endevoured to ensure that the number of interactions that people have through these events is roughly consistent with data on face-to-face interactions from sources such as the 2017 International Social Survey Programme for New Zealand (ISSP2017)[10], and POLYMOD[11]. ISSP2017[10] found a heavy tailed distribution in number of contacts, with 3% reporting more than 100 contacts in a typical workday. POLYMOD[11] found that respondents had an average of around 15 contacts per day, and that ≈30% of those were outside of home, school, or work. Combining these findings gives us an average number of community contacts for a week, outside of home, work, or school, of around 10–20. Specifically, in our network we assume that the number of interaction events that people participate in follows a Poisson distribution with a mean of 3 events per person. We then use a power-law (heavy-tailed) distribution[9] for event sizes (number of people interacting in that 'event'), with a mean of size 5.

People can participate in both local and long-range events. Local participation means that individuals with a dwelling in a particular TA attend events in that same TA, while long-range participation means that individuals leave the TA where they live to attend events in other TAs, mixing with locals. We specify 5% of links as long-range. This percentage is a rough estimate of the number of people who leave their home TA in a typical week (not school holidays) from DataVentures mobility reports[12]. Future work is needed to validate this percentage, and adjust to match data on travel by different demographic characteristics. In order to determine which TA individuals will 'travel' to, we use 2017 cellphone movement data on the relative probabilities of long-range movements between different pairs of TAs from Qrious 2017[13].

While our current assumptions are informed by these previous data on interactions, there is still much work to be done to implement more complexity in the community layer. This would include introducing variation in contacts by geographic and demographic characteristics (i.e., ethnicity, sex, age), better implementing structural elements, such as repeated connections between individuals across different contexts, and representing close contacts through e.g. visits to friends and family, separately from random contacts through e.g. public transport or shopping.

Contagion spread

We implement a stochastic model of infection dynamics on the above interaction network. Each individual in NZ is represented by a node in this network and additional *group* nodes are used to represent the different contexts through which individuals can interact and transmit infection. Group nodes are also used to calculate other changes in state such as quarantine status. The essential components of our model are Markovian; that is, there is no history dependence. Thus, in contrast to some branching process models, the information required to advance the system from one time step to the next is contained in the current state (statuses of all individuals in the network). We also allow dependence on some additional global state variables. For example, transition rates between states may depend on the time at which a policy intervention becomes active. For the key processes, the time between events is random and determined from the rate parameters that, in turn, can depend on the instantaneous state of the whole network.

The Epidemics on Networks (EoN) Python package[14] formed the basis for our initial contagion model, but we have significantly customised and extended this package to allow for additional functionality in our modelling. Some of these modifications are inspired by similar models and simulation frameworks developed recently for the spread of COVID-19 in other countries[15–17]. We use the Gillespie algorithm[18] to simulate our dynamics. This algorithm is a so-called exact algorithm for simulating realisations from a collection of independent *transition* processes with rates (or *hazards*) of the form:

$$\text{Probability of transition } i \text{ per unit time} = h_i(\text{system state}) \tag{1}$$

for $i = 1, \ldots, n$. In this context, a transition consists of one or more individuals changing their state; for example, a susceptible individual becoming exposed due to an encounter with an infected individual. This simulation approach was popularised in the stochastic chemical kinetics literature, but also has a long history of applications to population dynamics[19] and is now standard in the network contagion literature[4].

The key inputs to the simulation algorithm are, thus, the contact network; the possible infection and quarantine states of individuals, and their initial states; and the rate functions for all possible transitions. We describe these possible states and rate functions in detail in the Technical Appendix.

Assumptions and Limitations

Contagion spread is known to be very dependent upon both the number and the nature of the contacts a person has. Both of these are distributed heterogeneously across populations. The contact network we have constructed for these simulations attempts to capture some of the most important of these effects but is limited in what it can reproduce by both the availability of data and available time.

Key points about the interaction network

We know from survey data that the number of contacts people have is heterogeneously distributed. For example, young people have more contacts, predominantly with other young people, while people of 65 have far fewer. We only partly account for this heterogeneity — young people tend to be in larger dwellings, and will have school interactions as well. Older people will be mostly in smaller dwellings, and will not have workplace or education interactions. This means that the total number of interactions (over all four types of interaction) varies with age, due to the non-community layers.

Limitations of the current network relevant in this scenario:

- We have no ethnicity structure within different sizes of dwelling, beyond preserving demographics at the level of SA2s. We know that this is incorrect, e.g. Pacific peoples are statistically much more likely to be in higher occupancy dwellings, than people of other ethnicities.
- No sex structure within dwellings. Our checks have shown that males and females are reasonably evenly distributed between dwelling types, with the exceptions of people older than 60 (more women than men are likely to live alone) and for prisons and military facilities, which we do not yet explicitly include.
- Currently, we do not have age and ethnicity structure built into our community layer interactions. That is, individuals are linked to community events homogeneously, without regard to their age, sex and ethnicity. This is unrealistic in several ways, as not everyone within a TA is equally likely to interact. For instance, people are more likely to be in contact with people of a similar age in social gatherings or much more likely to have casual contact interactions (such as shopping) with people who live near them, within a TA.

In addition, such a complex simulation model involves numerous parameters governing the basic epidemic dynamics, many of which are hard to determine. The model structure can only ever be an imperfect reflection of the true disease progression dynamics. As such we emphasise that all estimates should be explicitly considered to be model-based, given our best current understanding, and will be more or less accurate depending on how well the model reflects different aspects of reality and/or how robust the model is to key assumptions.

Key simulation settings

The results presented in this report reflect a scenario with no intervention, with seedings reflecting re-emergence of COVID-19 infections in the Auckland region at the end of July/start of August 2020. Full details of our simulation are given in the Technical Appendix. Here, we highlight some key assumptions.

Firstly, in the absence of firm information about how the initial seeding for the re-emergence may have occurred — for example from a border or managed isolation and quarantine (MIQ) worker, or from environmental transmission — we assume that the demographic details of the true seed case are the similar to those of the first detected cases of community transmission. We hence seed with cases in a dwelling that is randomly selected from a set of dwellings of the same size, age structure, and SA2 as that of the first confirmed cases, as well as one infected student and one infected teacher at each of two schools matching the profile of the cases also reported in early August. These cases are seeded in the 'Exposed' state.

The main parameter that varies systematically across simulations is the transmission rate (β). This represents the probability per unit time of a transmission event, given a contact between a susceptible and infected individual. The values used are shown in Table 1. This parameter is a key input for determining R_0 (or and/or R_{eff}) in compartment-based — i.e. non-network — models. The values chosen here were calibrated against generation times found to be a good fit for models of COVID-19 transmission, as described in the Technical Appendix, and are also consistent with typical values used in compartment models. Note that the transmission rates denoted as 'Lower' and 'Higher' are relatively tight bounds on the 'Medium' or 'Typical' transmission case. That is, they should not be treated as best case or worst case scenarios. For each value of the transmission rate, we run 50 different simulations.

Transmission rate (β, units: per day)	Description
0.4	Lower (20% lower)
0.5	Medium (typical value)
0.6	Higher (20% higher)

Table 1. Transmission rates considered in simulations.

Though R_{eff} is not a basic input of our model, we can post-process our overall results to provide estimates of it[20]. In the present work we post-process our results using the simple approximation[21,22] $R_{\text{eff}} \approx r \times$generation time + 1 to provide an indication of this, where r is the observed exponential epidemic growth rate. This leads to estimates of about 1.60, 1.72, and 1.81, respectively, for the β values in Table 1 and our other simulation settings.

Other key assumptions include:

- Infected individuals get split into either asymptomatic or symptomatic, and if symptomatic into *mild* or *severe* cases.
- Asymptomatic and presymptomatic infections are assumed to have a reduced infectiousness and a lower testing rate due to the lack of symptoms. The literature regarding infectiousness of asymptomatic and presymptomatic individuals is mixed, with some more recent results suggesting that it may be almost as high as for symptomatic individuals. This would increase the number of infected individuals above the numbers reported here,
- To capture the behaviour of New Zelaanders in the period prior to August 11th 2020, contact tracing is turned off, symptomatic people keep going to work/school, and only confirmed cases self-isolate.
- We assume a reasonably high rate of people getting tested if they are moderate or severely symptomatic. Consequently, $\sim$70% of infected individuals with moderate to severe symptoms would receive a positive test during the normal progression of their illness. Note: this is close to a 100% testing rate, given RT-PCR false negative rates of $\sim$20% at best, and depending on the timing of testing relative to symptom onset[23]
- We assume a lower rate of people getting tested if 'mildly' symptomatic such that $\sim$20% would receive a positive test result during the normal progression of their illness
- A large number of contagion parameters are set based on past modelling and best guesses from available data, as described in the Technical Appendix.
- In this model, we assume a common basic transmission rate (per interaction) regardless of context, though susceptibility varies by age bands as described in the Technical Appendix.

Findings

Number of cases at the time of detection

In the following graphs we only show the results for our best estimate of the transmission rate ($\beta = 0.5$), and show values for all 50 simulations alongside the median and inter-quartile (IQ) range. We include tables for key values at selected times, including for the higher and lower transmission rates. Because of the stochastic nature of virus transmission, and the heterogeneous nature of the network, we see a high level of variability between runs.

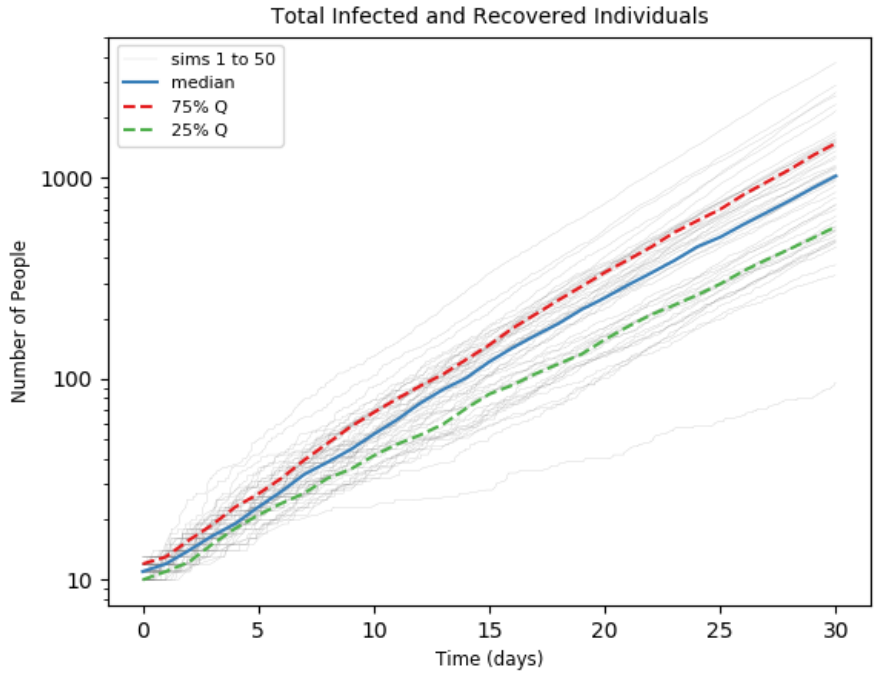

Figure 1. Total cumulative number of cases through time *on a log scale* after seeding cases in all family members in one dwelling at day 0, and one infected student and one infected teacher at each of two schools matching the profile of the cases reported early August. Light grey lines show the results for each of the 50 simulations, which had cases seeded in different dwellings but of the same size (between 5 and 9 people) in the same area of Auckland as the first detected cases, as well as different schools but the same number of infected cases (one student and one teacher in each of two schools). Median is shown in solid blue, with inter-quartile range in the heavy dashed lines. Medium transmission rate of $\beta = 0.5$.

Figure 1 shows the total (cumulative) number of cases in Aotearoa New Zealand for the first 30 days after the seeding cases. This is the number of cases with a low to moderate level of testing and no interventions; specifically, no contact tracing and no self-isolation (except for confirmed cases).

It is most likely (and was subsequently confirmed) that the detected cases of 11 August were not the first cases in the transmission tree, but we do know that they were the first *detected* cases, and that when they became infected there were multiple other concurrent cases. In order to estimate how many cases we might expect there to be at the time of first detection (11 August) we look at how long it took until the first detected case in our simulations. However, as our simulations begin by seeding with more than a single exposed case, we expect our times until first case detection to be shorter than if only one seed case was used. Assuming our seeding cases are second generation cases, we can expect that the case or cases that infected them to have been infected an additional five days earlier. The estimate of five days is based on the generation times as calibrated in the Appendix.

We find that for a medium transmission rate and a moderate assumed level of symptomatic cases getting tested, the first case would get confirmed in testing a mean of 7 days after our initial seeding cases (median 7, inter-quartile range of [4,10]). Assuming our seeding cases are second generation, this would put the timings from first case to detection at about 1–2 weeks. Based on a first detection of 11 August and a time to detection of about $7 + 5 = 12$ days, this puts the earlier case (or cases) at about 31 July, which is consistent with subsequent contact tracing information. Table 2 shows the time to detection for higher and lower transmission rates. We note that there is no significant difference in the distributions of time to first confirmation for

different transmission rates in these sets of realisations, but there is high variability between runs.

Without interventions, the first infected person outside the Auckland region would be confirmed in testing a mean of 17 days after initial seeding cases (median 17, IQ range of [13,21]). This suggests that the much increased rates of testing and contact tracing brought forward the detection of cases outside the Auckland region by around a week, relative to the scenario of unchanged testing rates and no contact tracing.

	Number of days after seeding until first confirmed case	
Transmission rates	All locations	Outside of Auckland
Lower	6 [4, 9]	18 [13, 23]
Medium	7 [4, 10]	17 [13, 21]
Higher	5 [4, 8]	15 [12, 18]

Table 2. Number of days after initial seeding until first confirmed case for various transmission levels and for both any location and outside of Auckland. Data shows median with inter-quartile range [25% – 75%] in brackets.

We find that without intervention, and with a medium transmission rate, at 7 days after the seeding cases our simulations have a mean of 29 total cases (median 27, inter-quartile range of [24, 32]). Without intervention, and with a typical transmission rate, at 17 days after the seeding cases our simulations have a mean of 141 cases (median 142, inter-quartile range of [93,178]). The values for higher and lower transmission rates are shown in Table 3. Assuming that our seeding corresponds to second generation infections, and that they were infected by cases from approximately 31 July, this latter set of numbers gives us an estimate of between 100–200 cases at approximately 21 August, 2020. We note here that this is only the estimate for cases from the specific seeded protocol used. In particular, if on the 4th of August there were more than the 9–13 initially infected (or, more precisely, exposed) cases that we seeded the model with, the later number of number of cases would also be higher.

	Number of infected cases	
Transmission rates	7 days after initial seeding	17 days after initial seeding
Lower	24 [21,26]	88 [67, 105]
Medium	27 [24, 32]	142 [93, 178]
Higher	36 [30, 42]	229 [177, 316]

Table 3. Number of infected cases (including unconfirmed cases) for various levels of transmission rate and for both 7 days and 17 days after initial seeding. Data shows median with inter-quartile range [25% – 75%] in brackets.

Spread to outside Auckland

Without intervention, and with a medium transmission rate, the first infected person outside the Auckland region occurs after a mean (and median) of 5 days after the seeding cases are first initialised. The first transmission outside Auckland occurs sooner with higher transmission rates, and later with lower transmission rates (Table 4), but there is high variability in this as it depends whether the infected individuals had contact links outside Auckland (Figure 2). At the estimated time of detection (7 days after the second generation seeding cases, and approximately 11 August), there is transmission to outside of Auckland in 76% of model runs for medium transmission. The majority of this transmission occurs to Waikato, Bay of Plenty, and Northland, as we would expect, since our links are weighted by the proportion of travel between Territorial Authorities in the Qrious 2017 dataset from cellphone movements[13].

	Days until first transmission outside Auckland	
Transmission rates	Mean	Median [LQ,UQ]
Lower	6	5 [4,8]
Medium	5	5 [3,7]
Higher	4	3 [2,5]

Table 4. Days until first transmission outside Auckland from seed cases

The proportion of total cases outside Auckland at 7 and 17 days after the seeding cases is shown in Figure 3. For a typical transmission rate, and without interventions, a mean of 6% and median of 4% of infections (IQ [2%, 8%]) occur outside of the Auckland region 7 days after the seeding cases. With no intervention, this has increased to a mean of 9% and median of 8% (IQ [5%,13%]) of total cases outside the Auckland region, 17 days after the seed cases (i.e. around the 21st of August).

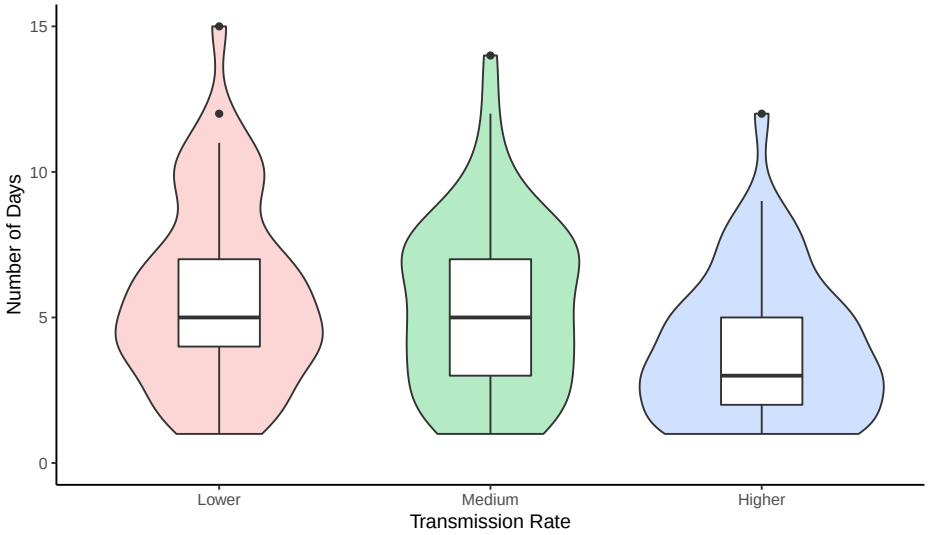

Figure 2. Day of first transmission out of Auckland for the three different transmission rates.

	Proportion of infected cases outside Auckland	
Transmission rates	7 days after initial seeding	17 days after initial seeding
Lower	0.05 [0.0, 0.1]	0.07 [0.04, 0.12]
Medium	0.04 [0.02, 0.08]	0.08 [0.05, 0.13]
Higher	0.05 [0.03, 0.07]	0.08 [0.07, 0.11]

Table 5. The proportion of total infected cases outside of Auckland 7 and 17 days after the first case, across the three different transmission rates. Figure 3 visualises the proportion of infected cases outside of Auckland for the medium transmission rate across 30 days.

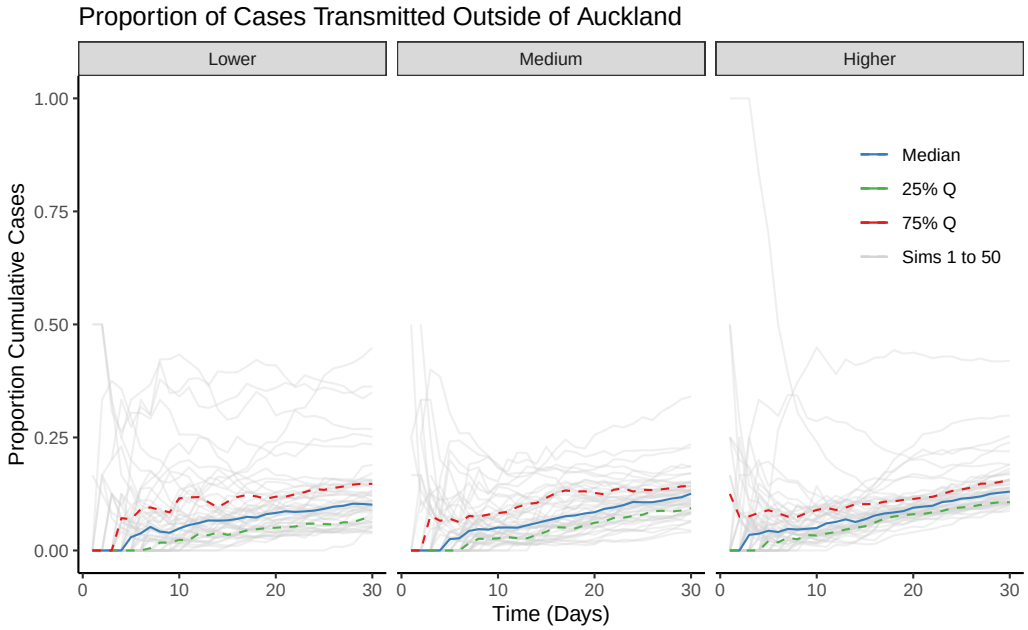

Figure 3. Proportion of all cases transmitted to someone living outside the Auckland region for each day after seeding cases in all family members in one dwelling, along with one student and one teacher at each of two schools, at day 0. Light grey lines show the results for each of the 50 simulations, which had cases seeded in different dwellings but of the same size (between 5 and 9 people) in the same area of Auckland as the first detected cases. Median is shown in solid blue, with inter-quartile range in the heavy dashed lines. Note that this proportion will differ very slightly from the total case count proportion, as it doesn't include the initial 9–13 seed cases. This effect becomes small later in time when there is a high case number.

Discussion

We have simulated the re-emergence and spread of COVID-19 using a stochastic, individual-based network model of all ≈5 million individuals in Aotearoa New Zealand. This reflects the detection of new cases in early August 2020, with an origin traced back to at least late July. The network includes interaction pathways, referred to as 'contexts' in the network, broken down into network 'layers' representing the structure of home, work, school, and community interactions. Under the 'no intervention' scenario we find that infections resulting from initial seedings that reflect the cases reported in early August are typically detected around one week after seeding. At that point the simulations estimate there would be around 20–40 other cases in the community. Assuming these cases are second generation cases, this in turn gives an indication that the source of these detected cases was likely a case in late July, 2020, though from an unknown source.

Without intervention, infection is quickly transmitted outside of the Auckland region. Simulations indicate that the first infected person outside of Auckland likely occurs around five days after the seeding cases of our simulation, (about 1–2 weeks after the assumed true index case of approx. 31 July). We therefore infer that spread outside of Auckland likely occurred before the first known cases were detected. Without contact tracing or other interventions, we would expect cases to be reported around one to two weeks after the first reported Auckland case. With intensive contact tracing this would likely be shorter, though we do not model this scenario in the present work. Again under the no intervention scenario, we expect that as of about 21st August we would expect to see around 100–200 cases in the community, with around 8–15 of those cases occurring outside of the Auckland region.

The results presented here have several limitations. For example, we know that the Auckland August cluster predominantly involved Pacific and Māori communities; however, as noted in the Assumptions and Limitations section of this report, interactions in the network are homogeneous with respect to ethnicity, except for spatial effects. As a consequence, the model has less ethnicity assortativity in the workplace and community layers than would be expected. This lack of assortativity limits the specificity of the results from the model for ethnicity. Furthermore, the lack of assortativity in the community layer in general means that our model could overestimate the rate of disease spread to communities beyond those who were initially infected. If there are differential rates of disease transmission in different communities, which is likely, though currently unquantified and not modelled, then this would affect the rate of spread and total number of cases. For example, if the detected cases are in a community with a higher rate of disease transmission than the data calibrated to, then the expected number of cases estimated here could be too low.

In addition to the above specific limitations, like all modelling studies, we make numerous additional simplifying assumptions, which may be more or less reflective of the actual real-world situation. Thus our results should, in general, be treated as a rough guide to the developing situation.

Acknowledgements

The authors acknowledge the support of Statistics New Zealand, ESR, and the Ministry of Health in supplying data in support of this work. The authors would also like to thank Shaun Hendy and Mike Plank for their useful comments on the manuscript; Patricia Priest, Amanda Kvalsvig, and Michael Baker for advice on epidemiological parameters; and NeSI for providing substantial computational resources and support. This work was funded by the New Zealand Ministry of Business, Innovation and Employment and Te Pūnaha Matatini, Centre of Research Excellence in Complex Systems. The Te Pūnaha Matatini Technical Review Group also provided substantial and helpful feedback and comments on this manuscript.

References

1. James, A., Hendy, S. C., Plank, M. J. & Steyn, N. Suppression and Mitigation Strategies for Control of COVID-19 in New Zealand. Tech. Rep., Te Pūnaha Matatini (2020). DOI: 10.1101/2020.03.26.20044677.

2. Plank, M. J. *et al.* A stochastic model for COVID-19 spread and the effects of Alert Level 4 in Aotearoa New Zealand. Tech. Rep., Te Pūnaha Matatini (2020).

3. Government, N. Z. COVID-19 alert system in New Zealand. https://covid19.govt.nz/alert-system/alert-system-overview/ (2020). Accessed 2020-08-29.

4. Kiss, I. Z., Miller, J. C. & Simon, P. L. *Mathematics of Epidemics on Networks: From Exact to Approximate Models* (Springer International Publishing, 2017).

5. Statistics New Zealand. 2018 Census. https://www.stats.govt.nz/2018-census/ (2018). Accessed: 2020-08-12.

6. Statistics New Zealand. Integrated Data Infrastructure. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/. Accessed: 2020-08-12.

7. Education Counts. School Rolls. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/ (2020). Accessed: 2020-08-12.
8. Vasques Filho, D. & O'Neale, D. R. J. Transitivity and degree assortativity explained: The bipartite structure of social networks. *Phys. Rev. E* **101**, 052305 (2020).
9. Muchnik, L. *et al.* Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Sci. reports* **3**, 1–8 (2013).
10. Li, E., Wu, I., and Milne, B. Methods and procedures for the 2017 International Social Survey Programme (ISSP) for New Zealand. https://figshare.com/articles/ISSP2017_Social_Networks_III/5405554?file=12732452 (2018).
11. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine* **5**, 1–1, DOI: 10.1371/journal.pmed.0050074 (2008).
12. DataVentures. School holidays domestic tourism report. https://reports.dataventures.nz/population/20200720-tourism-new-zealand-school-holidays-report-20th-july-2020.html (2020). Accessed 2020-08-29.
13. QriousNZ. Qrious inter-regional ground travel data. https://www.transport.govt.nz/mot-resources/research-papers/inter-regional-ground-travel-data-from-qrious/ (2019). Accessed: 2020-08-10.
14. Miller, J. C. & Ting, T. EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *J. Open Source Softw.* **4**, 1731, DOI: 10.21105/joss.01731 (2019).
15. SEIRS+ Model Framework. https://github.com/ryansmcgee/seirsplus. Accessed: 2020-08-12.
16. OpenABM-Covid19 model. https://github.com/BDI-pathogens/OpenABM-Covid19. Accessed: 2020-08-03.
17. Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M. & Prokopenko, M. Modelling transmission and control of the COVID-19 pandemic in australia. *arXiv preprint* (2020). arXiv:2003.10218.
18. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
19. Kendall, D. G. An artificial realization of a simple "Birth-and-Death" process. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **12**, 116–119 (1950).
20. Breban, R., Vardavas, R. & Blower, S. Theory versus data: how to calculate r0? *PLoS One* **2**, e282 (2007).
21. Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect Dis Model.* **5**, 129–141 (2020).
22. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).
23. Kucirka, L., Lauer, S., Laeyendecker, O., Boon, D. & Lessler, J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based sars-cov-2 tests by time since exposure. *Annals Intern. Medicine* **173**, 262–267, DOI: https://doi.org/10.7326/M20-1495 (2020).
24. Ferretti, L. *et al.* Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368** (2020).
25. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (covid-19) infections. *Int. J. Infect. Dis.* **93**, 284–286, DOI: https://doi.org/10.1016/j.ijid.2020.02.060 (2020).
26. Bi, Q. *et al.* Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. *The Lancet Infect. Dis.* **20**, 911–919 (2020).

Appendix

Interaction network

In order to effectively simulate COVID-19 on a network, it is necessary to first have a suitable interaction network that gives some sort of approximation of the person-to-person interactions through which it is that disease transmission occurs. Typically, constructing and then working with such a network is hampered by two problems: firstly, it is difficult to gather empirical data at sufficient scale and of sufficient quality to construct a contact network for even a small country; and secondly, even a relatively small country the size of Aotearoa New Zealand, with $\approx 5 \times 10^6$ inhabitants, has $\approx 25 \times 10^{12}$ possible connections, which can make for computationally hefty simulations.

We have endeavoured to address both these issues in constructing an ensemble of *multilayer bipartite* networks for Aotearoa New Zealand. We create a separate layer for each main context in which interactions occurs. The current version of our network has four layers, one each for dwellings, schools, and workplaces, which are constructed from data within the Statistics New Zealand IDI[6] and other official data sources, and a 'community' layer which captures all other interactions using knowledge of the significant topological features of other real-world interaction networks. A schematic of our network is shown in Figure 4. The same individuals persist in all layers of the network, but this multilayer structure helps to organise the different interaction types and allow for construction in parallel and for using context dependent transmission rates in the different layers. In Figure 4 we also show the *bipartite* nature of the network. Specifically that there are two distinct types of nodes: individuals and 'groups' or interaction contexts. Individuals are only connected to groups, and groups are only connected to individuals. The individual to individual interactions of traditional contact networks, necessary for contagion spread, occur *through* the group nodes, allowing us to modify the transmission rates (or turn them off entirely) by groups node type and details, e.g. industry sector of the workforce.

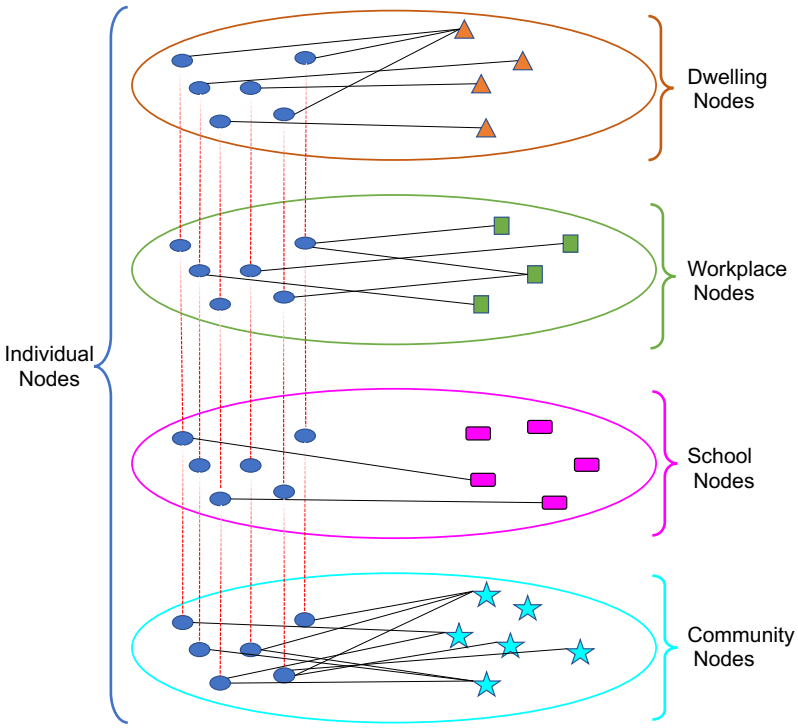

Figure 4. Diagram of the multilayer bipartite interaction network showing individual nodes on the left, which persist (are the same people) in all the layers of the network. The current version of the model uses four layers, one each for dwellings, workplaces, schools, and community.

State diagram

In the basic disease transmission (first letter), people begin as susceptible, S, transition to exposed, E, if infected. Then they subsequently transition into either an asymptomatic case, mild case, or moderate/severe case, depending on probabilities derived from their age band. For asymptomatic cases, there are no further state changes until recovery. But for mild and moderate/severe cases, they individuals enter a presymptomatic state, (B and P, respectively), before becoming symptomatic — states M and I, respectively. The exposed state is not infectious (related to incubation time), while the states A, B, P, M and I are infectious, but only the states M and I have symptoms. For the moderate/severe state, there is rate at which a proportion of cases will become hospitalised, H, and then a further proportion of those who will progress to need critical care, C. Finally, all cases progress either to recovered, R, or die, D. Note, recovered cases simply means that individuals are no longer infectious. They may still have ongoing symptoms and not be fully recovered, but we do assume that from this state they will no longer be hospitalised with COVID. Any of the post-exposure states (A, B, P, M, R, and I) have both an 'unconfirmed' (second letter U) and a 'confirmed' (second letter C) version. Individuals can transition from an unconfirmed to a confirmed state through the process of being tested and receiving a positive test result. Once confirmed, individuals are assumed to be isolated in their own home. Due to assumption of no testing for people without symptoms and no contact tracing in the current report, the transition rate for asymptomatic and presymptomatic individuals to being a 'confirmed' case, and thus being isolated, is set to zero. All possible individual states of our model are shown in Figure 5.

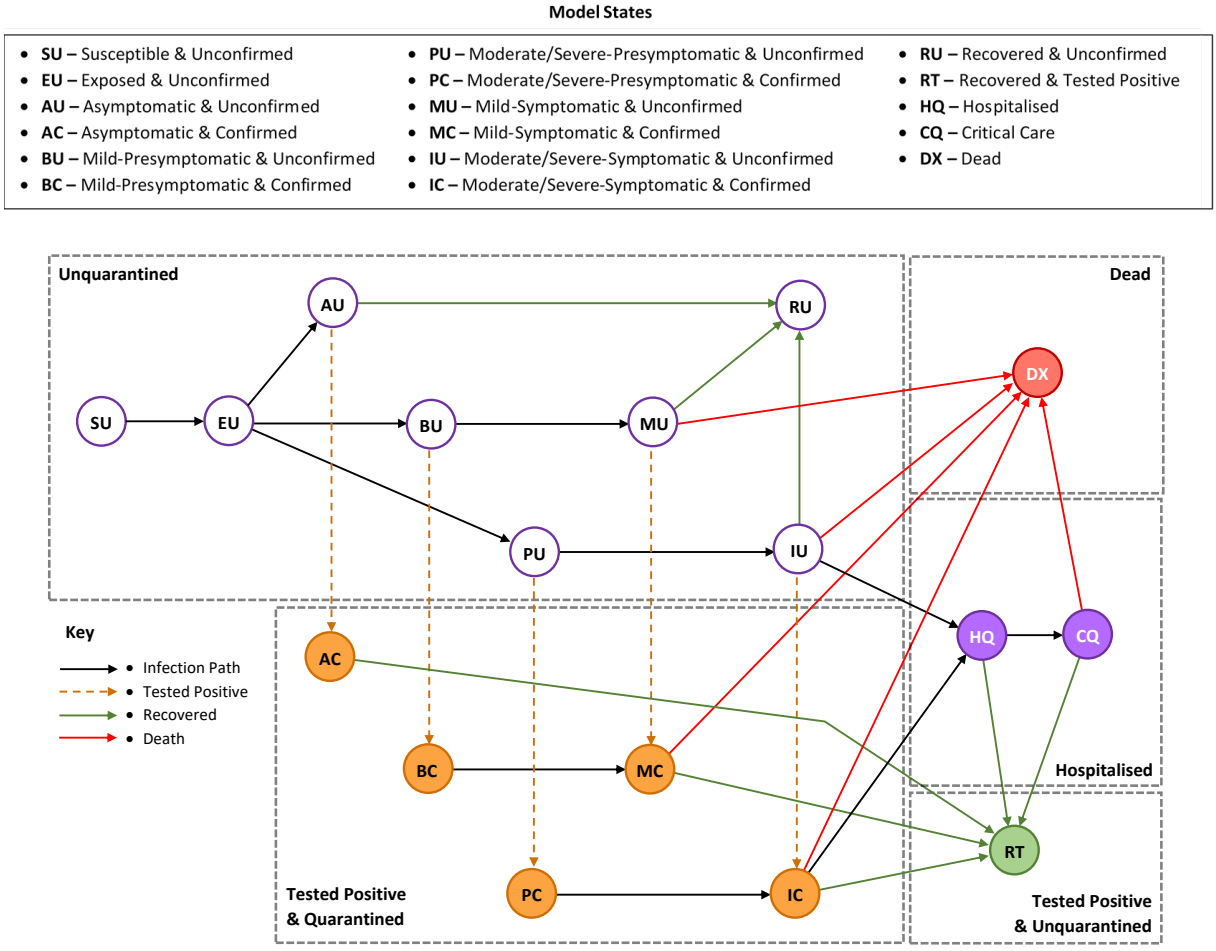

Figure 5. State and Transition Diagram for Network Contagion Model. We note that, for the parameter values used here, there is no transition from AU→AC, PU→PC, or BU→BC due to no testing of people without symptoms and no contact tracing, and there is no transition from MU→DX, MC→DX, IU→DX, or IC→DX due to the assumption that all life-threatening cases would be hospitalised first.

Model parameters

The following parameters are used in the model rate functions described below.

Table 6. Parameter table

Symbol	Value	Units	Description	Source
α_A	0.1	per day	Rate at which infected individuals recover for asymptomatic infections (**A** $\rightarrow$ **R**)	Estimated from Fraser Group parameters *
α_{CD}	0.083	per day	Fatality rate for infected individuals hospitalised in critical condition (**C** $\rightarrow$ **D**)	Estimated from Fraser Group parameters*
α_{CR}	0.053	per day	Rate at which individuals hospitalised in critical condition recover (**C** $\rightarrow$ **R**)	Estimated from Fraser Group parameters*
α_{HR}	0.111	per day	Rate at which hospitalised individuals recover (**H** $\rightarrow$ **R**)	Estimated from Fraser Group parameters*
α_{ID}	0.091	per day	Fatality rate for infected individuals with moderate/severe infections (**I** $\rightarrow$ **D**)	Estimated from Fraser Group parameters*
α_{IR}	0.083	per day	Rate at which infected individuals recover for moderate/severe infections (**I** $\rightarrow$ **R**)	Estimated from Fraser Group parameters*
α_{MD}	0.091	per day	Fatality rate for infected individuals with mild infections (**M** $\rightarrow$ **D**)	Estimated from Fraser Group parameters*
α_{MR}	0.083	per day	Rate at which infected individuals recover for mild infections (**M** $\rightarrow$ **R**)	Estimated from Fraser Group parameters*
β	0.5	per day	Transmission coefficient	Calibrated using generation times. See Fig. 6
ϕ_{age}	0.467, 0.917, 1.071, 1.5	–	Age structured susceptibility weightings for age bands 0–14, 15–29, 30–64 and 65+ respectively	Estimated from Fraser Group parameters[16]
δ_C	0.5	per day	Symptom progression rate for critical hospitalisation (**H** $\rightarrow$ **C**)	Not well known. Affects number in **H** and **C**, but not the number of new infections.
δ_H	0.2	per day	Symptom progression rate for hospitalisations (**I** $\rightarrow$ **H**)	Not well known. Affects number of new infections, as hospitalised people are effectively isolated.
δ_I	0.286	per day	Symptom progression rate for moderate/severe infections (**P** $\rightarrow$ **I**)	Calibrated from exposure to symptom onset times[1,2], and γ.
δ_M	0.286	per day	Symptom progression rate for mild infections (**B** $\rightarrow$ **M**)	Calibrated from exposure to symptom onset times[1,2], and γ.
ε_A	0.33	–	Relative infectivity of asymptomatic individuals	Estimated from Fraser Group parameters[16]

(continued on next page)

* All α values are derived from the total time in the Fraser report[16] after estimating γ and δ values

Table 6. (continued from prev)

Symbol	Value	Units	Description	Source
ε_B	0.5	–	Relative infectivity of presymptomatic mild individuals	Earlier TPM modelling[1,2]. More up to date estimates required.
ε_C	0.0	–	Relative infectivity of critically hospitalised individuals	Model assumption
ε_H	0.0	–	Relative infectivity of hospitalised individuals	Model assumption
ε_M	0.72	–	Relative infectivity of mild symptomatic individuals	Fraser Group parameters[16]. More up to date estimates required.
ε_P	0.5	–	Relative infectivity of presymptomatic moderate/severe individuals	Earlier TPM modelling[1,2]. More up to date estimates required.
η_C	0.05, 0.05, 0.1, 0.5	per day	Age structured rates at which hospitalised cases becoming critically hospitalised for age bands 0–14, 15–29, 30–64 and 65+ respectively ($\mathbf{H} \rightarrow \mathbf{C}$)	Estimated from Fraser Group parameters[16] from the UK.
η_H	0.003, 0.012, 0.222, 0.4	per day	Age structured rates of hospitalisation for age bands 0–14, 15–29, 30–64 and 65+ respectively ($\mathbf{I} \rightarrow \mathbf{H}$)	Estimated from Fraser Group parameters[16] from the UK.
γ	1	per day	Latent period rate, rate of becoming infectious from contracting infection ($\mathbf{E} \rightarrow \mathbf{P}$)	Existing TPM modelling[1,2]
μ_C	0.3, 0.417, 0.571, 0.88	per day	Age structured mortality rates of critically hospitalised individuals for age bands 0–14, 15–29, 30–64 and 65+ respectively	Estimated from Fraser Group parameters[16] from the UK.
μ_I	0.0	–	Mortality probability for moderate/severely infected individuals	Assume deaths only occur from hospital and critical care
μ_M	0.0	–	Mortality probability for mildly infected individuals	Assume deaths only occur from hospital and critical care.
ω	0.01	–	Relative risk of infection from confirmed cases in isolation	Related to compliance and ability to self-isolate. Not well known.
p_A	0.434, 0.391, 0.283, 0.215	–	Age structured proportions of infectious cases that are asymptomatic for age bands 0–14, 15–29, 30–64 and 65+ respectively	Estimated from Fraser Group parameters[16]
p_M	0.551, 0.583, 0.601, 0.518	–	Age structured proportions of infectious cases that are mild for age bands 0–14, 15–29, 30–64 and 65+ respectively	Estimated from Fraser Group parameters[16]

(continued on next page)

Table 6. (continued from prev)

Symbol	Value	Units	Description	Source
θ_0	0.0	per day	Rate of testing positive per unit time for asymptomatic and presymptomatic individuals	Assume that only symptomatic individuals would get tested
θ_I	0.2	per day	Rate of testing positive per unit time for moderate/severe symptomatic individuals	Policy and behavioural variable
θ_M	0.02	per day	Rate of testing positive per unit time for mildly symptomatic individuals	Policy and behavioural variable

We caution that while some parameter values have been taken directly from previous compartment modelling efforts by Te Pūnaha Matatini[1,2] and the Fraser Group[16], a number of parameters are difficult to determine and subject to large uncertainties. We will need to review, in particular, the parameters related to testing and contact tracing/isolation before running simulations which include these effects.

Generation times

In our network setting the key transmission parameter is the transmission rate β. In order to relate our transmission parameter to other models we can calibrate it using the so-called generation time (time from infection of the index case to infection of secondary cases, with no contact tracing or case isolation). We calibrated against typical values in the literature[24–26] to ensure appropriate transmission rates. The generation times for different transmission rates are shown in Figure 6. Values of $\beta = 0.4 - 0.6$ give a reasonable correspondence to the literature.

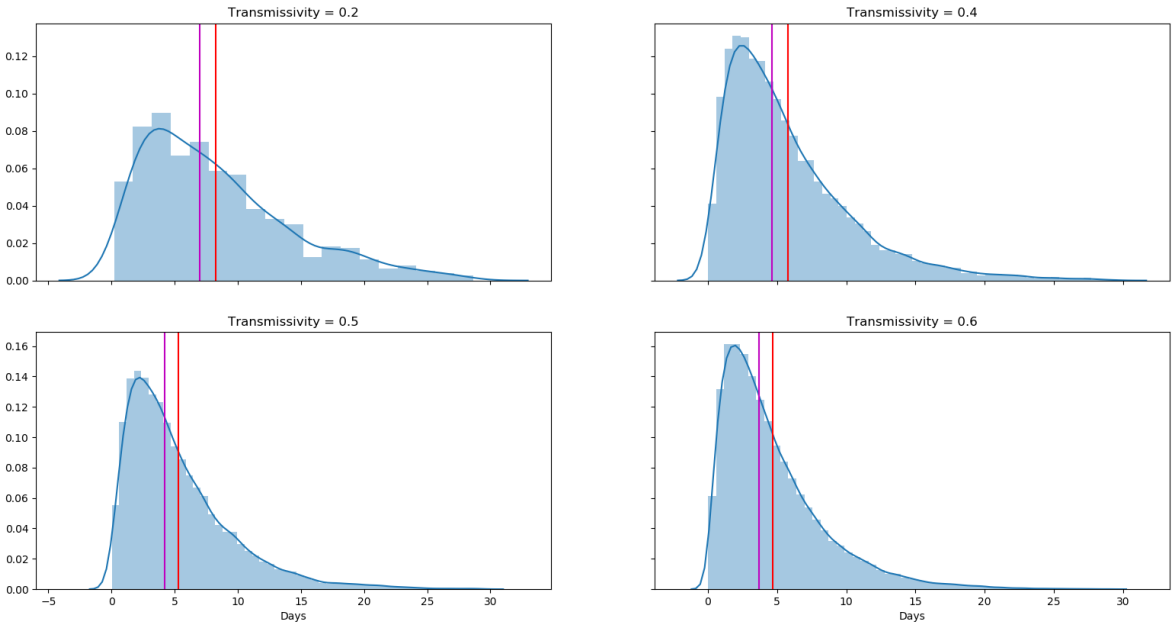

Figure 6. Histogram of Generation Times based on transmission rate β shown in blue, with the vertical lines showing the median (purple) and mean (red). Note the median is always lower than the mean (further to the left).

Rate Functions of the Model

Here we express the rate functions for transitions in our model, such that

$$\text{probability of transition } i \text{ per unit time} = h_i(\text{system state}) = r(\text{transition } i)$$

Each process in the present model can be associated with a specified node having a given state; we have one copy of each of the processes described below for each node in the specified state. A single node can be, and typically will be, associated with multiple processes. We include rate functions for contact tracing and isolation, though these are not used in the present set of simulations.

We use the notation $|X|$ to denote the number of neighbours (of the specified node) with state X, and $|X+Y|$ to denote the number of neighbours with state X or Y. The symbol N represents all neighbours such that $|N|$ is the number of neighbours. We will also use the notation $|X|_{hh}$ for the count of neighbours that are of state X that also share a dwelling with the node and $|X|_{\cancel{hh}}$ for those that do not share a dwelling.

For compactness, we will express all the transitions out of a state (of the specified node) as a single equation, and label the output state. The basic rate functions are given in the set of equations (2) below.

$$\begin{aligned}
r(SU \to X) &= \underbrace{\beta\phi_{age}\frac{\begin{pmatrix}\varepsilon_A|AU|+\varepsilon_B|BU|+\varepsilon_M|MU|+\varepsilon_P|PU|+|IU|+\varepsilon_H|HQ|+\varepsilon_C|CQ| \\ +\varepsilon_A|AC|_{hh}+\varepsilon_B|BC|_{hh}+\varepsilon_M|MC|_{hh}+\varepsilon_P|PC|_{hh}+|IC|_{hh} \\ +\omega(\varepsilon_A|AC|_{\cancel{hh}}+\varepsilon_B|BC|_{\cancel{hh}}+\varepsilon_M|MC|_{\cancel{hh}}+\varepsilon_P|PC|_{\cancel{hh}}+|IC|_{\cancel{hh}})\end{pmatrix}}{|N|}}_{\to EU} \\
r(EU \to X) &= \underbrace{p_A\gamma}_{\to AU}+\underbrace{p_M\gamma}_{\to BU}+\underbrace{(1-p_A)(1-p_M)\gamma}_{\to PU} \\
r(AU \to X) &= \underbrace{\alpha_A}_{\to RU}+\underbrace{\theta_0}_{\to AC} \\
r(BU \to X) &= \underbrace{\delta_M}_{\to MU}+\underbrace{\theta_0}_{\to BC} \\
r(MU \to X) &= \underbrace{\mu_M\alpha_{MD}}_{\to DX}+\underbrace{(1-\mu_M)\alpha_{MR}}_{\to RU}+\underbrace{\theta_M}_{\to MC} \\
r(PU \to X) &= \underbrace{\delta_I}_{\to IU}+\underbrace{\theta_0}_{\to PC} \\
r(IU \to X) &= \underbrace{\mu_I\alpha_{ID}}_{\to DX}+\underbrace{\eta_H\delta_H}_{\to HQ}+\underbrace{(1-\mu_I)(1-\eta_H)\alpha_{IR}}_{\to RU}+\underbrace{\theta_I}_{\to IC} \\
r(AC \to X) &= \underbrace{\alpha_A}_{\to RT} \\
r(BC \to X) &= \underbrace{\delta_M}_{\to MC} \\
r(MC \to X) &= \underbrace{\mu_M\alpha_{MD}}_{\to DX}\underbrace{(1-\mu_M)\alpha_{MR}}_{\to RT} \\
r(PC \to X) &= \underbrace{\delta_I}_{\to IC} \\
r(IC \to X) &= \underbrace{\mu_I\alpha_{ID}}_{\to DX}+\underbrace{\eta_H\delta_H}_{\to HQ}+\underbrace{(1-\mu_I)(1-\eta_H)\alpha_{IR}}_{\to RT} \\
r(HQ \to X) &= \underbrace{\eta_C\delta_C}_{\to CQ}+\underbrace{(1-\eta_C)\alpha_{HR}}_{\to RT} \\
r(CQ \to X) &= \underbrace{(1-\mu_c)\alpha_{CR}}_{\to RT}+\underbrace{\mu_c\alpha_{CD}}_{\to DX}
\end{aligned} \tag{2}$$

Statistics New Zealand Disclaimer

The contact network used for these simulations relies heavily on statistics derived from the statistics New Zealand Integrated Data Infrastructure. The results in this analysis are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand. The opinions, findings, recommendations and conclusions expressed in this analysis are those of the authors, not Statistics NZ. Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this analysis have been confidentialised to protect these groups from identification. Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz. The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit-record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.