RESEARCH ARTICLE

Model-free estimation of COVID-19 transmission dynamics from a complete outbreak

Alex James[1,2]*, **Michael J. Plank**[1,2], **Shaun Hendy**[2,3], **Rachelle N. Binny**[2,4], **Audrey Lustig**[2,4], **Nic Steyn**[2,3]

1 School of Maths and Stats, University of Canterbury, Christchurch, New Zealand, **2** Te Pūnaha Matatini, Auckland, New Zealand, **3** Dept of Physics, University of Auckland, Auckland, New Zealand, **4** Manaaki Whenua, Lincoln, New Zealand

* alex.james@canterbury.ac.nz

Abstract

New Zealand had 1499 cases of COVID-19 before eliminating transmission of the virus. Extensive contract tracing during the outbreak has resulted in a dataset of epidemiologically linked cases. This data contains useful information about the transmission dynamics of the virus, its dependence on factors such as age, and its response to different control measures. We use Monte-Carlo network construction techniques to provide an estimate of the number of secondary cases for every individual infected during the outbreak. We then apply standard statistical techniques to quantify differences between groups of individuals. Children under 10 years old are significantly under-represented in the case data. Children infected fewer people on average and had a lower probability of transmitting the disease in comparison to adults and the elderly. Imported cases infected fewer people on average and also had a lower probability of transmitting than domestically acquired cases. Superspreading is a significant contributor to the epidemic dynamics, with 20% of cases among adults responsible for 65–85% of transmission. Subclinical cases infected fewer individuals than clinical cases. After controlling for outliers serial intervals were approximated with a normal distribution (μ = 4.4 days, σ = 4.7 days). Border controls and strong social distancing measures, particularly when targeted at superspreading, play a significant role in reducing the spread of COVID-19.

Introduction

Between 26 February and 22 May 2020, New Zealand recorded 1499 confirmed and probable cases of COVID-19 (Fig 1). Of these, 575 had a recent history of overseas travel. With the exception of international arrivals, which are subject to mandatory 14-day quarantine, no community cases were reported from 22 May to 10 August. This means that New Zealand has a high probability of having eliminated community transmission during this period.

Progressively stricter border measures were introduced from mid-March to mid-April. From 16 March, all international arrivals to New Zealand were required to go into 14 days

Town, Chief Science Advisor, Ministry of Health, Wellington, New Zealand. Email: ian.town@health. govt.nz.

self-isolation. On 20 March, the border was closed to non-citizens and non-residents. From 10 April, all international arrivals were required to go into 14 day mandatory government-managed quarantine.

On 19 March, gatherings of more than 100 people were banned and, on 21 March, a four-level alert level system was introduced. Alert level 1 was effectively business as usual but with border restrictions. New Zealand entered alert level 2 on 21 March, level 3 on 23 March, and level 4 on 26 March. Alert level 4 included strict stay-at-home measures, and all non-essential business and schools were closed. The country remained at alert level 4 until 28 April, when it moved to level 3 allowing partial school openings and non-essential businesses to operate under strict non-contact guidelines. On 14 May, New Zealand moved to alert level 2 and most schools and businesses reopened but large gatherings were still banned. On 9 June, the country moved back to alert level 1 and a strict testing programme for international arrivals was established to complement the ongoing 14-day mandatory quarantine.

New Zealand also has a widespread testing regime that quickly encouraged testing of anyone with even minor symptoms. Up to 5 June (2 weeks after the last reported domestic case on 22 May), 291,994 tests were performed (58 tests per 1000 people) with a positivity rate of around 0.4% (Ministry of Health, 2020). Since mid-June, all international arrivals are tested on day three and day twelve, and are not permitted to leave their 14-day mandatory quarantine before returning a negative test result.

Mathematical models of COVID-19 transmission and control measures have estimated that the effective reproduction number R_{eff} decreased from around 1.8 prior to alert level 4 to around 0.35 during alert level 4 [1]. However, these results may depend on modelling assumptions, for example the assumed generation time distribution, the delay between infection, symptom onset and testing, and the effects of case-targeted control.

In this paper, we examine the epidemic transmission tree in New Zealand using detailed case data and contact tracing information. This is an unusual dataset because it describes a closed outbreak in which there are no ongoing transmission chains, and many cases are epidemiologically linked to a likely index case. Because not all cases have a potential index case identified and some cases have multiple potential index cases, the transmission tree is not completely or uniquely defined. We reconstruct multiple instances of the transmission tree using a Monte-Carlo technique. This approach assigns a randomly selected index case to each secondary case based on the dates of symptom onset, cluster identification, and the results of
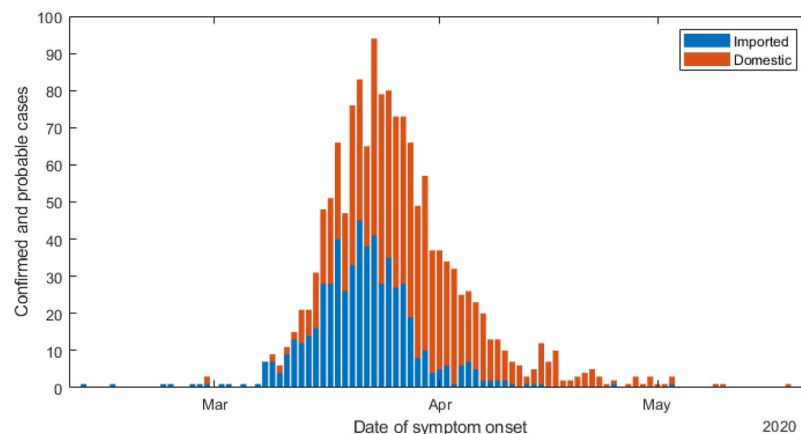
Fig 1. Daily new cases of COVID-19 in New Zealand from 26 February to 31 May 2020, split into cases with a recent history of international travel (blue) and those without (orange).

https://doi.org/10.1371/journal.pone.0238800.g001

contact tracing investigations. We assume a fixed proportion of infections were subclinical and undetected because they were asymptomatic, paucisymptomatic, or for some other reason.

From this suite of generated transmission trees, we calculate the individual reproduction number R_i for each infected individual and explore the distribution of R_i and the secondary attack rate at different points in the epidemic timeline. We explore how these quantities vary for different age groups and types of case, imported or domestic, and investigate the effect of early case isolation. We also investigate the degree of superspreading in the transmission dynamics. Although some assumptions are required to allow for missing or noisy data, our results are not model-dependent, unlike most estimates of effective reproduction number for COVID-19 [2–4]. Our sample size is larger than previous studies that have estimated the secondary attack rate (e.g. Bi, Wu [5], Ganyani, Kremer [6], Cheng, Jian [7]). This allows for more reliable estimates and for an investigation of how secondary attack rate varies with age. We also reconstruct the entire transmission tree rather than just estimating of the serial interval distribution. This provides a richer source of information about the transmission dynamics, allowing a more complete set of epidemiological parameters to be explored.

Methods

Anonymised data were supplied by the Ministry of Health, New Zealand. The dataset included the baseline demographic characteristics of age and ethnicity. The project was given ethics approval by the Ministry of Health, New Zealand who waived the requirement for informed consent. Of the 1499 confirmed and probable cases, 575 had a recent history of overseas travel and were considered to be imported cases. The remaining 924 cases were considered to be infected within New Zealand and are referred to as domestic cases. Following contact tracing interviews, 636 of the domestic cases were each assigned up to five potential index cases. The remaining domestic cases did not have a potential index case identified. 627 of the domestic cases were linked to one of 18 significant clusters. Of the cases linked to clusters, 146 had no potential index case identified. For 1444 cases, the number of close contacts, obtained via case recall at a contact tracing interview, was recorded. The number of close contacts varied widely between individuals, e.g. 753 cases had 2 or fewer contacts and 26 cases had more than 50. 31 cases were recorded as being asymptomatic at the time of testing. All cases recorded age.

We constructed a potential instance of the transmission tree by randomly assigning a symptom onset date to each of the i cases using the date recorded by case recall, D_i, (if available) with a random offset, $D_i = D_i + X_i$ where $X_i \sim N(\mu = 0, \sigma = 1)$. For the 38 cases with no onset date recorded (31 asymptomatic, 7 symptomatic) we assigned a randomly generated onset date using the reported date and a gamma distribution fitted to the data on time from onset to reporting, Γ ($\mu = 6.7, \sigma = 5.4$) days.

In addition to the 31 recorded asymptomatic cases, each generated transmission tree included additional subclinical cases so that the total proportion of subclinical cases was p_{asymp} = 0.33 on average. These additional cases were randomly sampled from the recorded cases with the symptom onset date randomly jittered $D_i = D_i + X_i$ where $X_i \sim N(\mu = 0, \sigma = 2)$. Model-generated subclinical cases had the same cluster status and potential index cases as the sampled case.

Of the domestic cases, 17 had a unique index case identified in the data. For the 288 cases with no index case, an index case was selected based on the implied serial interval, i.e. the time between symptom onset in the index case and in the secondary case. The serial interval SI can be expressed as

$$SI = GT - I_1 + I_2$$

where *GT* is the generation time (time from infection of index case to infection of secondary case), and I_1 and I_2 are the incubation periods of the index and secondary case respectively. We assume that the generation time $GT \sim Weibull(\mu = 5.05, \sigma = 1.93)$ days [8] and the incubation period $I_j \sim \Gamma(\mu = 5.5, \sigma = 2.28)$ days [9] and use this to calculate the probability density function for the serial interval $f(SI)$. Note that although the generation time must be positive, there is a positive probability of a negative serial interval. When index cases were assigned randomly no attempt was made to infer missing cases as part of transmission chains.

For the 619 domestic cases with $n > 1$ potential index cases, the serial interval SI_j corresponding to each potential index case $j = 1,...n$ was calculated. Potential index case j was selected as the index case with probability $P_j = f(SI_j) / \sum_{j=1}^{n} f(SI_j)$. For cases with no potential index case identified, all cases in the same cluster (or the no cluster group if appropriate) were assumed to be potential index cases and the same random selection method was used. The probability P_j defined above was assumed to be independent of whether the potential cases was clinical or subclinical.

Assigning an index case to each secondary case produces a single instance of the epidemic transmission tree. We then repeated this Monte Carlo process to produce $M = 500$ independent instances of the transmission tree. For each instance, we calculated the number of secondary cases resulting from each reported case (the reproduction number R_{ij} for individual i in transmission tree j) and the associated symptom onset date, D_{ij}.

Incidence by age

The original case data, including symptomatic and asymptomatic cases, was split by imported status and time period. The reported statistic is the probability of being in each age bracket within that group and the 95% binomial confidence interval. NZ population statistics are shown as benchmarks.

Resulting serial interval

The resulting serial interval distribution across all *M* transmission trees was fitted with a normal distribution (Matlab 2018b, `fitdist`).

Time varying effective reproduction number

We calculated the mean (or case [10]) reproduction number for all clinical individuals, split by imported status and day of onset (which is approximately equivalent to peak infectiousness) in transmission tree *j*. The reported statistic is the median and 95% range over the *M* transmission trees.

Reproduction number and superspreading

We fitted a negative binomial distribution (Matlab 2018b, `fitdist`) to the effective reproduction numbers in each transmission tree separately for clinical individuals of different groups (age and domestic/imported status) and with onset date D_{ij} in different time periods for each transmission tree. This gave an estimate of the mean effective reproduction number $-R_j$ for that group of cases in the jth transmission tree, and the associated dispersion or superspreading parameter k_j [11]. We also calculated the probability of causing no secondary infections or more than 5 secondary infections for each tree. The reported statistic is the median and 95% range over the *M* transmission trees. If the variance of the sample was less than the mean a negative binomial could not be fitted so a Poisson distribution was used instead (i.e. $k = \infty$).
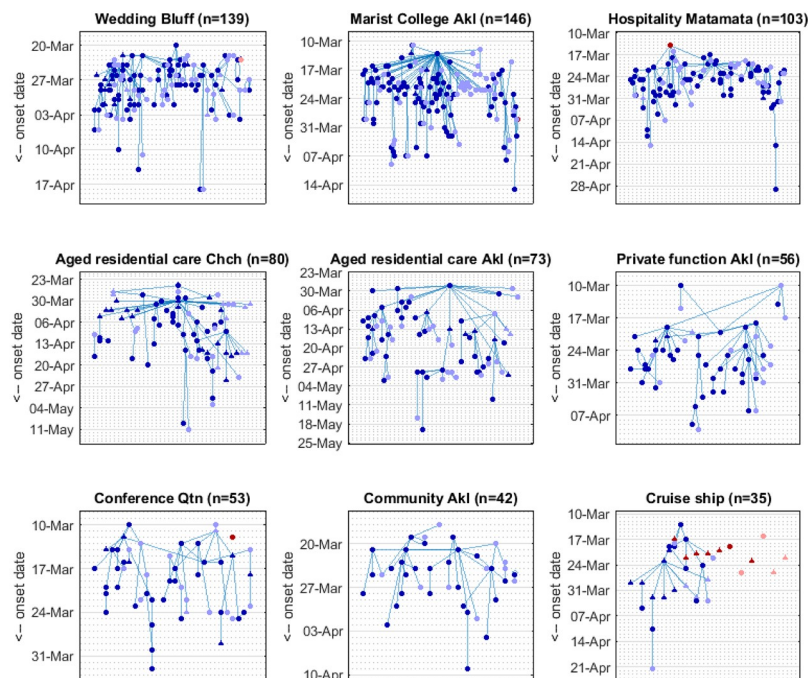
Fig 2. Reconstruction of New Zealand's nine largest clusters. Each panel shows one instance of the Monte Carlo method for reconstructing the epidemic transmission tree with cases plotted by date of symptom onset. Red = imported case; blue = domestic case; dark colours = clinical; light colours = subclinical; squares = age under 10 years; circles = age 10–65 years; triangles = age over 65 years. The title of each panel describes the setting and location (Akl = Auckland; Chch = Christchurch; Qtn = Queenstown) associated with each cluster; n is the total number of cases in the reconstructed tree including actual cases and simulated (missing) subclinical cases.

https://doi.org/10.1371/journal.pone.0238800.g002

Secondary attack rate

Secondary attack rates of clinical individuals were calculated for each group in each transmission tree by pooling the total number of secondary cases and the total number of contacts [12].

$$SAR = \frac{\sum secondary\ cases}{\sum close\ contacts}.$$

The reported statistic is the median and 95% range over the M transmission trees.

Results

One instance of the reconstruction of New Zealand's nine largest clusters is shown in Fig 2. These clusters were associated with settings typical of international patterns, including a high school, a wedding, hospitality venues, aged residential care facilities, and a conference. The age distribution of cases varies across these clusters (S1 Fig). Although one of New Zealand's largest clusters was associated with a school (Marist College cluster), more than half of the cases (55/96) in this cluster were in individuals over 20 years old. Of the 41 cases under 20, 35 were aged 10–20 years with only 6 confirmed cases in under 10s. In the two largest clusters associated with aged residential care facilities, there was as expected an over-representation of cases over 80 years old. However, the majority of cases in both these clusters (36/56 and 44/51 respectively) were under 65 years old, suggesting that staff and visitors are a more important driver of transmission than residents in aged residential care.
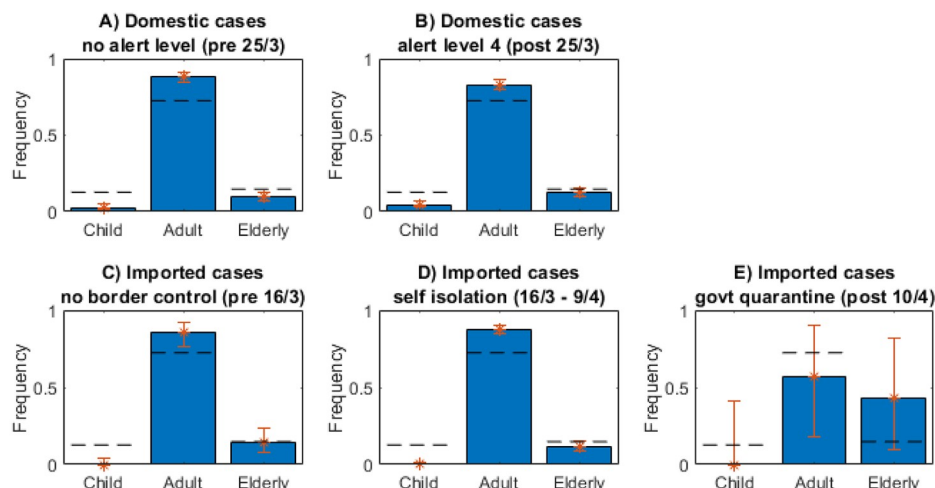
Fig 3. Across all categories of case children are under-represented. Frequency of cases for the different types of case aged: children under 10 years; adults 10–65 years; elderly over 65 years. Dashed lines show New Zealand population. Error bars show the binomial 95% confidence interval.

Incidence by age

We investigated COVID-19 incidence by age by splitting the clinical cases into three age groups: under 10 years, 10–65 years, and over 65 years. We consider domestic and imported cases separately and further split the data by significant time periods: before and after alert level 4 was introduced (26 March) for the domestic cases; and during three distinct periods of border control for imported cases: no restrictions (before 16 March), self-isolation (16 March to 9 April) and government-mandated quarantine (after 9 April).

Children under 10 are significantly under-represented in the case data relative to the New Zealand population demographic of 12.7% (Fig 3). The domestic cases (Fig 3A and 3B) contain relatively more children under 10 than the imported cases and these infections predominantly occurred during alert level 4 (2.6% of cases were in children before 26 March, 4.4% on or after 26 March). During alert level 4, schools and day care centres were closed so children are likely to have spent more time in family settings where attack rates are higher [5] resulting in a higher incidence for this group. Of the 576 imported cases (Fig 3C–3E), only 2 were in children under 10 years old. Both these cases were connected with an imported adult case with the same travel history and symptom onset 4 to 7 days prior. However, it is likely that children are under-represented in international travellers generally.

Serial interval distribution

The distribution of serial intervals (Fig 4) resulting from the Monte Carlo reconstruction contained a substantial number of outliers. It is likely these correspond to cases that were infected by an undetected case, or the index case was wrongly assigned either in the data or in the reconstructed tree. Including all serial intervals from the reconstructed tree, the best fit normal distribution had mean 5.0 days and standard deviation 5.7 days. There are some very long (> 20 days) serial intervals in the dataset, these are likely due to missing data as the method does not attempt to directly infer missing cases as part of transmission chains. Excluding these outliers (serial intervals longer than 20 days) resulted in mean 4.4 days and standard deviation
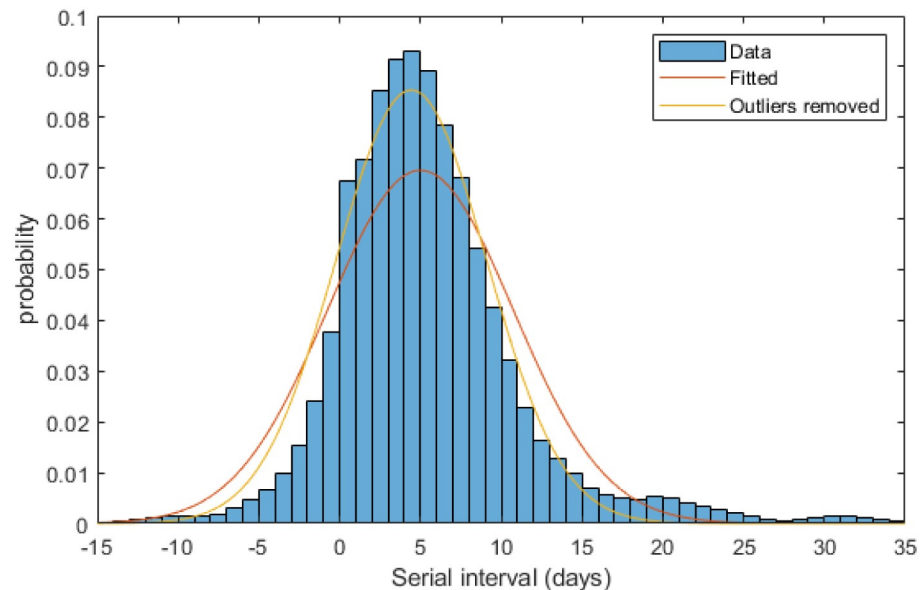
Fig 4. Serial interval distribution with fitted distributions. Red curve shows the normal distribution fitted to all data ($\sim N(\mu = 5.0, \sigma = 5.7)$); yellow curve shows the normal distribution fitted to data excluding serial intervals longer than 20 days ($\sim N(\mu = 4.4, \sigma = 4.7)$).

https://doi.org/10.1371/journal.pone.0238800.g004

4.7 days, excluding negative outliers (< -10 days) made no appreciable difference. This is comparable with previous estimates of the mean serial interval [13, 14]. When allowing for negative serial intervals in their Monte Carlo reconstruction method, Ganyani, Kremer [6] found the serial interval to be 3.86 ± 4.76 days (mean ± std. dev.) for Singapore and 2.90 ± 4.88 days for Tianjin, China. These estimates for the mean serial interval are shorter than ours, but the finding that a significant proportion of serial intervals were negative is consistent. The data we analysed and that of Ganyani, Kremer [6] were collected in the presence of case isolation and contact tracing. It is likely these interventions will reduce the mean serial interval, to varying extents, by suppressing transmission late in the infectious period. Our results reflect the serial interval distribution in the presence of these interventions in the New Zealand context.

Time-varying effective reproduction number

For domestic cases, the average number of secondary cases infected by a clinical case before New Zealand entered alert level 4 on 26 March was $R_{eff} = 1.47$ (382 index cases) (Fig 5A). During and after alert level 4 (26 March onwards), this dropped to $R_{eff} = 0.69$ (521 index cases). In the period after alert level 4, the vast majority of cases were detected as part of contact tracing efforts and were isolated swiftly, often before symptom onset. The reproduction number for domestic cases showed two small peaks during the alert level 4 period. The first coincided with Easter weekend (10–13 April). The second coincides with a long weekend (25–27 April) immediately preceding the pre-announced easing of restrictions on 28 April.

For imported cases (Fig 5B), the number of secondary infections per case was much lower throughout the outbreak. Before self-isolation for international arrivals was introduced on 16 March, $R_{eff} = 0.70$ (119 index cases). While self-isolation was in place from 17 March to 9 April, $R_{eff} = 0.42$ (439 index cases). After mandatory government-managed quarantine was introduced on 10 April, $R_{eff} = 0$ (7 cases).
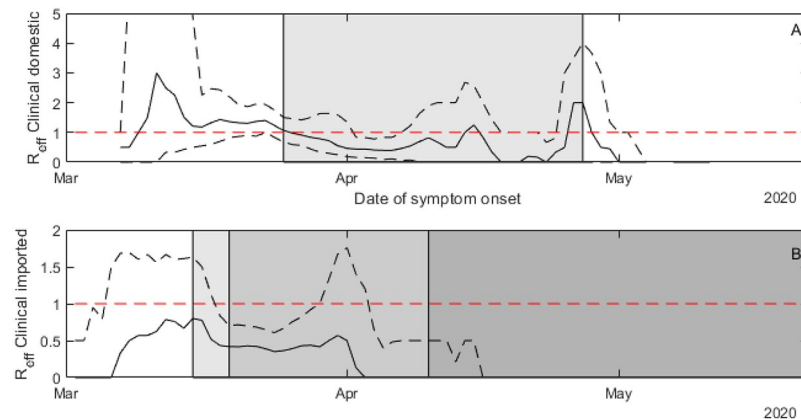
Fig 5. Increases in alert level and improved border measures resulted in decreases in secondary transmissions.
Expected value of R_{eff} each day for clinical domestic cases (A) and imported cases (B). Solid line show the median from 500 instances of the reconstructed transmission tree, dashed lines are the 95% range. Imported cases had a lower effective reproduction number than domestic cases. The date of secondary case exposure was assigned to be the date of symptom onset in the index case. Note that 10–13 April was the long Easter weekend and 27 April was a public holiday in New Zealand. The timing of the end of alert level four was announced on 22 April. Shading in (A) shows period of alert level 4 (26 March to 27 April). Shading in (B) shows progressively stricter border restrictions introduced 16 March, 20 March and 10 April.

https://doi.org/10.1371/journal.pone.0238800.g005

Effect of age on reproduction number and superspreading

Before the move to alert level 4, over half of all domestic cases resulted in at least one secondary case (Fig 5A). The proportion of cases resulting in at least one secondary case was similar for all age groups (Table 1). The expected number of secondary cases was R_{eff} = 0.87 for the under 10 years age group, R_{eff} = 1.49 for the 10 to 65 years age group, and R_{eff} = 1.51 for the over 65 years age group. Although children under 10 were equally likely to infect at least one person, adults tended to infect more people than children under 10 did. Cases among adults and older people also had a significant chance of being a superspreader (which we defined as infecting more than 5 people): 6% in the 10 to 65 year age group, and 7% in the over 65 age group.

During alert level 4 (Fig 5B, Table 1), R_{eff} dropped below 1 for the under 10 and the 10 to 65 age groups, but stayed above 1 in over 65 group. This may be due to the over-representation of cases in aged care facilities in the later stages of the epidemic, where it may be more difficult to limit interactions. The proportion of superspreaders decreased in adults during alert level 4 but not in the elderly. There were no recorded superspreaders under 10 years old at any stage in the epidemic.

Overall, there were 29 superspreaders, of which 21 had symptom onset before alert level 4 began, and the remaining 8 had symptom onset during alert level 4. Of these, the 6 most recent events were all related to clusters of cases at aged care facilities.

As seen previously, imported cases overall had a lower R_{eff} than domestic cases (Fig 6C–6E). Before the introduction of any border measures, many of these cases would have been tourists and other visitors who are less likely to have the close community contacts that local residents have. Throughout the outbreak, imported cases in the over 65 years age group had a higher R_{eff} than those aged 10–65. Imported cases in children, of which there were only 2 (<1%), had the lowest likelihood of transmitting the disease even accounting for both these cases happening during the mandatory quarantine period. There were 5 superspreaders amongst the imported cases, all of which occurred during the self-isolation period between 9 and 19 March.

Table 1. Summary statistics for the New Zealand epidemic by age and type of case.

Group	Time period	N	$E(R_i)$	k	$P(R_i > 5)$	$P(R_i = 0)$
Clinical cases (domestic)						
Child	no alert level (pre 25/3)	10	0.87 (0.33–1.38)	3.17 (0.32-∞)	0.00 (0.00–0.03)	0.49 (0.30–0.75)
Adult	no alert level (pre 25/3)	336	1.49 (1.41–1.56)	0.70 (0.60–0.81)	0.06 (0.05–0.06)	0.45 (0.43–0.48)
Elderly	no alert level (pre 25/3)	36	1.51 (1.14–1.81)	0.50 (0.30–0.84)	0.07 (0.04–0.09)	0.50 (0.41–0.60)
Child	alert level 4 (post 25/3)	23	0.25 (0.09–0.44)	0.80 (0.13-∞)	0.00 (0.00–0.01)	0.82 (0.70–0.92)
Adult	alert level 4 (post 25/3)	433	0.63 (0.58–0.67)	0.41 (0.32–0.50)	0.01 (0.01–0.02)	0.68 (0.66–0.71)
Elderly	alert level 4 (post 25/3)	65	1.27 (1.08–1.49)	0.23 (0.15–0.35)	0.07 (0.05–0.08)	0.65 (0.58–0.72)
Clinical cases (imported)						
Child	no control (pre 16/3)	0				
Adult	no control (pre 16/3)	78	0.63 (0.48–0.84)	0.39 (0.23–0.87)	0.01 (0.00–0.03)	0.69 (0.62–0.75)
Elderly	no control (pre 16/3)	13	0.87 (0.60–1.23)	0.37 (0.13–1.24)	0.03 (0.01–0.05)	0.64 (0.47–0.78)
Child	self isolation (16/3–9/4)	2	0.00 (0.00–1.00)	Poisson	0.00 (0.00–0.00)	1.00 (0.37–1.00)
Adult	self isolation (16/3–9/4)	410	0.46 (0.40–0.51)	0.44 (0.31–0.63)	0.00 (0.00–0.01)	0.73 (0.70–0.76)
Elderly	self isolation (16/3–9/4)	55	0.28 (0.18–0.42)	1.06 (0.22-∞)	0.00 (0.00–0.00)	0.78 (0.69–0.87)
Child	govt quarantine (post 10/4)	0				
Adult	govt quarantine (post 10/4)	4	0.00 (0.00–0.50)	Poisson	0.00 (0.00–0.00)	1.00 (0.61–1.00)
Elderly	govt quarantine (post 10/4)	3	0.00 (0.00–0.33)	Poisson	0.00 (0.00–0.00)	1.00 (0.72–1.00)
Subclinical cases						
All	no alert level (pre 25/3)	11	0.70 (0.18–1.40)	1.11 (0.15-∞)	0.00 (0.00–0.06)	0.60 (0.36–0.88)
All	alert level 4 (post 25/3)	20	0.52 (0.25–0.83)	0.29 (0.10–2.05)	0.01 (0.00–0.03)	0.75 (0.63–0.86)

Columns show: the number of cases in each group (N); the expected number of secondary infections caused by an individual in this group ($E(R_i)$); the dispersion coefficient of a fitted negative binomial distribution (k); the probability of an individual infecting more than 5 people ($P(R_i > 5)$) and the probability of an individual not infecting anyone ($P(R_i = 0)$) according to the fitted distribution. Numbers show the median (95% range) across 500 instances of reconstructed transmission tree.

https://doi.org/10.1371/journal.pone.0238800.t001

The degree of superspreading in the transmission dynamics can be quantified by the variance in the distribution of individual case reproduction numbers $-R_i$. This can be measured by the dispersion parameter k of a negative binomial distribution fitted to the data [11, 15]. The smaller the value of k, the more heavy-tailed the distribution of secondary cases, and therefore the more superspreading is driving the transmission process. The results show that older age groups tend to be have more superspreading than younger ones (Table 1). This could be influenced partly by the number of cases transmitted in aged residential care facilities. Among the 10–65 and over 65 year age groups, the estimated values of k range from 0.27 to 0.70. This means that 20% of cases are responsible for between 65% and 85% of transmission.

Secondary attack rate

In all case groups, children under 10 years old had a lower secondary attack rate than the older age groups (Fig 7). For domestic cases in the under 10 and the 10–65 age groups, there was very little difference in secondary attack rate before and during alert level 4, despite contacts being largely confined to the home setting which has a higher attack rate [5]. The secondary attack rate for elderly people did show a large increase during alert level 4. This likely reflects the number of cases that occurred in aged care facilities rather than private households.

Overall, the secondary attack rate for imported cases was much lower than for domestic cases, particularly during the period before introduction of border measures. This is likely due to this group including a high proportion of visiting tourists for whom most contacts are in a
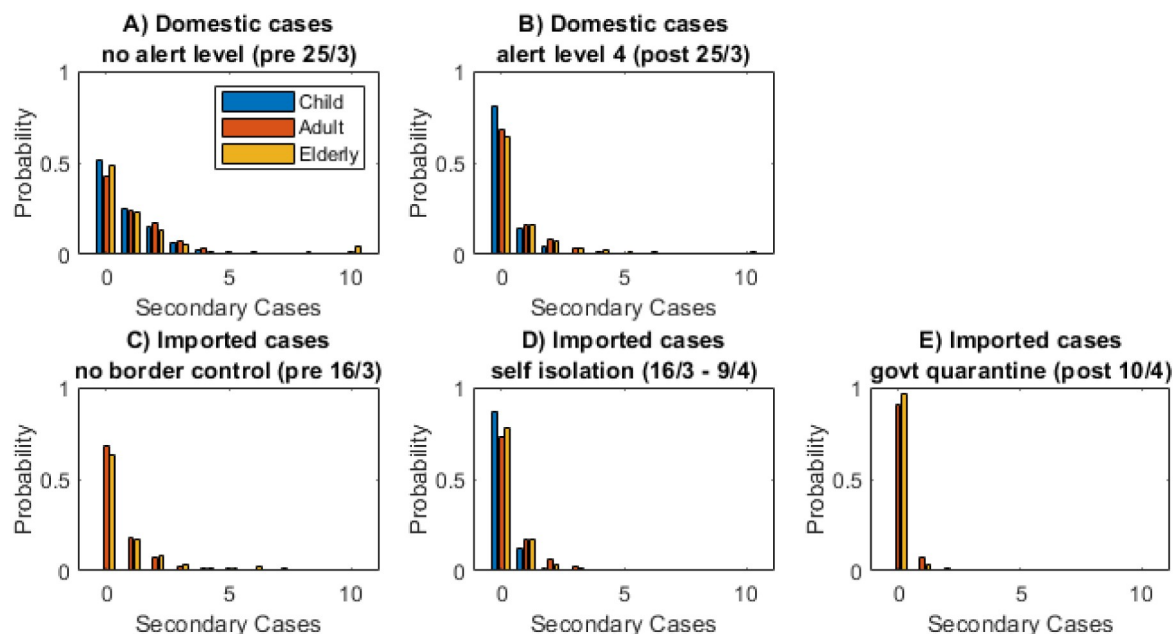
Fig 6. Children under 10 tended to infect fewer people and were less likely to be superspreaders. Distribution of individual case reproduction numbers for domestic and imported cases in each age group: under 10 years (blue), 10 to 65 years (red), over 65 years (yellow).

casual rather than home or work setting. The secondary attack rate for imported cases increased after the self-isolation measures were introduced, possibly reflecting the environment and make-up of these groups. However, the secondary attack rate dropped significantly during the mandatory quarantine period.

The lack of a clear effect of control measures on secondary attack rates suggests that reduction in the reproduction number was achieved primarily by reducing the number of close contacts rather than by reducing the likelihood of transmission when close contacts did occur. However this should be seen in light of the inevitable reporting biases of individuals who may be reluctant to report higher numbers of contacts during periods of high alert level restrictions.

Subclinical cases

Of the 31 subclinical cases (21 domestic, 10 imported) recorded as asymptomatic, 30 were in adults and older people. Before the move to alert level 4, these cases had R_{eff} of approximately 50% of the adult domestic clinical cases (Table 1). During alert level 4, this dropped slightly further but the drop was not as significant as or the clinical cases during this time.

In most cases results for subclinical individuals are not included. There were very few subclinical individuals in the dataset so statistics derived only from the recorded individuals are not reliable. Including the randomly generated, i.e. missing, subclinical individuals can give misleading information. The results either mirror the clinical data, e.g. Figs 3, 4 and 7, or lose heterogeneity e.g. Fig 6. Some results are given in Table 1 but these limitations should be considered.

Discussion

We analysed the transmission dynamics of COVID-19 using a comprehensive dataset for a completed outbreak. Our analysis investigated the role of age and type of case (imported or
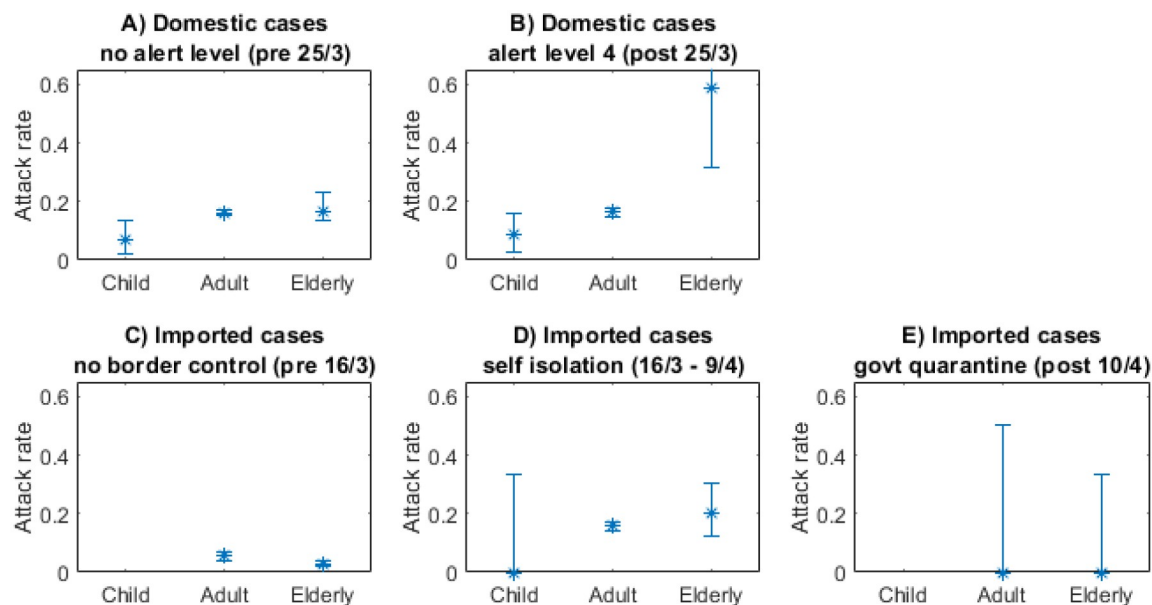
Fig 7. Children under 10 had a lower secondary attack rate across all case groups. Secondary attack rate for each group by age.

domestic) on reproduction number, secondary attack rate, likelihood of superspreading. We also investigated the effect of various control measures, including population-wide restrictions, and self-isolation or quarantine of international arrivals.

We found that children under 10 years contribute less to the spread of COVID-19 than other age groups. Children under 10 infected fewer people on average, had a lower secondary attack rate, and were less likely to be superspreaders. This is consistent with existing evidence on the role of children in COVID-19 transmission [16, 17].

The finding that superspreading is a significant contributor to transmission is consistent with existing empirical and modelling studies [18–20]. Our results show that among adults 20% of cases are responsible for between 65% and 85% of transmission. This suggests that interventions targeting superspreaders or superspreading events may be particularly effective in reducing the spread of COVID-19. These may include restrictions on gathering size, particularly in closed environments or crowded spaces [18].

In the dataset we used to reconstruct epidemic transmission tree, the index case is either missing or not uniquely identified in some cases. It is also likely there were subclinical cases that were undetected. We therefore used Monte Carlo methods to reconstruct missing cases and missing transmission routes (nodes and links respectively in the epidemic tree), though without inferring missing cases as part of transmission chains directly. It is possible that phylodynamic data on the genetic sequence of the virus recovered from positive test results could be used to reduce uncertainty in this step and improve the accuracy of the reconstructed tree.

Supporting information

S1 Fig. Age distribution of cases in New Zealand's nine largest clusters. The black line shows the age distribution of the New Zealand population (number of cases that would be expected if the cluster followed the same age distribution as the New Zealand population). The title of each panel describes the setting and location (Akl = Auckland; Chch = Christchurch;

Qtn = Queenstown) associated with each cluster; n is the total number of recorded cases in the cluster.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Alex James, Michael J. Plank, Shaun Hendy.

Data curation: Rachelle N. Binny, Nic Steyn.

Funding acquisition: Shaun Hendy.

Investigation: Alex James, Michael J. Plank.

Methodology: Alex James, Michael J. Plank, Rachelle N. Binny, Audrey Lustig.

Project administration: Shaun Hendy.

Resources: Nic Steyn.

Writing – original draft: Alex James, Michael J. Plank, Shaun Hendy.

Writing – review & editing: Alex James, Michael J. Plank, Shaun Hendy, Rachelle N. Binny, Audrey Lustig, Nic Steyn.

References

1. Binny RN, Hendy SC, James A, Lustig A, Plank MJ, Steyn N. Effect of Alert Level 4 on effective reproduction number: review of international COVID-19 cases. medRxiv. 2020.

2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of travel medicine. 2020. https://doi.org/10.1093/jtm/taaa021 PMID: 32052846

3. Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T, et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. 2020.

4. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC medicine. 2020; 18:1–10.

5. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, 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 Infectious Diseases. 2020. https://doi.org/10.1016/S1473-3099(20)30287-5 PMID: 32353347

6. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Eurosurveillance. 2020; 25(17):2000257.

7. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA internal medicine. 2020. https://doi.org/10.1001/jamainternmed.2020.2020 PMID: 32356867

8. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020; 368(6491). https://doi.org/10.1126/science.abb6936 PMID: 32234805

9. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine. 2020; 172(9):577–82. https://doi.org/10.7326/M20-0504 PMID: 32150748

10. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PloS one. 2007; 2(8):e758. https://doi.org/10.1371/journal.pone.0000758 PMID: 17712406

11. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005; 438(7066):355–9. https://doi.org/10.1038/nature04153 PMID: 16292310

12. Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020. https://doi.org/10.1016/S1473-3099(20)30471-0 PMID: 32562601

13. Tindale L, Coombe M, Stockdale JE, Garlock E, Lau WYV, Saraswat M, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. MedRxiv. 2020.

14. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed cases. Emerging infectious diseases. 2020; 26(6):1341. https://doi.org/10.3201/eid2606.200357 PMID: 32191173

15. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PloS one. 2007; 2(2):e180. https://doi.org/10.1371/journal.pone.0000180 PMID: 17299582

16. Li W, Zhang B, Lu J, Liu S, Chang Z, Cao P, et al. The characteristics of household transmission of COVID-19. Clinical Infectious Diseases. 2020.

17. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science. 2020. https://doi.org/10.1126/science.abb8001 PMID: 32350060

18. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International journal of infectious diseases. 2020. https://doi.org/10.1016/j.ijid.2020.02.060 PMID: 32145466

19. Adam D, Wu P, Wong J, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. 2020.

20. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Research. 2020; 5(67):67. https://doi.org/10.12688/wellcomeopenres.15842.3 PMID: 32685698