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

Authors: A. James[1,2*] PhD, M. Plank[1,2] PhD, S. Hendy[2,3] PhD, R. Binny[2,4] PhD, A. Lustig[2,4] PhD, N. Steyn[2,3] BSc (Hons).

***Corresponding author:**

Email: alex.james@canterbury.ac.nz

Phone: +64 3 369 2425

Address: School of Maths and Stats, University of Canterbury, Private Bag 4800, Christchurch, NZ.

Affiliations:

[1] University of Canterbury, New Zealand

[2] Te Pūnaha Matatini, New Zealand.

[3] University of Auckland, New Zealand.

[4] Manaaki Whenua, Lincoln, New Zealand.

Abstract:

Background

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.

Method

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.

Findings

Children under 10 years old are significantly under-represented in the case data. Children infected fewer people on average and had a lower secondary attack rate in comparison to adults and the elderly. Imported cases infected fewer people on average and had a lower secondary attack rate 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. Asymptomatic cases infected fewer individuals than clinical cases. Serial intervals are approximately normally distributed ($\mu = 5.0$ days, $\sigma = 5.7$ days). Early isolation and quarantine of cases reduced secondary transmission rates.

Interpretation

Border controls and strong social distancing measures, particularly when targeted at superspreading, play a significant role in reducing the spread of COVID-19.

Funding

Te Pūnaha Matatini, the New Zealand Centre of Research Excellence in complex systems. New Zealand Ministry of Business, Innovation and Employment.

Research in context

Evidence before this study

The basic reproduction number for COVID-19 is between 2 and 4, although there is substantial individual variation in the level of infectiousness and number of secondary cases. There is evidence that superspreading represents a significant component of COVID-19 transmission. Many estimates of key epidemiological parameters are either based on modelling assumptions or come from incomplete data that may be subject to sampling bias.

Added value of this study

Our study provides model-free estimates of key epidemiological parameters for COVID-19 from a comprehensive dataset on a complete outbreak. These include time-varying effective reproduction number, age-dependent secondary attack rates, serial interval, and the relative contribution of superspreading to transmission. Our results are less sensitive to modelling assumptions than many existing estimates and from a more complete dataset.

Implications of all the available evidence

Superspreading is a significant contribution to overall transmission of COVID-19 with 20% of cases among adults responsible for 65-85% of transmission. Children under 10 years old are less infectious than adults on average and less likely to be superspreaders. Asymptomatic infections tend to infect fewer people than symptomatic cases. Population-wide social distancing, particularly when targeted at superspreading, and early isolation are effective ways of reducing the spread of COVID-19.

Main text

Between 26 February and 22 May 2020, New Zealand recorded 1499 confirmed and probable cases of COVID-19 (Figure 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 cases have been reported since 22 May. This means that New Zealand has a high probability of having eliminated community transmission during this period.

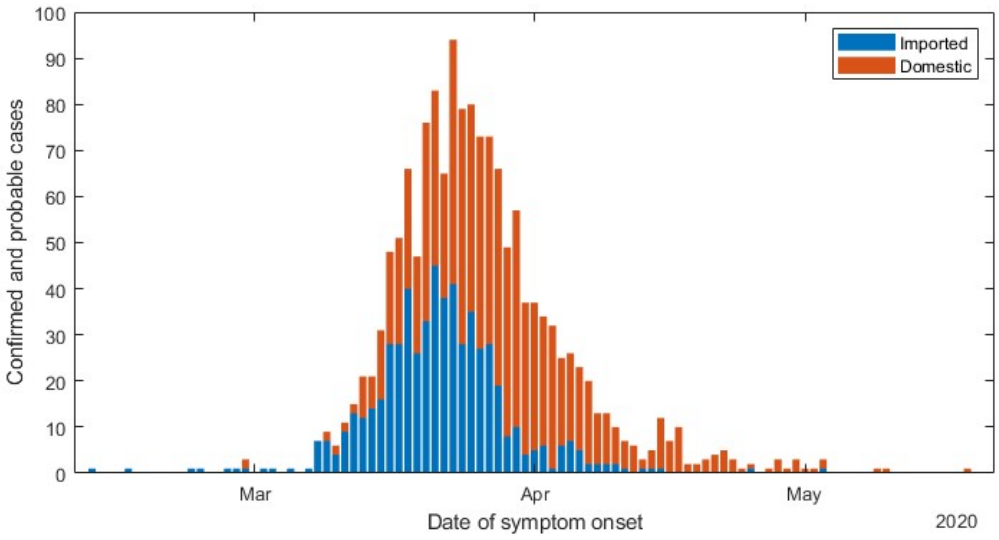


Figure 1: Daily new cases of COVID-19 in New Zealand from 26 February to 31 May 2020, split into cases with a recent history of international travel (blue) and those without (orange).

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 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 contact tracing investigations. We assume a fixed proportion of infections were 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 3,4]. 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

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, $\Gamma(\mu = 6.7, \sigma = 5.4)$ days.

In addition to the 31 recorded asymptomatic cases, each generated transmission tree included additional subclinical or asymptomatic cases so that the total proportion of asymptomatic 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 asymptomatic 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.

For the 619 domestic cases with $n > 1$ potential index cases, the serial interval SI_j corresponding to each potential index case $j = 1, \ldots n$ 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 symptomatic or asymptomatic.

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 clinical 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.

Time varying effective reproduction number We calculated the mean reproduction number for all clinical individuals, split by imported status with onset that day (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 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 $\bar{R}_j$ for that group of cases in the jth transmission tree, and the associated dispersion or superspreading parameter k_j[10]. 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 a negative binomial could not be fitted to a particular instance a Poisson distribution was used instead (i.e. $k = \infty$).

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 [11]

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

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

Effect of early isolation We used a general linear model to predict $\overline{R_\iota}$, the expected reproduction number for each clinical individual across all j trees, accounting for days between the original recorded symptom onset and isolation, i.e. number of days for which the individual was infectious and not isolated or quarantined. Initial exploration showed no significant difference between adults and elderly so these groups were combined, children (under 10 years) were significantly different. Imported or domestic status was significant ($p < 0.05$). The number of contacts variable was also explored but this was strongly correlated with the number of infectious days which was a better predictor variable judged by the p-value for each coefficient (Matlab 2018b, stepwiseglm). The final model was

$$\log(\overline{R_\iota}) \sim (symptom\ onset - isolation)^2 + Is\ child + Is\ imported.$$

In the regression cases with $\overline{R_\iota} = 0$ (i.e. with no secondary cases in any of the M instances) were offset to half the minimum value of $\overline{R_\iota}$ in the dataset. The results are presented graphically for an individual with the expected number of contacts.

Results

One instance of the reconstruction of New Zealand's nine largest clusters is shown in Figure 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 (Supplementary Figure S1). 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.

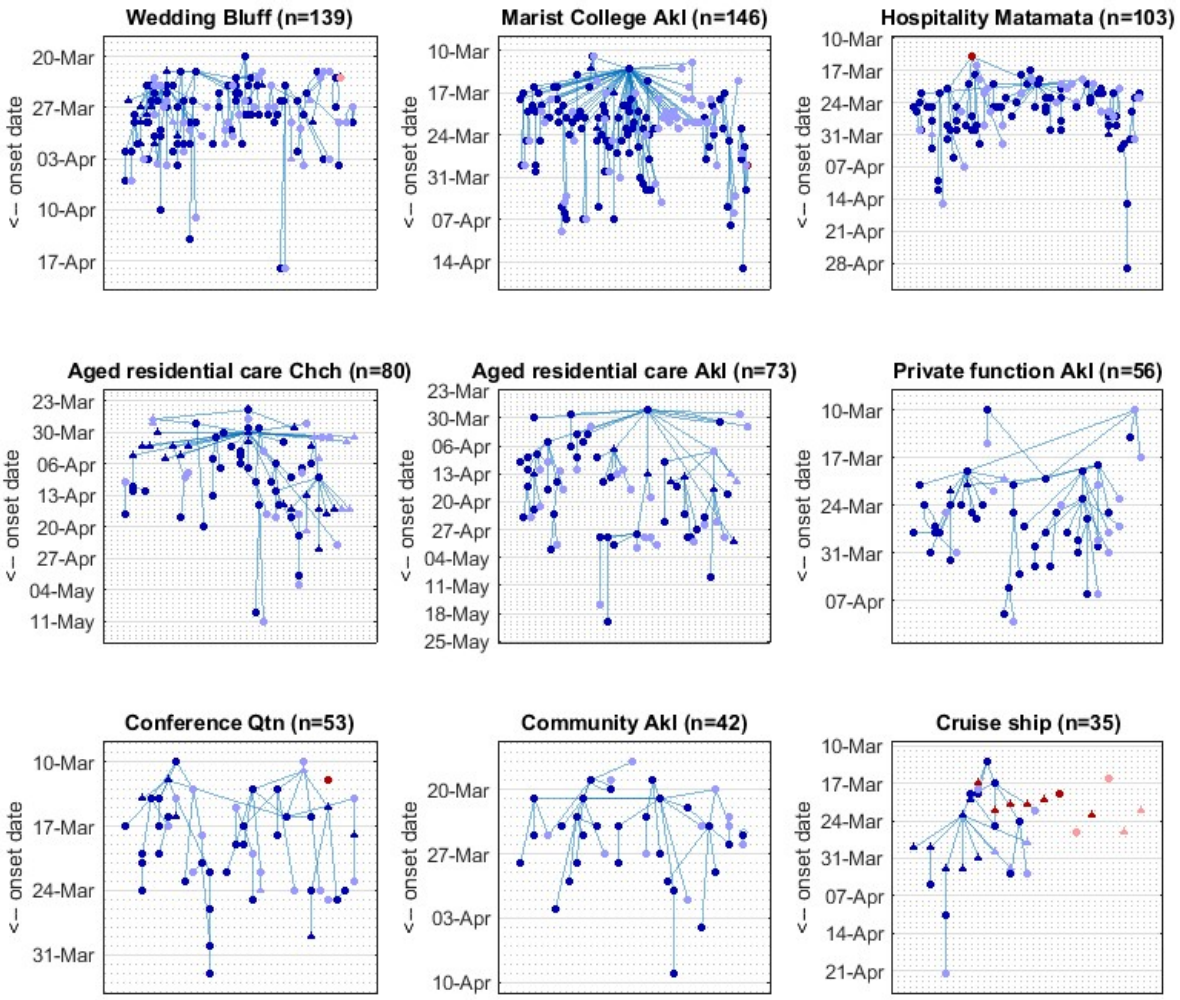


Figure 2: Reconstruction of New Zealand's nine largest clusters. Each panel shows one instance of the Monte Carlo method for reconstructing the epidemic transmission tree with cases plotted by date of symptom onset. Red = imported case; blue = domestic case; dark colours = symptomatic; light colours = asymptomatic; 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) asymptomatic cases.

251 **Incidence by age**

252 We investigated COVID-19 incidence by age by splitting the clinical cases into three
253 age groups: under 10 years, 0-65 years, and over 65 years. We consider domestic
254 and imported cases separately and further split the data by significant time periods:
255 before and after alert level 4 was introduced (26 March) for the domestic cases; and
256 during three distinct periods of border control for imported cases: no restrictions
257 (before 16 March), self-isolation (16 March to 9 April) and government-mandated
258 quarantine (after 9 April).

259 Children under 10 are significantly under-represented in the case data relative to the
260 New Zealand population demographic of 12.7% (Figure 3). The domestic cases (Fig
261 3 A,B) contain relatively more children under 10 than the imported cases and these
262 infections predominantly occurred during alert level 4 (2.6% of cases were in children
263 before 26 March, 4.4% on or after 26 March). During alert level 4, schools and day
264 care centres were closed so children are likely to have spent more time in family
265 settings where attack rates are higher [5] resulting in a higher incidence for this group.
266 Of the 576 imported cases (Fig 3 C-E), only 2 were in children under 10 years old.
267 Both these cases were connected with an imported adult case with the same travel
268 history and symptom onset 4 to 7 days prior.

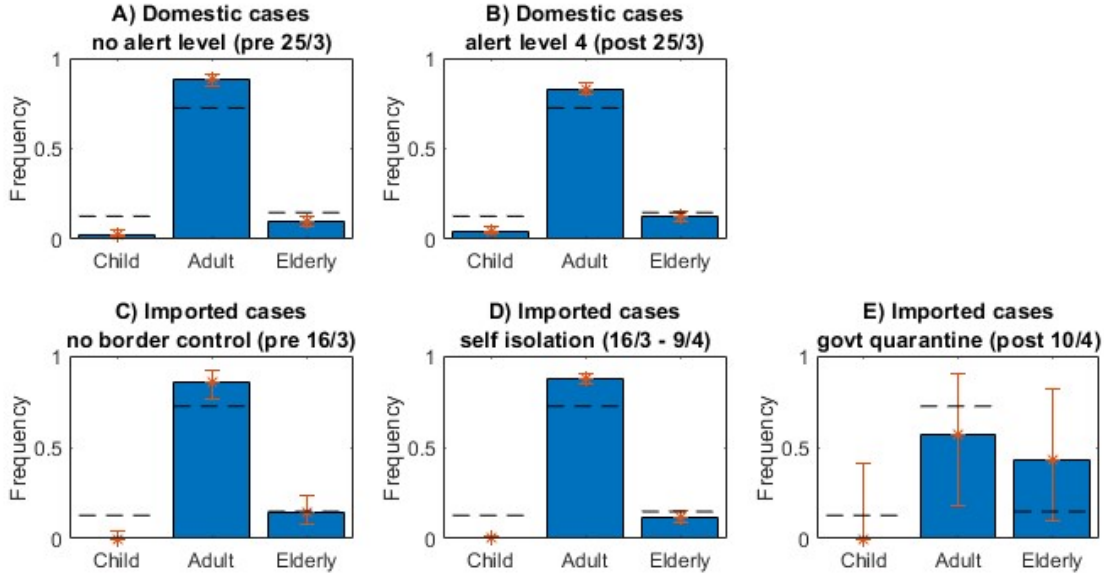


Figure 3: Across all categories of case children are under-represented. Frequency of cases for the different types of case aged: children under 10 years; adults 10-65 years; elderly over 65 years. Dashed lines show New Zealand population. Error bars show the binomial 95% confidence interval.

Serial interval distribution

The distribution of serial intervals (Figure 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. Excluding outliers (serial intervals longer than 20 days) resulted in mean 4.4 days and standard deviation 4.7 days. This is comparable with previous estimates of the mean serial interval [12,13]. 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.

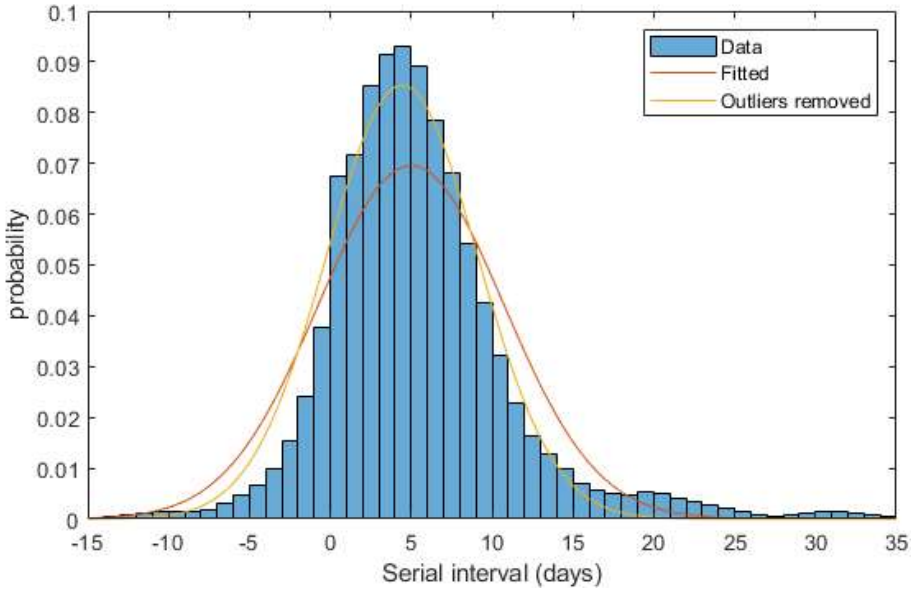


Figure 4. Serial interval distribution with fitted distributions. Red curve shows the normal distribution fitted to all data ($\sim N(\mu = 5.0, \sigma = 5.7)$); yellow curve shows the normal distribution fitted to data excluding serial intervals longer than 20 days ($\sim N(\mu = 4.4, \sigma = 4.7)$).

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) (Figure 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 (Figure 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, $R_{eff} = 0.08$ (7 cases).

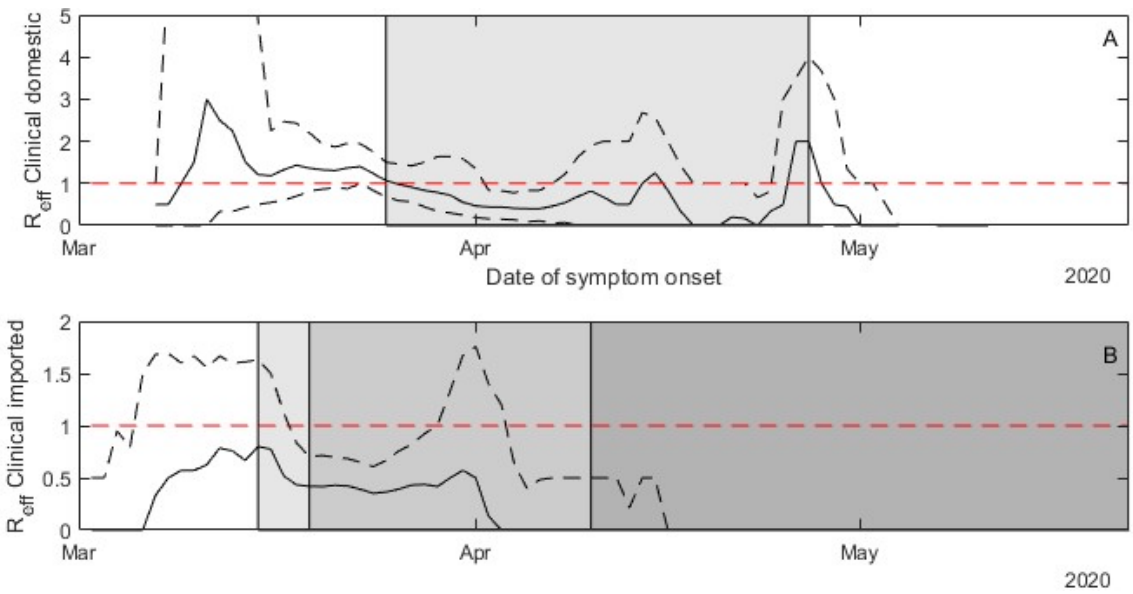


Figure 5: Increases in alert level and improved border measures resulted in decreases in secondary transmissions. Expected value of R_{eff} each day for clinical domestic cases (A) and imported cases (B). Solid line show the median from 500 instances of the reconstructed transmission tree, dashed lines 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.

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 (Figure 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.

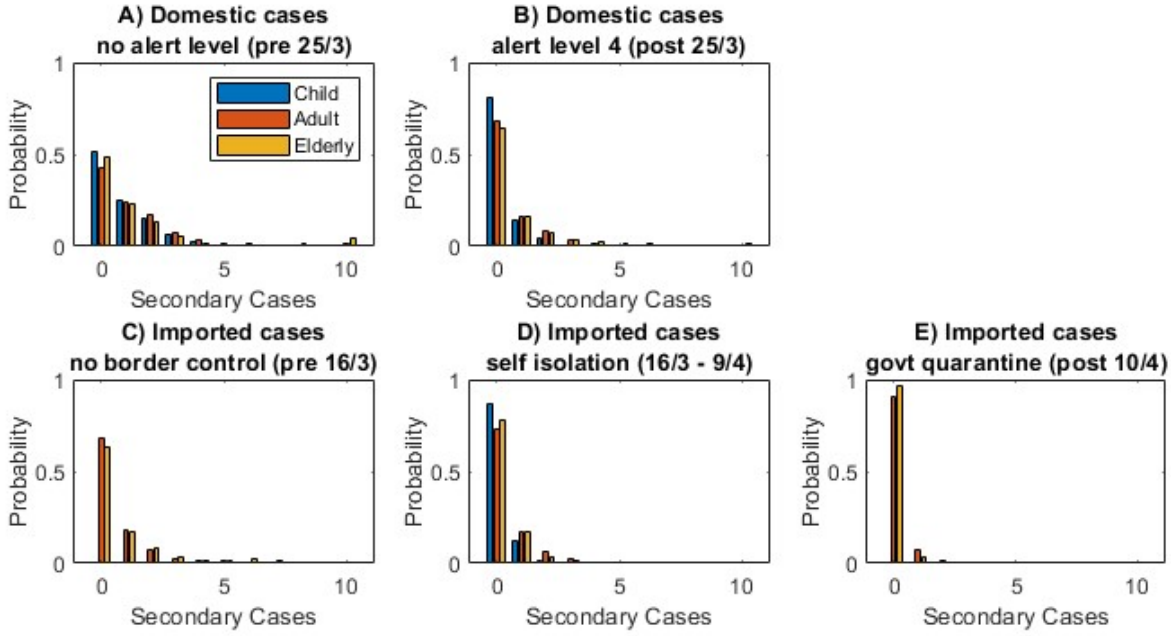


Figure 6: Children under 10 tended to infect fewer people and were less likely to be superspreaders. Distribution of individual reproduction numbers for domestic and imported cases in each age group: under 10 years (blue), 10 to 65 years (red), over 65 years (yellow).

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 (Figure 6C-E). 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

356 cases in the over 65 years age group had a higher R_{eff} than those aged 10 – 65.
357 Imported cases in children, of which there were only 2 (<1%), had the lowest likelihood
358 of being an index case even accounting for both these cases happening during the
359 mandatory quarantine period. There were 5 superspreaders amongst the imported
360 cases, all of which occurred during the self-isolation period between 9 and 19 March.

361 The degree of superspreading in the transmission dynamics can be quantified by the
362 variance in the distribution of individual reproduction numbers $\overline{R_i}$. This can be
363 measured by the dispersion parameter k of a negative binomial distribution fitted to
364 the data [10,14]. The smaller the value of k, the more heavy-tailed the distribution of
365 secondary cases, and therefore the more superspreading is driving the transmission
366 process. The results show that older age groups tend to be have more superspreading
367 than younger ones (Table 1). This could be influenced partly by the number of cases
368 transmitted in aged residential care facilities. Among the 10-65 and over 65 year age
369 groups, the estimated values of k range from 0.27 to 0.70. This means that 20% of
370 cases are responsible for between 65% and 85% of transmission.

371 **Secondary attack rate**

372 In all case groups, children under 10 years old had a lower secondary attack rate than
373 the older age groups (Figure 7). For domestic cases in the under 10 and the 10–65
374 age groups, there was very little difference in secondary attack rate before and during
375 alert level 4, despite contacts being largely confined to the home setting which has a
376 higher attack rate [5]. The secondary attack rate for elderly people did show a large
377 increase during alert level 4. This likely reflects the number of cases that occurred in
378 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 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.

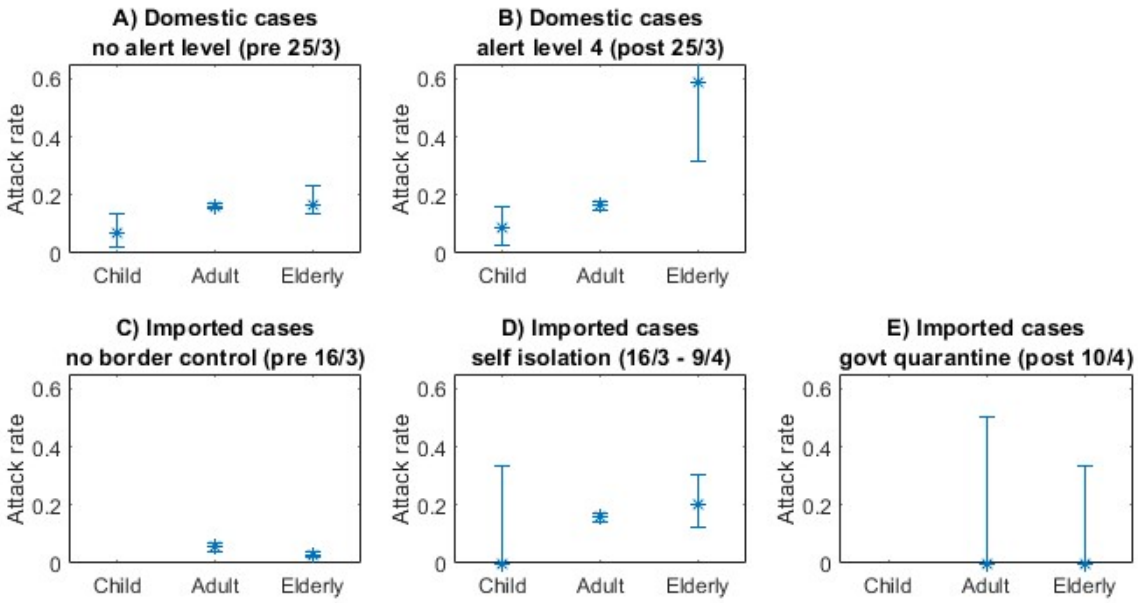


Figure 7: Children under 10 had a lower secondary attack rate across all case groups. Secondary attack rate for each group by age.

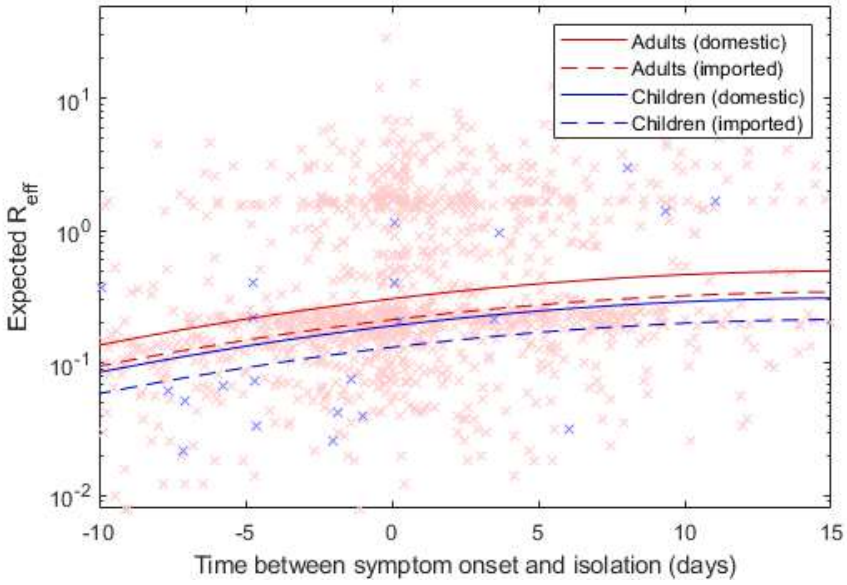


Figure 8: Early isolation substantially decreases secondary transmission and children and imported cases have a significantly lower transmission rate compared to adults and domestic cases. The expected reproduction for each individual across all 500 trees (red crosses adults and elderly, blue crosses under 10 year olds) against number of infectious days measured as time from symptom onset to isolation. Lines are the output of a regression model for each class of individual.

Effect of case isolation

Figure 8 shows the general linear model for the relationship between the individual reproduction number $\bar{R}_i$ and the time between symptom onset and isolation. This shows that cases that were isolated earlier, relative to symptom onset, tended to infect fewer people. Accounting for overseas status and number of infectious days before isolation or quarantine children infect, on average, 62% the individuals that adults and the elderly infect ($p = 0.09$). Imported cases infect 70% of the number of secondary cases in comparison to domestic cases ($p = 10^{-5}$).

Asymptomatic cases

Of the 31 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.

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)
Asymptomatic/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)

Table 1: Summary statistics for the New Zealand epidemic by age and type of case. 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 ($\boldsymbol{P(R_i > 5)}$) and the probability of an individual not infecting anyone ($\boldsymbol{P(R_i = 0)}$) according to the fitted distribution. Numbers show the median (95% range) across 500 instances of reconstructed transmission tree.

425 **Discussion**

426 We analysed the transmission dynamics of COVID-19 using a comprehensive dataset
427 for a completed outbreak. Our analysis investigated the role of age and type of case
428 (imported or domestic) on reproduction number, secondary attack rate, likelihood of
429 superspreading. We also investigated the effect of various control measures, including
430 case isolation, population-wide restrictions, and self-isolation or quarantine of
431 international arrivals.

432 We found that children under 10 years contribute less to the spread of COVID-19 than
433 other age groups. Children under 10 infected fewer people on average, had a lower
434 secondary attack rate, and were less likely to be superspreaders. This is consistent
435 with existing evidence on the role of children in COVID-19 transmission [15,16].

436 The finding that superspreading is a significant contributor to transmission is
437 consistent with existing empirical and modelling studies [17 18 19]. Our results show that
438 among adults 20% of cases are responsible for between 65% and 85% of
439 transmission. This suggests that interventions targeting superspreaders or
440 superspreading events may be particularly effective in reducing the spread of COVID-
441 19. These may include restrictions on gathering size, particularly in closed
442 environments or crowded spaces [17].

443 In the dataset we used to reconstruct epidemic transmission tree, the index case is
444 either missing or not uniquely identified in some cases. It is also likely there were
445 asymptomatic or paucisymptomatic cases that were undetected. We therefore used
446 Monte Carlo methods to reconstruct missing cases and missing transmission routes
447 (nodes and links respectively in the epidemic tree). It is possible that phylodynamic
448 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.

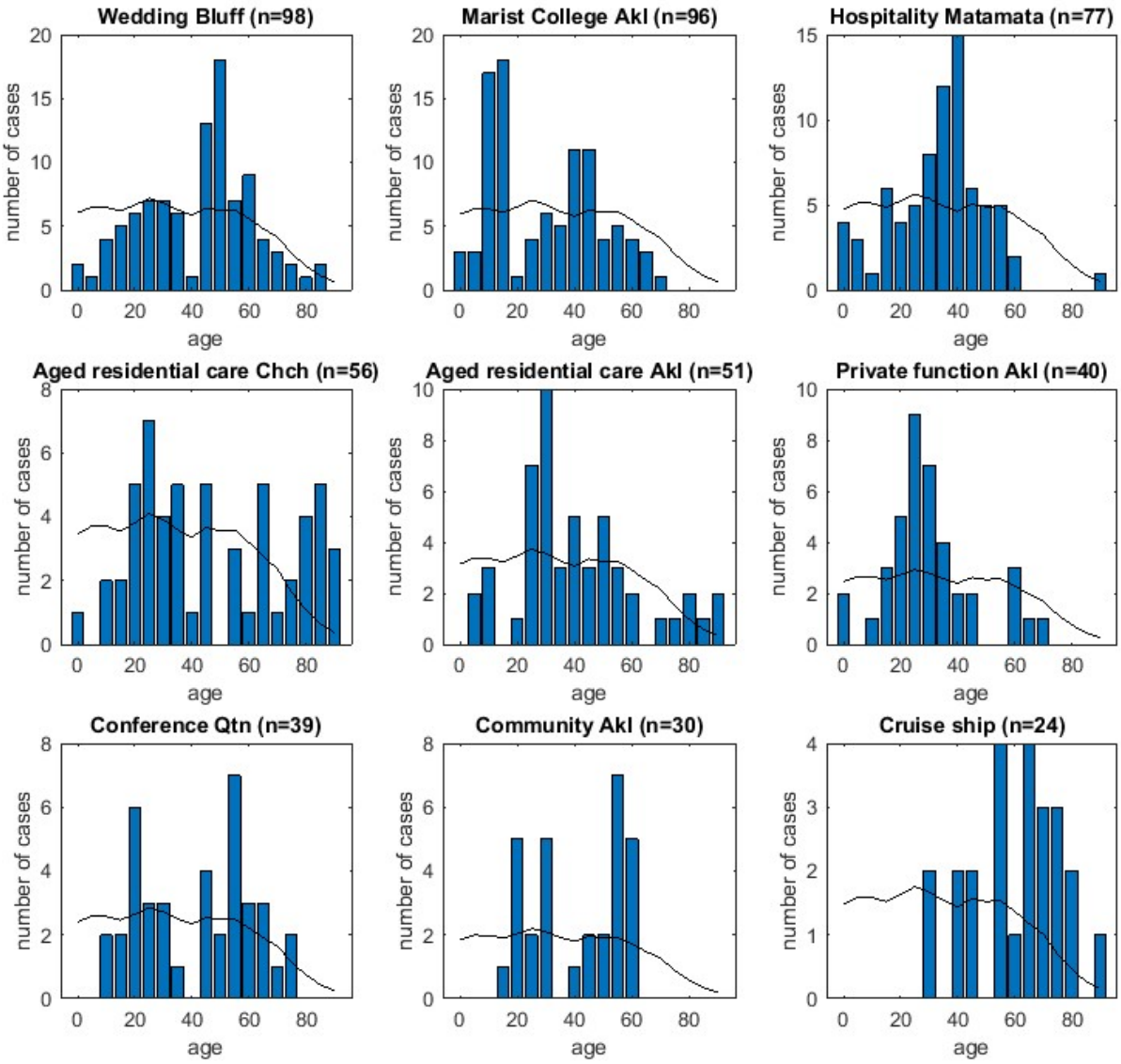
Acknowledgements

The authors are grateful to Annette Nesdale for discussions about the data analysed in this study. We acknowledge the support of StatsNZ, ESR, and the Ministry of Health in supplying data in support of this work. This work was funded by the Ministry of Business, Innovation and Employment and Te Pūnaha Matatini, New Zealand's Centre of Research Excellence in complex systems.

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.
3. Flaxman S, Mishra S, Gandy A, 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, 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, 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.
6. Ganyani T, Kremer C, Chen D, 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.
8. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; **368**(6491).
9. Lauer SA, Grantz KH, Bi Q, 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.
10. 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.
11. Jing Q-L, Liu M-J, Zhang Z-B, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *The Lancet Infectious Diseases* 2020.
12. Tindale L, Coombe M, Stockdale JE, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *MedRxiv* 2020.
13. 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.
14. 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.
15. Li W, Zhang B, Lu J, et al. The characteristics of household transmission of COVID-19. *Clinical Infectious Diseases* 2020.
16. Zhang J, Litvinova M, Liang Y, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 2020.
17. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *International journal of infectious diseases* 2020.
18. Adam D, Wu P, Wong J, et al. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. 2020.
19. 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.

Supplementary Material



Supplementary Figure S1. 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.