

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

3

4 **Authors:** A. James<sup>1,2\*</sup> PhD, M. Plank<sup>1,2</sup> PhD, S. Hendy<sup>2,3</sup> PhD, R. Binny<sup>2,4</sup> PhD, A.  
5 Lustig<sup>2,4</sup> PhD, N. Steyn<sup>2,3</sup> BSc (Hons).

6

7 **\*Corresponding author:**

8 Email: alex.james@canterbury.ac.nz

9 Phone: +64 3 369 2425

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

12

13 **Affiliations:**

14 <sup>1</sup> University of Canterbury, New Zealand

15 <sup>2</sup> Te Pūnaha Matatini, New Zealand.

16 <sup>3</sup> University of Auckland, New Zealand.

17 <sup>4</sup> Manaaki Whenua, Lincoln, New Zealand.

18 **Abstract:**

19 New Zealand had 1499 cases of COVID-19 before eliminating transmission of the  
20 virus. Extensive contact tracing during the outbreak has resulted in a dataset of  
21 epidemiologically linked cases. This data contains useful information about the  
22 transmission dynamics of the virus, its dependence on factors such as age, and its  
23 response to different control measures.

24 We use Monte-Carlo network construction techniques to provide an estimate of the  
25 number of secondary cases for every individual infected during the outbreak. We then  
26 apply standard statistical techniques to quantify differences between groups of  
27 individuals.

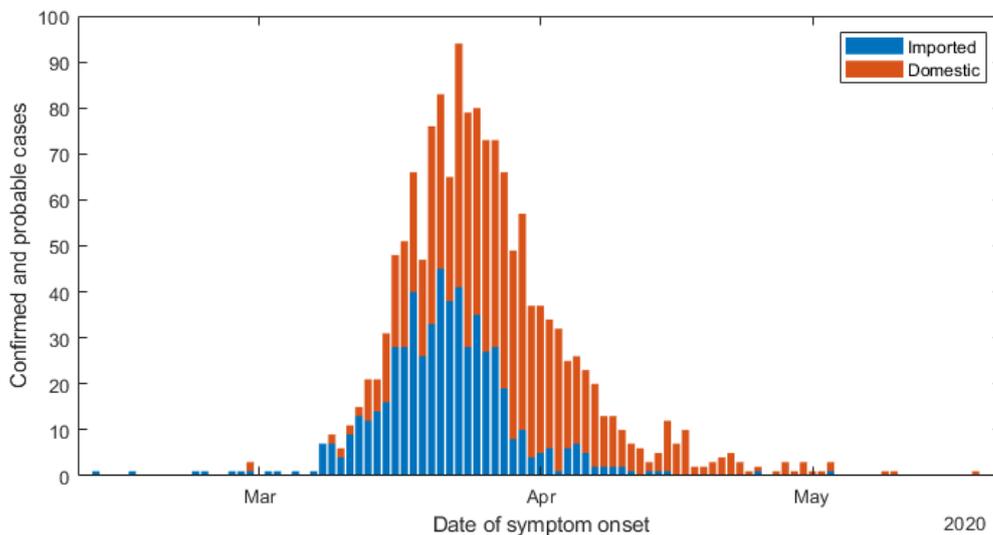
28 Children under 10 years old are significantly under-represented in the case data.  
29 Children infected fewer people on average and had a lower secondary attack rate in  
30 comparison to adults and the elderly. Imported cases infected fewer people on  
31 average and had a lower secondary attack rate than domestically acquired cases.  
32 Superspreading is a significant contributor to the epidemic dynamics, with 20% of  
33 cases among adults responsible for 65-85% of transmission. Asymptomatic cases  
34 infected fewer individuals than clinical cases. Serial intervals are approximately  
35 normally distributed ( $\mu = 5.0$  days,  $\sigma = 5.7$  days). Early isolation and quarantine of  
36 cases reduced secondary transmission rates.

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

39

## 40 Introduction

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



47  
48 **Figure 1:** Daily new cases of COVID-19 in New Zealand from 26 February to 31 May  
49 2020, split into cases with a recent history of international travel (blue) and those  
50 without (orange).

51  
52 Progressively stricter border measures were introduced from mid-March to mid-April.  
53 From 16 March, all international arrivals to New Zealand were required to go into 14  
54 days self-isolation. On 20 March, the border was closed to non-citizens and non-  
55 residents. From 10 April, all international arrivals were required to go into 14 day  
56 mandatory government-managed quarantine.

57 On 19 March, gatherings of more than 100 people were banned and, on 21 March, a  
58 four-level alert level system was introduced. Alert level 1 was effectively business as  
59 usual but with border restrictions. New Zealand entered alert level 2 on 21 March, level  
60 3 on 23 March, and level 4 on 26 March. Alert level 4 included strict stay-at-home  
61 measures, and all non-essential business and schools were closed. The country  
62 remained at alert level 4 until 28 April, when it moved to level 3 allowing partial school  
63 openings and non-essential businesses to operate under strict non-contact guidelines.  
64 On 14 May, New Zealand moved to alert level 2 and most schools and businesses  
65 reopened but large gatherings were still banned. On 9 June, the country moved back  
66 to alert level 1 and a strict testing programme for international arrivals was established  
67 to complement the ongoing 14-day mandatory quarantine.

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

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

80 In this paper, we examine the epidemic transmission tree in New Zealand using  
81 detailed case data and contact tracing information. This is an unusual dataset because  
82 it describes a closed outbreak in which there are no ongoing transmission chains, and  
83 many cases are epidemiologically linked to a likely index case. Because not all cases  
84 have a potential index case identified and some cases have multiple potential index  
85 cases, the transmission tree is not completely or uniquely defined. We reconstruct  
86 multiple instances of the transmission tree using a Monte-Carlo technique. This  
87 approach assigns a randomly selected index case to each secondary case based on  
88 the dates of symptom onset, cluster identification, and the results of contact tracing  
89 investigations. We assume a fixed proportion of infections were undetected because  
90 they were asymptomatic, paucisymptomatic, or for some other reason.

91 From this suite of generated transmission trees, we calculate the individual  
92 reproduction number  $R_i$  for each infected individual and explore the distribution of  $R_i$   
93 and the secondary attack rate at different points in the epidemic timeline. We explore  
94 how these quantities vary for different age groups and types of case, imported or  
95 domestic, and investigate the effect of early case isolation. We also investigate the  
96 degree of superspreading in the transmission dynamics. Although some assumptions  
97 are required to allow for missing or noisy data, our results are not model-dependent,  
98 unlike most estimates of effective reproduction number for COVID-19 [2] [3, 4]. Our  
99 sample size is larger than previous studies that have estimated the secondary attack  
100 rate (e.g. Bi, Wu (5), Ganyani, Kremer (6), Cheng, Jian (7)). This allows for more  
101 reliable estimates and for an investigation of how secondary attack rate varies with  
102 age. We also reconstruct the entire transmission tree rather than just estimating of the  
103 serial interval distribution. This provides a richer source of information about the

104 transmission dynamics, allowing a more complete set of epidemiological parameters  
105 to be explored.

## 106 **Methods**

107 Of the 1499 confirmed and probable cases, 575 had a recent history of overseas travel  
108 and were considered to be imported cases. The remaining 924 cases were considered  
109 to be infected within New Zealand and are referred to as domestic cases. Following  
110 contact tracing interviews, 636 of the domestic cases were each assigned up to five  
111 potential index cases. The remaining domestic cases did not have a potential index  
112 case identified. 627 of the domestic cases were linked to one of 18 significant clusters.  
113 Of the cases linked to clusters, 146 had no potential index case identified. For 1444  
114 cases, the number of close contacts, obtained via case recall at a contact tracing  
115 interview, was recorded. The number of close contacts varied widely between  
116 individuals, e.g. 753 cases had 2 or fewer contacts and 26 cases had more than 50.  
117 31 cases were recorded as being asymptomatic at the time of testing. All cases  
118 recorded age.

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

125 In addition to the 31 recorded asymptomatic cases, each generated transmission tree  
126 included additional subclinical or asymptomatic cases so that the total proportion of  
127 asymptomatic cases was  $p_{asympt} = 0.33$  on average. These additional cases were

128 randomly sampled from the recorded cases with the symptom onset date randomly  
129 jittered  $D_i = D_i + X_i$  where  $X_i \sim N(\mu = 0, \sigma = 2)$ . Model-generated asymptomatic cases  
130 had the same cluster status and potential index cases as the sampled case.

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

$$135 \quad SI = GT - I_1 + I_2$$

136 where  $GT$  is the generation time (time from infection of index case to infection of  
137 secondary case), and  $I_1$  and  $I_2$  are the incubation periods of the index and secondary  
138 case respectively. We assume that the generation time  $GT \sim Weibull(\mu = 5.05, \sigma =$   
139  $1.93)$  days[8] and the incubation period  $I_j \sim \Gamma(\mu = 5.5, \sigma = 2.28)$  days[9] and use this to  
140 calculate the probability density function for the serial interval  $f(SI)$ . Note that although  
141 the generation time must be positive, there is a positive probability of a negative serial  
142 interval.

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

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

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

160 **Resulting serial interval** The resulting serial interval distribution across all  $M$   
161 transmission trees was fitted with a normal distribution.

162 **Time varying effective reproduction number** We calculated the mean reproduction  
163 number for all clinical individuals, split by imported status with onset that day (which is  
164 approximately equivalent to peak infectiousness) in transmission tree  $j$ . The reported  
165 statistic is the median and 95% range over the  $M$  transmission trees.

166 **Reproduction number and superspreading** We fitted a negative binomial  
167 distribution to the effective reproduction numbers in each transmission tree separately  
168 for clinical individuals of different groups (age and domestic/imported status) and with  
169 onset date  $D_{ij}$  in different time periods for each transmission tree. This gave an  
170 estimate of the mean effective reproduction number  $\bar{R}_j$  for that group of cases in the  
171  $j^{\text{th}}$  transmission tree, and the associated dispersion or superspreading parameter  
172  $k_j[10]$ . We also calculated the probability of causing no secondary infections or more  
173 than 5 secondary infections for each tree. The reported statistic is the median and

174 95% range over the  $M$  transmission trees. If a negative binomial could not be fitted to  
175 a particular instance a Poisson distribution was used instead (i.e.  $k = \infty$ ).

176 **Secondary attack rate** Secondary attack rates of clinical individuals were calculated  
177 for each group in each transmission tree by pooling the total number of secondary  
178 cases and the total number of contacts [11]

179 
$$SAR = \frac{\sum secondary\ cases}{\sum close\ contacts}$$

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

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

191 
$$\log(\bar{R}_i) \sim (symptom\ onset - isolation)^2 + Is\ child + Is\ imported.$$

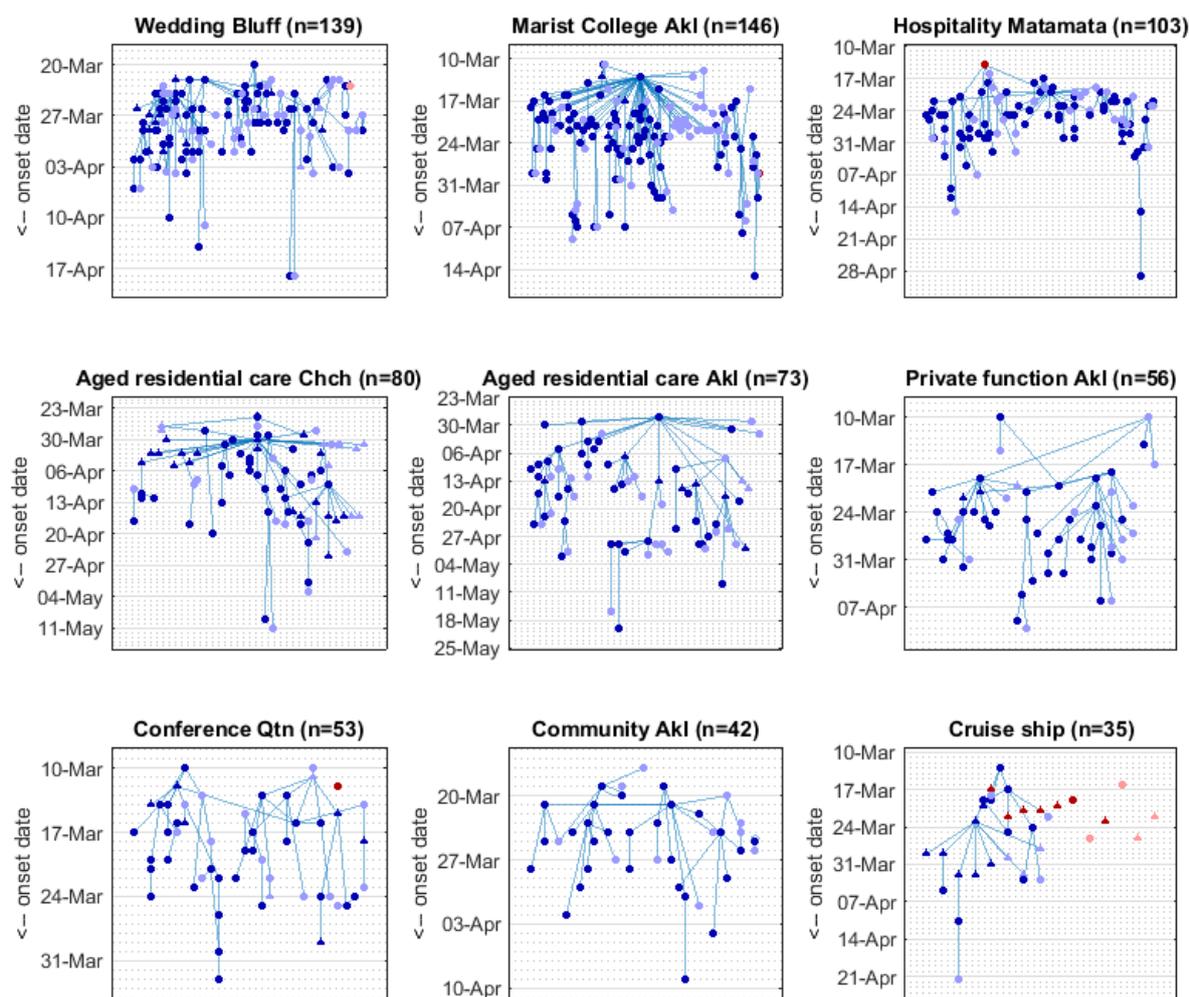
192 In the regression cases with  $\bar{R}_i = 0$  (i.e. with no secondary cases in any of the  $M$   
193 instances) were offset to half the minimum value of  $\bar{R}_i$  in the dataset. The results are  
194 presented graphically for an individual with the expected number of contacts.

195

## 196 **Results**

197 One instance of the reconstruction of New Zealand's nine largest clusters is shown in  
198 Figure 2. These clusters were associated with settings typical of international patterns,  
199 including a high school, a wedding, hospitality venues, aged residential care facilities,  
200 and a conference. The age distribution of cases varies across these clusters  
201 (Supplementary Figure S1). Although one of New Zealand's largest clusters was  
202 associated with a school (Marist College cluster), more than half of the cases (55/96)  
203 in this cluster were in individuals over 20 years old. Of the 41 cases under 20, 35 were  
204 aged 10–20 years with only 6 confirmed cases in under 10s. In the two largest clusters  
205 associated with aged residential care facilities, there was as expected an over-  
206 representation of cases over 80 years old. However, the majority of cases in both  
207 these clusters (36/56 and 44/51 respectively) were under 65 years old, suggesting that  
208 staff and visitors are a more important driver of transmission than residents in aged  
209 residential care.

210



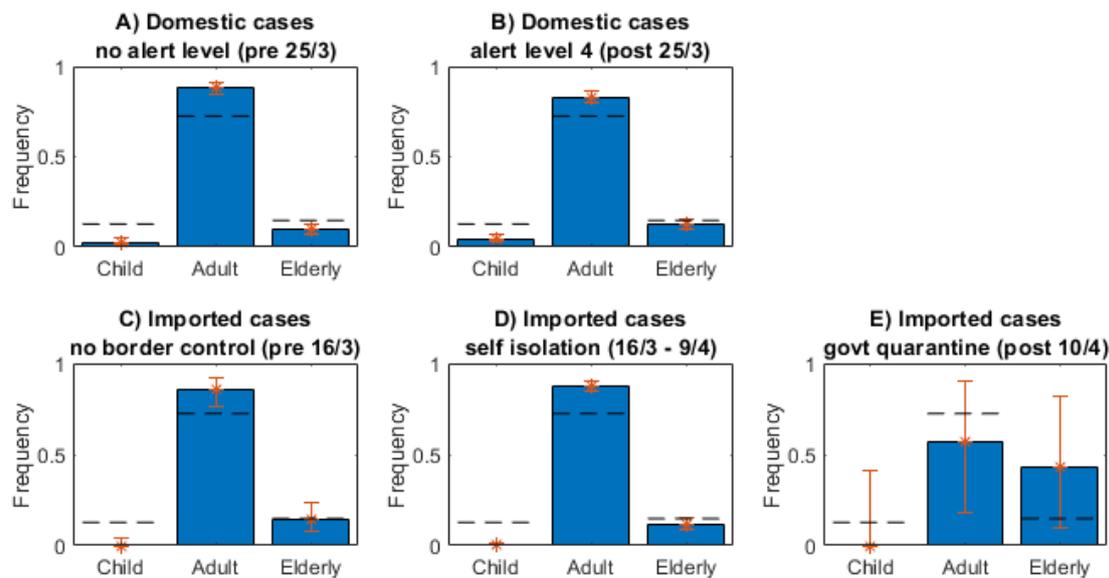
211

212 **Figure 2: Reconstruction of New Zealand's nine largest clusters.** Each panel  
213 shows one instance of the Monte Carlo method for reconstructing the epidemic  
214 transmission tree with cases plotted by date of symptom onset. Red = imported case;  
215 blue = domestic case; dark colours = symptomatic; light colours = asymptomatic;  
216 squares = age under 10 years; circles = age 10-65 years; triangles = age over 65  
217 years. The title of each panel describes the setting and location (Akl = Auckland; Chch  
218 = Christchurch; Qtn = Queenstown) associated with each cluster; n is the total number  
219 of cases in the reconstructed tree including actual cases and simulated (missing)  
220 asymptomatic cases.

## 221 **Incidence by age**

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

229 Children under 10 are significantly under-represented in the case data relative to the  
230 New Zealand population demographic of 12.7% (Figure 3). The domestic cases (Fig  
231 3 A,B) contain relatively more children under 10 than the imported cases and these  
232 infections predominantly occurred during alert level 4 (2.6% of cases were in children  
233 before 26 March, 4.4% on or after 26 March). During alert level 4, schools and day  
234 care centres were closed so children are likely to have spent more time in family  
235 settings where attack rates are higher [5] resulting in a higher incidence for this group.  
236 Of the 576 imported cases (Fig 3 C-E), only 2 were in children under 10 years old.  
237 Both these cases were connected with an imported adult case with the same travel  
238 history and symptom onset 4 to 7 days prior.



239

240 **Figure 3: Across all categories of case children are under-represented.**

241 Frequency of cases for the different types of case aged: children under 10 years;

242 adults 10-65 years; elderly over 65 years. Dashed lines show New Zealand population.

243 Error bars show the binomial 95% confidence interval.

244

### 245 Serial interval distribution

246 The distribution of serial intervals (Figure 4) resulting from the Monte Carlo

247 reconstruction contained a substantial number of outliers. It is likely these correspond

248 to cases that were infected by an undetected case, or the index case was wrongly

249 assigned either in the data or in the reconstructed tree. Including all serial intervals

250 from the reconstructed tree, the best fit normal distribution had mean 5.0 days and

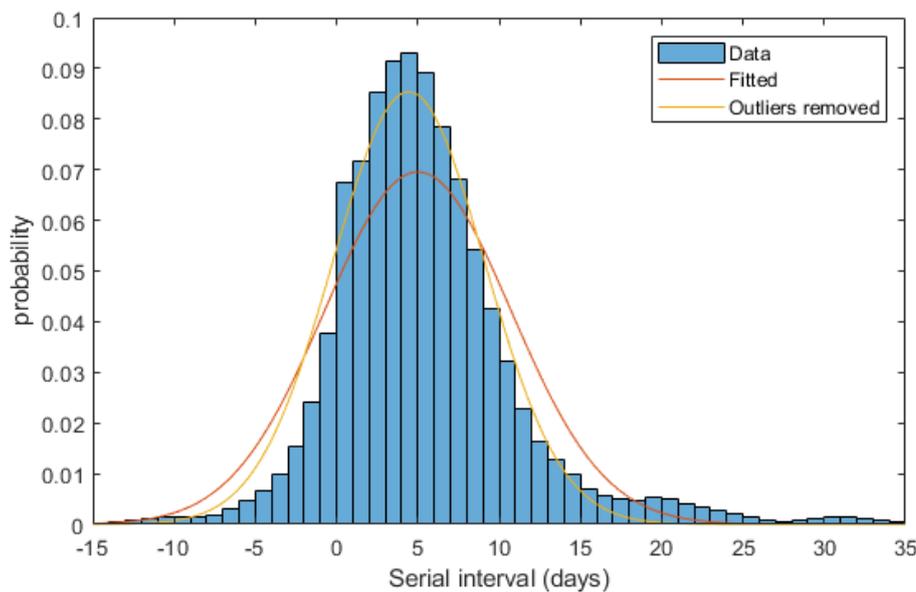
251 standard deviation 5.7 days. Excluding outliers (serial intervals longer than 20 days)

252 resulted in mean 4.4 days and standard deviation 4.7 days. This is comparable with

253 previous estimates of the mean serial interval [12, 13]. When allowing for negative

254 serial intervals in their Monte Carlo reconstruction method, Ganyani, Kremer (6) found

255 the serial interval to be  $3.86 \pm 4.76$  days (mean  $\pm$  std. dev.) for Singapore and  $2.90 \pm$   
256  $4.88$  days for Tianjin, China. These estimates for the mean serial interval are shorter  
257 than ours, but the finding that a significant proportion of serial intervals were negative  
258 is consistent. The data we analysed and that of Ganyani, Kremer (6) were collected in  
259 the presence of case isolation and contact tracing. It is likely these interventions will  
260 reduce the mean serial interval, to varying extents, by suppressing transmission late  
261 in the infectious period. Our results reflect the serial interval distribution in the  
262 presence of these interventions in the New Zealand context.



263

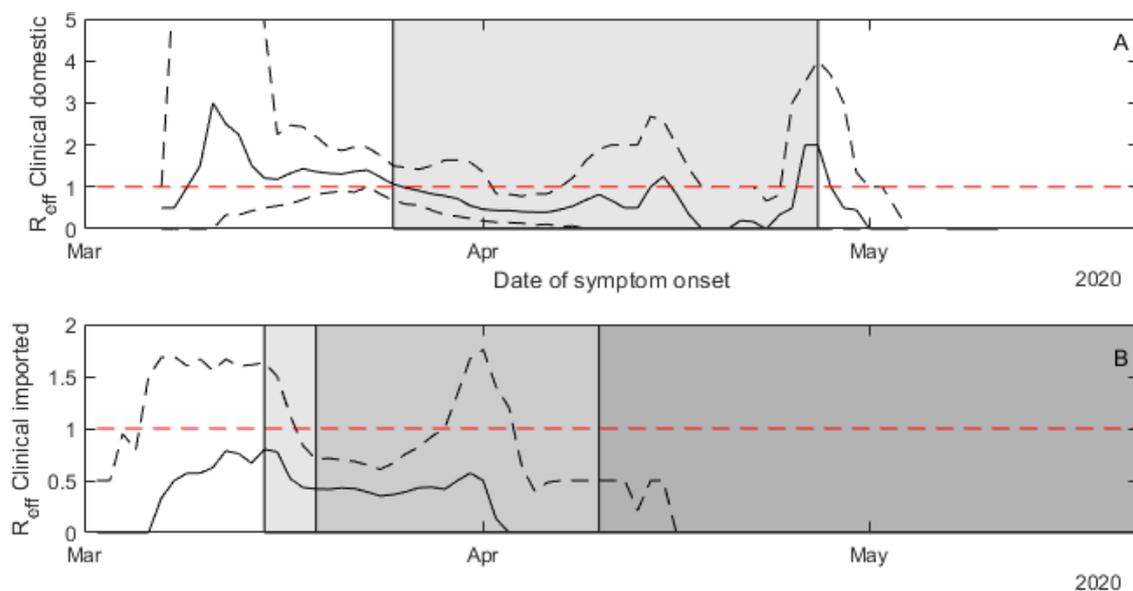
264 **Figure 4. Serial interval distribution with fitted distributions.** Red curve shows the  
265 normal distribution fitted to all data ( $\sim N(\mu = 5.0, \sigma = 5.7)$ ); yellow curve shows the  
266 normal distribution fitted to data excluding serial intervals longer than 20 days ( $\sim N(\mu =$   
267  $4.4, \sigma = 4.7)$ ).

#### 268 **Time-varying effective reproduction number**

269 For domestic cases, the average number of secondary cases infected by a clinical  
270 case before New Zealand entered alert level 4 on 26 March was  $R_{eff} = 1.47$  (382

271 index cases) (Figure 5A). During and after alert level 4 (26 March onwards), this  
272 dropped to  $R_{eff} = 0.69$  (521 index cases). In the period after alert level 4, the vast  
273 majority of cases were detected as part of contact tracing efforts and were isolated  
274 swiftly, often before symptom onset. The reproduction number for domestic cases  
275 showed two small peaks during the alert level 4 period. The first coincided with Easter  
276 weekend (10-13 April). The second coincides with a long weekend (25-27 April)  
277 immediately preceding the pre-announced easing of restrictions on 28 April.

278 For imported cases (Figure 5B), the number of secondary infections per case was  
279 much lower throughout the outbreak. Before self-isolation for international arrivals was  
280 introduced on 16 March,  $R_{eff} = 0.70$  (119 index cases). While self-isolation was in  
281 place from 17 March to 9 April,  $R_{eff} = 0.42$  (439 index cases). After mandatory  
282 government-managed quarantine was introduced on 10 April,  $R_{eff} = 0.08$  (7 cases).



283

284 **Figure 5: Increases in alert level and improved border measures resulted in**  
285 **decreases in secondary transmissions.** Expected value of  $R_{eff}$  each day for clinical  
286 domestic cases (A) and imported cases (B). Solid line show the median from 500  
287 instances of the reconstructed transmission tree, dashed lines are the 95% range.

288 Imported cases had a lower effective reproduction number than domestic cases. The  
289 date of secondary case exposure was assigned to be the date of symptom onset in  
290 the index case. Note that 10-13 April was the long Easter weekend and 27 April was  
291 a public holiday in New Zealand. The timing of the end of alert level four was  
292 announced on 22 April. Shading in (A) shows period of alert level 4 (26 March to 27  
293 April). Shading in (B) shows progressively stricter border restrictions introduced 16  
294 March, 20 March and 10 April.

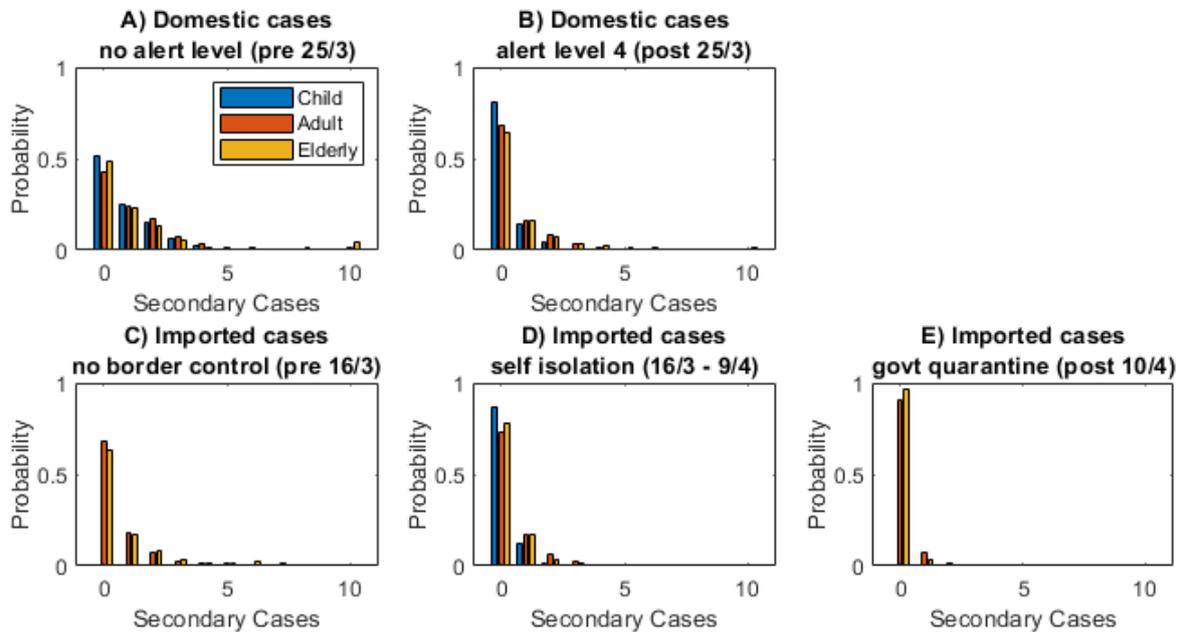
295

### 296 **Effect of age on reproduction number and superspreading**

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

307 During alert level 4 (Fig. 5B, Table 1),  $R_{eff}$  dropped below 1 for the under 10 and the  
308 10 to 65 age groups, but stayed above 1 in over 65 group. This may be due to the  
309 over-representation of cases in aged care facilities in the later stages of the epidemic,  
310 where it may be more difficult to limit interactions. The proportion of superspreaders

311 decreased in adults during alert level 4 but not in the elderly. There were no recorded  
312 superspreaders under 10 years old at any stage in the epidemic.



313

314 **Figure 6: Children under 10 tended to infect fewer people and were less likely to**  
315 **be superspreaders.** Distribution of individual reproduction numbers for domestic and  
316 imported cases in each age group: under 10 years (blue), 10 to 65 years (red), over  
317 65 years (yellow).

318

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

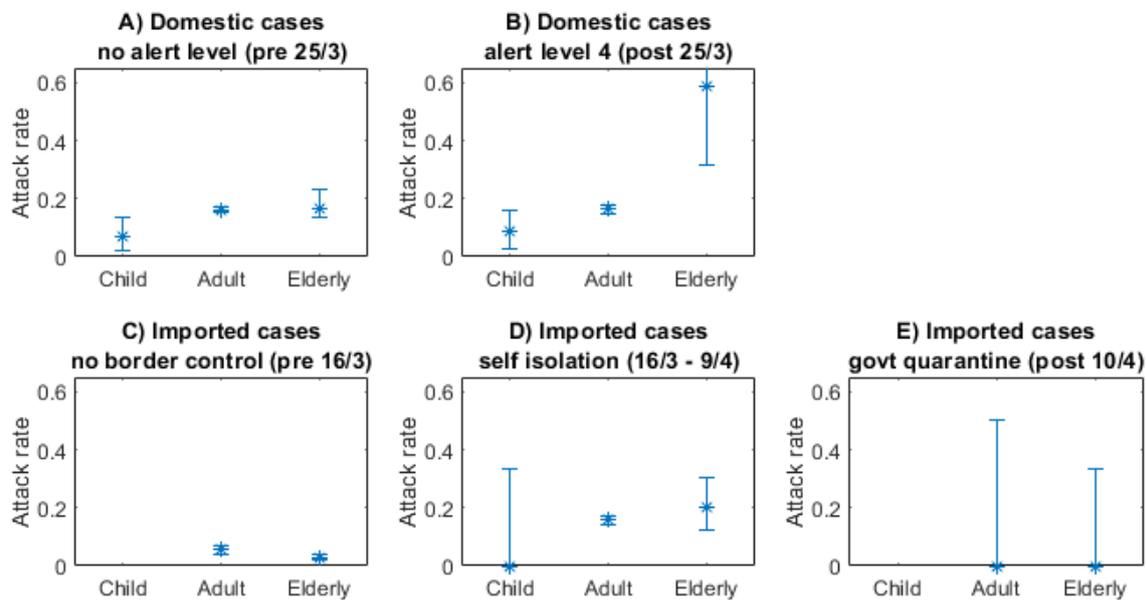
322 As seen previously, imported cases overall had a lower  $R_{eff}$  than domestic cases  
323 (Figure 6C-E). Before the introduction of any border measures, many of these cases  
324 would have been tourists and other visitors who are less likely to have the close  
325 community contacts that local residents have. Throughout the outbreak, imported

326 cases in the over 65 years age group had a higher  $R_{eff}$  than those aged 10 – 65.  
327 Imported cases in children, of which there were only 2 (<1%), had the lowest likelihood  
328 of being an index case even accounting for both these cases happening during the  
329 mandatory quarantine period. There were 5 superspreaders amongst the imported  
330 cases, all of which occurred during the self-isolation period between 9 and 19 March.  
331 The degree of superspreading in the transmission dynamics can be quantified by the  
332 variance in the distribution of individual reproduction numbers  $\bar{R}_i$ . This can be  
333 measured by the dispersion parameter  $k$  of a negative binomial distribution fitted to  
334 the data [10, 14]. The smaller the value of  $k$ , the more heavy-tailed the distribution of  
335 secondary cases, and therefore the more superspreading is driving the transmission  
336 process. The results show that older age groups tend to have more superspreading  
337 than younger ones (Table 1). This could be influenced partly by the number of cases  
338 transmitted in aged residential care facilities. Among the 10-65 and over 65 year age  
339 groups, the estimated values of  $k$  range from 0.27 to 0.70. This means that 20% of  
340 cases are responsible for between 65% and 85% of transmission.

#### 341 **Secondary attack rate**

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

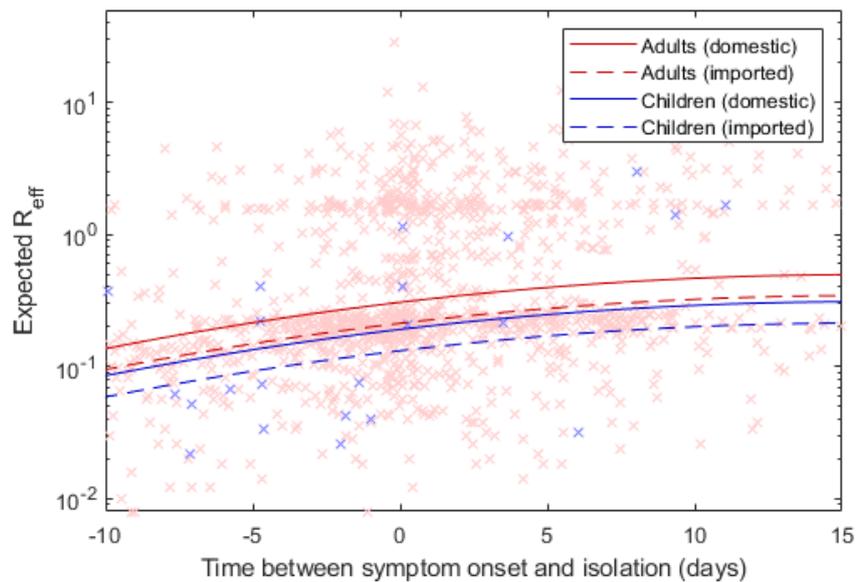
349 Overall, the secondary attack rate for imported cases was much lower than for  
350 domestic cases, particularly during the period before introduction of border measures.  
351 This is likely due to this group including a high proportion of visiting tourists for whom  
352 most contacts are in a casual rather than home or work setting. The secondary attack  
353 rate for imported cases increased after the self-isolation measures were introduced,  
354 possibly reflecting the environment and make-up of these groups. However, the  
355 secondary attack rate dropped significantly during the mandatory quarantine period.  
356 The lack of a clear effect of control measures on secondary attack rates suggests that  
357 reduction in the reproduction number was achieved primarily by reducing the number  
358 of close contacts rather than by reducing the likelihood of transmission when close  
359 contacts did occur.



360

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

363



364

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

371

### 372 **Effect of case isolation**

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

380

381 **Asymptomatic cases**

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

387

Group	Time period	N	$E(R_i)$	k	$P(R_i > 5)$	$P(R_i = 0)$
<b>Clinical cases (domestic)</b>						
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)
<b>Clinical cases (imported)</b>						
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)
<b>Asymptomatic/subclinical cases</b>						
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)

388 **Table 1: Summary statistics for the New Zealand epidemic by age and type of**  
389 **case.** Columns show: the number of cases in each group (N); the expected number  
390 of secondary infections caused by an individual in this group ( $E(R_i)$ ); the dispersion  
391 coefficient of a fitted negative binomial distribution (k); the probability of an individual  
392 infecting more than 5 people ( $P(R_i > 5)$ ) and the probability of an individual not infecting  
393 anyone ( $P(R_i = 0)$ ) according to the fitted distribution. Numbers show the median (95%  
394 range) across 500 instances of reconstructed transmission tree.

## 395 Discussion

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

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

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

413 In the dataset we used to reconstruct epidemic transmission tree, the index case is  
414 either missing or not uniquely identified in some cases. It is also likely there were  
415 asymptomatic or paucisymptomatic cases that were undetected. We therefore used  
416 Monte Carlo methods to reconstruct missing cases and missing transmission routes  
417 (nodes and links respectively in the epidemic tree). It is possible that phylodynamic  
418 data on the genetic sequence of the virus recovered from positive test results could

419 be used to reduce uncertainty in this step and improve the accuracy of the  
420 reconstructed tree.

421

## 422 **Acknowledgements**

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

428

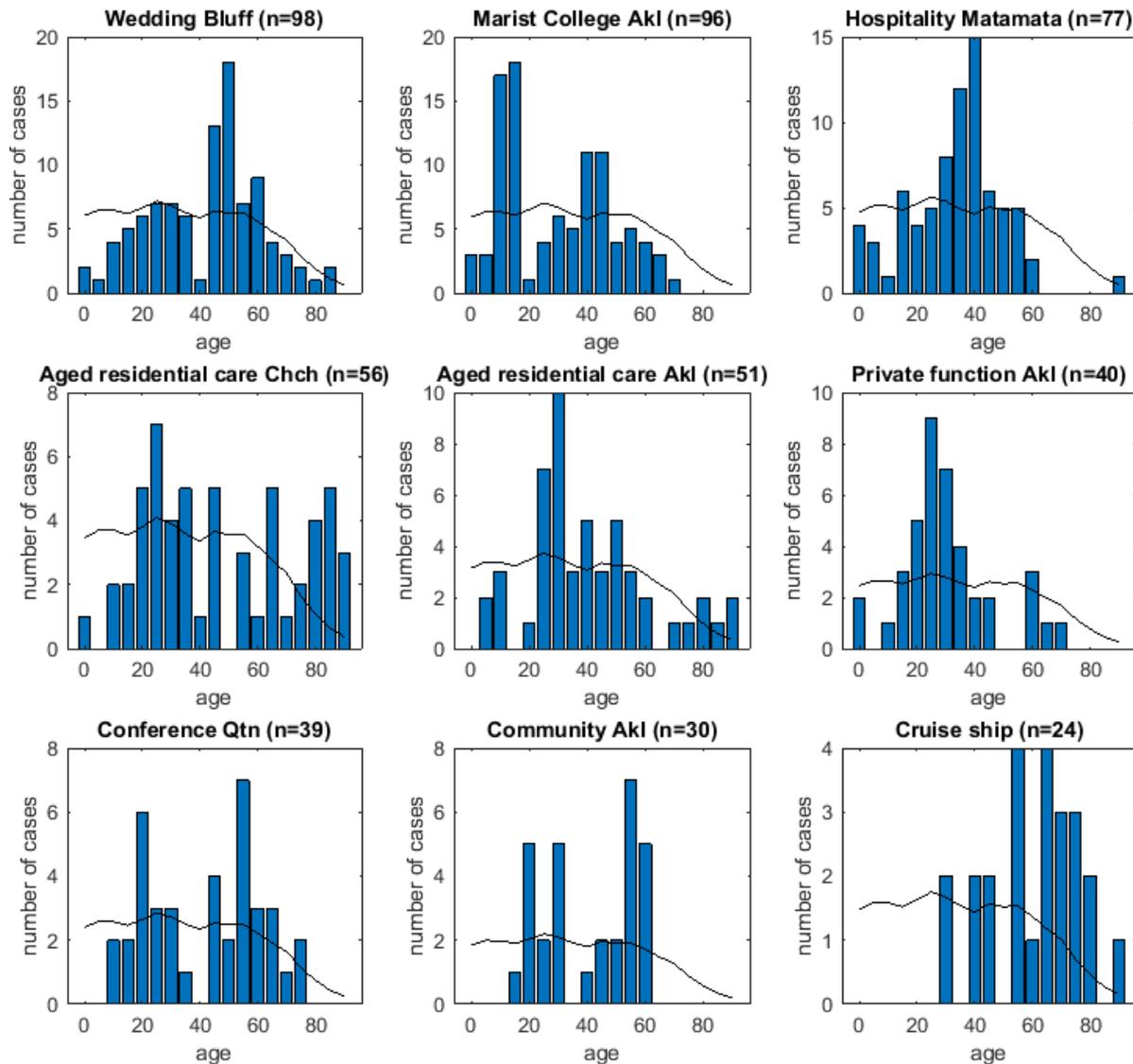
429

## 430 References

- 431 1. Binny RN, Hendy SC, James A, Lustig A, Plank MJ, Steyn N. Effect of Alert Level 4 on effective  
432 reproduction number: review of international COVID-19 cases. medRxiv. 2020.
- 433 2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher  
434 compared to SARS coronavirus. Journal of travel medicine. 2020.
- 435 3. Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T, et al. Report 13: Estimating the  
436 number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11  
437 European countries. 2020.
- 438 4. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact  
439 of physical distance measures on the transmission of COVID-19 in the UK. BMC medicine. 2020;18:1-  
440 10.
- 441 5. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in  
442 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet  
443 Infectious Diseases. 2020.
- 444 6. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation  
445 interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020.  
446 Eurosurveillance. 2020;25(17):2000257.
- 447 7. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H. Contact tracing assessment of  
448 COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after  
449 symptom onset. JAMA internal medicine. 2020.
- 450 8. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-  
451 CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368(6491).
- 452 9. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of  
453 coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and  
454 application. Annals of internal medicine. 2020;172(9):577-82.
- 455 10. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual  
456 variation on disease emergence. Nature. 2005;438(7066):355-9.
- 457 11. Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household secondary attack  
458 rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. The  
459 Lancet Infectious Diseases. 2020.
- 460 12. Tindale L, Coombe M, Stockdale JE, Garlock E, Lau WYV, Saraswat M, et al. Transmission  
461 interval estimates suggest pre-symptomatic spread of COVID-19. MedRxiv. 2020.
- 462 13. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly  
463 reported confirmed cases. Emerging infectious diseases. 2020;26(6):1341.
- 464 14. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion  
465 parameter for highly overdispersed data, with applications to infectious diseases. PloS one.  
466 2007;2(2):e180.
- 467 15. Li W, Zhang B, Lu J, Liu S, Chang Z, Cao P, et al. The characteristics of household transmission  
468 of COVID-19. Clinical Infectious Diseases. 2020.
- 469 16. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns  
470 shape the dynamics of the COVID-19 outbreak in China. Science. 2020.
- 471 17. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19)  
472 infections. International journal of infectious diseases. 2020.
- 473 18. Adam D, Wu P, Wong J, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading  
474 potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong.  
475 2020.
- 476 19. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19  
477 transmission using outbreak sizes outside China. Wellcome Open Research. 2020;5(67):67.



480 **Supplementary Material**



481

482 Supplementary Figure S1. Age distribution of cases in New Zealand's nine largest  
483 clusters. The black line shows the age distribution of the New Zealand population  
484 (number of cases that would be expected if the cluster followed the same age  
485 distribution as the New Zealand population). The title of each panel describes the  
486 setting and location (Akl = Auckland; Chch = Christchurch; Qtn = Queenstown)  
487 associated with each cluster; n is the total number of recorded cases in the cluster.