

1 **Predicting elimination of evolving virus variants**

2 Elliott Hughes^{1,3}, Rachele Binny^{2,3}, Shaun Hendy^{3,4}, Alex James^{1,3}

- 3 1. School of Mathematics and Statistics University of Canterbury, New Zealand.
4 2. Manaaki Whenua, Lincoln, New Zealand.
5 3. Te Pūnaha Matatini: the Centre for Complex Systems and Networks, New Zealand.
6 4. Dept of Physics, University of Auckland, New Zealand.

7
8 Keywords: SARS-CoV-2, Covid-19, Branching process, virus evolutionary tree, selection pressure

9
10 **Abstract**

11 As the SARS-CoV-2 virus spreads around the world new variants are appearing regularly. Although some
12 countries have achieved very swift and successful vaccination campaigns, on a global scale the vast majority
13 of the population is unvaccinated and new variants are proving more resistant to the current set of vaccines.
14 We present a simple model of disease spread which includes the evolution of new variants and varying vaccine
15 effectiveness to these new strains. We show that rapid vaccine updates to target new strains are more
16 effective than slow updates and containing spread through non-pharmaceutical interventions is vital whilst
17 these vaccines are delivered. Finally when measuring the key model inputs, e.g. the rate at which new
18 mutations and variants of concern emerge, is difficult we show how an observable model output, the number
19 of new variants which have been seen, is strongly correlated with the probability the virus is eliminated.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

A Centre of Research Excellence hosted by the University of Auckland



20 Introduction

21 The recent emergence of multiple SARS-CoV-2 variants poses a risk to the global efforts to control
22 the COVID-19 pandemic. In the short term the increased transmissibility of some variants raises the
23 risk of uncontrolled outbreaks (Davies et al., 2021; Tegally et al., 2020). In the long term, the
24 increasing genetic diversity of the virus may result in reduced vaccine efficacy due to immune escape,
25 i.e. variants may mutate so that antibodies which target other variants cannot successfully bond to
26 viral particles from the new variant (Atlani-Duault, Lina, Chauvin, Delfraissy, & Malvy, 2021; Callaway,
27 2021). Many models for COVID-19 spread have assumed that only one variant will be active at any
28 given time (James, Plank, Binny, Hannah, et al., 2020; James, Plank, Binny, Lustig, et al., 2020), a
29 simplifying assumption which is no longer accurate. Here we construct a simple stochastic branching
30 process model for multi-strain epidemics of COVID-19. By simulating outbreaks under different
31 assumptions, we suggest some possible dynamics for genetic diversity and propose observable
32 measures to assess the effect of these multiple strains on long term elimination.

33 Genomic data of the Sars-CoV-2 virus has been widely shared worldwide. Although RNA viruses in
34 general tend to mutate quickly, coronaviruses in particular mutate much slower, likely due to a proof-
35 reading enzyme that corrects copying mistakes which are potentially fatal to the virus. Although 12,000
36 mutations were spotted in Sars-CoV-2 within a year of its emergence almost all of these appeared to
37 have almost no consequence on the virus's transmission rate (Callaway, 2020).

38 The mutation rate of a virus can be thought of on many levels. At the fastest level there are small
39 genetic mutations which happen at almost every infection and can be used to track individual
40 infection routes by public health departments. Conversely, there are the much-publicised Variants of
41 Concern (VoCs), highly significant mutations seen across many individuals (Centers for Disease
42 Control and Prevention (CDC), 2021; World Health Organization, 2021). For a new strain to be
43 classified as a Variant of Concern there must be evidence of impact on diagnosis, treatments or
44 vaccines, increased transmissibility or increased severe disease. Our model includes only impact on
45 vaccines and increased transmissibility, and like reality not all new variants that appear in the model
46 will have these characteristics, most variants are simply small changes of very little consequence. As
47 of May 2020, the CDC had recognised five Variants of Concern (VoC) for Covid-19 which were all first
48 detected in regions that had high infection rates (B.1.1.7 - UK, P.1 – Brazil, B.1.351 – South Africa,
49 B.1.427 and B.1.429 – California, USA). Conversely there are many regions that have had very high
50 infection rates and no VoCs have been recognised (e.g. France, Italy, Spain, Russia). This could be in

51 part due to low rates of genetic sequencing but it is likely that no VoCs have arisen in many of these
52 areas despite the high transmission rates.

53 Although the number of VoCs is small, the number of mutations seen in widespread genetic testing
54 is much larger. Within a year of the virus emerging there were many distinct viral clades in global
55 circulation (Hamed, Elkhatib, Khairalla, & Noreddin, 2021). However, it was after 13 months in Jan
56 2021 that the first VoC was identified, B.1.1.7, which rapidly became the main variant in circulation in
57 Europe (Planas et al., 2021). The current literature on the genetic diversity of SARS-CoV-2 focuses on
58 reconstructing past genetic trees rather than forecasting the long-term number of active
59 antigenically distinct variants e.g. Nie et al. (2020). While these analyses provide useful information,
60 as vaccines have begun to be distributed it has become crucial to determine the risk of an
61 antigenically distinct variant emerging.

62 Multi-strain influenza outbreaks have been modelled using both branching processes and
63 compartmental models (Kucharski, Andreasen, & Gog, 2016). Branching processes are inherently
64 more computationally intensive than compartmental approaches and as a result they are more often
65 used to find the probability of a new variant becoming established than to predict long-term genetic
66 diversity e.g. Bedford, Rambaut, and Pascual (2012). Compartmental models can be broadly
67 separated into 'history-based' and 'status-based' approaches. History-based models attempt to
68 capture the infection history of each individual, while status-based models only record the current
69 immune status of individuals. While history-based models contain more detail, a naïve history-based
70 model for n strains will have at least $2n$ variables (Kucharski et al., 2016) and may be computationally
71 intensive as a result. Various attempts have been made to reduce the number of variables necessary
72 for a history-based model, but these either require certain assumptions to be met and/or have
73 increased analytical complexity. Status-based models are simpler and have been used to model the
74 dynamics of other RNA viruses, e.g. Koelle, Kamradt, and Pascual (2009) or Koelle, Khatri, Kamradt,
75 and Kepler (2010), but often lack realistic details. For a summary of the advantages and
76 disadvantages of various approaches, see Kucharski et al. (2016).

77 Table S1 gives details of the administration and uptake of vaccines in different countries, and the
78 incidence of each VoC (as a proportion of all sequenced infected cases) . As of early April all countries
79 with a significant number of infections had significant numbers of infections by at least one VoC.
80 Between March and April 2021 there was a clear rise in the number of infections caused by new
81 variants of SARS-CoV-2, in particular the B.1.1.7, P.1 and B.1.617.1 variants. Vaccination rates as of

82 April 2021 varied between <0.1% of the population daily, e.g. New Zealand, and over 0.8% of the
83 population daily, e.g. UK and USA. The proportion of the population who are likely to be vaccinated
84 ranges from less than 50% (USA) to over 75% (NZ, Brazil).

85 Our model tracks virus mutations at a rate somewhere between the two extremes, we assume there
86 is a small probability the virus can mutate each time a new person is infected. Although in reality
87 mutations happen within a single person our model assumes each individual has single variant which
88 does not change over time and the resulting secondary case is infected by the new mutation. The
89 mutation rate parameter is intended to track significant mutations where the transmission rate of
90 the virus changes significantly (either increasing or decreasing) rather than each individual genetic
91 mutation. A new variant will have a different transmission rate to the parent variant it evolved from
92 but this rate may be smaller and hence the variant quickly vanishes.

93 The model includes selection pressure as new variants are more likely to gain traction and spread to
94 many individuals if they have a higher transmission rate than other variants, and are able to evade
95 any immunity the population has already built up from infection by or vaccination against other
96 variants.

97 We explore the model and give results on how the probability of elimination changes as the speed at
98 which vaccines are rolled out is increased for different cross-immunity and mutation rate scenarios.
99 Finally, we discuss how to estimate the probability of elimination in a system where the key
100 parameters are not known.

101

102 **Model**

103 The model consists of four separate parts: disease transmission, virus evolution, immunity and
104 vaccination. The disease transmission model is based on a simple branching process. As each
105 infection occurs there is a small chance the virus mutates into a new variant. The virus evolution
106 model builds a simple tree of these variants tracking how closely related they are. The immunity
107 model predicts immunity to infection depending on how closely related the virus variants are. Finally
108 we consider a number of immunisation strategies and assess their effect on the epidemic spread.

109 *Disease transmission model* We consider a population of N_p individuals. Each realisation starts with
110 N_0 infected individuals who expose, on average, R_1 individuals each over D_{Inf} days. N_0 is chosen to

111 be sufficiently high such that in a fully-susceptible population almost all realisations will result in an
112 epidemic wave rather than elimination by chance. New exposures to an infected individual occur
113 uniformly over that individual's infectious period of D_{Inf} days. The actual number of exposures
114 caused by an infectious individual is Poisson distributed with expected value R . The population is
115 well-mixed, i.e. all individuals are equally likely to come into contact and infect each other. An
116 exposed individual will become infected provided they are not currently infected and are not immune
117 to that variant of the virus. There is no latent period, once an individual is infected they start infecting
118 others immediately. Individuals can acquire immunity (after infection or vaccination) so the number
119 of susceptible individuals in the population varies over time, however the model assumes there are
120 no fatalities so the population size N_p remains constant over time. Once an individual is immune to a
121 particular strain they remain immune forever, i.e. there is no waning immunity.

122 *Virus evolution model* When an individual is infected they are usually infected with the same virus
123 variant as the index case from whom they were infected. However, there is a small
124 probability, p_{mutate} , that they will be infected with a new variant of the virus. Each new variant has
125 its own transmission rate, R_j , which is the expected number of new cases caused by a single individual
126 infected with variant j in a fully susceptible population, i.e. no individual has immunity to variant j .
127 The transmission rate of a new variant, R_j , depends on the transmission rate of the parent variant, R_p

$$128 \quad R_j \sim \max(R_p + R_\sigma X_j, R_{max})$$

129 where X_j is a normally distributed random variable, $X_j \sim N(0,1)$. We cap R_j at a maximum rate R_{max} .
130 This assumes that transmission rates will be self-limiting as social distancing measures imposed
131 either by individuals or at a population level will stop very high transmission rates e.g. when
132 transmission rates are high, non-pharmaceutical interventions are likely to be implemented to curb
133 spread and public perception of risk increases.

134 Each new variant is a new node of the virus tree model and edges connect new variants with their
135 parent variant. The model produces a non-binary, rooted tree which allows multiple edges from a
136 single node as a variant begets many different mutations. The distance d_{ij} between nodes
137 representing variants i and j tells us how closely genetically related the two variants are, e.g. if two
138 variants share the same parent variant they are distance 2 apart on the virus tree.

139 *Immunity model* Once an individual has been infected with variant j of the virus they naturally acquire
140 full immunity to that variant. However, they may still become infected by a different stain of the virus.
141 The probability that an individual who has been infected with variant j is immune to variant i is I_{ji} .

142 I_{ji} is a decreasing function of, d_{ji} , the distance between nodes j and i in the virus tree. If two variants
 143 are very closely related there is a high probability that infection by one variant will confer immunity
 144 to the other. If variants are more distantly related the probability is much lower. Parameters are
 145 chosen so that if two variants are more than distance 6 from each other on the tree then there is a
 146 zero probability of immunity.

147 Although the immunity matrix, I , gives a probability of being immune to a particular variant, a
 148 particular individual's immunity to a given variant is a binary measure, i.e. if $I_{ji} = 0.5$ then 50% of
 149 individuals who have been infected with variant j are fully immune to variant i and 50% are fully
 150 susceptible. Immunity across variants is independent, e.g. catching variant A may make one person
 151 immune to variant B but not variant C, whereas it may confer immunity to both variants B and C in
 152 another individual.

Description	Parameter
Population size	$N_p = 10,000$
Initial number of infected individuals	$N_0 = 10$
Initial variant transmission	$R_1 = 1.2$
Maximum transmission (high and low transmission)	High: $R_{max} = 1.8$ Low: $R_{max} = 1.4$
Transmission variation of new variants	$R_\sigma = 0.1$
Number of infectious days	$D_{Inf} = 14$
Immunity rate at distance $d = 1, \dots, 6$ (high and low cross immunity)	High: $I(d) = 1, 0.9, 0.5, 0.3, 0.1, 0.05$ Low: $I(d) = 0.5, 0.45, 0.25, 0.15, 0.05, 0.025$
Virus mutation rate	$p_{mutate} = 0.01$
First day of vaccinations	$D_1 = 350$

153 Table 1: Fixed parameter values.

154 *Vaccinations* If a vaccine is developed to give immunity to a particular variant then, like previous
 155 infection with that variant, it is presumed to be 100% effective against that variant after a single shot.
 156 It also confers partial immunity against other variants following the same rules as the natural
 157 immunity model. Vaccination begins on day $D_1 = 350$ and fraction f_v individuals are vaccinated each
 158 day. When a vaccine is updated the old vaccine is discontinued immediately and all subsequent
 159 vaccinations use the vaccine developed against the new variant. Individuals are vaccinated in a
 160 priority order, those who have been infected are equally likely to be vaccinated as those that have

161 not. Initially we use a Continuous vaccination strategy i.e. when the last person on the priority list has
162 been vaccinated the list resets and returns to the first individual till either the entire population has
163 received that vaccine or the vaccine is next updated.

164 The model is run in time steps of one day. Results are averages taken from 1000 simulations at each
165 set of parameter values. All fixed parameter values are given in Table 1.

166

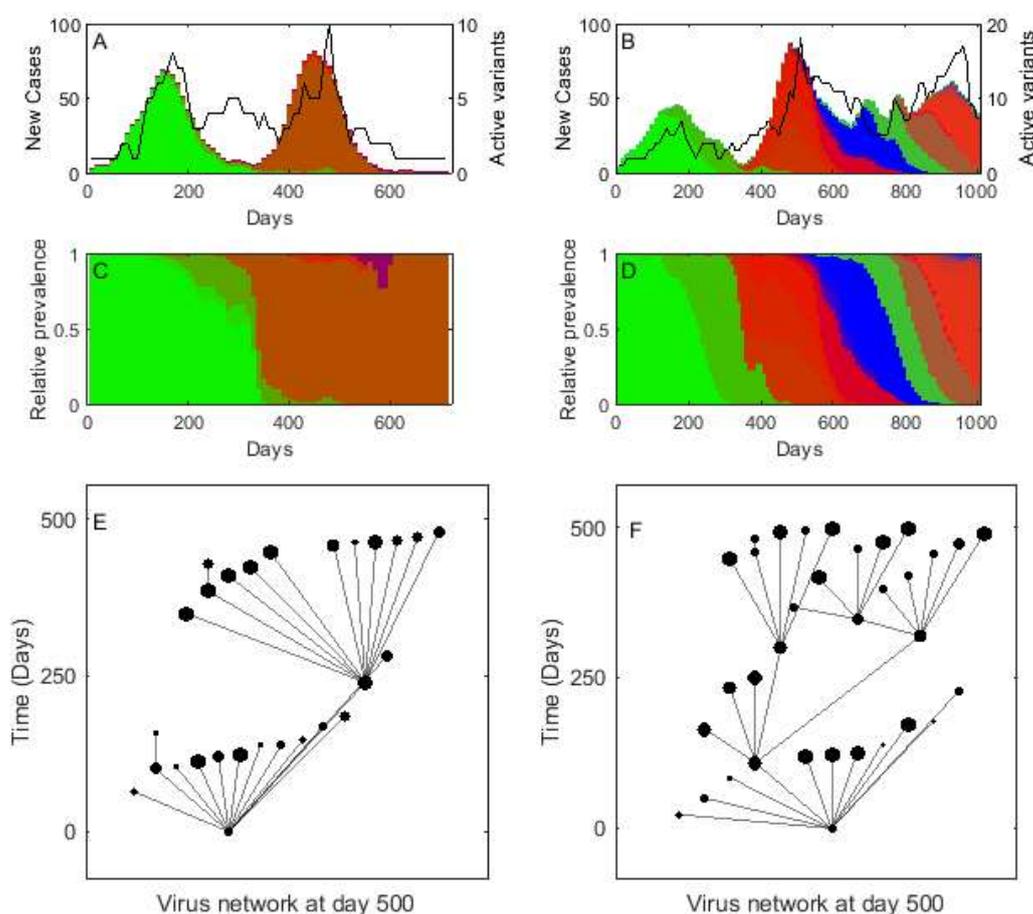
167 **Results - No vaccines available**

168 We start by considering results from the model assuming no vaccine is available. The raw model
169 output is a time series of the number of individuals infected with each virus variant and the tree of the
170 evolving virus. Each model realisation shows one of two distinct types of behaviour: either the disease
171 is self-limiting, the population reaches herd immunity, all variants die out and the disease is
172 eliminated; or the disease undergoes variant runaway where there are multiple variants in circulation,
173 and the virus is never eliminated as new variants appear faster than herd immunity to each variant is
174 achieved. This second situation is similar to the current global circulation of influenza viruses. Due to
175 the stochasticity in the model it is possible to see both types of behaviour at a single set of parameter
176 values.

177 Figure 1 shows example output for each of these behaviours. In both of these examples there is no
178 vaccination and herd immunity can only be reached through infection. In the first realisation (Figure
179 1 A, C, E) after approximately 600 days the virus has been eliminated. There were 2 distinct waves of
180 infection, peaking at days 150 and 440. Both peaks were dominated by a single virus variant though
181 other variants were active. The number of active variants, i.e. variants with at least one infectious
182 individual, peaks shortly after the peak in the number of infectious cases (Fig 1A). The virus tree shows
183 that the original variant produced 13 new variants. It was the last of these that appeared around day
184 250 that dominated the second wave and produced 12 new variants. The size of the node shows the
185 transmission rate of each variant.

186 Initially, the second realisation (Figure 1 B, D, F) follows a similar pattern, the first wave peaking at
187 day 150 is dominated by a single variant, the second wave peaking around day 450 is also dominated
188 by a single variant. But as this second wave peaks more new variants become active and these
189 variants are more distantly related than in the first realisation so there is less cross immunity between

190 variants. Here after the second wave the disease becomes more widespread with a large number of
 191 active variants and elimination through natural herd immunity is no longer possible.



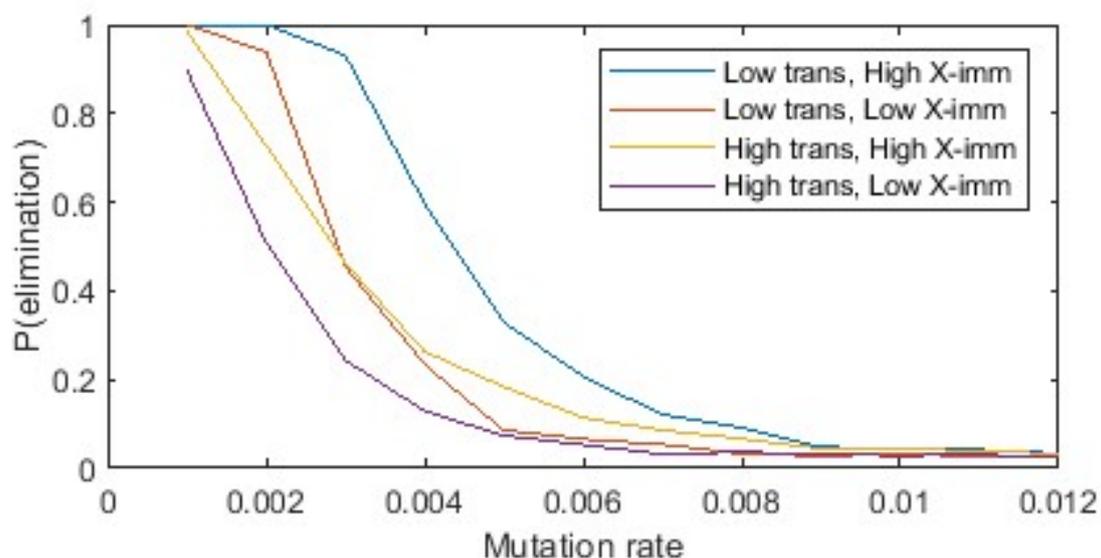
192

193 **Figure 1: Model realisations either eliminate the virus or continue indefinitely with multiple**
 194 **variants.** A and B show the number of new cases each day (averaged over a week, coloured bars) in
 195 two realisations of the model. Each virus variant is given a different bar colour. The black line shows
 196 the number of active variants (i.e. variants which had new infections) each week. C and D repeat the
 197 bars of A and B but scaled by the total number of new cases that week, to give the relative prevalence
 198 of each variant. E and F show the tree of variants at day 500, vertical position shows the time the
 199 variant appeared, node size shows the transmission rate R_j of each variant. The realisation in A, C and
 200 E eliminates after 650 days due to herd immunity. The realisation in B, D and F continues indefinitely.
 201 Bar colour is chosen randomly and is not related to any properties of the variant. Parameters as in
 202 Table 1 for high cross-immunity between strains and low transmission $p_{mutate} = 0.005$.

203

204 Extensive numerical exploration across many sets of parameter values shows that over 95% of
205 realisations that eliminated have less than 200 variants in total. This was used to give a numerical
206 definition of elimination, i.e. any realisation that reached 200 variants was presumed to continue
207 indefinitely without elimination. In the scenarios below, simulations are run until either elimination is
208 achieved or 200 variants have arisen whichever is sooner.

209 Figure 2 shows the probability of elimination with no vaccination for a range of mutation rates under
210 the four transmission and cross immunity scenarios. In all scenarios as the mutation rate increases
211 and new variants are more likely to appear the probability of elimination by herd immunity caused by
212 infection decreases. This probability is higher if new variants are able to reach higher transmission
213 levels, i.e. R_{max} is greater as there are fewer non-pharmaceutical interventions in place to keep
214 transmission rates low. Similarly, if infection provides only a small probability of immunity to an
215 evolved variant, i.e. there is low cross-immunity between variants, the probability of elimination is
216 lower.



217

218 **Figure 2: The probability of elimination decreases with mutation rate.** The probability of
219 elimination with no vaccine strategy. High transmission rates and low cross-immunity result in a
220 lower probability of elimination. Calculated from 1000 realisations at each mutation rate, parameters
221 as in Table 1.

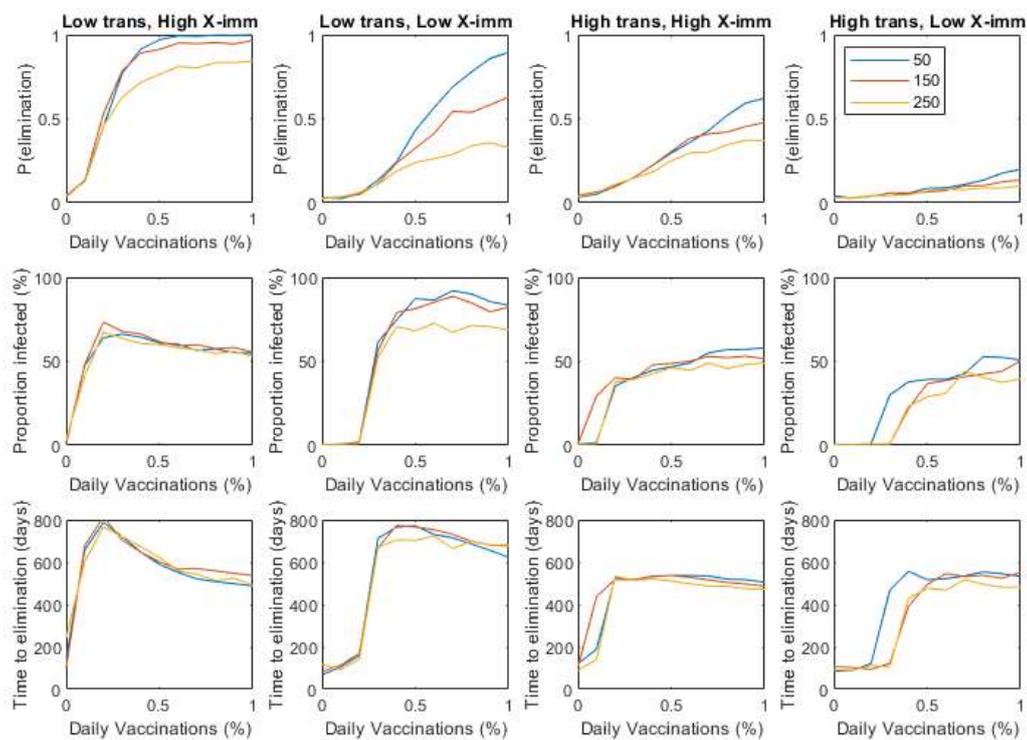
222

223 Results – Vaccines available

224 We now analyse the model results when a vaccine is available and can be updated to target new and
225 emerging variants. We assume the first vaccine is available 350 days after the first infections and
226 updated every D_{update} days after that. We consider a simple rolling vaccination programme where
227 there is only one vaccine available at any time. Vaccine delivery is constrained by logistics, i.e. there
228 is a maximum number of vaccine doses that can be administered each day. Every vaccine is 100%
229 effective against the variant it was developed for and the immunity for other variants follows the
230 immunity model. There is 100% vaccine take up, i.e. every individual in the population will take the
231 vaccine if offered. We fix the mutation rate at 0.01 mutations per infection so with no vaccine the
232 probability of elimination is almost zero. New vaccines are targeted at the variant that is most
233 prevalent (i.e. has the most active cases) on the day the new vaccine begins delivery (i.e. there is no
234 vaccine development lag), and that has not been a vaccine target before. Each day a fraction f_V of
235 the population is vaccinated.

236 *Vaccine development speed*

237 Figure 3 shows the results from each scenario for different vaccination rates, i.e. the fraction of the
238 population that is vaccinated each day, and different updating rates. As the vaccination rate increases
239 the probability of elimination also increases. Fast vaccination updates every 50 days are more
240 effective in achieving elimination than slower updates of every 250 days. This improvement with
241 update rate is most pronounced when the probability of achieving elimination is high. If the
242 probability of achieving elimination is low, usually due to a slow vaccination rate, then more regular
243 updating of the vaccine for new variants has only a marginal affect at best. High transmission rates
244 have a very detrimental effect on the elimination probability. Given that the virus is eliminated, the
245 proportion of people infected and the time taken to eliminate is not significantly affected by the
246 vaccine update rate but is affected by a very slow vaccine delivery rate.

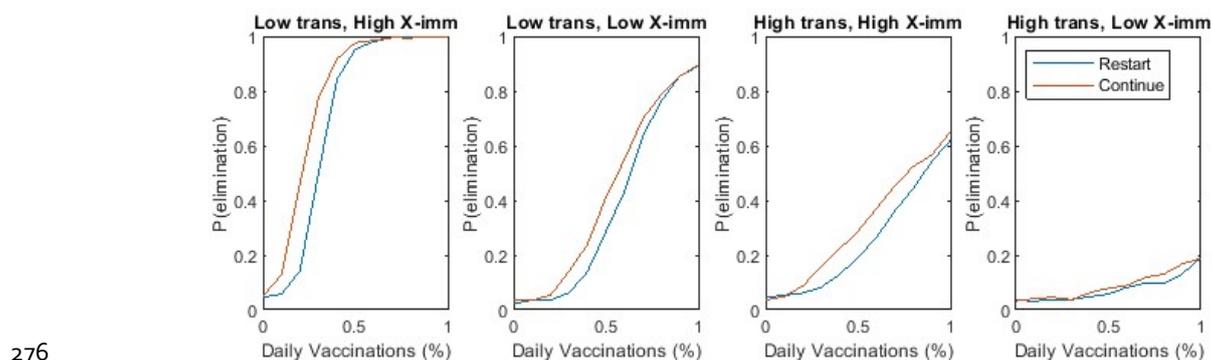


247

248 **Figure 3: Faster delivery and faster updating of vaccines leads to a higher probability of**
 249 **elimination.** The probability of elimination for a range of vaccination delivery rates, given as fraction
 250 of the population vaccinated each day, for three different vaccine updating rates (50, 150 and 250
 251 days) across three different infection scenarios. Also the proportion of the population infected and
 252 time to elimination given elimination is achieved. Parameters are as in Table 1.

253 *Vaccine priority* - We assume that there is a priority order for vaccine delivery. This order is decided on
 254 external factors such as job status (e.g. frontline workers and medical staff) or probability of severe
 255 disease. We assume these factors do not depend on an individual's transmission rate, i.e. individual's
 256 with higher transmission rates are not given higher priority, so do not affect the model presented
 257 here. If vaccinations are constrained by logistics, a simple question is: If a vaccine is updated before
 258 the entire population has been vaccinated who should receive the new vaccine? Should the
 259 programme continue and vaccinate previously unvaccinated individuals with the updated vaccine or
 260 should it reset back to the highest priority individuals, leaving some individuals completely
 261 unvaccinated but the higher priority individuals are more fully protected. Vaccine priority is the same
 262 for each vaccine.

263 In the Continuous strategy presented in Figure 3, the updated vaccine is given to the next individual
264 on the priority list. This means lower priority individuals will likely receive the new vaccine first but
265 every individual in the population will be vaccinated before the first individual receives their second
266 vaccination, i.e. the higher priority individuals will only receive a vaccine update after all the other
267 individuals have been vaccinated. We compare this to a Restart strategy where each new vaccine is
268 given to individuals at the start of the priority list first, i.e. high priority individuals will receive every
269 new vaccine as soon as it is available, while low priority individuals may never be vaccinated. Figure 4
270 shows that in all four transmission and cross-immunity scenarios the probability of elimination is
271 slightly higher using a Continuous strategy and ensuring all individuals have been vaccinated rather
272 than returning to the high priority individuals. This strategy may result in the highest probability of
273 elimination but it would take a structured population model, not used here, to assess whether this
274 strategy resulted in an increased risk to higher priority individuals who may have a higher mortality
275 rate or likelihood of severe infection needing hospitalisation.

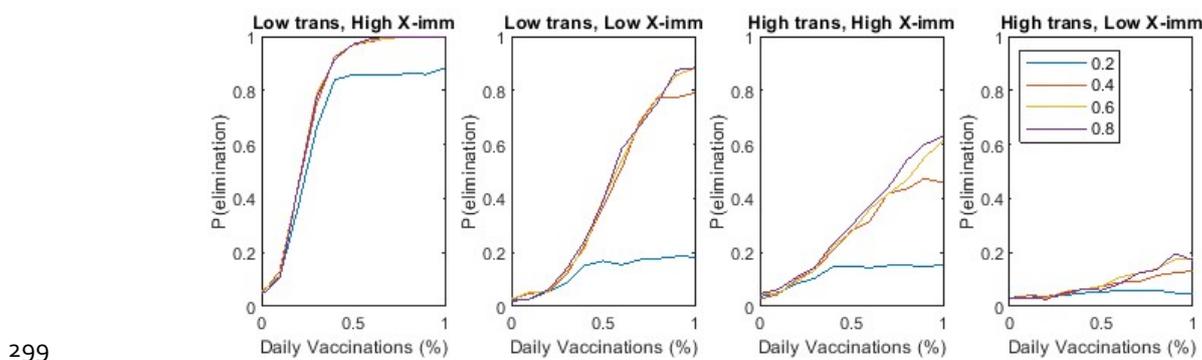


277 **Figure 4: A continuous vaccination programme where updated vaccines are given to**
278 **unvaccinated individuals, rather than revaccinating high priority individuals, gives a slightly**
279 **higher probability of elimination in most scenarios.** The probability of elimination for two different
280 vaccination priority strategies as new vaccines are developed: Continuous – give the newly updated
281 vaccination to the next person on the priority list to ensure everyone receives a vaccination; Restart
282 – give the newly developed vaccine to the first person the priority list, i.e. some people get many
283 vaccines and some get none at all.

284 *Vaccine uptake rate and effectiveness*

285 Previous results assume that all individuals who are offered a vaccine will take it whereas Table S1
286 shows that in most countries uptake rates fall between 40% and 70%. As these rates are for the adult
287 population and paediatric vaccines are often developed later than adult ones the actual uptake rate

288 across the entire population may be lower still, especially in countries with a relatively young
289 population. For example, in Israel almost 30% of the population are under 16 and ineligible for current
290 vaccines (United Nations, 2019). In the low transmission scenario the maximum transmission rate of
291 a variant is capped at $R_j = 1.3$ so using a very simple estimate of the herd immunity threshold
292 (Murray, 2007) $H = 1 - 1/R_0$ implies that provided 24% of the population was vaccinated there
293 would be no epidemic. The corresponding threshold for the high transmission scenario
294 (maximum $R_j = 1.8$) is 44%. Figure 5 shows the effect of uptake rate on the probability of
295 elimination. If the uptake rate is significantly higher than the simple herd immunity threshold it has
296 very little difference on the probability of elimination, however once the uptake becomes closer to
297 the threshold the very high probability of elimination previously seen when the vaccine delivery is
298 fast drops significantly.



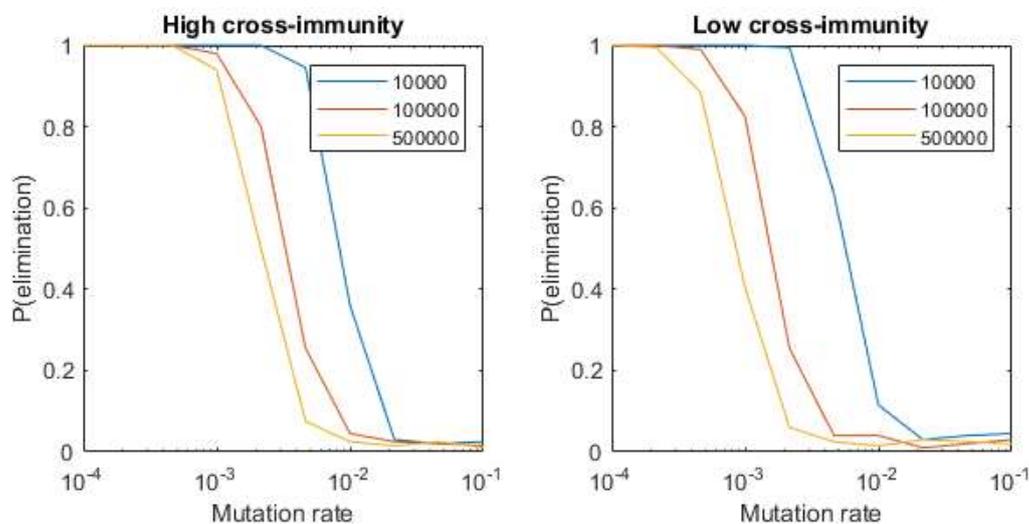
299
300 **Figure 5: For higher vaccination rates if the uptake rate falls below the herd immunity threshold**
301 **the probability of elimination drops. At lower vaccination rates, where the probability of**
302 **elimination is always low it makes little difference.** The probability of elimination at different daily
303 vaccination rates for different vaccine uptake rates (20%, 40%, 60% and 80% of those who are
304 offered the vaccine) in the four different scenarios.

305

306 *Population size*

307 The results so far are shown for a relatively small population of only 10,000 individuals. Whilst this is
308 computational cheaper it is less realistic. For the same reasons the results are constrained to relatively
309 low maximum transmission rates. Figure 6 shows the output for a more realistic set of parameter
310 values with population sizes up to 500,000, a higher maximum transmission rate $R_{max} = 3$, vaccine
311 uptake rate of 60%, delivery rate of 1% of the population each day and vaccines are updated every

312 150 days. The threshold mutation rate at which elimination becomes considerably less likely
 313 decreases with population size but otherwise the qualitative results are very similar.



314
 315 **Figure 6: As the population size increases the model shows similar results with a slower mutation**
 316 **rate.** The probability of elimination for different mutation rates for three different population sizes
 317 and two levels of cross-immunity.

318
 319 *Predicting elimination with hard to estimate parameters*

320 A key predictor of elimination is the rate at which significant new variants appear, p_{mutate} . In practice
 321 predicting this rate and the level of cross immunity gained by infection or vaccination to particular
 322 strains is somewhere between tricky and impossible. Instead, using Monte-Carlo simulations across
 323 a wide range of parameter values (population size $N_p \sim U(10^5, 10^6)$, mutation rate
 324 $p_{mutate} \sim U(10^{-4}, 10^{-2})$ and cross immunity $I(1) \sim U(0.1, 0.5)$) we predict the probability of
 325 elimination, $P(elim)$, conditional on the number of significant variants that have been observed after
 326 300 days. We define a significant variant as one that has infected at least 100 individuals, i.e. it would
 327 likely have been spotted during standard levels of genetic sequencing. We use a general linear model
 328 of the form

$$\text{logit}(P(\text{elim})) \sim V_{300} + \log(N_p) + \log(p_{mutate}) + \log(X_{imm})$$

329 where V_{300} is the number of significant variants seen in that realisation during the first 300 days,
 330 p_{mutate} is the mutation rate, N_p is the population size and X_{imm} is the level of cross-immunity defined
 331 as a linear scaling on all the elements of $I(d)$ in Table 2. Using all available predictor variables results
 332

333 in a highly accurate prediction of elimination probability (area under the receiver operator curve
 334 $AUC = 0.97$, Table 3). However, more encouragingly, using a simpler model with only the single
 335 predictor V_{300} gives a very good prediction of elimination probability ($AUC = 0.94$, Table 3).

V_{300}	$I(1)$	$\log(N_p)$	$\log(p_{mutate})$	AUC
-1.06 (10^{-165})				0.91
	2.15(10^{-21})			0.57
		-0.56 (10^{-28})		0.60
			-2.15 (10^{-255})	0.94
-0.92 (10^{-68})	8.86 (7.98e-41)	-1.11 (10^{-27})	-1.95 (10^{-115})	0.97

336 Table 3: **The observable variable V_{300} , the number of variants seen after 300 days, is a very good**
 337 **single predictor of elimination probability.** Coefficients (and p-values) of regression models
 338 predicting the number of significant variants and the probability of elimination.

339 The single variable model predicts that with no additional significant variants after 300 days, i.e.
 340 $V_{300} = 1$, there is an 86% probability of elimination. With 1 additional significant variant this falls to
 341 68%. Using the more detailed predictor model shows that for any given number of additional
 342 significant variants the probability of elimination will decrease for bigger populations, lower cross-
 343 immunity and faster mutation rates.

344 Here we used the number of variants that arise in the first 300 days after the virus emerged. Very
 345 similar results can be seen for other time periods. When more than two or three significant new
 346 strains of the virus have emerged the probability of elimination drops significantly.

347 *Sensitivity analysis*

348 We used a numerical definition of elimination where any realisation that reached 200 variants was
 349 assumed to continue indefinitely. This was checked extensively throughout the study and was seen
 350 to be correct in over 97% of realisations even for higher population sizes.

351 We have assumed a 100% vaccine effectiveness rate for the strain for which the vaccine was
 352 developed. Although we have not changed this explicitly the results of doing so would look very
 353 similar to the results for a lower uptake rate. For example the results for a 50% effective vaccine
 354 delivered at a rate of 1% of the population each day would be very similar to those for a 100% effective
 355 vaccine with an uptake rate of 50% delivered at a speed of 0.5% of the population each day.

356 We have assumed that the first vaccine to be developed takes approximately one year of
357 development time. Explorations show that improving this initial development speed would increase
358 the probability of elimination overall but the qualitative differences in results for changing other
359 parameters would not change significantly.

360 Changing the disease transmission parameters, e.g. the number of infectious days does not change
361 the overall results but does give a corresponding change in the timescale. Significant changes come
362 with changing the transmission rates but these are explored through changes to R_{max} , the maximum
363 transmission rate.

364 Although cross immunity was varied in the results we assumed a linear scaling of cross-immunity
365 across all distances in the virus tree. If cross-immunity remains high for closely related variants but
366 decreases more rapidly with the network distance between variants this has a similar effect as lower
367 overall cross-immunity.

368 Overall, the key results, in particular the usefulness of using the number of new variants after 300
369 days as a predictor of elimination, do not change if other model parameters are varied.

370 Discussion

371 Our model shows that viruses which mutate into multiple new variants need fast vaccine delivery to
372 be contained. Frequent vaccine updates to target new strains also increase the chance of containing
373 the virus. Keeping transmission rates low through non-pharmaceutical interventions such as mask-
374 wearing, social distancing and contact tracing will also improve the probability the disease is
375 eradicated. Our model suggests that a continuous vaccination roll-out programme, where updated
376 vaccines are given to unvaccinated individuals, rather than revaccinating high priority individuals,
377 may slightly increase the probability of elimination. However, this prediction warrants further
378 investigation using population-structured models to assess the risk this would pose to vulnerable
379 individuals (e.g. frontline workers or those at higher risk of severe disease).

380 Under high vaccination rates, our results highlight the importance of ensuring that vaccine uptake
381 rates are kept above the herd immunity threshold to avoid a considerable reduction in elimination
382 probability. However, at lower vaccination rates this factor has less impact on the probability of
383 elimination, which is low regardless.

384 Measuring the parameters needed to quantify models of multiple strains and predict the likelihood
385 of long term elimination or even containment is somewhere between tricky and impossible. Instead
386 we have presented a more observable quantity, the number of new variants that have caused a
387 significant number of infections after a defined period of time, which is a good single predictor of the
388 probability of elimination. As of May 2021 there had only been 5 countries that had seen a significant
389 new variant emerge despite many countries having high infection rates. There has only been one
390 nation that could claim to have had more than one new variant arise, i.e. India with very closely
391 related strains of B.1.617. However, with each new strain that emerges the likelihood of long term
392 containment decreases, especially as strains with very high transmission rates emerge.

393 However, when more than two or three new variants have spread to a significant number of people
394 the probability of eliminating the disease drops significantly. At this point the long term outcome is
395 similar to the current global influenza situation where the dominant viral strains change regularly,
396 vaccines need to be constantly updated to stop the more severe outcomes and elimination is no
397 longer an option.

398

399

400 **Acknowledgements**

401 This work was funded by the New Zealand Ministry of Business, Innovation and Employment and Te
402 Pūnaha Matatini, Centre of Research Excellence in Complex Systems.

403 References

- 404 Atlani-Duault, L., Lina, B., Chauvin, F., Delfraissy, J.-F., & Malvy, D. (2021). Immune evasion means we need a
405 new COVID-19 social contract. *The Lancet Public Health*, *6*(4), e199-e200.
- 406 Bedford, T., Rambaut, A., & Pascual, M. (2012). Canalization of the evolutionary trajectory of the human
407 influenza virus. *BMC biology*, *10*(1), 1-12.
- 408 Callaway, E. (2020). The coronavirus is mutating-does it matter? *Nature*, *585*(7824), 174-177.
- 409 Callaway, E. (2021). Fast-spreading COVID variant can elude immune responses. *Nature*, *589*(7843), 500-501.
- 410 Centers for Disease Control and Prevention (CDC). (2021). SARS-CoV-2 Variant Classifications and Definitions.
411 Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#Concern>
- 412 Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., . . . Washburne, A. D. (2021).
413 Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science*, *372*(6538).
- 414 Hamed, S. M., Elkhatib, W. F., Khairalla, A. S., & Noreddin, A. M. (2021). Global dynamics of SARS-CoV-2 clades
415 and their relation to COVID-19 epidemiology. *Scientific reports*, *11*(1), 1-8.
- 416 James, A., Plank, M. J., Binny, R. N., Hannah, K., Hendy, S. C., Lustig, A., & Steyn, N. (2020). A structured model
417 for COVID-19 spread: modelling age and healthcare inequities. *medRxiv*.
- 418 James, A., Plank, M. J., Binny, R. N., Lustig, A., Hannah, K., Hendy, S., & Steyn, N. (2020). A structured model
419 for COVID-19 spread: modelling age and healthcare inequities. *Mathematical Medicine and Biology*.
- 420 Koelle, K., Kamradt, M., & Pascual, M. (2009). Understanding the dynamics of rapidly evolving pathogens
421 through modeling the tempo of antigenic change: influenza as a case study. *Epidemics*, *1*(2), 129-137.
- 422 Koelle, K., Khatri, P., Kamradt, M., & Kepler, T. B. (2010). A two-tiered model for simulating the ecological and
423 evolutionary dynamics of rapidly evolving viruses, with an application to influenza. *Journal of The Royal*
424 *Society Interface*, *7*(50), 1257-1274.
- 425 Kucharski, A. J., Andreasen, V., & Gog, J. R. (2016). Capturing the dynamics of pathogens with many strains.
426 *Journal of mathematical biology*, *72*(1), 1-24.
- 427 Murray, J. D. (2007). *Mathematical biology: I. An introduction* (Vol. 17): Springer Science & Business Media.
- 428 Nie, Q., Li, X., Chen, W., Liu, D., Chen, Y., Li, H., . . . Zai, J. (2020). Phylogenetic and phylodynamic analyses of
429 SARS-CoV-2. *Virus research*, *287*, 198098.
- 430 Planas, D., Bruel, T., Grzelak, L., Guivel-Benhassine, F., Staropoli, I., Porrot, F., . . . Bishop, E. (2021). Sensitivity
431 of infectious SARS-CoV-2 B. 1.1. 7 and B. 1.351 variants to neutralizing antibodies. *Nature medicine*,
432 *27*(5), 917-924.
- 433 Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., . . . Msomi, N. (2020).
434 Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-
435 CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*.
- 436 United Nations, Department of Economic and Social Affairs, Population Division. (2019). World Population
437 Prospects 2019.
- 438 World Health Organization. (2021). COVID-19 weekly epidemiological update, 18 May 2021.

439