

1
2
3
4
5
6
7

**Wrong person, place and time: viral load and contact network structure predict
SARS-CoV-2 transmission and super-spreading events**

Ashish Goyal¹, Daniel B. Reeves¹, E. Fabian Cardozo-Ojeda¹, Joshua T. Schiffer^{1,2,3*†}, Bryan
T. Mayer^{1†}

¹ Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center

² Department of Medicine, University of Washington, Seattle

³ Clinical Research Division, Fred Hutchinson Cancer Research Center

† These authors contributed equally to the work.

Corresponding author: Joshua T. Schiffer, jschiffe@fredhutch.org

One Sentence Summary: We developed a coupled within-host and between-host mathematical model to identify viral shedding levels required for transmission of SARS-CoV-2 and influenza, and to explain why super-spreading events occur more commonly during SARS-CoV-2 infection.

8
9

10 **Abstract**

11 SARS-CoV-2 is difficult to contain because most transmissions occur during the pre-
12 symptomatic phase of infection. Moreover, in contrast to influenza, while most SARS-CoV-2
13 infected people do not transmit the virus to anybody, a small percentage secondarily infect large
14 numbers of people. We designed mathematical models of SARS-CoV-2 and influenza which link
15 observed viral shedding patterns with key epidemiologic features of each virus, including
16 distributions of the number of secondary cases attributed to each infected person (individual R_0)
17 and the duration between symptom onset in the transmitter and secondarily infected person
18 (serial interval). We identify that people with SARS-CoV-2 or influenza infections are usually
19 contagious for fewer than two days congruent with peak viral load several days after infection,
20 and that transmission is unlikely below a certain viral load. SARS-CoV-2 super-spreader events
21 with over 10 secondary infections occur when an infected person is briefly shedding at a very
22 high viral load and has a high concurrent number of exposed contacts. The higher predisposition
23 of SARS-CoV-2 towards super-spreading events is not due to its 1-2 additional weeks of viral
24 shedding relative to influenza. Rather, a person infected with SARS-CoV-2 exposes more people
25 within equivalent physical contact networks than a person infected with influenza, likely due to
26 aerosolization of virus. Our results support policies that limit crowd size in indoor spaces and
27 provide viral load benchmarks for infection control and therapeutic interventions intended to
28 prevent secondary transmission.

29 Introduction

30

31 The SARS-CoV-2 pandemic is an ongoing tragedy that has caused 700,000 deaths and
32 massively disrupted the global economy. The pandemic is rapidly expanding in the United States
33 and is re-emerging focally in many countries that had previous success in limiting its spread.¹

34 Two features have proven challenging in containing outbreaks. First, most transmissions
35 occur during the pre-symptomatic phase of infection.² Underlying this observation is a highly
36 variable incubation period, defined as time between infection and symptom onset, which often
37 extends beyond an infected person's peak viral shedding.³

38 Second, there is substantial over-dispersion of the basic reproduction number (R_0) for an
39 individual infected with SARS-CoV-2,⁴ meaning that most infected people do not transmit at all,
40 while a minority may transmit to dozens of people, with the average, population R_0 achieving a
41 high enough level (>1) to allow exponential growth of cases in the absence of an effective
42 intervention.⁵ Approximately 10-20% of infected people account for 80% of SARS-CoV-2
43 transmissions.^{4,6} Super-spreader events, in which the duration of contact between a single
44 transmitter and large number of secondarily infected people is often limited to hours, are well
45 documented.^{7,8} This pattern is not evident for influenza which has more homogeneous individual
46 transmissions numbers.^{9,10} Differing shedding kinetics between the two viruses might explain
47 this distinction; SARS-CoV-2 is often present intermittently in the upper airways for many
48 weeks,^{11,12} while influenza is rarely shed for more than a week.¹³ Alternatively, SARS-CoV-2
49 aerosolization may predispose to wider exposure networks given the presence of an infected
50 person in a crowded indoor space.

51 Viral load is recognized as a strong determinant of transmission risk. For influenza, the
52 dose of viral exposure is related to the probability of infection in human challenge studies,¹⁴ and
53 early treatment reduces household transmission.^{15,16} Household shedding of human herpesvirus-6
54 is closely linked to subsequent infection in newborns,¹⁷ and infants shedding high levels of
55 cytomegalovirus in the oropharynx predictably transmit the virus back to their mothers.¹⁸

56 The epidemiology of viral infections can also be perturbed by biomedical interventions
57 that lower viral load at mucosal transmission surfaces. Reduction of genital herpes simplex virus-
58 2 shedding with antiviral treatments decreases probability of transmission.¹⁹ Suppressive
59 antiretroviral therapy (ART) for HIV virtually eliminates the possibility of partner-to-partner
60 sexual transmission and has limited community transmission dramatically.^{20,21}

61 These concepts are relevant for SARS-CoV-2 infection and require urgent attention as the
62 pandemic continues to wreak havoc. Early therapies that lower peak viral load may reduce the
63 severity of COVID-19 but may also decrease the probability of transmission and of super-
64 spreader events.²² Similarly, the effectiveness of policies such as limiting mass gatherings, and
65 enforcing mask use can be directly evaluated by their ability to reduce exposure viral load and
66 transmission risk.²³ Here we developed a transmission simulation framework to capture the
67 contribution of viral load to observed epidemiologic transmission metrics for influenza and
68 SARS-CoV-2 and used this approach to explain why SARS-CoV-2 is predisposed to super-
69 spreading events.

70 **Results**

71

72 **Overall approach.** We designed a series of steps to estimate the viral load required for SARS-
73 CoV-2 and influenza transmission, as well as conditions required to explain the observed over-
74 dispersion of secondary infections (*individual R0*) and frequent super-spreader events associated
75 with SARS-CoV-2 but not influenza. This process included within-host modeling of viral loads,
76 simulations of exposures and possible transmissions based on various transmission dose response
77 curves, testing of various parameter sets against epidemiologic data and exploratory analyses
78 with the best fitting model (**Fig S1**).

79

80 **Within-host mathematical model of SARS CoV-2 shedding.** First, we used our previously
81 developed within-host mathematical model (equations in the **Methods**),²⁴ to generate plausible
82 viral load patterns in the upper airway of an infected person or *transmitter* who could potentially
83 transmit the virus to others (**Fig 1, Fig S2a**). Briefly, the model captures observed upper airway
84 viral kinetics from 25 people from four different countries.²⁵⁻²⁸ Key observed features include an
85 early viral peak followed by a decelerating viral clearance phase, which in turn leads to a
86 temporary plateau at a lower viral load, ultimately followed by rapid viral elimination. Our
87 model captures these patterns by including a density dependent term for early infected cell
88 elimination and a nonspecific acquired immune term for late infected cell elimination.

89 One limitation of our model is that only half of study participants provided longitudinal
90 viral load data from the very early days of infection when COVID-19 is often asymptomatic.
91 Therefore, the model's output is most reliable for later time points. In particular, we have
92 somewhat limited information on viral expansion rate and duration of peak shedding. To impute

93 possible variability, we generated a set of heterogeneous shedding curves in which the viral
94 upslope, the downslope of viral load after peak and the viral load during plateau phase were
95 varied (**Fig S2b**). Overall, the model generated several distinct patterns of infection: rapid
96 elimination after the initial peak, a prolonged plateau phase with a low viral load, and a
97 prolonged plateau phase with higher viral load. We simulated the transmission model with and
98 without imputed heterogeneity.

99

100 ***Transmission dose response curves.*** We defined an *exposure event* in very specific biologic
101 terms as a discrete event consisting of sufficient contact in time and space between a transmitter
102 and one or more uninfected persons (*exposure contacts*) to allow for the possibility of a
103 successful transmission. We next designed hundreds of dose response curves which separately
104 predict contagiousness (CD curves) and infectiousness (ID curves) at a certain viral dose given
105 an exposure contact. *Contagiousness* is defined as the viral load dependent probability of passage
106 of virus-laden droplets or airborne particles from the airways of a potential transmitter to the
107 airway of an exposure contact. *Infectiousness* is defined as the viral load dependent probability
108 of transmission given direct airway exposure to virus in an exposure contact. *Transmission risk*
109 is the product of these two mechanistic probabilities derived from the ID and CD curves and
110 results is a transmission dose (TD) response curve. Each CD or ID curve is defined by its ID50
111 (λ) or viral load at which contagion or infection probability is 50% (**Fig S2c**), as well as its slope
112 (α) (**Fig S2d**).²⁹ The TD50 is defined as viral load at which there is 50% transmission
113 probability. We assumed equivalent curves for contagiousness and infectiousness for model
114 fitting purposes. We also considered a simpler model with only a single TD curve (for
115 *infectiousness*) and obtained qualitatively similar results (**Supplement and Methods**). Our

116 model includes the possibility that increasing viral load is not a key determinant of transmission
117 when $\alpha=0.01$ (**Fig 2d**).

118

119 ***Exposure contact rate simulations.*** We introduced heterogeneity of exposure contact rates
120 among possible transmitters by randomly selecting from a gamma distribution defined by mean
121 number of exposure contacts per day (θ) and a scaling factor (ρ) that controls daily variability
122 (**Fig S3**).

123

124 ***Transmission simulations.*** For each defined exposure contact, viral load in the transmitter was
125 sampled and transmission risk was then identified based on the product of the CD and ID curves,
126 or the TD curve (**Fig S2e, f; Fig 1**). Based on these probabilities, we stochastically modeled
127 whether a transmission occurred for each exposure contact. This process was repeated when
128 there were multiple possible exposure events within a given discretized time interval and the
129 total number of exposures and transmissions within that interval was calculated.

130 For each successful transmission, we assumed that it takes τ days for the first infected
131 cell to produce virus. To inform simulated values of *serial interval* (SI or time between symptom
132 onset in the secondarily infected and transmitter), we randomly selected the *incubation period*
133 (IP), for both the transmitter and the newly infected person, from a gamma distribution based on
134 existing data (**Fig S4a**).^{3,30} Incubation period was defined as time from infection to the time of
135 the onset of symptoms, where the mean incubation for SARS-CoV-2 is 5.2 days compared to 2
136 days for influenza.^{3,9,30}

137

138 **Model fitting.** In order to identify the parameter set that best recapitulated the observed data, we
139 then simulated several hundred thousands of parameter sets with ~250 possible TD curves
140 defined by ID50 and CD50 (λ) and slope (α), along with ~180 combinations of the mean
141 exposed contact rate per day (θ) and associated variance parameter (ρ), and values of $\tau \in$
142 [0.5, 1, 2, 3] days. We aimed to identify the parameter set that best recapitulated the following
143 features of the observed epidemiologic and individual-level data for SARS-CoV-2: mean R_0
144 across individuals ($R_0 \in [1.4, 2.5]$),^{3,4,6,31,32} mean serial interval across individuals (SI \in
145 [4.0, 4.5]),^{3,31,33} cumulative distribution functions of individual R_0 ,^{4,6,34-36} and cumulative
146 distribution functions of serial intervals derived from SARS-CoV-2 transmission pair studies that
147 were conducted early during the pandemic,³¹ prior to any confounding influence of social
148 distancing measures. Here, we define *individual R_0* as the total number of secondary
149 transmissions from the transmitter in a fully susceptible population (**Methods**). We further
150 checked the closeness of the solved ID curve with the observed relationship between viral RNA
151 and infectious virus levels from a longitudinal cohort of infected people.³⁷

152
153 **Influenza modeling.** Next, we performed equivalent analyses for influenza to explain the lower
154 frequency of observed super-spreader events with this infection. Influenza viral kinetics were
155 modelled using a previously data-validated model.³⁸ Incubation periods for influenza are lower
156 and less variable than for SARS-CoV-2 and were randomly selected for each simulation of the
157 model using a gamma distribution (**Fig S4b**).³⁹ We again fit the model to: mean R_0 across
158 individuals ($R_0 \in [1.1, 1.5]$),⁴⁰⁻⁴² mean serial interval (SI $\in [2.9, 4.3]$),⁹ cumulative distribution
159 functions of individual R_0 corresponding to the 2008-2009 influenza A H1N1 pandemic with

160 mean $R_0=1.26$ and dispersion parameter= 2.36 in the negative binomial distribution, and
161 cumulative distribution functions of serial intervals.^{9,10,40}

162

163 ***Model-predicted individual R_0 and serial intervals for SARS-CoV-2 infection.*** A single model
164 parameter set ($[\alpha, \lambda, \tau, \theta, \rho] = [0.8, 10^7, 0.5, 4, 40]$) most closely reproduces empirically
165 observed individual R_0 and serial interval histograms (**Fig 2a, c**) and cumulative distribution
166 functions (**Fig 2b, d**). Despite assuming that each infected person sheds at a high viral load for a
167 period of time (**Fig 1, Fig S2b**), the model captures the fact that $\sim 75\%$ of 10,000 simulated
168 transmitters do not infect any other people and that each increase in the number of possible
169 transmissions is associated with a decreasing probability (**Fig. 2a**).

170 SARS-CoV-2 viral load was recently measured with viral RNA levels and mapped to
171 concurrent level of infectious virus by dividing by approximately 25.³⁷ We divided observed
172 viral RNA levels at each exposure contact by 25, and noted that the modeled ID curve closely
173 recapitulates predicted quantitative viral culture level (**Fig S5**).

174 The model also generates super-spreader events with 10,000 simulated transmissions
175 (**Fig. 2b**). If super-spreaders are defined as those who produce at least 5 secondary infections, we
176 estimate that $\sim 10\%$ of all infected people and $\sim 35\%$ of all transmitters are super-spreaders. If
177 super-spreaders are defined as those who produce at least 10 secondary infections, we estimate
178 that $\sim 6\%$ of all infected people and $\sim 25\%$ of all transmitters are super-spreaders. If super-
179 spreaders are defined as those who produce at least 20 secondary infections, we estimate that
180 $\sim 2.5\%$ of all infected people and $\sim 10\%$ of all transmitters are super-spreaders. If super-spreaders
181 are defined as those producing ≥ 5 , ≥ 10 , or ≥ 20 secondary infections, the contribution to all
182 secondary infections is estimated at $\sim 85\%$, $\sim 70\%$, or $\sim 44\%$, respectively (**Table 1**).

183 The model also recapitulates the high variance of the serial interval observed within
184 SARS-CoV-2 transmission pairs, including negative values observed in the data (**Fig 2c, d**). We
185 next project *generation time*, defined as the period between when an individual becomes infected
186 and when they transmit the virus, for all transmission pairs and identify that the mean serial
187 interval (4.4 days) provides an accurate approximation of mean generation time. However, the
188 variance of generation time is considerably lower and by definition does not include negative
189 values. A majority of generation times fell between 4 and 7 days, compared to -5 to 12 days for
190 the serial interval (**Fig 2e**).

191
192 ***Viral load thresholds for SARS-CoV-2 transmission.*** The optimized ID curve has an ID50 of
193 10^7 viral RNA copies and a moderately steep slope (**Fig 3a**). The TD50 for SARS-CoV-2 is
194 slightly higher at $10^{7.5}$ viral RNA copies (**Fig 3a**). To assess the impact of these parameters on
195 transmission, we performed simulations with 10,000 transmitters and concluded that
196 transmission is very unlikely ($\sim 0.00005\%$) given an exposure to an infected person with an upper
197 airway viral load of $<10^4$ SARS-CoV-2 RNA copies, and unlikely ($\sim 0.002\%$) given an exposure
198 to an infected person with a viral load of $<10^5$ SARS-CoV-2 RNA copies. On the other hand,
199 transmission is much more likely (39%) given an exposure to an infected person who is shedding
200 $>10^7$ SARS-CoV-2 RNA copies, and 75% given an exposure to an infected person with a viral
201 load of $>10^8$ SARS-CoV-2 RNA copies. We obtain similar results (not shown) when we solve
202 our model using the assumption of homogeneous viral load trajectories as in **Fig S2a**.

203
204 ***Narrow duration of high infectivity during SARS-CoV-2 infection.*** We next plotted the
205 probability of infection given an exposure to a transmitter. Under multiple shedding scenarios,

206 the window of high probability transmission is limited to time points around peak viral load, and
207 some heterogeneity in regard to peak infectivity is noted between people (**Fig 3b-d**). In general,
208 infected persons are likely to be most infectious (i.e., above TD50) for a ~0.5-1.0-day period
209 between days 2 and 6 after infection. We therefore conclude that the observed wide variance in
210 serial interval (**Fig 2c**) results primarily from the possibility of highly discrepant incubation
211 periods between the transmitter and infected person, rather than wide variability in shedding
212 patterns across transmitters.

213

214 **Requirements for SARS CoV-2 super-spreader events.** The solved value for exposed contact
215 network heterogeneity (ρ) is 40 indicating high variability in day-to-day exposure contact rates
216 (**Fig S3d**) with a high average number of exposed contacts per day ($\theta=4$). We generated a heat
217 map from our TD curve to identify conditions required for super-spreader events which included
218 viral load exceeding 10^7 SARS CoV-2 RNA copies and a high number of daily exposure
219 contacts per day. We observe an inflection point between 10^6 and 10^7 SARS CoV-2 RNA copies
220 where large increases in the number of daily exposure contacts have a more limited impact on
221 increasing the number of transmissions from a single person (**Fig 4a**). The exposure contact
222 network occasionally results in days with ≥ 150 exposure contacts per day, which may allow an
223 extremely high number of secondary infections from a single person (**Fig 4a**).

224 We next plotted transmission events simulated on a daily basis over 30 days since
225 infection from 10,000 transmitters according to viral load at exposure and number of exposure
226 contacts on that day (**Fig 4b**). Secondary transmissions to only 1-3 people occurred almost
227 exclusively with daily numbers of exposure contacts below 10 with any exposure viral load
228 exceeding 10^6 RNA copies or with higher numbers of exposure contacts per day and viral loads

229 exceeding 10^5 RNA copies. Massive super-spreader events with over 50 infected people almost
230 always occurred at viral loads exceeding 10^7 RNA copies / day with high levels of concurrent
231 exposure contacts (**Fig 4b**).

232 We next identified that over 50% of secondary infections were associated with a
233 transmitter who has a high number of exposed contacts (11-100 per day) and a viral load
234 exceeding 10^6 RNA copies (**Fig 4c**), which is the mechanistic underpinning of why ~70% of all
235 secondary infections arose from transmitters who produced more than 10 secondary infections
236 (**Table 1**).

237
238 ***Model predicted individual R_0 and serial intervals for influenza infection.*** A single model
239 parameter set most closely reproduced empirically observed histograms and cumulative
240 distribution functions for individual R_0 and serial intervals for influenza: $(\alpha, \lambda, \tau, \theta, \rho) = (0.7,$
241 $10^{5.5}, 0-0.5, 4, 1)$. ID50 values for influenza are lower than SARS CoV-2, but a direct
242 comparison cannot be made because tissue culture infectious dose (TCID) has been more
243 commonly used for measurements of influenza viral load, whereas viral RNA is used for SARS-
244 CoV-2. Nevertheless, TCID is a closer measure of infectious virus and it is thus reasonable that
245 ID50 based on TCID for influenza would be ~30-fold lower than ID50 based on total viral RNA
246 (infectious and non-infectious virus) for SARS-CoV-2.³⁷

247 The other notable difference is a considerably lower ρ value for influenza (**Fig S3b**),
248 denoting much less heterogeneity in the number of exposure contacts per person while the
249 average daily exposure contact was the same for both viruses (4 per day). The model captures the
250 fact that 40% of influenza infected people do not transmit to anyone else and that each increase
251 in the number of individual transmissions is associated with a lower probability (**Fig. 5a**).

252 Relative to SARS-CoV-2, super-spreader events involving 5 or more people are predicted to be
253 5-fold less common overall and 10-fold less common among transmitters (~2% of all infected
254 people and ~3% of transmitters) (**Fig. 5b, Table 1**). Super-spreaders defined as those infecting
255 ≥ 5 individuals contribute to only ~10% to all transmissions (**Table 1**).

256 The model also recapitulates the lower variance of serial interval for influenza relative to
257 SARS-CoV-2 (**Fig 5c, d**). We next identified that the mean and variance of the serial interval
258 provide good approximations of the mean and variance for generation time. A majority of
259 generation times fell between 2 and 6 days (**Fig 5e**).

260
261 ***Viral load thresholds for influenza transmission.*** Based on the optimized TD curve for
262 influenza (**Fig 6a**), we next plotted the probability of infection given an exposure to an infected
263 person. The TD50 for influenza is $10^{6.1}$ TCID/mL. Under various shedding scenarios, the
264 window of high probability transmission is limited to time points around peak viral load (**Fig 6b-**
265 **d**). In general, infected persons are likely to be most infectious (i.e., above TD50) for a ~0.5-1.0
266 day period.. The observed narrow variance in serial interval (**Fig 5c**) results primarily from the
267 narrow range of incubation periods within the transmitter and secondarily infected person, as
268 well as the limited variability in shedding patterns across transmitters.

269
270 ***Determinants of influenza individual R_0 .*** We generated a heat map from our TD curve to
271 identify conditions governing influenza transmission to multiple people including viral load
272 exceeding 10^6 influenza TCID and a high number of exposure contacts per day. The contact
273 network never results in days with more than 15 exposure contacts per day, which severely limits

274 the possible number of transmissions from a single person relative to SARS-CoV-2 (**Fig 7a,**
275 **S3b**).

276 We plotted transmission events simulated on a daily basis over 30 days since infection
277 from 10,000 transmitters according to viral load at exposure and number of exposure contacts on
278 that day (**Fig 7b**). Secondary transmissions to fewer than 5 people accounted for 90% of
279 infections (**Table 1**) and occurred with fewer than 10 daily exposure contacts and exposure viral
280 loads exceeding 10^4 TCID. Small scale super-spreader events with 5-10 infected people almost
281 always occurred at viral loads exceeding 10^5 TCID with 5-10 concurrent exposure contacts (**Fig**
282 **7b**).

283 We next identified that over 50% of infections were associated with a transmitter who
284 had fewer than 10 exposure contacts per day and a viral load exceeding $10^{4.5}$ TCID (**Fig 7c**),
285 which is why no infected person ever transmitted to more than 10 other people (**Table 1**).

286
287 *Differing exposed contact distributions, rather than viral kinetics, explain SARS CoV-2 super-*
288 *spreader events.* We sought to explain why SARS-CoV-2 has a more over-dispersed distribution
289 of individual R_0 relative to influenza. To assess viral kinetics as a potential factor, we
290 comparatively plotted transmission risk per exposure contact as a function of time since infection
291 in 10,000 transmitters for each virus. The median per contact transmission risk is slightly higher
292 for influenza; however, 75% and 95% transmission risks are marginally higher for SARS-CoV-2
293 compared to influenza with slightly higher peak transmission risk, and a longer tail of low
294 transmission risk beyond 7 days (**Fig 8a**). The transmission risk was considerably higher for the
295 25% of simulated SARS-CoV-2 infections with the highest viral loads, suggesting that a
296 substantial subset of infected people may be more pre-disposed to super-spreading. When plotted

297 as time since onset of symptoms the variability in transmission potential is considerably larger
298 for persons with high SARS-CoV-2 viral load, owing to the variable incubation period of this
299 virus (**Fig 8b**).

300 The median duration of shedding over infectivity thresholds was short and nearly
301 equivalent for both viruses. For SARS-CoV-2 and influenza, median [range] time above ID10
302 was 2.7 [0, 7] and 2.4 [1.6, 3.7] days respectively; median time above ID25 was 1.7 [0, 3] and
303 1.5 [0, 2.2] days respectively; median time above ID50 was 0.8 [0, 1.3] and 0 [0, 1.3] days
304 respectively; median time above ID75 was 0 [0, 0.4] and 0 [0, 0] days respectively; median time
305 above ID90 was 0 [0, 0] and 0 [0, 0] days respectively. ID10, ID25 and ID50 values are more
306 variable across SARS-CoV-2 simulations due to a minority of trajectories with prolonged
307 moderate viral loads.

308 For SARS-CoV-2 and influenza, median [range] time above TD10 was 1.4 [0, 2.5] and
309 1.2 [0, 2.0] days respectively; median time above TD25 was 0.8 [0, 1.3] and 0.3 [0, 1.3] days
310 respectively; median time above TD50 was 0 [0, 0.5] and 0 [0, 0.4] days respectively; median
311 time above TD75 was 0 [0, 0] and 0 [0, 0] days respectively. TD10, TD25 and TD50 values are
312 more variable across SARS-CoV-2 simulations due to a minority of trajectories with prolonged
313 moderate viral loads (**Fig 8c**).

314 We next plotted the frequency of exposure contacts per day for both viruses and noted a
315 higher frequency of days with no exposed contacts (**Fig 8d**), but also a higher frequency of days
316 with more than 10 exposure contacts (**Fig 8e**) for SARS-CoV-2 relative to influenza, despite an
317 equivalent mean number of daily exposure contacts. To confirm that this distribution drives the
318 different observed distributions of individual R_0 values (**Fig 8f**), we simulated SARS-CoV-2
319 infection with an assumed $\rho=1$ and generated a distribution of individual R_0 similar to that of

320 influenza (**Fig S6a**). Similarly, we simulated influenza infection with an assumed $\rho=40$ and
321 generated a distribution of individual R_0 similar to that of SARS-CoV-2 (**Fig S6b**). Under all
322 scenarios, predicted distributions of serial interval (**Fig 8g, Fig S6**) and generation time (**Fig 8h,**
323 **Fig S6**) were unchanged by shifts in the exposed contact network.

324

325 *Projections of targeted physical distancing.* Physical distancing is a strategy to decrease R_0 . We
326 simulated a decrease in the contact rate uniformly across the population and noted a decrease in
327 population R_0 (**Fig S7a**) as well the percent of infected people who will transmit (**Fig 7b**) and
328 become super-spreaders (**Fig S7c-d**). An approximately 40% decrease in the average exposed
329 contact rate decreased R_0 below 1 (**Fig S6a**). We further investigated whether lowering contact
330 rate among larger groups only, in particular by banning exposure events with a high number of
331 exposure contacts, could control the epidemic. We identify that limiting exposure contacts to no
332 more than 5 per day is nearly equivalent to limiting exposure contacts altogether and that only a
333 small decrease in mean exposure contact rate can achieve $R_0 < 1$ if exposure events with less than
334 20 contacts are eliminated (**Fig S8**).

335

336 *Pre-symptomatic transmission and super-spreading risk.* Much of the highest transmission risk
337 for SARS-CoV-2 exists in the pre-symptomatic phase (**Fig8b**) which explains why 62% of
338 simulated transmissions occurred in the pre-symptomatic phase for SARS-CoV-2, compared to
339 10% for influenza. Similarly, 62% and 21% of SARS-CoV-2 and influenza super-spreader
340 events with secondary transmissions ≥ 5 and 39% of SARS-CoV-2 super-spreader events with
341 secondary transmissions $R_0 \geq 10$ fell in the pre-symptomatic period.

342

343 Discussion

344 Our results demonstrate that SARS-CoV-2 shedding kinetics are directly linked to the
345 virus' most fundamental epidemiologic properties. First, we identify a transmission dose
346 response curve which specifies that a nasal viral load below 10^5 RNA copies is unlikely to
347 commonly result in transmission. For SARS-CoV-2, this threshold is consistent with the overall
348 rarity of positive cultures at these levels.³⁷ We also predict a relatively steep TD curve such that
349 transmission becomes much more likely when shedding exceeds 10^8 viral RNA copies and there
350 is an exposure contact between an infected person and susceptible person. The amount of viral
351 RNA can be roughly converted to an estimate of viral quantity by culture which approximates
352 infectiousness. Our results therefore have relevance for dosing of SARS-CoV-2 in human
353 challenge experiments that are being considered for vaccine trials.

354 While the duration of shedding for SARS-CoV-2 is often three weeks or longer,^{11,12} we
355 predict that the duration of shedding above thresholds required for a moderate probability of
356 transmission per contact is much shorter, often less than half a day, and is comparable to that of
357 influenza. While transmission after the first week of infection is quite rare, our model is
358 consistent with the observation that transmissions commonly occur during the pre-symptomatic
359 phase of infection,² given the highly variable incubation period associated with SARS-CoV-2.

360 The observed high heterogeneity in serial interval is attributable almost entirely to the
361 variable nature of the incubation period, rather than transmission occurring extremely late after
362 infection. While our estimate for mean generation time is equivalent to that of mean serial
363 interval, it is notable that the range of SARS-CoV-2 serial intervals is much wider than the range
364 of generation times. This result is evident even though we built substantial heterogeneity into our
365 viral shedding curves beyond that observed in the somewhat limited existing shedding data.

366 The finding of limited duration of SARS-CoV-2 infectivity has practical implications.
367 First, considerable resources are being used in hospitals and skilled nursing facilities to isolate
368 patients with persistent SARS-CoV-2 shedding. We propose that a low nasal viral load,
369 particularly during late infection, need not justify full patient isolation procedures in the absence
370 of aerosolizing procedures. This observation could save substantial hospital resources and
371 valuable isolation beds during subsequent waves of infection. Similar considerations are relevant
372 for employees wishing to return to work. Our results also suggest that time since first positive
373 test may be predictive of lack of contagion, though more viral load kinetic studies will be needed
374 to confirm the existing observation that viral loads after a week of infection are usually low and
375 associated with negative viral cultures.³⁷ Finally, our conclusions are supportive of rapid, less
376 sensitive assays which are more likely to detect infection at periods of contagion.⁴³

377 Many of these conclusions, including specific viral load thresholds for transmission, a
378 steep dose response curve and a maximum 2-day duration of contagion within an infected
379 individual are equally relevant for influenza infection. One important difference is that
380 incubation periods for influenza are far less variable which means that at the individual level, the
381 serial interval is much more likely to be predictive of the generation time.

382 Another finding is that SARS-CoV-2 super-spreading events are dependent on a large
383 number of exposure contacts during the relatively narrow 1-2 days window during which a ~25%
384 subset of infected people is shedding at extremely high levels above the TD50. Because we
385 predict that super-spreader potential may be somewhat of a generalized property of infection,
386 rather than a characteristic of a tiny subset of infected people, this result also has practical
387 implications. A common experience during the pandemic has been early identification of a
388 cluster of infected people within a specific confined environment such as a senior living home,

389 crowded work environment, athletic team, or restaurant. Our results demonstrate that newly
390 diagnosed people within small clusters may be past the peak of their super-spreading potential.
391 At this stage, many more infections have often been established and drastic quarantine
392 procedures should be considered. Other undiagnosed, pre-symptomatic infected people may have
393 super-spreader potential while the known infected person is no longer contagious, highlighting
394 the importance of effective contact tracing.

395 At the prevention level, school opening and work opening strategies should focus on
396 severely limiting the possible number of exposure contacts per day. Where large numbers of
397 exposure contacts are unavoidable, mandatory masking policies, perhaps with N95 masks that
398 may more significantly lower exposure viral loads should be considered.²³

399 Influenza infection is much less predisposed to super-spreader events than SARS-CoV-2.
400 Yet, influenza shedding at levels above those required for a high probability of transmission
401 occurs with only slightly lower frequency. Therefore, the markedly different probability of
402 super-spreader events between the two viruses is unlikely to relate to different viral host kinetics,
403 despite the fact that the overall duration of SARS-CoV-2 shedding exceeds duration of influenza
404 shedding often by more than two weeks.

405 Rather, our analysis suggests that the exposure contact networks of SARS-CoV-2
406 transmitters are highly variable relative to those of influenza. One possible explanation
407 underlying this finding is that SARS-CoV-2 is more predisposed to airborne transmission than
408 influenza.⁴⁴ Here our precise definition of an exposure contact (sufficient contact between a
409 transmitter and an uninfected person to potentially allow transmission) is of high relevance. Our
410 result suggests that a SARS-CoV-2 infected person in a crowded, poorly ventilated room, may
411 generate more exposure contacts than an influenza infected person in the same room, likely

412 based on wider dispersal and / or longer airborne survival of the virus. Thus, our results suggest a
413 possible downstream quantitative effect of airborne transmission on SARS-CoV-2 epidemiology.
414 Another possibly important variable is that pre-symptomatic transmission, which is a common
415 feature of SARS-CoV-2 may predispose to multiple transmissions. This prediction reinforces
416 current public health recommendation to avoid crowded indoor spaces with poor air
417 recirculation.

418 On the other hand, a much higher proportion of SARS-CoV-2 infected people than
419 influenza infected people do not transmit at all. This result lacks a clear mechanistic explanation
420 but may imply that aerosolization occurs only in a subset of infected people. One theoretical
421 explanation is that high viral load shedding in the pre-symptomatic phase is defined by lack of
422 cough or sneeze leading to limited spatial diffusion of virus. Alternatively, it is also possible that
423 a proportion of infected people never shed virus at high enough viral loads to allow efficient
424 transmission. This possibility speaks to the need for more quantitative viral load data gathered
425 during the initial stages of infection.

426 Age cohort structure differs between the two infections, with a lower proportion of
427 observed pediatric infections for SARS-CoV-2. If adults have more high exposure events than
428 children, then this could also explain super-spreader events. We are less enthusiastic about this
429 hypothesis. First, SARS-CoV-2 super-spreader events have occurred in schools and camps and
430 would likely be more common in the absence of widespread global school closures in high
431 prevalence regions. Second, a sufficient proportion of influenza cases occur in adults to rule out
432 the presence of frequent large super-spreading events in this population.

433 Our analysis has important limitations. First, exposure contacts were assumed to be
434 homogeneous and we do not capture the volume of the exposing aerosol or droplet. For instance,

435 if a large-volume droplet contains ten times more viral particles than an aerosol droplet, then the
436 exposure could be dictated by this volume as well as the viral load of the potential transmitter. It
437 is possible that under rare circumstances with extremely high-volume exposures, even persons
438 with extremely low viral loads may transmit. Second, based on the quality of available data, we
439 fit our models for SARS-CoV-2 and influenza to viral RNA and viral culture respectively.
440 Existing data suggest that kinetics of viral RNA and culture are similar during both infections,
441 with culture having lower sensitivity to detect virus.³⁷ Third, our intra-host model of SARS-
442 CoV-2 was fit to heterogeneous data with different sampling techniques and PCR assays.²⁴
443 Moreover, R0 estimates have varied across the globe. Our estimates of TD50 are necessarily
444 imprecise based on available data and should serve only as a conservative benchmark. Most
445 importantly, we cannot rule out the possibility that a small minority of infected people shed at
446 sufficient levels for transmission for much longer than has been observed to date. Finally,
447 contagiousness could have different dose response dynamics than viral load dependent
448 infectiousness and may require investigation in the future upon the availability of
449 epidemiologically relevant additional data.

450 In conclusion, fundamental epidemiologic features of SARS-CoV-2 and influenza
451 infections can be directly related to viral shedding patterns in the upper airway as well as the
452 nature of exposure contact networks. We contend that this information should be leveraged for
453 more nuanced public health practice in the next phase of the pandemic.

454 **Methods**

455

456 **SARS-CoV-2 within-host model.** To simulate SARS-CoV-2 shedding dynamics, we employed
457 our previously-described viral infection model.²⁴ In this model, susceptible cells (S) after coming
458 into contact with SARS-CoV-2 (V) become infected at rate βVS . The infected cells (I) produce
459 new virus at a per-capita rate π . The model also includes the clearance of infected cells in two
460 ways: (1) by an innate response with density dependent rate δI^k ; and (2) an acquired response
461 with rate $\frac{mE^r}{E^r + \phi^r}$ mediated by SARS-CoV-2-specific effector cells (E). The clearance mediated by
462 innate immunity depends on the infected cell density and is controlled by the exponent k . The
463 Hill coefficient r parameterizes the nonlinearity of the second response and allows for rapid
464 saturation of the killing. Parameter ϕ defines the effector cell level by which killing of infected
465 cells by E is half maximal.

466 In the model, SARS-CoV-2-specific effector cells rise after 2 stages from precursors cells
467 (M_1 and M_2). The first precursor cell compartment (M_1) proliferates in the presence of infection
468 with rate ωIM_1 and differentiates into the effector cell at a per capita rate q during the next
469 intermediate stage. Finally, effector cells die at rate δ_E . The model is expressed as a system of
470 ordinary differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta VS \\ \frac{dI}{dt} &= \beta VS - \delta I^k I - m \frac{E^r}{E^r + \phi^r} I \\ \frac{dV}{dt} &= \pi I - \gamma V \\ \frac{dM_1}{dt} &= \omega I M_1 - q M_1 \\ \frac{dM_2}{dt} &= q(M_1 - M_2) \\ \frac{dE}{dt} &= q M_2 - \delta_E E\end{aligned}$$

471

472 We assumed $S(0) = 10^7$ cells/mL, $I(0) = 1$ cells/mL, $V(0) = \frac{\pi I(0)}{c}$ copies/mL, $M_1(0) = 1$,

473 $M_2(0) = 0$ and $E_0 = 0$.

474 When we introduce simulated heterogeneity in cases of SARS-CoV-2 (by increasing
475 the standard deviation of the random effects of parameters β by 20, δ by 2, k by 2 and π by 5 in
476 the original distribution from²⁴), some of the viral shedding curves suggest that viral shedding
477 could continue for long period (over 6 weeks). Indeed, while median viral shedding duration has
478 been estimated at 12-20 days, shedding for many months is also observed commonly.⁴⁵ We
479 assumed that viral loads after day 20 drop to a exposure-level viral load level (i.e., $V(0)$) as most
480 viral shedding observed after this point is transient and at an extremely low viral load.⁴⁶ The
481 population distribution of parameters to simulate artificial SARS-CoV-2 viral shedding dynamics
482 is provided in **Table S1**.

483

484 ***Influenza within-host model.*** To simulate viral shedding dynamics of influenza viral, we employ
485 a model³⁸ that is a simplified version of the viral dynamics model presented for SARS-CoV-2.
486 This model assumes $k = 0$ and $m = 0$ and can be expressed as a system of ordinary differential
487 equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta VS \\ \frac{dI}{dt} &= \beta VS - \delta I \\ \frac{dV}{dt} &= \pi I - \gamma V\end{aligned}$$

488 Following this model,³⁸ we assumed $S(0) = 4 \times 10^8$ cells/mL, $I(0) = 1$ cells/mL, $V(0) = \frac{\pi I(0)}{c}$
489 copies/mL. To simulate artificial influenza viral shedding dynamics, we assumed the population
490 distribution of parameters $\text{Log}_{10}(\beta)$, $\text{Log}_{10}(\pi)$, $\text{Log}_{10}(\gamma)$ and $\text{Log}_{10}(\delta)$ are -4.56 (0.17), -
491 1.98 (0.14), 0.47 (0.03) and 0.60 (0.06), respectively.

492
493 **Dose-response model.** For both viruses, to estimate the infectiousness $P_t[V(t)]$ (response) based
494 on viral loads $V(t)$ (dose), we employed the function, $P_t[V(t)] = \frac{V(t)^\alpha}{\lambda^\alpha + V(t)^\alpha}$. Here, λ is the
495 infectivity parameter that represents the viral load that corresponds to 50% infectiousness and
496 50% contagiousness, and α is the Hill coefficient that controls the sharpness in the dose-response
497 curve.

498
499 **Transmission Model and Reproduction number.** Our transmission model assumes that only
500 some contacts of an infected individual with viral load dependent infectiousness are physically
501 exposed to the virus (defined as exposure contacts), that only some exposure contacts have virus
502 passaged to their airways (contagiousness) and that only some exposed contacts with virus in
503 their airways become secondarily infected (successful secondary infection). Contagiousness and
504 infectiousness are then treated as viral load dependent multiplicative probabilities with
505 transmission risk for a single exposure contact being the product. Contagiousness is considered
506 to be viral load dependent based on the concept that a transmitter's dispersal cloud of virus is

507 more likely to prove contagious at higher viral load, which is entirely separate for considerations
508 of viral infectivity within the airway once a virus contacts the surface of susceptible cells.

509 We next assume that the total exposed contacts within a time step (η_{Δ_t}) is gamma
510 distributed, i.e. $\eta_{\Delta_t} \sim \Gamma\left(\frac{\theta}{\rho}, \rho\right) \Delta_t$, using the average daily contact rates (θ) and the dispersion
511 parameter (ρ). To obtain the true number of exposure contacts with airway exposure to virus, we
512 simply multiply the contagiousness of the transmitter with the total exposed contacts within a
513 time step (i.e., $\zeta_t = \eta_{\Delta_t} P_t$).

514 Transmissions within a time step are simulated stochastically using time-dependent viral
515 load to determine infectiousness (P_t). Successful transmission is modelled stochastically by
516 drawing a random uniform variable ($U(0,1)$) and comparing it with infectiousness of the
517 transmitter. In the case of successful transmission, the number of secondary infections within
518 that time step (T_{Δ_t}) is obtained by the product of the infectiousness (P_t) and the number of
519 exposure contacts drawn from the gamma distribution (ζ_t). In other words, the number of
520 secondary infections for a time step is $T_{\Delta_t} = Ber(P_t) P_t \eta_{\Delta_t}$. If we disregard contagiousness by
521 assuming $P_t = 1$ in ζ_t , we identify that there are little to no differences on overall results other
522 than the emergent TD curve and optimal parameter set describing dose-response curve and
523 exposed contact network, which no longer agrees as closely with in vitro probability of positive
524 virus culture (**Fig S5**).³⁷

525 We obtain the number of secondary infections from a transmitter on a daily basis noting
526 that viral load, and subsequent risk, does not change substantially within a day. We then summed
527 up the number of secondary infections over 30 days since the time of exposure to obtain the
528 individual reproduction number, i.e. $R_0 = \sum_{\Delta_t} T_{\Delta_t}$.

529

530 **Serial interval and generation time.** We further assume that upon successful infection, it takes τ
531 days for the virus to move within-host, reach infection site and produce the first infected cell.
532 To calculate serial interval (time between the onset of symptoms of transmitter and secondarily
533 infected person), we sample the incubation period for both transmitter and secondarily infected
534 person from a gamma distribution with a shape described in the **Fig S4**.^{3,30} In cases in which
535 symptom onset in the newly infected person precedes symptom onset in the transmitter, the serial
536 interval is negative; otherwise, serial interval is non-negative. Similarly, we calculate generation
537 time as the difference between the time of infection of transmitter and the time of infection of
538 secondarily infected person.

539

540 **Fitting procedure.** To estimate the values of unknown parameters in cases of SARS-CoV-2, we
541 performed a grid search comprehensively exploring a total of ~500,000 combinations of 5
542 parameters taking the following values,

- 543 (i) $\tau \in [0.5, 1, 2, 3]$ days,
544 (ii) $\alpha \in [0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0]$
545 (iii) $\lambda \in [10^0, 10^{0.5}, 10^{1.0} \dots, 10^8]$
546 (iv) $\theta \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0, 20.0, 50.0]$.
547 (v) $\rho \in [0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 5.0, 10.0,$
548 $20.0, 30.0, 40.0, 50.0, 75.0, 100, 200, 500]$.

549 The parameter sets of $(\lambda, \tau, \alpha, \theta)$ were simulated for 1000 infected individuals to determine how
550 well each set generates the summary statistics of mean R0, mean SI and the R0 histograms by
551 following a procedure explained in **Fig S1** and below:

552 **Step A:**

- 553 1. Simulate viral load $V(t)$ of 1,000 simulated infected individuals using **Eq. 1**
- 554 2. For each combination of $(\lambda, \tau, \alpha, \theta, \rho)$
- 555 a. For each time step Δ_t
- 556 i. Compute $P_t[V(t); \lambda, \alpha]$
- 557 ii. Draw $\eta_{\Delta_t} \sim \Gamma\left(\frac{\theta}{\rho}, \rho\right) \Delta_t$
- 558 iii. Calculate $T_{\Delta_t} = Ber(P_t)P_t\eta_{\Delta_t}$
- 559 b. Calculate $R_0 = \sum_{\Delta_t} T_{\Delta_t}$
- 560 i. Check if calculated mean R_0 is in the range:^{3,31}
- 561 c. Calculate Serial Interval based on τ and incubation period
- 562 i. Check if calculated SI is in the range in:^{3,31,33}

563 **Step B:**

- 564 1. If the parameter combination in Step A satisfy the criteria, then
- 565 i. Compute RSS for the obtained R_0 and histogram from:^{4,6,34,36} [Ref]
- 566

567 We visually checked whether our dose-response curve matched the observed probability
568 of positive virus culture.³⁷ We assumed that viral loads derived from positive culture³⁷ can be
569 considered equivalent to viral loads in the within-host model if divided by a positive integer. We
570 found this positive integer to be 25 (**Fig S5**).

571 We performed a global sensitivity analysis to identify which parameter variability
572 accounted for fit to different components of the data. Only narrow ranges of λ permitted close fit
573 to the mean of R_0 and distribution functions of individual R_0 (**Fig S9**), while a specific value for
574 α was necessary to fit to mean serial interval and distribution functions of individual R_0 (**Fig**

575 **S9)**. Only narrow ranges of θ permitted close fit to the mean of R0 and distribution functions of
576 individual R0 (**Fig S10**), while a specific value for ρ was necessary to fit to distribution functions
577 of individual R0 (**Fig S10**).

578 To obtain TD50 (λ_T) based on ID50 (λ), we use the relation

$$\frac{1}{\left(\left(\frac{10^\lambda}{V}\right)^\alpha + 1\right)^2} = \frac{1}{\left(\frac{10^{\lambda_T}}{V}\right)^{\alpha_T} + 1} = 0.5$$

579 From solving the second half ($\frac{1}{\left(\frac{10^{\lambda_T}}{V}\right)^{\alpha_T} + 1} = 0.5$), we get

$$V = 10^{\lambda_T}$$

580 Substituting $V = 10^{\lambda_T}$ in the first-half, we have

$$\frac{1}{\left(\left(\frac{10^\lambda}{10^{\lambda_T}}\right)^\alpha + 1\right)^2} = 0.5$$

581 Or, $\left(\left(\frac{10^\lambda}{10^{\lambda_T}}\right)^\alpha + 1\right)^2 = 2$

582 Or, $\left(\frac{10^\lambda}{10^{\lambda_T}}\right)^\alpha = \sqrt{2} - 1$

583 Or, $10^{\lambda_T \alpha} = \frac{10^{\lambda \alpha}}{\sqrt{2} - 1}$

584 Or, $\lambda_T = \lambda + \frac{0.38}{\alpha}$

585 **References**

586

587 1 <https://coronavirus.jhu.edu/map.html>.

588 2 He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat*
589 *Med* **26**, 672-675, doi:10.1038/s41591-020-0869-5 (2020).

590 3 Ganyani, T. et al. Estimating the generation interval for coronavirus disease (COVID-19)
591 based on symptom onset data, March 2020. *Euro Surveill* **25**, doi:10.2807/1560-
592 7917.ES.2020.25.17.2000257 (2020).

593 4 Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19
594 Working Group, Abbott, S., Kucharski, A. & Funk, S. Estimating the overdispersion in
595 COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* **5**,
596 doi:10.12688/wellcomeopenres.15842.3 (2020).

597 5 Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the
598 effect of individual variation on disease emergence. *Nature* **438**, 355-359,
599 doi:10.1038/nature04153 (2005).

600 6 Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their
601 close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis*,
602 doi:10.1016/S1473-3099(20)30287-5 (2020).

603 7 L., H., P., D., I., C. & al., e. High SARS-CoV-2 Attack Rate Following Exposure at a
604 Choir Practice — Skagit County, Washington, March 2020. . *MMWR Morb Mortal Wkly*
605 *Rep* **69:606–610**. (2020).

606 8 Park, S. Y. et al. Coronavirus Disease Outbreak in Call Center, South Korea. *Emerg*
607 *Infect Dis* **26**, 1666-1670, doi:10.3201/eid2608.201274 (2020).

- 608 9 Cowling, B. J., Fang, V. J., Riley, S., Malik Peiris, J. S. & Leung, G. M. Estimation of
609 the serial interval of influenza. *Epidemiology* **20**, 344-347,
610 doi:10.1097/EDE.0b013e31819d1092 (2009).
- 611 10 Brugger, J. & Althaus, C. L. Transmission of and susceptibility to seasonal influenza in
612 Switzerland from 2003 to 2015. *Epidemics* **30**, 100373,
613 doi:10.1016/j.epidem.2019.100373 (2020).
- 614 11 Qi, L. et al. Factors associated with the duration of viral shedding in adults with COVID-
615 19 outside of Wuhan, China: a retrospective cohort study. *Int J Infect Dis* **96**, 531-537,
616 doi:10.1016/j.ijid.2020.05.045 (2020).
- 617 12 Cao, B. et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-
618 19. *N Engl J Med*, doi:10.1056/NEJMoa2001282 (2020).
- 619 13 Pawelek, K. A. et al. Modeling within-host dynamics of influenza virus infection
620 including immune responses. *PLoS Comput Biol* **8**, e1002588,
621 doi:10.1371/journal.pcbi.1002588 (2012).
- 622 14 Memoli, M. J. et al. Validation of the wild-type influenza A human challenge model
623 H1N1pdMIST: an A(H1N1)pdm09 dose-finding investigational new drug study. *Clin*
624 *Infect Dis* **60**, 693-702, doi:10.1093/cid/ciu924 (2015).
- 625 15 Pebody, R. G. et al. Use of antiviral drugs to reduce household transmission of pandemic
626 (H1N1) 2009, United Kingdom. *Emerg Infect Dis* **17**, 990-999,
627 doi:10.3201/eid1706.101161 (2011).
- 628 16 Goldstein, E. et al. Oseltamivir for treatment and prevention of pandemic influenza
629 A/H1N1 virus infection in households, Milwaukee, 2009. *BMC Infect Dis* **10**, 211,
630 doi:10.1186/1471-2334-10-211 (2010).

- 631 17 Mayer, B. T. et al. Estimating the Risk of Human Herpesvirus 6 and Cytomegalovirus
632 Transmission to Ugandan Infants from Viral Shedding in Saliva by Household Contacts.
633 *Viruses* **12**, doi:10.3390/v12020171 (2020).
- 634 18 Boucoiran, I. et al. Nonprimary Maternal Cytomegalovirus Infection After Viral
635 Shedding in Infants. *Pediatr Infect Dis J* **37**, 627-631,
636 doi:10.1097/INF.0000000000001877 (2018).
- 637 19 Corey, L. et al. Once-daily valacyclovir to reduce the risk of transmission of genital
638 herpes. *N Engl J Med* **350**, 11-20, doi:10.1056/NEJMoa035144 (2004).
- 639 20 Rodger, A. J. et al. Risk of HIV transmission through condomless sex in serodifferent gay
640 couples with the HIV-positive partner taking suppressive antiretroviral therapy
641 (PARTNER): final results of a multicentre, prospective, observational study. *Lancet* **393**,
642 2428-2438, doi:10.1016/S0140-6736(19)30418-0 (2019).
- 643 21 Cohen, M. S. et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N*
644 *Engl J Med* **375**, 830-839, doi:10.1056/NEJMoa1600693 (2016).
- 645 22 Schiffer, J. T., Johnston, C., Wald, A. & Corey, L. An Early Test-and-Treat Strategy for
646 Severe Acute Respiratory Syndrome Coronavirus 2. *Open Forum Infect Dis* **7**, ofaa232,
647 doi:10.1093/ofid/ofaa232 (2020).
- 648 23 Leung, N. H. L. et al. Respiratory virus shedding in exhaled breath and efficacy of face
649 masks. *Nat Med* **26**, 676-680, doi:10.1038/s41591-020-0843-2 (2020).
- 650 24 Goyal, A., Cardozo-Ojeda, E. & Schiffer, J. Potency and timing of antiviral therapy as
651 determinants of duration of SARS CoV-2 shedding and intensity of inflammatory
652 response. *medRxiv* **2020.04.10.20061325**, doi:10.1101/2020.04.10.20061325 (2020).

- 653 25 Wölfel, R. et al. Virological assessment of hospitalized patients with COVID-2019.
654 Nature **581**, 465-469, doi:10.1038/s41586-020-2196-x (2020).
- 655 26 Lescure, F. X. et al. Clinical and virological data of the first cases of COVID-19 in
656 Europe: a case series. Lancet Infect Dis **20**, 697-706, doi:10.1016/S1473-3099(20)30200-
657 0 (2020).
- 658 27 Young, B. E. et al. Epidemiologic Features and Clinical Course of Patients Infected With
659 SARS-CoV-2 in Singapore. JAMA, doi:10.1001/jama.2020.3204 (2020).
- 660 28 Kim, J. Y. et al. Viral Load Kinetics of SARS-CoV-2 Infection in First Two Patients in
661 Korea. J Korean Med Sci **35**, e86, doi:10.3346/jkms.2020.35.e86 (2020).
- 662 29 Brouwer, A. F., Weir, M. H., Eisenberg, M. C., Meza, R. & Eisenberg, J. N. S. Dose-
663 response relationships for environmentally mediated infectious disease transmission
664 models. PLoS Comput Biol **13**, e1005481, doi:10.1371/journal.pcbi.1005481 (2017).
- 665 30 Lauer, S. A. et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From
666 Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med **172**,
667 577-582, doi:10.7326/M20-0504 (2020).
- 668 31 Du, Z. et al. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases.
669 Emerg Infect Dis **26**, 1341-1343, doi:10.3201/eid2606.200357 (2020).
- 670 32 World Health Organization. Statement on the meeting of the International Health
671 Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus
672 (2019-nCoV). (2020).
- 673 33 Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus
674 (COVID-19) infections. Int J Infect Dis **93**, 284-286, doi:10.1016/j.ijid.2020.02.060
675 (2020).

- 676 34 Zhang, Y., Li, Y., Wang, L., Li, M. & Zhou, X. Evaluating Transmission Heterogeneity
677 and Super-Spreading Event of COVID-19 in a Metropolis of China. *Int J Environ Res*
678 *Public Health* **17**, doi:10.3390/ijerph17103705 (2020).
- 679 35 Dillon, A. et al. Clustering and superspreading potential of severe acute respiratory
680 syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. PREPRINT (Version
681 1) available at Research Square, doi:10.21203/rs.3.rs-29548/v1 (2020).
- 682 36 Miller, D. et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into
683 and within Israel. medRxiv, 2020.2005.2021.20104521,
684 doi:10.1101/2020.05.21.20104521 (2020).
- 685 37 van Kampen, J. J. A. et al. Shedding of infectious virus in hospitalized patients with
686 coronavirus disease-2019 (COVID-19): duration and key determinants. medRxiv,
687 2020.2006.2008.20125310, doi:10.1101/2020.06.08.20125310 (2020).
- 688 38 Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G. & Perelson, A. S. Kinetics of
689 influenza A virus infection in humans. *J Virol* **80**, 7590-7599, doi:10.1128/JVI.01623-05
690 (2006).
- 691 39 Lessler, J. et al. Outbreak of 2009 pandemic influenza A (H1N1) at a New York City
692 school. *N Engl J Med* **361**, 2628-2636, doi:10.1056/NEJMoa0906089 (2009).
- 693 40 Opatowski, L. et al. Transmission characteristics of the 2009 H1N1 influenza pandemic:
694 comparison of 8 Southern hemisphere countries. *PLoS Pathog* **7**, e1002225,
695 doi:10.1371/journal.ppat.1002225 (2011).
- 696 41 Cowling, B. J. et al. The effective reproduction number of pandemic influenza:
697 prospective estimation. *Epidemiology* **21**, 842-846, doi:10.1097/EDE.0b013e3181f20977
698 (2010).

- 699 42 Roberts, M. G. & Nishiura, H. Early estimation of the reproduction number in the
700 presence of imported cases: pandemic influenza H1N1-2009 in New Zealand. PLoS One
701 6, e17835, doi:10.1371/journal.pone.0017835 (2011).
- 702 43 Larremore, D. B. et al. Test sensitivity is secondary to frequency and turnaround time for
703 COVID-19 surveillance. medRxiv, 2020.2006.2022.20136309,
704 doi:10.1101/2020.06.22.20136309 (2020).
- 705 44 van Doremalen, N. et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared
706 with SARS-CoV-1. N Engl J Med, doi:10.1056/NEJMc2004973 (2020).
- 707 45 Widders, A., Broom, A. & Broom, J. SARS-CoV-2: The viral shedding vs infectivity
708 dilemma. Infect Dis Health, doi:10.1016/j.idh.2020.05.002 (2020).
- 709 46 Huang, C.-G. et al. Relative COVID-19 viral persistence and antibody kinetics. medRxiv,
710 2020.2007.2001.20143917, doi:10.1101/2020.07.01.20143917 (2020).
- 711

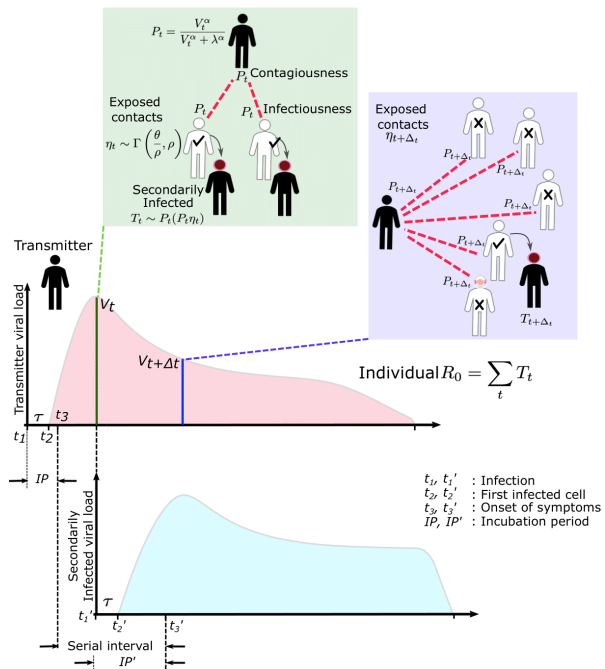


Fig 1. SARS-CoV-2 and influenza transmission model schematic. In the above cartoon, the transmitter has 2 exposure events at discrete timepoints resulting in 7 total exposure contacts and 3 secondary infections. Transmission is more likely at the first exposure event due to higher exposure viral load. To model this process, the timing of exposure events and number of exposed contacts is governed by a random draw from a gamma distribution which allows for heterogeneity in number of exposed contacts per day (**Fig S3**). Viral load is sampled at the precise time of each exposure event. Probability of transmission is identified based on the product of two dose curves (**Fig S2C, D**) which capture contagiousness (probability of viral passage to an exposure contact's airway) and infectiousness (probability of transmission given viral presence in the airway). Incubation period (**Fig S4**) of the transmitter and secondarily infected person is an input into each simulation and is depicted graphically. Individual R_0 is an output of each simulation and is defined as the number of secondary infections generated by an infected individual. Serial interval is an output of each simulated transmission and is depicted graphically.

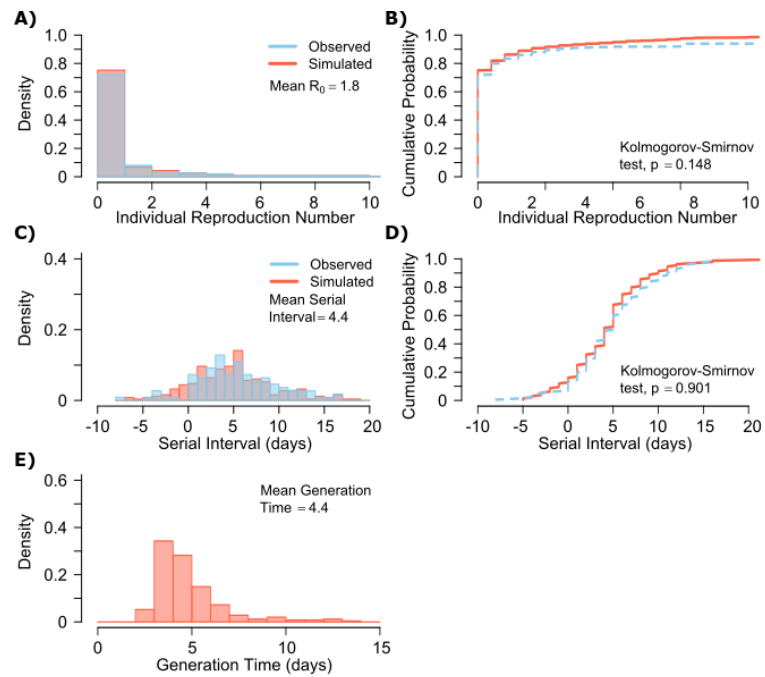


Fig 2. SARS-CoV-2 transmission model fit. **A.** Simulated and actual frequency histograms of individual R_0 values, **B.** Simulated and actual cumulative distribution of individual R_0 values. **C.** Simulated and actual frequency histograms of individual serial intervals, **D.** Simulated and actual cumulative distribution of individual serial intervals. **E.** Frequency distribution of simulated generation times.

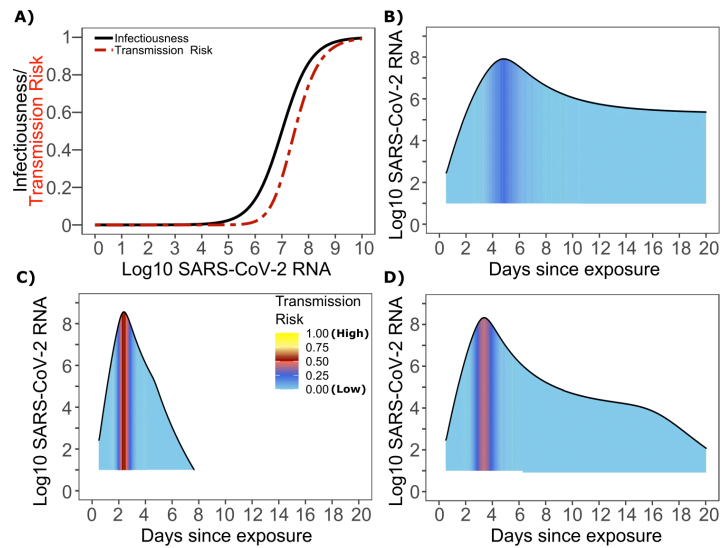


Fig 3. SARS-CoV-2 transmission probability as a function of shedding. **A.** Optimal infectious dose (ID) response curve (infection risk = P_i) and transmission dose (TD) response curve (transmission risk = $P_i * P_t$) curves for SARS-CoV-2. Transmission probability is a product of two probabilities, contagiousness and infectiousness (**Fig 1**). **B-D.** Three simulated viral shedding curves. Heat maps represent risk of transmission at each shedding timepoint given an exposed contact with an uninfected person at that time.

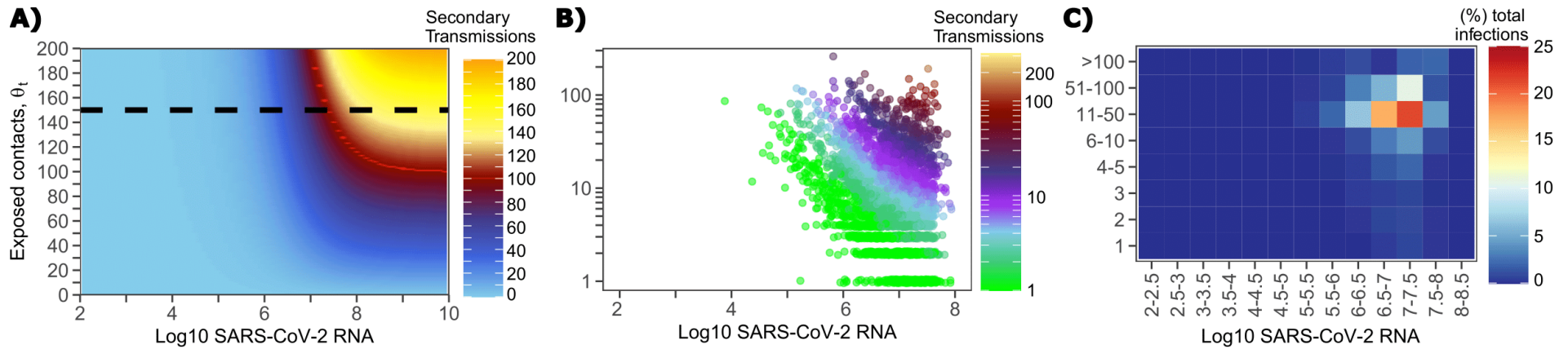


Fig 4. Conditional requirements for SARS-CoV-2 superspreading events. **A.** Heatmap demonstrating the maximum number of feasible secondary infections per day from a transmitter given an exposure viral load on log10 scale (x-axis) and number of exposed contacts per day (y-axis). The exposed contact network allows a maximum of 150 exposed contacts per day (black dotted line) which is sufficient for multiple transmissions from a single person per day. **B.** 10,000 simulated transmitters followed for 30 days. The white space is a parameter space with no transmissions. Each dot represents the number of secondary transmissions from a transmitter per day. Input variables are log10 SARS-CoV-2 on the start of that day and number of contact exposures per day for the transmitter. There are 1,154,001 total exposure contacts and 15,992 total infections. **C.** 10,000 simulated infections with percent of infections due to exposure viral load binned in intervals of 0.5 intervals on log10 scale (x-axis) and number of exposed contacts (y-axis).

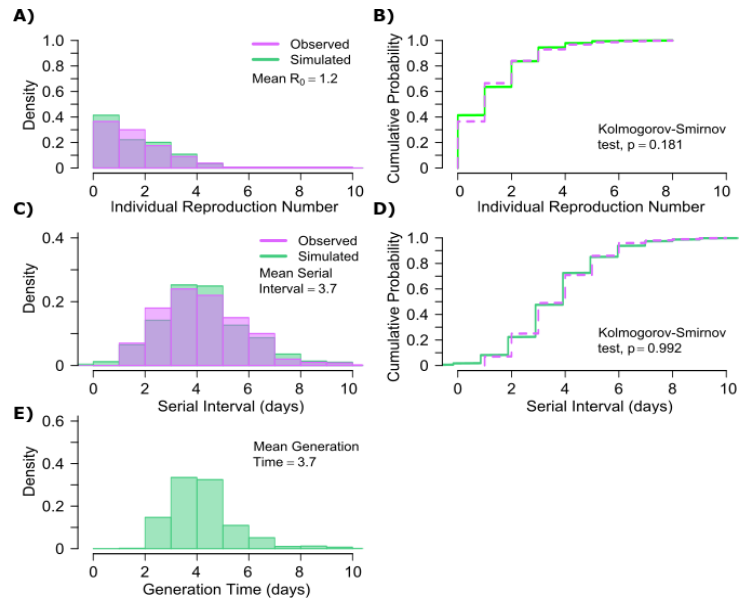


Fig 5. Influenza transmission model fit. **A.** Simulated and actual frequency histograms of individual R_0 values, **B.** Simulated and actual cumulative distribution of individual R_0 values. **C.** Simulated and actual frequency histograms of individual serial intervals, **D.** Simulated and actual cumulative distribution of individual serial intervals. **E.** Frequency distribution of simulated generation times.

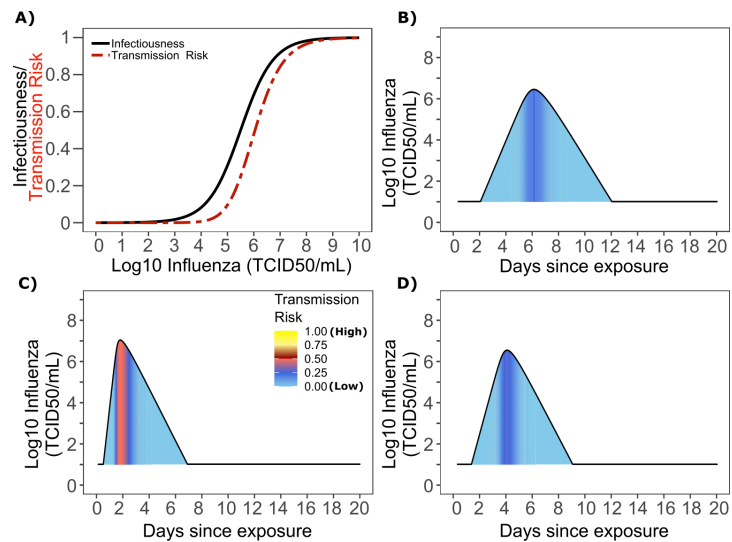


Fig 6. Influenza transmission probability as a function of shedding.
A. Optimal infectious dose (ID) response curve (infection risk = P_i) and transmission dose (TD) response curve (transmission risk = $P_i * P_t$) curves for influenza. Transmission probability is a product of two probabilities, contagiousness and infectiousness (**Fig 1**). **B-D.** Three simulated viral shedding curves. Heat maps represent risk of transmission at each shedding timepoint given an exposed contact with an uninfected person at that time.

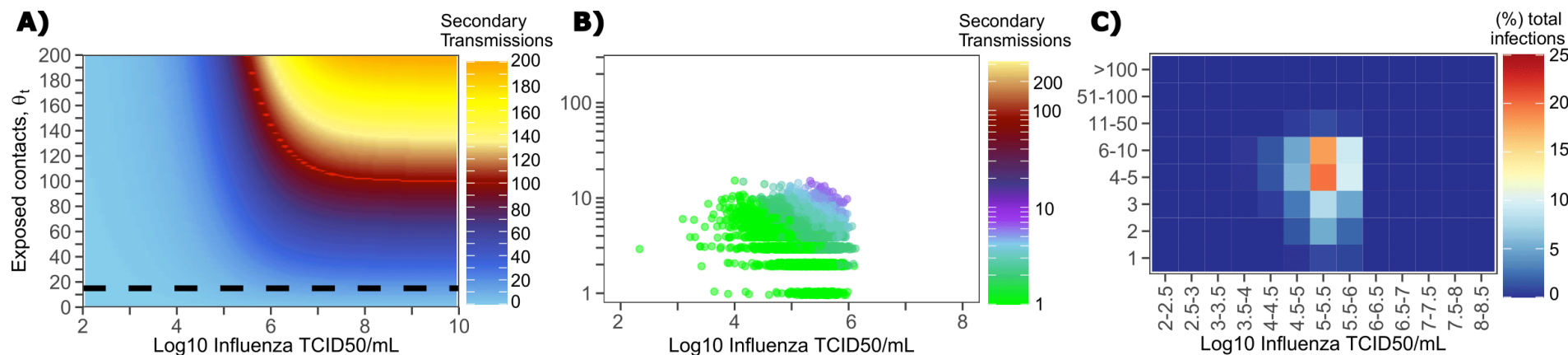


Fig 7. Conditional requirements for influenza super spreading events. **A.** Heatmap demonstrating the maximum number of secondary infections per day feasible from a transmitter given an exposure viral load on log10 scale (x-axis) and number of exposed contacts per day (y-axis). The exposed contact network allows a maximum of 15 exposed contacts per day (black dotted line) which is not sufficient for more than 15 transmissions from a single person per day. **B.** 10,000 simulated transmitters followed for 30 days. The white space is a parameter space with no transmissions. Each dot represents the number of secondary transmissions from a transmitter per day. Input variables are log10 influenza TCID on the start of that day and number of contact exposures per day for the transmitter. There are 1,239,984 total exposure contacts and 11,141 total infections. **C.** 10,000 simulated infections with percent of infections due to exposure viral load binned in intervals of 0.5 intervals on log10 scale (x-axis) and number of exposed contacts (y-axis).

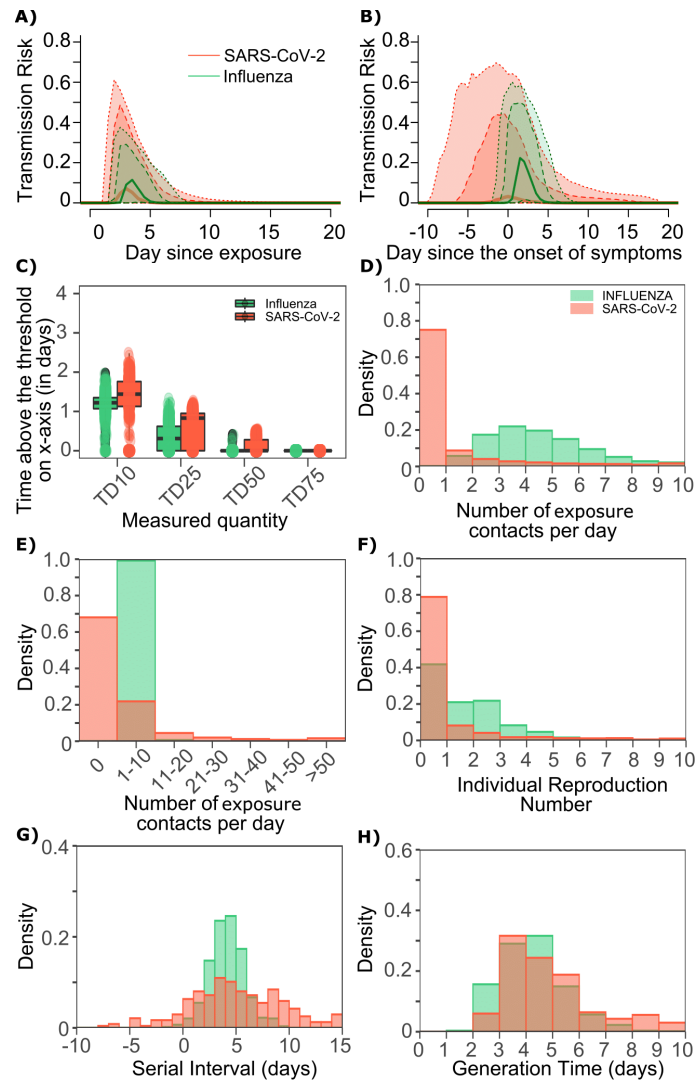
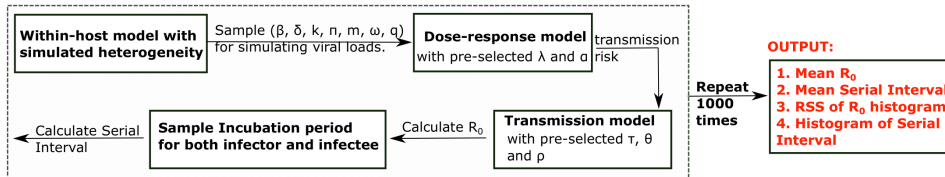


Fig 8. Differing transmission contact distributions, rather than viral kinetics explain SARS CoV-2 super spreader events. **A.** Simulated transmission risk dynamics for 10,000 infected persons with SARS-CoV-2 and influenza. Solid line is median transmission risk. Dark, dotted line is transmission risk of 75th percentile viral loads, and light dotted line is transmission risk of 95th percentile viral loads. **B.** Same as **A** but plotted as transmission risk since onset of symptoms. Highest transmission risk for SARS-CoV-2 is pre-symptoms and for influenza is post symptoms. **C.** Boxplots of duration of time spent above TD10, TD25, TD50, TD75 and TD90 for 10,000 simulated SARS-CoV-2 and influenza shedding episodes. TD10, TD25, TD50, TD75 and TD90 are viral loads at which transmission probability is 10%, 25%, 50%, 75% and 90% respectively. The midlines are median values, boxes are interquartile ranges (IQR), and datapoints are outliers. Superimposed probability distributions of: **D & E.** number of transmission contacts per day, **F.** individual R_0 , **G.** serial interval and **H.** generation time for influenza and SARS-CoV-2.

A) Calculating Mean R_0 , Mean Serial Interval and histogram of R_0



B) Finding parameter sets

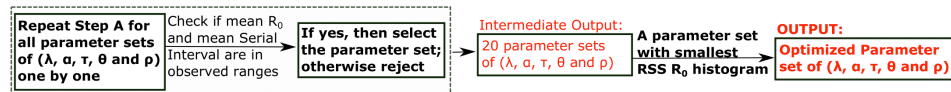


Fig S1. Mathematical model workflow.

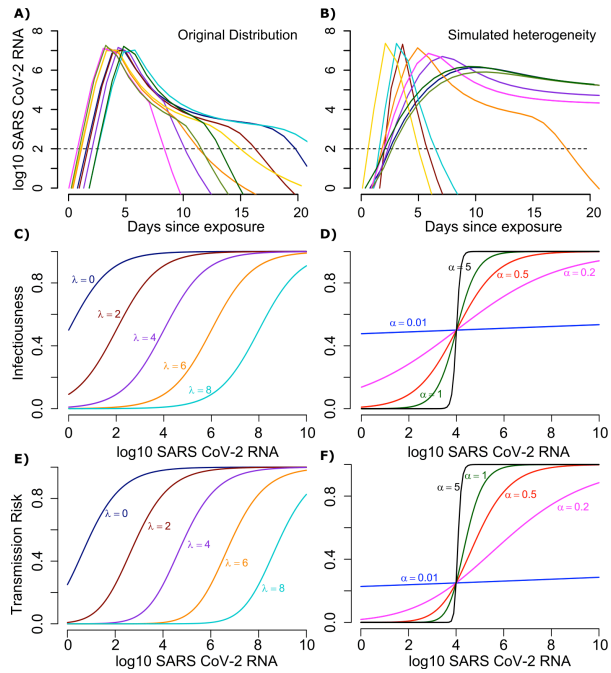


Fig S2. Mathematical model of SARS-CoV-2 transmission dynamics. **A.** Simulated viral load shedding tracings of possible transmitters. **B.** Simulated viral load shedding with imputed heterogeneity. **C.** Simulated infection dose (ID) response curves with variance in infectivity (ID50) and **D.** dose response slopes. **E.** Simulated transmission dose (TD) response curves with variance in infectivity (TD50) and **F.** dose response slopes. The TD response curve is a product of the infection and contagion dose response curves.

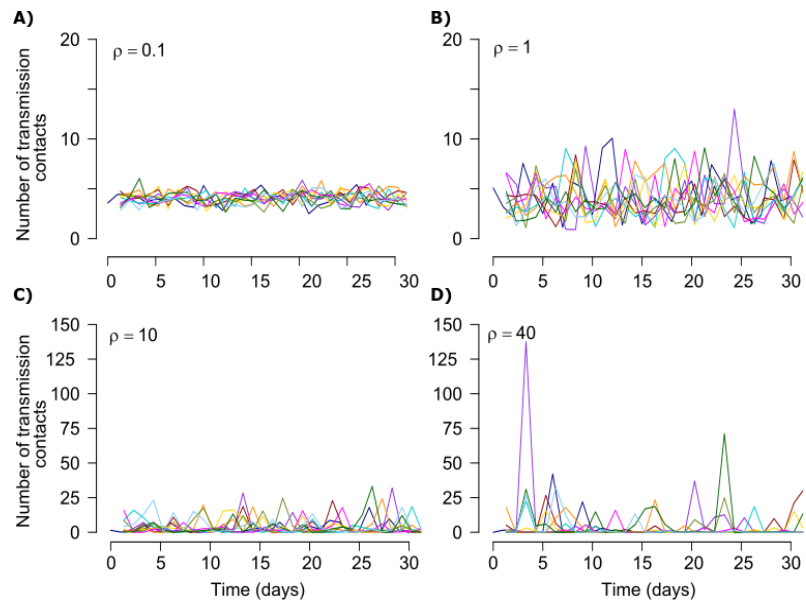


Fig S3. Stochastic simulations of exposed contact frequency for varying dispersion (ρ). The average number of exposed contacts is 4 per day in each example with imputed daily heterogeneity based on an elevated value of ρ from a gamma distribution $\sim \Gamma(4/\rho, \rho)$.

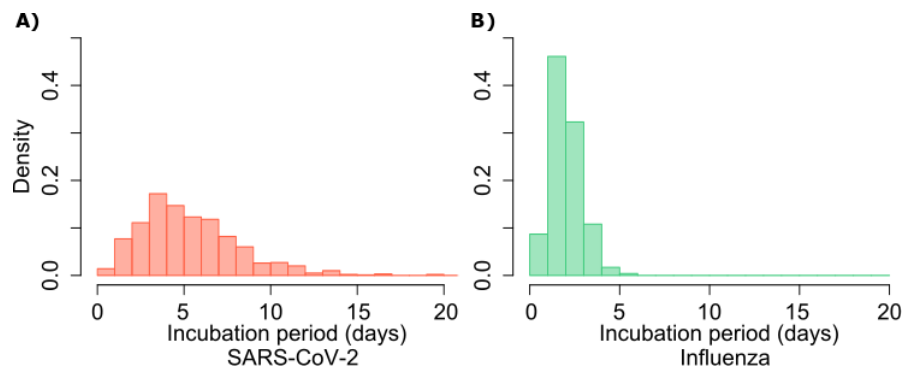


Fig S4. Gamma distribution functions of incubation periods. A. SARS-CoV-2 (mean 5.2 days, shape parameter =3.45 and rate =0.66) and **B.** influenza (mean 2 days, shape parameter=6.25 and scale parameter=0.32).

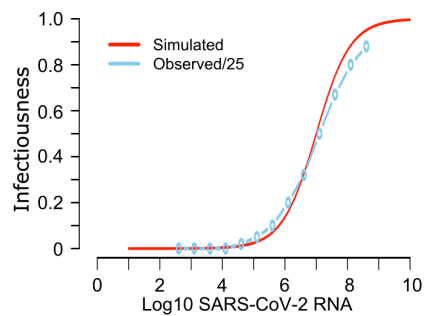


Fig S5. Mathematical model recapitulation of relationship between SARS-CoV-2 viral load and viral culture. In a clinical study, quantitative viral culture was ~25-fold lower than viral RNA measurement by PCR (<https://www.medrxiv.org/content/10.1101/2020.06.08.20125310v1>). We identify high similarity between observed viral RNA level divided by 25 and model predicted infectiousness shown here with the ID curve..

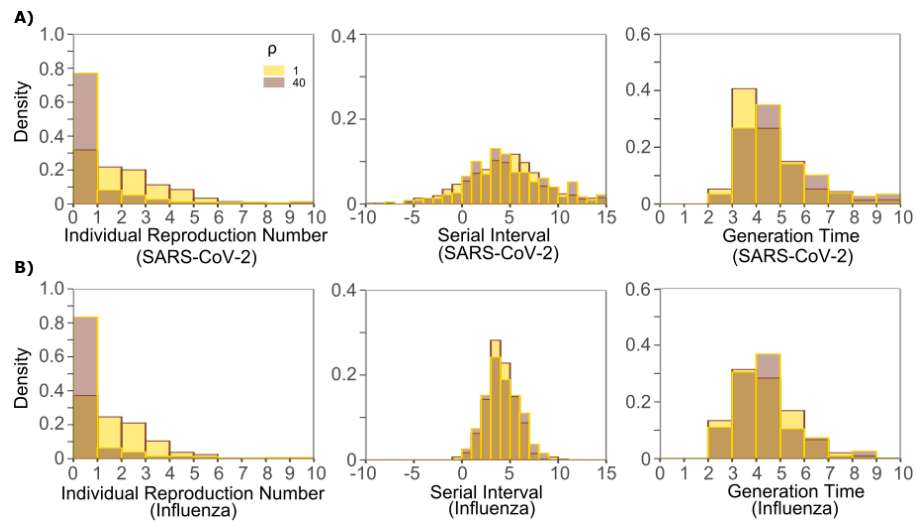


Fig S6. Impact of changes in contact network heterogeneity on individual R_0 , serial interval, and generation time. A. SARS-CoV-2, and **B.** influenza. Lowering exposed contact network heterogeneity to levels observed with influenza decreases SARS-CoV-2 individual R_0 over-dispersion. Increasing exposed contact network heterogeneity to levels observed with SARS-CoV-2 increases influenza R_0 over-dispersion. Neither change impacts observed serial interval or estimate generation time.

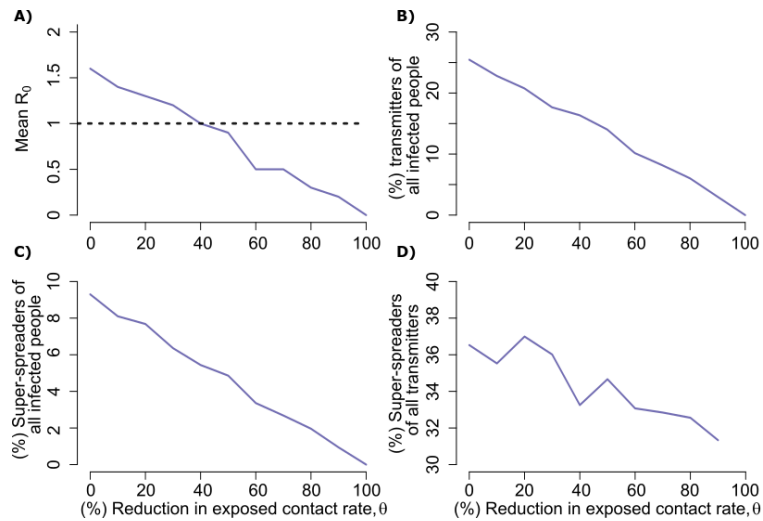


Fig S7. Potential impact of population physical distancing on SARS-CoV-2 epidemiology. **A.** Mean reproductive number **B.** Percent transmitters of all infected people **C.** Percent super-spreaders (individual $R_0 > 5$) of all infected people **D.** Percent super spreaders of all transmitters. Transmitters are defined as infected people who generate at least one secondary infection.

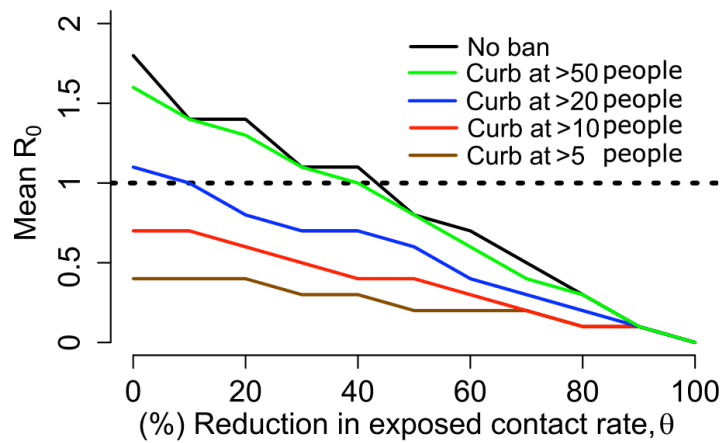


Fig S8. Potential impact of enhanced physical distancing only within high exposure contact networks on SARS-CoV-2 epidemiology. Simulations assume limitation of exposed contacts only among daily exposures of more than 5, 10, 20 or 50 people. Mean reproductive number decreases below one with only marginal decreases in overall rate of exposure contacts when contacts are limited to fewer than 20 people.

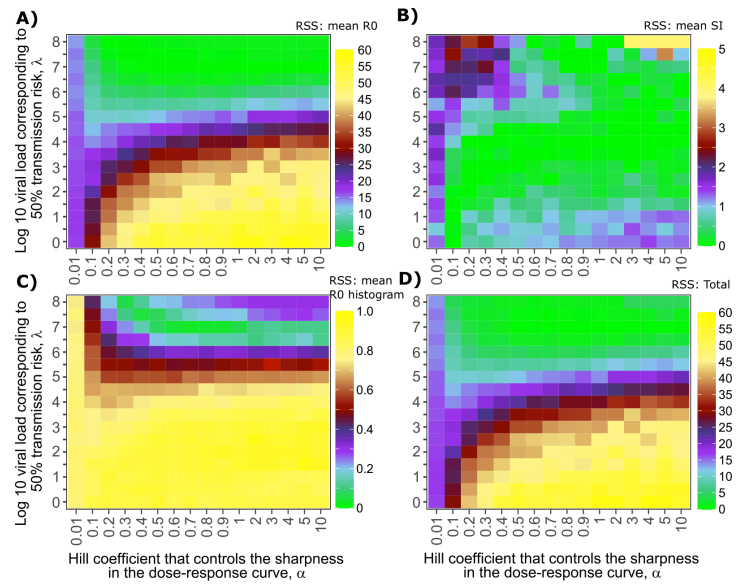


Fig S9. Sensitivity analysis of transmission curve parameter for model fit to SARS-CoV-2 data. Effects of varying transmission curve slope (x-axis) and TD50 for infectiousness (y-axis) on fit to **A.** Mean R_0 , **B.** Mean serial interval, **C.** Cumulative distribution function of individual R_0 , and **D.** Sum of Errors in A, B and C.

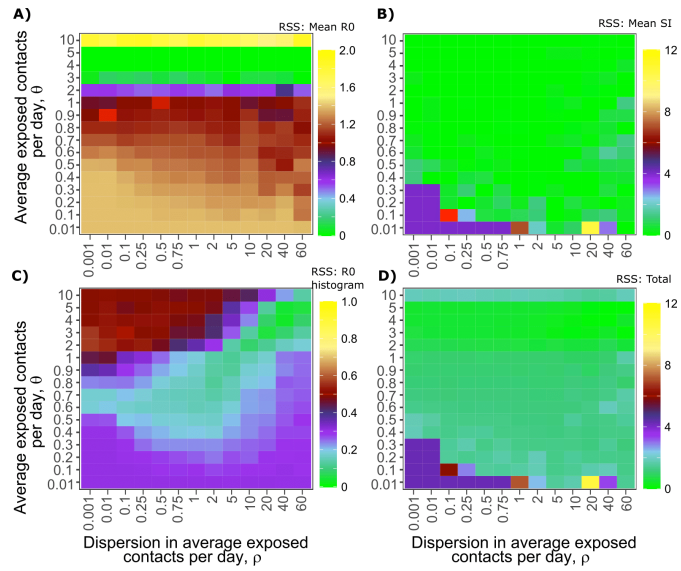


Fig S10. Sensitivity analysis of contact network structure for model fit to SARS-CoV-2 data. Effects of dispersion parameter (x-axis) and average exposed contacts per day (y-axis) on fit to **A.** Mean R0, **B.** Mean serial interval, **C.** Cumulative distribution function of individual R0, and **D.** Sum of Errors in A, B and C.

Super-spreader definitions	SARS-CoV-2			Influenza		
	All infected people	All transmitters	Contribution of super-spreaders to all transmissions	All infected people	All transmitters	Contribution of super-spreaders to all transmissions
Individual $R_0 \geq 5$	~10%	~35%	~85%	~2%	~3%	~10%
Individual $R_0 \geq 10$	~6%	~25%	~70%	~0%	~0%	~0%
Individual $R_0 \geq 20$	~2.5%	~10%	~44%	~0%	~0%	~0%

Table 1: Prevalence of super-spreaders among transmitters, and contribution of super-spreading events to all SARS-CoV-2 and influenza transmissions. Estimates are from 10,000 simulations.

$\text{Log}_{10}\beta$ (virions ⁻¹ day ⁻¹)	δ (day ⁻¹ cells ^{-k})	k (-)	$\text{Log}_{10}\pi$ (log ₁₀ day ⁻¹)	m (day ⁻¹ cells ⁻¹)	$\text{Log}_{10}\omega$ (day ⁻¹ cells ⁻¹)
-7.23	3.13	0.08	2.59	3.21	-4.55
0.2	0.02	0.02	0.05	0.33	0.01

Table S1: Population parameter estimates for simulated SARS-CoV-2 viral shedding dynamics. Parameters are from (doi: <https://doi.org/10.1101/2020.04.10.20061325>).¹³ The top row is the fixed effects (mean) and the bottom row is the standard deviation of the random effects. We also fixed $r=10$, $\delta E=1/\text{day}$, $q=2.4 \times 10^{-5}/\text{day}$ and $c=15/\text{day}$.