

1 **Clustering and superspreading potential of severe acute respiratory syndrome**
2 **coronavirus 2 (SARS-CoV-2) infections in Hong Kong**

3

4 Dillon C. Adam^{1,2}, Peng Wu¹, Jessica Y. Wong¹, Eric H. Y. Lau¹, Tim K. Tsang¹, Simon
5 Cauchemez³, Gabriel M. Leung^{1*}, Benjamin J. Cowling^{1*}

6

7 **Joint senior authors*

8

9 **Affiliations:**

10 1. WHO Collaborating Centre for Infectious Disease Epidemiology and Control,
11 School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong
12 Kong, Hong Kong Special Administrative Region, China.

13 2. Biosecurity Program, Kirby Institute, University of New South Wales, Sydney,
14 Australia.

15 3. Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Paris,
16 France.

17

18 **Corresponding author:**

19 Peng Wu, School of Public Health, Li Ka Shing Faculty of Medicine, The University
20 of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong

21 Tel: +852 3917 6746; Email: pengwu@hku.hk

22

23 Word count (abstract): 138

24 Word count (main text, excluding methods): 2,861

25

26 Key words: coronavirus; COVID-19; superspreading; transmission; public health

27

28 **Superspreading events have characterised previous epidemics of severe acute**
29 **respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory**
30 **syndrome coronavirus (MERS-CoV) infections. Using contact tracing data, we**
31 **identified and characterized SARS-CoV-2 clusters in Hong Kong. Given a**
32 **superspreading threshold of 6-8 secondary cases, we identified 5-7 probable**
33 **superspreading events and evidence of substantial overdispersion in**
34 **transmissibility, and estimated that 20% of cases were responsible for 80% of local**
35 **transmission. Among terminal cluster cases, 27% (45/167) ended in quarantine.**
36 **Social exposures produced a greater number of secondary cases compared to**
37 **family or work exposures ($p < 0.001$) while delays between symptom onset and**
38 **isolation did not reliably predict the number of individual secondary cases or**
39 **resulting cluster sizes. Public health authorities should focus on rapid tracing and**
40 **quarantine of contacts, along with physical distancing to prevent superspreading**
41 **events in high-risk social environments.**

42

43 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel beta-
44 coronavirus that was first identified in patients in China in December 2019 and
45 January 2020, and has since caused over four million confirmed human infections
46 worldwide and more than 300,000 deaths (1). Disease caused by SARS-CoV-2 is
47 termed coronavirus disease 2019 (COVID-19) and was declared a global pandemic
48 on March 11, 2020 (2). Epidemics of severe acute respiratory syndrome coronavirus
49 (SARS-CoV) in 2002-03 and Middle-East respiratory syndrome coronavirus (MERS-

50 CoV) since 2013 have been notable for the occurrence of superspreading events
51 (SSEs) (3-6). SSEs are generally defined as outbreaks in which a small number of
52 cases infect a large number of secondary cases well-above the expected average (7).
53 For SARS-CoV, with an basic reproductive number (R_0) of around 3, (8), greater than
54 eight or 10 secondary cases have been suggested to constitute SSEs (9, 10), while for
55 MERS-CoV, SSEs have reportedly involved up to 82 secondary cases (11). For SARS-
56 CoV-2, the R_0 has been estimated between 2 and 3 (12, 13), and the degree to which
57 SSEs are involved in transmission of COVID-19 remains unclear. Here, we used
58 contact tracing data to identify and describe clusters of SARS-CoV-2 infection in
59 Hong Kong and estimate the degree of variance or overdispersion in
60 transmissibility, and therefore the potential for SARS-CoV-2 SSEs.

61

62 **RESULTS**

63 *Cases and Clusters of SARS-CoV-2 in Hong Kong*

64 As of 28 April 2020, there have been a total of 1,037 laboratory-confirmed cases of
65 SARS-CoV-2 infection in Hong Kong, and one probable case based on clinical and
66 epidemiological features. The first case of SARS-CoV-2 infection in Hong Kong was
67 confirmed on January 23, 2020, imported from Hubei, China (Onset January 18, 2020).
68 The first local cases without known travel history were confirmed on February 4,
69 2020 among a family cluster with symptom onsets ranging from January 22, 2020 to
70 February 4, 2020 indicating potentially undocumented community transmission had
71 occurred in mid-January 2020 prior to the first travel restrictions implemented from

72 January 25, 2020 (Figure 1). Stable numbers of sporadic and cluster cases were
73 detected until early March when a substantial increase in the number of imported
74 cases was observed (Figure 1). Cases continued to rise until a total ban on non-
75 resident entry and mandatory 14-day quarantine for all arrivals was implemented. A
76 rise in the number of local cases including clusters initiated by local or imported
77 cases saw the implementation of various physical distancing measures such as
78 restrictions on gatherings and bar closures to prevent community transmission
79 (Figure 1).

80
81 Overall the majority (51.9%; 539/1,038) of SARS-CoV-2 infections in Hong Kong have
82 been associated with at least one of 135 known clusters. The median cluster size was
83 two and the largest involved 106 local cases. The remaining 38.9% (210/539) of
84 cluster cases solely involved imported cases where no onward local transmission
85 could be identified. Most clusters comprised only imported cases (82/135; 60.7%),
86 followed by clusters initiated by an imported case (30/135; 22.2%) and clusters
87 initiated by a local case (23/135; 17.0%). Among the 499 sporadic cases not linked to
88 any cluster, 90.0% of infections were acquired overseas (449/499), while the
89 remaining 50 (10.0%) had no history of travel and could not be linked to any other
90 case either as a source or subsequent generation. Of cluster cases, 225 (41.7%;
91 225/539) belonged to clusters initiated by another local case, compared to 19.3%
92 (104/539) of cluster cases which belonged to clusters initiated by an imported case.
93 The composition of clusters is described further in Supplementary Table 1.

94

95 *Transmission Pairs & Individual variation in SARS-CoV-2 transmission*

96 From the 539 infections that occurred in clusters, all 210 solely imported cases
97 (210/539; 38.9%) were excluded from subsequent analysis due to uncertainties
98 concerning transmission within each cluster whilst overseas. Within the remaining
99 53 clusters initiated by a local or imported infection, 245 (245/329; 74.5%) could be
100 linked resulting in 171 unique infector-infectee transmission pairs with 94 unique
101 infectors. The age of distribution of infectors and infectees was not significantly
102 different (Supplementary Figure 1A) however a positive trend by age can be seen in
103 the transmission pair matrix (Supplementary Figure 1B).

104

105 Figure 2A shows the empirical serial distribution between all infector-infectee pairs
106 and fitted normal and lognormal distributions. The median serial interval was 4
107 days (IQR 3 – 8 days), and the mean of the fitted normal distribution was 5.6 days
108 (standard deviation 4.3 days). We observed at least eight instances of likely pre-
109 symptomatic transmission where symptom onset of the infectee preceded that of the
110 infector by one day (N=2) or occurred on the same day (N=6). Thirty-four unique
111 infectors (35/94; 37.2%) were linked to two or more secondary cases, and the largest
112 number of individual secondary cases was 11. From the empirical offspring
113 distribution and fitted negative binomial distribution shown in Figure 2B, we
114 estimated an observed reproductive number (R) of 0.58 (95% CI: 0.45 – 0.71) and
115 dispersion parameter (k) of 0.45 (95% CI: 0.31 – 0.76). Given a superspreading

116 threshold of 6-8 secondary cases, we directly observed 2-4 SSEs where the source
117 was known. Likelihood analysis based on final cluster sizes increased the estimate of
118 R to 0.75 (95% CI: 0.6 – 0.96) and decreased estimates of k to 0.37 (95% CI: 0.16 –
119 1.16). Based on these estimates we determined a coefficient of variation of 2.5 and
120 inferred that approximately 20% of SARS-CoV-2 infections are responsible for 80%
121 of all transmission events in Hong Kong (Table 1). Epoch analyses by wave (before
122 or after March 1, 2020) showed similar estimates of R and k (Supplementary Table 2)
123 from their respective offspring distributions (Supplementary Figure 2). A peak
124 however in the number of index cases per day could be seen in wave two
125 (Supplementary Figure 3). Parameter estimates and distributions for all fitted
126 distributions are shown in Supplementary Tables 3 & 4, and Supplementary Figures
127 4 & 5.

128

129 *Chains of SARS-CoV-2 transmission in Hong Kong*

130 The largest local cluster was 106 cases and was traced back to multiple social
131 exposures among a collection of bars across Hong Kong (Figure 3A). Evidence
132 suggested this “bar and band” cluster originated in Lan Kwai Fong among a few
133 staff and customers before being spread to additional venues by a number of
134 musicians, however the source and chains of transmission between many bar cases
135 could not be determined from epidemiological data. The earliest reported exposure
136 was reported on March 7th among two customers who later presented with
137 symptoms on March 11th (Supplementary Figure 6A). The earliest symptom onset

138 was among two staff members on March 10th & 11th, however neither case was
139 confirmed until March 24th and 25th and neither reported their history of exposure at
140 the bar during this time. Given the serial intervals between these four cases and later
141 cases (Supplementary Figure 6A), it is possible that an undetected source may have
142 initially infected these staff and customers on March 7, before being spread to the
143 musicians. The earliest onset among the musicians was on March 17, with most
144 subsequently affected customers reporting exposures between March 17 and March
145 20 (Supplementary Figure 6C), which is suggestive of at least one or more probably
146 SSEs within this exposure period. The bars were voluntarily closed from March 23
147 for cleaning prior to mandated closures from April 3. Of the 73 primary bar cases, 39
148 customers, 20 staff, and 14 musicians were infected, while the remaining 33
149 infections were secondary, tertiary or quaternary contacts (Supplementary Figure
150 6A). This single outbreak accounts for 10.2% (106/1,038) of all cases in Hong Kong
151 regardless of the source, but 30.4% of all local cases acquired in Hong Kong
152 (106/349). Evidence of asymptomatic transmission was determined in a single case (a
153 staff member at the bar) who was placed in quarantine before their wife
154 subsequently presented with symptoms and tested positive. Supplementary Figure
155 6B shows the age distribution of all cases associated with this cluster.

156

157 Figure 3B describes a cluster of 21 cases linked to a wedding. Ten cases resulted from
158 a previous social exposure (SSE), of which four cases subsequently attended the
159 wedding. Individual transmission pairs within the wedding could not be

160 determined, however there were at least seven secondary infections, and therefore
161 another potential SSE, and an additional two tertiary cases among family members
162 of the wedding guests. A final potential SSE of undetermined origin was associated
163 with religious activities at a local temple and resulted in 11 primary cases and 18
164 recorded cases total (Figure 3C). Cases reported multiple exposures over a number
165 of days such that a single point source exposure was unlikely. Six secondary cases
166 were linked via family exposures. The last case who worked at the temple was also
167 infected however remained asymptomatic. It is unknown if this case was the source
168 of the temple exposures or was infected by the undetermined source/s.

169

170 Figure 3D shows all other local cluster cases and chains of SARS-CoV-2 transmission
171 in Hong Kong. In total, 51 subsequent cases were quarantined. Among terminal
172 transmission cases, 26.9% (45/167) ended in quarantine (excluding five quarantined
173 bars cases whose terminal transmission position was ambiguous; Figure 4A;
174 Supplementary Table 5). The odds that a terminal case ended in quarantine was
175 16.97 (95% CI: 2.27 – 126.69), while the probability that a case was terminal and
176 quarantined was 21.0% (45/214) compared to 57.0% (122/214) for cases who were
177 terminal but not quarantined (Supplementary Table 5). Transmission within families
178 was the most frequent event among all recognised transmission pairs (96/171; 56.1%)
179 followed by transmission in social settings (55/171; 32.2%) and at work (20/171;
180 11.1%). However, there was very strong evidence that the number of individual
181 secondary cases resulting from social exposures was significantly higher compared

182 to work or family exposures ($\mu = 2.89$ vs. 1.36; $p < 0.001$, negative binomial regression
183 model).

184

185 *Relationship between delayed case isolation and potential COVID-19 cluster size*

186 Figure 4A shows the median individual delay from symptom onset-to-
187 confirmation/isolation among SARS-CoV-2 infections in Hong Kong by cluster
188 membership on the basis of size. We found that the delay from symptom onset to
189 confirmation among cluster cases did not predict the overall cluster size although if
190 excluding the two largest clusters there was a slight increase in onset-to-isolation
191 delays in some of the larger clusters (Figure 4A). Delay from symptom-onset to
192 isolation among infectors similarly was not associated with the number of secondary
193 cases or the overall size of each cluster ($p = 0.955$, linear regression; Figure 4B).

194

195 **DISCUSSION**

196 Public health measures have successfully suppressed transmission of SARS-CoV-2
197 infections in Hong Kong with an estimated reproductive number below 1 (Table 1;
198 $R=0.58$, 95%: 0.45 – 0.71) compared to estimates of the basic reproductive number R_0
199 of 2-3 (12, 13), and our estimate varied only marginally between phase one (Jan-Feb)
200 and phase two of the epidemic in Hong Kong (Mar-Apr) (Supplementary Table 2;
201 $R=0.60$, 95%: 0.35 – 0.94 and $R=0.56$, 95%: 0.43 – 0.70). Thirty-three percent (349/1,038)
202 of all SARS-CoV-2 infections in have been acquired within Hong Kong, either within
203 clusters or as sporadic local cases occurring through limited community

204 transmission that has been controlled through a range of public health measures
205 (14). Our results indicate 5-7 probable SSEs in Hong Kong, and substantial
206 individual heterogeneity in the transmissibility of SARS-CoV-2 infection (Table 1;
207 $k=0.45$, 95%: 0.30 – 0.72) and therefore potential for future superspreading, albeit less
208 than SARS-CoV and MERS-CoV ($k=0.16$ and 0.26 respectively) (15, 16). Super-
209 spreading events pose considerable challenges for local control as they can quickly
210 overwhelm public health contact tracing capacity, and because most infected
211 persons will generate few secondary infections, while a small fraction can generate
212 many (Figure 2B). Indeed, we estimated that approximately 20% of cases were
213 responsible for 80% of all SARS-CoV-2 transmission in Hong Kong (Table 1). These
214 results however should be interpreted in the context of constrained community
215 transmission given moderate levels of physical distancing currently practiced in
216 Hong Kong, including school closures, some adults working at home, cancellation of
217 mass gatherings, as well as improved hygiene and universal mask wearing (17).
218 Recent findings from Shenzhen, China estimated comparable levels of
219 overdispersion using contact tracing data ($k=0.58$) (18), however other studies
220 utilizing global datasets have estimated even greater potential for SARS-CoV-2
221 superspreading ($k=0.1$ given a global R of 2.5) and suggest as little as 10% of cases
222 could account for 80% of all transmission (19). Such degrees of overdispersion
223 however can be advantageous to disease control efforts if interventions can
224 effectively target the core high-risk group responsible for the majority of
225 transmission (20, 21).

226

227 Superspreading is considered a function of both variations in individual
228 transmissibility and individual susceptibility or exposure. Our results show that the
229 number of individual secondary cases was significantly higher within social settings
230 such as bars and restaurants compared to family or work exposures ($p < 0.001$). This is
231 certainly due to the greater numbers of contacts expected in such settings. Social
232 exposures are therefore at an increased risk for SARS-CoV-2 transmission and likely
233 constitute the core behavioural risk factor for SSEs. Targeted interventions should
234 therefore focus on reducing extreme numbers of social contacts at high-risk venues
235 such as bars, nightclubs and restaurants, which also appear at increased risk of SSE
236 (22), either via closures or physical distancing policies, both of which currently
237 remain implemented in Hong Kong (17). Models have shown however that
238 heterogeneity in exposure may also have disproportionate effects on herd immunity
239 (23), including SARS-CoV-2 (24, 25). Because some individuals are at greater risk of
240 exposure, achieving herd immunity in this core risk group may therefore have
241 suppressive effects on local transmission more generally, potentially reducing the
242 otherwise necessary duration and intensity of more disruptive community-wide
243 physical distancing policies. Given a coefficient of variation of 2.5 from our study
244 (Table 1) this roughly corresponds to an adjusted herd immunity threshold of 20%
245 (25), which is consistent with our estimate that 20% of individuals are responsible for
246 80% of transmission in Hong Kong. This compares to a herd-immunity threshold of
247 60-70% via vaccine-acquired immunity. The real-world effects of naturally acquired

248 herd immunity however should be further studies and measured with serology
249 along with the progressive easing of restrictions.

250

251 Previous modelling has suggested that reduced delays between symptom onset and
252 isolation are important indicators for the control of SARS-CoV-2 outbreaks (20),
253 however this did not appear to play a major role in Hong Kong: in our analysis,
254 delay from symptom onset to isolation were not associated with the number of
255 individual secondary cases, nor the final cluster size (Figure 4A and 4B). In contrast,
256 for SARS-CoV various epidemics in 2003, due to delayed viral shedding among
257 cases (highest approximately seven days after symptom onset) (26, 27), delays in
258 case isolation adversely affected disease control efforts (28). For COVID-19,
259 confirmation and isolation of symptomatic cases will have a limited effect on
260 reducing overall SARS-CoV-2 transmission in the community unless done very
261 quickly, noting the growing body of evidence of substantial transmission during the
262 pre-symptomatic and early symptomatic period (29-31). In Hong Kong, the average
263 delay to isolation was 5.6 days for cluster cases (Figure 4), and by this time most
264 onwards transmission had perhaps already occurred. The most important public
265 health measures therefore are likely to be case identification followed by rapid and
266 parallel (e.g. before contacts are confirmed as cases) contact tracing and quarantine.
267 Some degree of intermittent physical distancing may also be required to suppress
268 potential local transmission from unidentified infections and pre-symptomatic

269 transmission, but must necessarily be balanced with the social, economic and
270 educational costs associated with such policies.

271

272 Overall our study has a few limitations. Primarily, because this study relies on the
273 completeness of case ascertainment and contact tracing data, any degree of
274 incompleteness could therefore bias our estimates. Further, because the source of 50
275 sporadic local infections could not be determined, we likely underestimated R based
276 on the empirical data alone. In fact, the expected difference between of R and k from
277 our observed estimates and likelihood model indicates the presence of some
278 incompleteness and therefore potential bias, however only marginally so (Table 1).
279 Secondly, our results also suggest a potential caveat: despite evidence supporting
280 rapid contact tracing and the effectiveness of quarantine (Figure 4), in our analysis,
281 most chains of transmission did not terminate in quarantine (Figure 3;
282 Supplementary Table 5). This however must be interpreted in the context of
283 moderate physical distancing in Hong Kong as highlighted previously (17), rather
284 than a suggestion that quarantine was not essential.

285

286 Overall there is substantial potential for SSEs in COVID-19, but less so than for SARS
287 and MERS. In the absence of non-pharmaceutical interventions such as physical
288 distancing implemented in Hong Kong, the potential for SSEs is likely greater than
289 observed in our study. Assuming local elimination is not possible, disease control
290 efforts should focus on rapid tracing and quarantine of confirmed contacts, along

291 with the implementation of physical distancing policies or closures targeting high-
292 risk social exposures such as bars, nightclubs and restaurants to prevent the
293 occurrence of SSEs. Given the long right-hand tail of the distribution of individual
294 reproductive numbers (Figure 2B), preventing SSEs would have a considerable effect
295 in reducing the overall reproductive number. In lieu of an effective vaccine, these
296 results have significant implications for the control of COVID-19 and public health
297 measures such as physical distancing and the relaxation of lockdowns around the
298 world.

299

300 **METHODS**

301 *Characterisation of clusters and chains of SARS-CoV-2 transmission*

302 Using case line lists and contact tracing data obtained from the Centre for Health
303 Protection (CHP) of the Department of Health in Hong Kong we characterised
304 clusters of SARS-CoV-2 infections and chains of transmission within clusters. Cases
305 of SARS-CoV-2 infection were laboratory confirmed via RT-PCR. Clusters were
306 defined as two or more confirmed infections with close contact prior to or following
307 symptom onset. Each cluster was characterised by the travel history of the index case
308 as either initiated by an imported case (i.e. index acquired SARS-CoV-2 infection
309 overseas), initiated by a local case or clusters of solely imported cases. Cases not
310 linked to any cluster were categorised as sporadic local or sporadic imported cases if
311 infection was acquired locally or overseas. We determined probable infector-infectee
312 transmission pairs and chains of transmission within clusters from reported contact

313 histories and symptom onset dates for all clusters initiated by an imported or local
314 case. Transmission pairs were characterised by the reported site of contact as either
315 family, social, work or during local travel such as on public transport and modelled
316 using a negative binomial log-linear model. Transmission between cases within
317 clusters of solely imported cases were not considered due to substantial ambiguities
318 concerning transmission within each cluster whilst overseas.

319

320 *Calculation of serial interval and empirical offspring distribution*

321 We calculated the median serial interval as the difference between the symptom
322 onset dates of each infector-infectee pair, excluding asymptomatic cases, and fitted
323 normal, lognormal, gamma and Weibull distributions using the R package
324 "fitdistrplus" and maximum-likelihood methods, excluding eight non-positive data
325 points for the latter three distributions. We generated the empirical offspring
326 distribution from the observed number of secondary cases per individual infector
327 and similarly fit negative binomial, geometric and Poisson distributions as before.
328 Cases terminal to the inferred chain of transmission and sporadic local cases were
329 considered to have zero secondary cases. We investigated the relationship between
330 delay in days from symptom onset-to-confirmation as a proxy for potential
331 infectiousness in the community and the number of secondary cases per infector by
332 linear regression. We compared each fit distribution using Akaike information
333 criterion (AIC) scores and calculated confidence intervals for parameter from 1000
334 bootstrapped replicates.

336 Individual variation in SARS-CoV-2 transmission

337 Following the approach described by Lloyd Smith et al. (15), estimates of the
 338 observed reproductive number (R) was determined from the mean of the negative
 339 binomial distribution fit to the empirical offspring distribution, and the degree of
 340 individual variation on transmission from the corresponding dispersion parameter k .
 341 We repeated this analysis for both wave one (January-February 2020) and wave two
 342 (March-April 2020), sub-setting by illness onset date of the index case. Due to
 343 potential biases affecting the empirical offspring distributions, we simultaneously
 344 implemented a likelihood-based branching process model to jointly infer R and k
 345 given the final size of all local clusters following Kucharski and Althaus (16) where
 346 the probability that an index case generates clusters n of size j is given by (32, 33):

347

$$348 \quad r_j = \frac{\Gamma(kj + (j - 1))}{\Gamma(kj)\Gamma(j + 1)} \frac{\left(\frac{R_0}{k}\right)^{j-1}}{\left(1 + \frac{R_0}{k}\right)^{kj+j-1}}$$

349 And the likelihood is:

$$350 \quad L = \prod_{j=1}^{\infty} r_j^{n_j}$$

351

352 For a given range of values for R (0.10 – 3.00) and k (0.01 – 55), the probability of an
 353 index case generating a cluster equal to or greater than j is given by:

354

355

$$p_j = 1 - \sum_{i=1}^{j-1} r_i$$

356

357

From the empirical and likelihood model estimates of R and k , we calculated the

358

coefficient of variation as the ratio of standard deviation of the corresponding

359

negative binomial distribution and the mean. Following from Endo et al. (19), given

360

parameters R and k , the expected proportion of cases responsible for 80% of

361

transmission in Hong Kong is given by:

362

363

$$1 - \text{P80\%} = \frac{1}{R_0} \int_0^X [x] \text{NB} \left([x]; k, \frac{k}{R_0 + k} \right) dx$$

364

where X satisfies:

365

$$1 - 0.8 = \frac{1}{R_0} \int_0^X [x] \text{NB} \left([x]; k, \frac{k}{R_0 + k} \right) dx$$

366

and:

367

$$\frac{1}{R_0} \int_0^X [x] \text{NB} \left([x]; k, \frac{k}{R_0 + k} \right) dx = \int_0^{X-1} [x] \text{NB} \left([x]; k + 1, \frac{k}{R_0 + k} \right) dx$$

368

369

Finally, as per Lloyd Smith et al. (15) given R_0 , the superspreading threshold is

370

calculated as the 99th-percentile of the Poisson(R_0) distribution where:

371

$\Pr(Z \leq Z^{(99)} | Z \sim \text{Poisson}(R_0)) = 0.01$. Therefore, with the global consensus of R_0 in

372

the range 2 – 3, we define the superspreading threshold for SARS-CoV-2 here as 6-8

373 individual secondary cases. All statistical analyses were performed in R version 4.0.0
374 (R Foundation for Statistical Computing, Vienna, Austria).

375

376

377 **ACKNOWLEDGMENTS**

378 This project was supported by a commissioned grant from the Health and Medical
379 Research Fund, Food and Health Bureau, Government of the Hong Kong Special
380 Administrative Region, and the Theme-based Research Scheme [Project No. T11-
381 712/19-N] of the Research Grants Council of the Hong Kong SAR Government.

382

383 **AUTHOR CONTRIBUTIONS**

384 All authors meet the ICMJE criteria for authorship. The study was conceived by
385 DCA and BJC. Data analyses were done by DCA, JYW and EHYL. DCA wrote the
386 first draft of the manuscript, and all authors provided critical review and revision of
387 the text and approved the final version.

388

389 **COMPETING INTERESTS STATEMENT**

390 BJC consults for Roche and Sanofi Pasteur. The authors report no other potential
391 conflicts of interest.

392

393 **REFERENCES**

- 394 1. World Health Organization. Coronavirus disease 2019 (COVID-19): situation
395 report, 100. 2020.
- 396 2. WHO Director-General's opening remarks at the media briefing on COVID-19 -
397 11 March 2020 Geneva, Switzerland: World Health Organization; [cited 2020
398 May 07]. Available from: [https://www.who.int/dg/speeches/detail/who-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)
399 [director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)
400 [march-2020](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020).
- 401 3. Wang SX, Li Y, Sun B, Zhang S, Zhao W, Wei M, et al. The SARS outbreak in a
402 general hospital in Tianjin, China—the case of super-spreader. *Epidemiology &*
403 *Infection*. 2006;134(4):786-91.
- 404 4. Kim K, Tandil T, Choi JW, Moon J, Kim M. Middle East respiratory syndrome
405 coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology,
406 characteristics and public health implications. *Journal of Hospital Infection*.
407 2017;95(2):207-13.
- 408 5. Cho SY, Kang J-M, Ha YE, Park GE, Lee JY, Ko J-H, et al. MERS-CoV outbreak
409 following a single patient exposure in an emergency room in South Korea: an
410 epidemiological outbreak study. *Lancet*. 2016;388(10048):994-1001.
- 411 6. Cowling BJ, Park M, Fang VJ, Wu P, Leung GM, Wu JT. Preliminary
412 epidemiological assessment of MERS-CoV outbreak in South Korea, May to
413 June 2015. *Euro Surveill*. 2015;20(25):7-13.
- 414 7. Galvani AP, May RM. Dimensions of superspreading. *Nature*.
415 2005;438(7066):293-5.
- 416 8. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory
417 syndrome reveal similar impacts of control measures. *American Journal of*
418 *Epidemiology*, 2004;160(6):509-16.
- 419 9. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, et al. Superspreading sars
420 events, Beijing, 2003. *Emerging infectious diseases*. 2004;10(2):256.

- 421 10. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory
422 syndrome reveal similar impacts of control measures. *American Journal of*
423 *Epidemiology*. 2004;160(6):509-16.
- 424 11. Cho SY, Kang J-M, Ha YE, Park GE, Lee JY, Ko J-H, et al. MERS-CoV outbreak
425 following a single patient exposure in an emergency room in South Korea: an
426 epidemiological outbreak study. *Lancet*. 2016;388(10048):994-1001.
- 427 12. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation
428 of the basic reproduction number of novel coronavirus (2019-nCoV) in China,
429 from 2019 to 2020: A data-driven analysis in the early phase of the outbreak.
430 *International Journal of Infectious Diseases*. 2020;92:214-7.
- 431 13. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive
432 number of novel coronavirus (COVID-19) and the probable outbreak size on
433 the Diamond Princess cruise ship: A data-driven analysis. *International Journal*
434 *of Infectious Diseases*. 2020;93:201-4.
- 435 14. Leung GM, Cowling BJ, Wu JT. From a Sprint to a Marathon in Hong Kong.
436 *New England Journal of Medicine*. 2020;382(18):e45.
- 437 15. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the
438 effect of individual variation on disease emergence. *Nature*. 2005;438(7066):355-
439 9.
- 440 16. Kucharski A, Althaus C. The role of superspreading in Middle East respiratory
441 syndrome coronavirus (MERS-CoV) transmission. *Euro surveillance*.
442 2015;20(25):pii= 21167.
- 443 17. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact
444 assessment of non-pharmaceutical interventions against coronavirus disease
445 2019 and influenza in Hong Kong: an observational study. *Lancet Public*
446 *Health*. 2020;5(5):e279-e88.
- 447 18. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission
448 of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a
449 retrospective cohort study. *Lancet Infectious Diseases*.

- 450 19. Endo A, null n, Abbott S, Kucharski A, Funk S. Estimating the overdispersion
451 in COVID-19 transmission using outbreak sizes outside China [version 1; peer
452 review: 1 approved]. Wellcome Open Research. 2020;5(67).
- 453 20. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al.
454 Feasibility of controlling COVID-19 outbreaks by isolation of cases and
455 contacts. *Lancet Global Health*. 2020;8(4):e488-e96.
- 456 21. Woolhouse ME, Dye C, Etard J-F, Smith T, Charlwood J, Garnett G, et al.
457 Heterogeneities in the transmission of infectious agents: implications for the
458 design of control programs. *Proceedings of the National Academy of Sciences*.
459 1997;94(1):338-42.
- 460 22. Lu J, Gu J, Li K, Xu C, Su W, Lai Z, et al. COVID-19 Outbreak Associated with
461 Air Conditioning in Restaurant, Guangzhou, China, 2020. *Emerg Infect Dis*.
462 2020;26(7).
- 463 23. Hill AN, Longini Jr IM. The critical vaccination fraction for heterogeneous
464 epidemic models. *Mathematical Biosciences*. 2003;181(1):85-106.
- 465 24. Britton T, Ball F, Trapman P. The disease-induced herd immunity level for
466 Covid-19 is substantially lower than the classical herd immunity level. *arXiv*
467 preprint arXiv:200503085. 2020.
- 468 25. Gomes MGM, Aguas R, Corder RM, King JG, Langwig KE, Souto-Maior C, et
469 al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the
470 herd immunity threshold. *medRxiv*. 2020.
- 471 26. Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, et al. Clinical
472 progression and viral load in a community outbreak of coronavirus-associated
473 SARS pneumonia: a prospective study. *Lancet*. 2003;361(9371):1767-72.
- 474 27. Pitzer VE, Leung GM, Lipsitch M. Estimating variability in the transmission of
475 severe acute respiratory syndrome to household contacts in Hong Kong, China.
476 *American Journal of Epidemiol*. 2007;166(3):355-63.
- 477 28. Li Y, Yu ITS, Xu P, Lee JHW, Wong TW, Ooi PL, et al. Predicting Super
478 Spreading Events during the 2003 Severe Acute Respiratory Syndrome

- 479 Epidemics in Hong Kong and Singapore. *American Journal of Epidemiology*.
480 2004;160(8):719-28.
- 481 29. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral
482 shedding and transmissibility of COVID-19. *Nature Medicine*. 2020:1-4.
- 483 30. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dorner L, et al.
484 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital
485 contact tracing. *Science*. 2020.
- 486 31. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al.
487 Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing
488 Facility. *New England Journal of Medicine*. 2020.
- 489 32. Blumberg S, Lloyd-Smith JO. Inference of R_0 and transmission heterogeneity
490 from the size distribution of stuttering chains. *PLoS Computational Biology*.
491 2013;9(5).
- 492 33. Nishiura H, Yan P, Sleeman CK, Mode CJ. Estimating the transmission
493 potential of supercritical processes based on the final size distribution of minor
494 outbreaks. *Journal of Theoretical Biology*. 2012;294:48-55.

495

496

497 **FIGURE LEGENDS**

498 **Figure 1.** Epidemic curve of daily cases of laboratory-confirmed SARS-CoV-2
499 infection in Hong Kong by symptom onset date and coloured by cluster category
500 (N=1,038). Important travel and community health interventions are indicated with
501 arrows. Asymptomatic cases are included here by date of confirmation.

502

503 **Figure 2: (A)** Empirical serial distribution of SARS-CoV-2 infections in Hong Kong
504 among symptomatic infector-infectee pairs with fitted normal (solid line) and
505 lognormal (dotted) distributions. The lognormal distribution was fitted excluding
506 observations ≤ 0 days. **(B)** Empirical offspring distribution of SARS-CoV-2 among
507 local and imported COVID-19 cases in Hong Kong (excluding clusters of cases
508 where infections had been acquired overseas) and a fitted negative binomial
509 distribution with $R=0.58$ and $k=0.45$.

510

511 **Figure 3.** Local chains of SARS-CoV-2 transmission in Hong Kong initiated by
512 another local case or an imported case. (A) Transmission network of the “bar and
513 band” cluster of cases. The source case of the initial exposure could not be
514 determined. (B) Transmission network associated with a single wedding exposure
515 subsequently linked to a preceding social gathering and local source exposure. (C)
516 Transmission network associated with a Buddhist temple in Hong Kong. Primary
517 cases report multiple distinct exposures to the temple over the course of many days.
518 It is suspected that an asymptomatic monk at the temple may have been the source

519 (indicated by the pink circle) however it cannot be known if they were exposed to
520 another case, possibly an environmental exposure. (D) All other clusters of SARS-
521 CoV-2 infections initiated by local and imported cases where the source and
522 transmission chain could be determined. *Footnote 1: An asymptomatic employee at one
523 of the bars was placed in quarantine on 26 March following recognition of the cluster and
524 tested positive on 4 April. Was transferred to hospital for observation but did not develop
525 symptoms. This employee's wife developed symptoms on 1 April and was confirmed to have
526 SARS-CoV-2 infection on 6 April.*

527

528 **Figure 4. (A)** Distribution of delays in days from symptom onset to isolation of cases
529 by cluster size. **(B)** Delays from symptom onset to isolation of index cases by the
530 number of secondary cases per index coloured by site of transmission (Dark Blue:
531 Family, Red: Social, Green: Local travel, Light Blue: Work)

532

533

534 **Table 1:** Estimated effective reproductive number, overdispersion parameter,
 535 coefficient of variation and proportion of infectious cases responsible for 80% of
 536 transmission.

Distribution/Model	R (95% CI)	k (95% CI)	Coefficient of variation (95% CI)	Prop80%* (95% CI)
Empirical Offspring	0.58 (0.45 – 0.71)	0.45 (0.30 – 0.72)	2.5 (1.8 – 3.3)	20% (15% – 25%)
Branching Process	0.74 (0.58 – 0.96)	0.33 (0.15 – 0.97)	2.6 (1.4 – 4.3)	17% (11% – 29%)

537 *Prop80% is the proportion of cases responsible for 80% of transmission