

1 Inference of SARS-CoV-2 generation times using UK household data

2 W.S. Hart^{1,*}, S. Abbott², A. Endo², J. Hellewell², E. Miller³, N. Andrews⁴, P.K. Maini¹, S.
3 Funk^{2,†}, R.N. Thompson^{5,6,†}

4 Affiliations:

5 ¹Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK.

6 ²Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and
7 Tropical Medicine, London, WC1E 7HT, UK.

8 ³Department of Infectious Disease Epidemiology, Faculty of Epidemiology & Population
9 Health, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.

10 ⁴Data and Analytical Sciences, Public Health England, London, UK.

11 ⁵Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.

12 ⁶Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research,
13 University of Warwick, Coventry, CV4 7AL, UK.

14 *Correspondence to: william.hart@keble.ox.ac.uk.

15 †These authors contributed jointly to this research

16

17 Abstract:

18 The distribution of the generation time (the interval between individuals becoming infected
19 and passing on the virus) characterises changes in the transmission risk during SARS-CoV-2
20 infections. Inferring the generation time distribution is essential to plan and assess public
21 health measures. We previously developed a mechanistic approach for estimating the
22 generation time, which provided an improved fit to SARS-CoV-2 data from January-March

23 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
2020 compared to existing models. However, few estimates of the generation time exist

24 based on data from later in the pandemic. Here, using data from a household study conducted
25 from March-November 2020 in the UK, we provide updated estimates of the generation time.
26 We consider both a commonly used approach in which the transmission risk is assumed to be
27 independent of when symptoms develop, and our mechanistic model in which transmission
28 and symptoms are linked explicitly. Assuming independent transmission and symptoms, we
29 estimated a mean generation time (4.2 days, 95% CrI 3.3-5.3 days) similar to previous
30 estimates from other countries, but with a higher standard deviation (4.9 days, 3.0-8.3 days).
31 Using our mechanistic approach, we estimated a longer mean generation time (6.0 days, 5.2-
32 7.0 days) and a similar standard deviation (4.9 days, 4.0-6.3 days). Both models suggest a
33 shorter mean generation time in September-November 2020 compared to earlier months.
34 Since the SARS-CoV-2 generation time appears to be changing, continued data collection
35 and analysis is necessary to inform future public health policy decisions.

36

INTRODUCTION

37 The generation time of a SARS-CoV-2 infector-infectee pair is defined as the period of time
38 between the infector and infectee each becoming infected [1–5]. The distribution of the
39 generation times of many infector-infectee pairs characterises the temporal profile of
40 infectiousness of an infected host (averaged over all hosts and normalised so that it represents
41 a valid probability distribution) [6]. Inferring the generation time distribution of SARS-CoV-
42 2 is important in order to predict the effects of non-pharmaceutical interventions such as
43 contact tracing and quarantine [7, 8]. In addition, the generation time distribution is widely
44 used in epidemiological models for estimating the reproduction number from case
45 notification data [6, 9–11] and is crucial for understanding the relationship between the
46 reproduction number and the epidemic growth rate [3, 6].

47

48 The SARS-CoV-2 generation time distribution has previously been estimated using data from
49 known infector-infectee transmission pairs [8, 12, 13] or entire clusters of cases [14–16].
50 These studies involved data [8, 14, 17–20] collected between December 2019 and April 2020,
51 almost all from countries in East and Southeast Asia (with the exception of four transmission
52 pairs from Germany and four from Italy in [8]). Evidence from January and February 2020 in
53 China suggested a temporal reduction in the mean generation time due to non-pharmaceutical
54 interventions [15]. Specifically, effective isolation of infected individuals is likely to have
55 reduced the proportion of transmissions occurring when potential infectors were in the later
56 stages of infection, thereby shortening the generation time [15]. Similarly, two other studies
57 found a decrease in the serial interval (the difference between symptom onset times of an
58 infector and infectee) [21] and an increase in the proportion of presymptomatic transmissions
59 [22] in China over the same time period, which can be attributed to symptomatic hosts being
60 isolated increasingly quickly over time.

61

62 Despite estimation of the SARS-CoV-2 generation time in Asia early in the pandemic,
63 relatively little is known about the generation time distribution outside Asia, and whether or
64 not any changes have occurred in the generation time since the early months of the pandemic.
65 In particular, we are aware of only one study in which the generation time was estimated
66 using data from the UK [23]. In that study [23], data describing symptom onset dates for 50
67 infector-infectee pairs, collected by Public Health England between January and March 2020
68 as part of the “First Few Hundred” case protocol [24, 25], were used to infer the generation
69 time distribution. However, since these transmission pairs mostly consisted of international
70 travellers and their household contacts, the authors concluded that their estimates of the
71 generation time may have been biased downwards due to enhanced surveillance and isolation
72 of these cases [23].

73

74 Here, we use data from a household study, conducted between March and November 2020, to
75 estimate the SARS-CoV-2 generation time distribution in the UK under two different
76 underlying transmission models. In the first model (the “independent transmission and
77 symptoms model”), a parsimonious assumption is made that the generation time and the
78 incubation period of the infector are independent (i.e., there is no link between the times at
79 which infectors transmit the virus and the times at which they develop symptoms), as has
80 often been employed in studies in which the SARS-CoV-2 generation time has been
81 estimated [8, 12–14, 23, 26] (Table 1). In the second model (the “mechanistic model”), we
82 use a mechanistic approach in which potential infectors progress through different stages of
83 infection, first becoming infectious before developing symptoms [12]. Infectiousness is
84 therefore explicitly linked to symptoms in the mechanistic model. By fitting separately to
85 data from three different time intervals within the study period, we explore whether or not

86 there was a detectable temporal change in the generation time distribution. Our study
87 represents the first analysis of the SARS-CoV-2 generation time outside Asia conducted
88 using data from after the earliest stages of the pandemic.

89

Study	Location	Time period	Mean generation time (days)	Standard deviation of generation times (days)
Ferretti et al. [8]	Various	December 2019-February 2020	5.0	1.9
Ganyani et al. [14]	Singapore	January-February 2020	5.20 (3.78-6.78)	1.72 (0.91-3.93)
Ganyani et al. [14]	China	January-February 2020	3.95 (3.01-4.91)	1.51 (0.74-2.97)
Hart et al. [12]	Various	December 2019-March 2020	5.57 (5.08-6.09)	2.32 (1.83-2.91)
Ferretti et al. [13]	Various	December 2019-March 2020	5.5	1.8
Challen et al. [23]	UK	January-March 2020	4.8 (4.3-5.41)	1.7 (1.0-2.6)

90

91 **Table 1. Previous SARS-CoV-2 generation time estimates.** Estimates of the mean and standard deviation of
92 the generation time distribution, obtained under the assumption of independent transmission and symptoms.
93 95% credible intervals are shown in brackets where available.

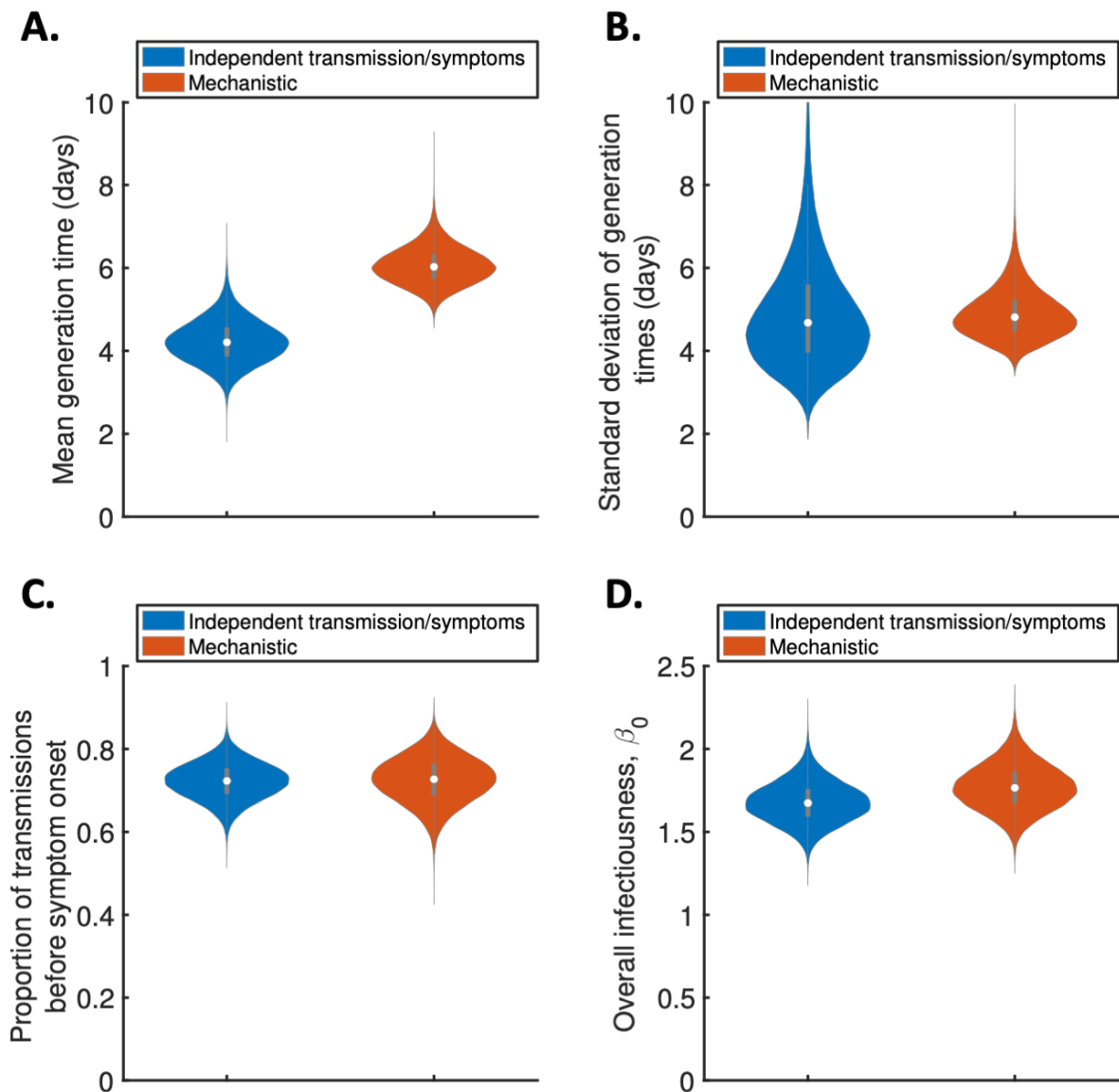
94

95 **RESULTS**

96 Inferring generation times from UK household data

97 We used data augmentation MCMC to fit two models of infectiousness (the independent
98 transmission and symptoms model and the mechanistic model) to data describing individuals'
99 infection status and symptom onset dates in 172 UK households (see Methods). For the two
100 fitted models, we calculated posterior estimates of the mean (Figure 1A) and standard

101 deviation of the generation time distribution (Figure 1B), in addition to the proportion of
102 transmissions occurring prior to symptom onset (among infectors who develop symptoms;
103 Figure 1C) and the overall infectiousness parameter, β_0 (see Methods; Figure 1D). Under the
104 commonly used independent transmission and symptoms model, we obtained a point estimate
105 (posterior mean) of 4.2 days (95% CrI 3.3-5.3 days) for the mean generation time (Figure 1A,
106 blue violin). This value is similar to a previous estimate using data from China in [14], and is
107 slightly lower than estimates for Singapore in [14] and for several countries (predominantly
108 in Asia) in [8] (Table 1), although those estimates lie within our credible interval. On the
109 other hand, our estimated standard deviation of 4.9 days (95% CrI 3.0-8.3 days; Figure 1B,
110 blue violin) is substantially higher than previous estimates (Table 1). Using our mechanistic
111 model, we obtained a higher estimate for the mean generation time of 6.0 days (95% CrI 5.2-
112 7.0 days; Figure 1A, red violin), and a similar estimate for the standard deviation (4.9 days,
113 95% CrI 4.0-6.3 days; Figure 1B, red violin), compared to those predicted by the independent
114 transmission and symptoms model. The two models gave similar posterior distributions for
115 the proportion of transmissions prior to symptom onset (Figure 1C), with point estimate
116 values of model parameters generating estimates of 0.72 (95% CrI 0.63-0.80) for the
117 independent transmission and symptoms model, and 0.73 (95% CrI 0.61-0.83) for the
118 mechanistic model.
119



120

121 **Figure 1. Comparison of posterior predictions.** Violin plots indicating posterior distributions of the mean
122 generation time (A), standard deviation of generation times (B), proportion of transmissions occurring prior to
123 symptom onset (among infectors who develop symptoms; C), and overall infectiousness parameter, β_0 (D). We
124 show results obtained both using a model in which infectiousness is assumed to be independent of when
125 symptoms develop (“independent transmission and symptoms model”, blue), and using the mechanistic
126 approach from [12] in which infectiousness is explicitly linked to symptoms (“mechanistic model”, red).

127

128 Posterior distributions and point estimates (posterior means) for all fitted model parameters
129 are shown in the Supplementary Material (Figures S1-S2 and Tables S2-S3). Since only the
130 likelihood with respect to augmented data was calculated in the MCMC procedure, explicit

131 comparisons of goodness of fit between the models were not readily available. However,
132 comparing model predictions of the difference between successive symptom onset dates with
133 the UK household data indicated that both models provided a similar fit to the data (Figure
134 S3).

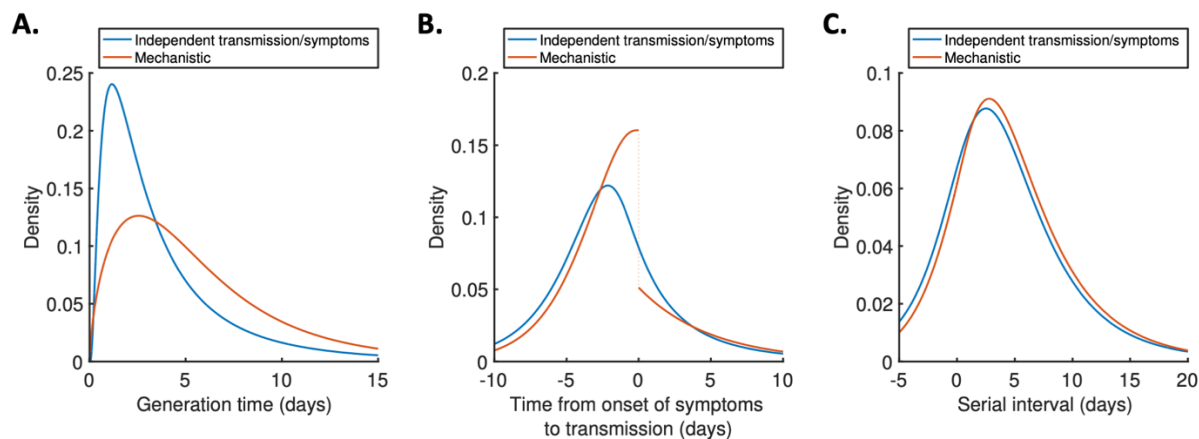
135

136 As described in Methods, the generation time distribution that we considered in Figure 1 (and
137 elsewhere, unless otherwise stated) corresponds to the normalised expected infectiousness
138 profile of a host at each time since infection. However, realised generation times are expected
139 to be shorter for infected individuals in small households compared to those in large
140 households. This is due to depletion of susceptible household members in small households
141 before longer generation times can be attained [6, 27]. As a result, we also predicted the mean
142 and standard deviation of realised generation times within the study households (Figure S4A-
143 B), accounting for the precise distribution of household sizes in the study. For both the
144 independent transmission and symptoms model and the mechanistic model, point estimates
145 for the mean (3.6 days and 4.9 days for the two models, respectively) and standard deviation
146 (3.8 days and 4.1 days) of realised household generation times were lower than those shown
147 in Figure 1. Since transmission then typically occurred earlier in the infector's course of
148 infection than shown in Figure 1, we predicted a higher proportion of presymptomatic
149 transmissions within the study households (Figure S4C) compared to the estimates in Figure
150 1C.

151

152 For both models, we then used point estimates of fitted model parameters to infer the
153 distributions of the generation time (Figure 2A), the time from onset of symptoms to
154 transmission (TOST; Figure 2B) and the serial interval (Figure 2C). The TOST distribution
155 (which characterises the relative expected infectiousness of a host at each time from symptom

156 onset, as opposed to from infection [13, 20, 26, 28, 29]) predicted using the mechanistic
157 model was more concentrated around the time of symptom onset compared to that obtained
158 using the independent transmission and symptoms model (Figure 2B), as was found in [12].
159 In contrast, the estimated serial interval distributions were similar for the two models (Figure
160 2C).

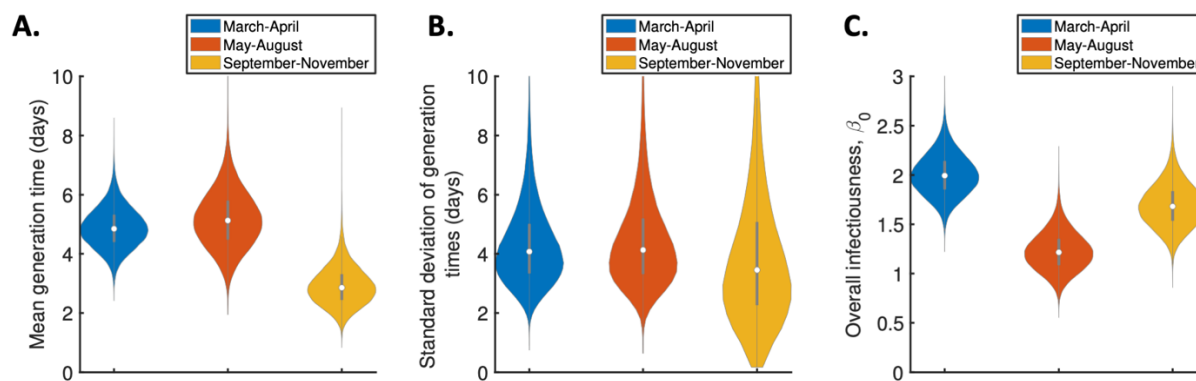


161
162 **Figure 2. Generation time, TOST and serial interval distributions.** Inferred generation time (A), TOST (B)
163 and serial interval (C) distributions for the two models, obtained using point estimate (posterior mean)
164 parameters. Note that the discontinuity in the red curve in Figure 2B occurs because two different transmission
165 rates were fitted for infectors in the presymptomatic and symptomatic periods. The reduction in transmission
166 following symptom onset can be attributed to changes in behaviour in response to symptoms [30].

167

168 Temporal variation in the generation time distribution

169 To explore whether or not the generation time distribution changed during the study period,
170 we separately fitted the independent transmission and symptoms model to the data from
171 households in which the index case was recruited in (i) March-April, (ii) May-August, or (iii)
172 September-November 2020 (Figure 3). We chose these time periods to ensure the numbers of
173 households recruited into the study during each interval were similar (Figure S5).



174

175 **Figure 3. Temporal changes in the generation time distribution.** Violin plots indicating posterior
176 distributions of the mean generation time (A), standard deviation of generation times (B), and overall
177 infectiousness parameter, β_0 (C), for the independent transmission and symptoms model fitted to data from
178 March-April (blue), May-August (red) or September-November 2020 (orange).

179

180 The results shown in Figure 3A suggest a lower mean generation time in September-
181 November (2.9 days, 95% CrI 1.8-4.3 days) compared to earlier months (4.9 days, 95% CrI
182 3.6-6.3 days, for March-April and 5.2 days, 95% CrI 3.4-7.2 days, for May-August). A
183 similar temporal reduction in the mean generation time was found when we instead fitted the
184 mechanistic model to the data from the three time intervals (Figure S6A). To explore the
185 lower estimated generation time for September-November further, we also fitted the
186 independent transmission and symptoms model to the data from each of these months
187 individually (Figure S7). The shorter estimated generation time compared to earlier in the
188 pandemic was consistent in each of the three months (Figure S7A).

189

190 Sensitivity analyses

191 In the independent transmission and symptoms model, we assumed that both the generation
192 time and incubation period followed lognormal distributions. The mean and standard
193 deviation of the generation time distribution were estimated by fitting the model to the
194 household transmission data. In the fitting procedure, we assumed that the incubation period

195 followed a lognormal distribution that was obtained in a previous meta-analysis [31]. In
196 contrast, we assumed in our mechanistic approach that each infection could be decomposed
197 into three gamma distributed stages (latent, presymptomatic infectious and symptomatic
198 infectious), so that the incubation period was also gamma distributed (with the same mean
199 and standard deviation as the lognormal distribution obtained in [31]). An expression for the
200 generation time distribution in the mechanistic model, which did not take a simple parametric
201 form, is given in the Supplementary Text. However, we conducted supplementary analyses in
202 which we instead assumed that either the generation time (Figure S8A-C) or incubation
203 period (Figure S8D-F) in the independent transmission and symptoms model was gamma
204 distributed. In both cases, we obtained similar results to those shown for that model in Figure
205 1.

206

207 In Figure S9, we relaxed the assumption of a fixed incubation period distribution, using the
208 confidence intervals obtained in [31] to account for uncertainty in the incubation period
209 distribution (Figure S9A-B). For both the independent transmission and symptoms model and
210 the mechanistic model, accounting for this uncertainty did not substantially affect posterior
211 estimates of either the mean (Figure S9C) or the standard deviation (Figure S9D) of the
212 generation time distribution.

213

214 In our main analyses, we assumed that household transmission was frequency-dependent, so
215 that the force of infection exerted by an infected household member on each susceptible
216 household member scaled with $1/n$, where n is the household size. In order to explore the
217 robustness of our results to this assumption, we considered alternative possibilities where
218 infectiousness scaled with $n^{-\rho}$, for $\rho = 0$ (density-dependent transmission) and $\rho = 0.5$
219 (Figure S10A-B). We also conducted an analysis in which the dependency, ρ , was estimated

220 alongside other model parameters (Figure S10C). We found that our estimates of the mean
221 and standard deviation of the generation time distribution were robust to the assumed value of
222 ρ (Figure S10A-B). However, when the value ρ was fitted (Figure S10C), we estimated a
223 value of 1.0 (95% CrI 0.6-1.5). This supported our assumption of frequency-dependent
224 transmission, although the credible interval was relatively wide. In addition, we considered
225 the possibility that infectiousness instead scales with $1/(n - 1)$, so that the infectious
226 individual is removed from the scaling in the force of infection $\beta(\tau)$, and again obtained
227 similar estimates of the mean and standard deviation of the generation time distribution
228 compared to those shown in Figure 1 (Figure S10D-E).

229
230 We also considered the sensitivity of our results to the assumed relative infectiousness of
231 asymptomatic infected hosts (Figure S11). In most of our analyses, we assumed that the
232 expected infectiousness of an infected host who remained asymptomatic throughout infection
233 was a factor $\alpha_A = 0.35$ times that of a host who develops symptoms, at each time since
234 infection [31]. However, similar estimates of the mean (Figure S11A) and standard deviation
235 (Figure S11B) of the generation time distribution were obtained when we instead assumed
236 $\alpha_A = 0.1$ or $\alpha_A = 1.27$ (these values corresponded to the lower and upper confidence bounds
237 obtained in [32]).

238
239 Finally, we explored the robustness of our results to the exclusion of hosts with unknown
240 infection status (see Methods), considering the extreme possibilities where these hosts were
241 instead assumed to have either all remained uninfected, or all become infected (Figure S12).
242 Although the estimates of β_0 were affected by this assumption (Figure S12C), the estimated
243 generation time distribution was robust to the assumed infection status of these hosts (Figure
244 S12A-B).

245

246

DISCUSSION

247 In this study, we estimated the generation time distribution of SARS-CoV-2 in the UK by
248 fitting two different models to data describing the infection status and symptom onset dates of
249 individuals in 172 households. The first of these models was predicated on an assumption
250 that transmission and symptoms are independent. While this assumption has often been made
251 in previous studies in which the SARS-CoV-2 generation time has been estimated [8, 14, 23,
252 33, 34], it is not an accurate reflection of the underlying epidemiology [26, 35]. Therefore, we
253 also considered a mechanistic model based on compartmental modelling, which was shown
254 in [12] to provide an improved fit to data from 191 SARS-CoV-2 infector-infectee pairs
255 compared to previously used models that have been used to estimate the generation time.
256 Here, infection times and the order of transmissions within households were unknown,
257 whereas in [12] the direction of transmission was assumed to be known for each infector-
258 infectee pair. For that reason, we used data augmentation MCMC to fit the two models to the
259 household data, in a similar fashion to previous studies of household influenza virus
260 transmission [27, 36, 37].

261

262 Under the model assuming independent transmission and symptoms, we estimated a mean
263 generation time of 4.2 days (95% CrI 3.3-5.3 days) and a standard deviation of 4.9 days (95%
264 CrI 3.0-8.3 days). The estimate of the mean generation time was comparable to previous
265 estimates obtained under this assumption using data from elsewhere [8, 13, 14] (Table 1). On
266 the other hand, while our credible interval for the standard deviation was wide, the estimates
267 obtained in those previous studies [8, 13, 14] all lay below our lower 95% credible limit of
268 3.0 days. One potential cause of this disparity is the difference in isolation policies for
269 symptomatic hosts between countries. In particular, the UK's policy of self-isolation may be

270 expected to lead to a longer-tailed generation time distribution compared to countries with a
271 policy of isolation outside the home, since under home isolation, some within-household
272 transmission is likely to occur even following isolation. Isolation outside the home was
273 commonplace in the East and Southeast Asian countries where the majority of the data
274 underlying the estimates in [8, 13, 14] were collected. Because we used only household
275 transmission data, this effect was likely to be particularly pronounced in our estimates.

276

277 Using the mechanistic model, we predicted a higher mean generation time of 6.0 days (95%
278 CrI 5.2-7.0 days) compared to the value estimated under the assumption of independent
279 transmission and symptoms. On the other hand, the inferred serial intervals for the
280 independent transmission and symptoms model and mechanistic model were similar (Figure
281 2C), with means of 4.2 days and 4.7 days, respectively. Temporal information in our
282 household transmission data consisted mostly of symptom onset dates, with very few hosts
283 testing positive before developing symptoms. Therefore, the variation in estimates of the
284 generation time between the models can be attributed to differences in the assumed
285 relationships between the generation time and serial interval under those models. For the
286 independent transmission and symptoms model, the generation time and serial interval
287 distributions have the same mean, as is commonly assumed to be the case [26]. However, this
288 was not true for the mechanistic model, in which hosts with longer presymptomatic infectious
289 periods generate (on average) a higher number of transmissions. As a result, under the
290 mechanistic model, a randomly chosen infection is likely to arise from an infector with a
291 longer than expected incubation period, thereby leading to a longer generation time than
292 serial interval (an analytical expression for the exact difference between the mean generation
293 time and serial interval for that model is derived in the Supplementary Material).

294

295 Our results did not indicate any clear difference in goodness of fit between the two models
296 (Figure S3). A range of factors should therefore be considered when deciding which of our
297 estimates of epidemiological parameters to use in subsequent analyses. Although any model
298 requires simplifying assumptions to be made, our mechanistic approach allows the standard
299 assumption of independent transmission and symptoms to be relaxed by providing a
300 mechanistic underpinning to the relationship between the times at which individuals display
301 symptoms and become infectious. Furthermore, as described above, this model was shown in
302 [12] to provide a better fit to another SARS-CoV-2 dataset than a model assuming
303 independence between transmission and symptoms (in [12], the simpler setting of
304 transmission pairs rather than households facilitated direct model comparison). On the other
305 hand, the independent transmission and symptoms model has the advantage of producing an
306 estimated generation time distribution with a simple parametric form. The choice of estimates
307 to use may also depend on precisely what the estimates are being used for. For example, the
308 generation time distribution inferred under the assumption of independent transmission and
309 symptoms may be better suited for use in some models for estimating the time-dependent
310 reproduction number, since those models often also implicitly involve the assumption that
311 transmission and symptoms are independent [10]. In contrast, the parameter estimates from
312 our mechanistic approach correspond naturally to parameters in compartmental epidemic
313 models.

314

315 By fitting separately to data from three different intervals within the study period (March-
316 November 2020), we investigated whether or not the generation time distribution in the UK
317 changed as the pandemic progressed. Our results indicate a lower mean generation time in
318 September-November compared to earlier months (Figure 3A). One possible explanation for
319 this is a higher proportion of time spent indoors in colder months leading to an increased

320 household transmission risk, particularly in the early stages of infection before symptoms
321 develop (since symptomatic infected hosts are still likely to self-isolate). While shorter
322 generation times may in general be over-represented in data collected from infector-infectee
323 pairs at times when case numbers are rising [8, 26, 38] (as was the case in September and
324 October [39, 40]), we estimated the mean generation time to be similar in November (when
325 national case numbers were mostly decreasing [39, 40]) compared to September and October
326 (Figure S7).

327

328 Our finding of a temporal decrease in the mean generation time during the study period
329 highlights the importance of obtaining up-to-date generation time estimates specific to the
330 location under study. Should variations in the generation time distribution occur and not be
331 accounted for, estimates of the time-dependent reproduction number may be incorrect [3, 41].
332 Specifically, if the mean generation time is shorter than assumed, then the true value of the
333 time-dependent reproduction number is closer to one than the inferred value [3] (and vice
334 versa).

335

336 The B1.1.7 SARS-CoV-2 variant [42], which has since become the dominant lineage in the
337 UK, was responsible for infections in only two households within our dataset. Therefore,
338 additional data are needed to quantify the impact of the emergence of this variant on the
339 generation time. A shorter mean generation time of the B1.1.7 variant has been hypothesised
340 as one possible contribution to the increased growth rate of cases in the UK in late 2020 and
341 early 2021, but a decrease in the generation time alone was not found to explain trends in UK
342 case data [43, 44]. Conversely, viral load data suggest a longer duration of viral shedding due
343 to infection with B1.1.7 compared to the original variant of SARS-CoV-2 [45]. If higher viral

344 loads lead to increased infectiousness [46–50], this may suggest a longer-tailed generation
345 time distribution for the B.1.1.7 variant.

346

347 One advantage of our approach is that we are able to include the contribution of
348 asymptomatic infected hosts to household transmission chains in our analyses. We showed
349 that our estimated generation time distribution was robust to the assumed relative
350 infectiousness of infected hosts who remained asymptomatic, α_A (Figure S11). Similarly,
351 while we assumed frequency-dependent household transmission in most of our analyses, we
352 found that the exact relationship between the household size and transmission rate had little
353 effect on our estimates of the mean and standard deviation of the generation time distribution
354 (Figure S10). We also considered estimating the exponent governing the dependency of
355 transmission on household size (Figure S10C). This supported our assumption of frequency-
356 dependent transmission. In previous studies of influenza transmission within households,
357 evidence has been found both in favour of [36] and against [51] frequency-dependent
358 transmission.

359

360 Our study has some limitations. Since we used household transmission data in our analyses,
361 the generation time for transmission outside the household may differ from our estimates.
362 Future extensions to our approach may account for the possibility that more than one
363 household member was infected in the same primary infection event or the potential for
364 multiple sequential introductions of the virus into a household [37]. Allowing for multiple
365 introductions may shorten estimates of the generation time, although any effect will
366 dependent significantly on the community prevalence and the number of contacts that
367 household members have with individuals in the community. In contrast, accounting for
368 potential co-primary infections is likely to lead to higher estimates of the generation time.

369 Other further work may include exploring heterogeneity in the generation time distribution
370 between individuals and/or households with different characteristics. This could involve, but
371 is not limited to, estimating the generation time distribution for individuals of different age,
372 sex and ethnicity.

373

374 In summary, we have inferred the SARS-CoV-2 generation time distribution in the UK using
375 household data and two different transmission models. A key output of this research is one of
376 the only estimates of the SARS-CoV-2 generation time outside Asia. Another crucial feature
377 of our analysis is that it was based on data from beyond the first few months of the pandemic.
378 Since this research suggests that the generation time may be changing, continued data
379 collection and analysis is of clear importance.

380

381

METHODS

Data

383 Data were obtained from a household study conducted by Public Health England in 172
384 households (with 603 household members in total) between March and November 2020. In
385 each household, an index case was recruited following a positive PCR test. The following
386 were then recorded for each household member:

- 387 ● The timing and outcome of (up to) two subsequent PCR tests.
- 388 ● The outcome of an antibody test (carried out for 90% of individuals).
- 389 ● Whether or not the household member developed symptoms.
- 390 ● The date of symptom onset (only for symptomatic individuals with a positive PCR or
391 antibody test).

392

393 In the study, all household members who tested positive in either a PCR or antibody test were
394 assumed to have been infected. Conversely, all individuals who tested negative for antibodies
395 (and where the two PCR tests were either negative or were not carried out) were assumed to
396 have remained uninfected, irrespective of symptom status. For 6% of the study cohort, no
397 antibody test was carried out and any PCR tests were negative. These hosts with unknown
398 infection status were excluded from our main analyses (but were counted in the household
399 size), although we also considered the sensitivity of our results to this assumption.

400

401 In two households, at least one household member developed symptoms 55-56 days prior to
402 the symptom onset date of the index case, with no other household members developing
403 symptoms (or returning a positive PCR or antibody test) between these dates. In contrast, the
404 maximum gap between successive symptom onset dates in the remaining households was 25
405 days. Data from these two households were excluded from our analyses, on the basis that the
406 virus was most likely introduced multiple times into these households.

407

408 Models

409 *General modelling framework*

410 We denote the force of infection exerted by a given infected host onto each susceptible
411 member of their household, at time τ days since infection, by $\beta(\tau)$, where we assumed

$$412 \quad \beta(\tau) = (\beta_0/n)f(\tau),$$

413 if the host develops symptoms, and

$$414 \quad \beta(\tau) = \alpha_A(\beta_0/n)f(\tau),$$

415 if the host remains asymptomatic. Here:

- 416 • β_0 represents the overall infectiousness of a host who develops symptoms.

417 • n is the household size. The scaling of $\beta(\tau)$ with $1/n$ corresponds to frequency-
418 dependent transmission, as assumed in [36, 52], although we considered alternative
419 possibilities where transmission was either density-dependent (without the scaling
420 factor $1/n$) or scaled with $1/n^{0.5}$ [51] in sensitivity analyses.

421 • $f(\tau)$ is the generation time distribution.

422 • α_A is the relative infectiousness of infected hosts who remain asymptomatic. We
423 assumed a value of 0.35 [32] in most of our analyses, although we also carried out
424 sensitivity analyses in which different values of α_A were considered.

425

426 Throughout, the normalised infectiousness profile, $f(\tau)$, is referred to as the generation time
427 distribution. However, realised generation times within a household may be shortened due to
428 the depletion of susceptible household members before longer generation times can be
429 attained [6, 27]. For example, if an infected host was equally infectious at two times since
430 infection, $\tau_1 < \tau_2$, then we would have $f(\tau_1) = f(\tau_2)$. However, because the number of
431 susceptible household members may decrease between these two times (i.e., either the host
432 under consideration, or another infected household member, may transmit the virus within the
433 household in the intervening time), then household transmission is in fact more likely to
434 occur in the household at the earlier time, τ_1 , when more susceptibles are available.

435 Therefore, we also predicted the mean and standard deviation of realised generation times
436 within the study households in the Supplementary Material (Figure S4).

437

438 We considered two different models of infectiousness, which are outlined below. Under each
439 model, expressions were derived in [12] for the generation time, time from onset of
440 symptoms to transmission (TOST) and serial interval distributions, in addition to the
441 proportion of transmissions occurring before symptom onset. These expressions are given in

442 the Supplementary Material (other than the generation time distribution and proportion of
443 presymptomatic transmissions for the independent transmission and symptoms model, which
444 are stated below).

445

446 *Independent transmission and symptoms model*

447 In this model, the infectiousness of an infected host (who does not remain asymptomatic
448 throughout infection; asymptomatic infected hosts are considered separately) at a given time
449 since infection, τ days, is assumed to be independent of whether or not the host has yet
450 developed symptoms – i.e., the generation time and incubation period are independent. In this
451 case, the generation time distribution, $f(\tau)$, was assumed to be the probability density
452 function of a lognormal distribution [37] (an alternative case of a gamma distributed
453 generation time is considered in Figure S8). The mean and standard deviation of this
454 distribution, in addition to β_0 , were estimated when we fitted the model to the household
455 transmission data.

456

457 Under the assumption of independent transmission and symptoms, the proportion of
458 transmissions occurring prior to symptom onset (among infectors who develop symptoms) is
459 given by [8, 53]

$$460 \int_0^{\infty} f(\tau)(1 - F_{inc}(\tau))d\tau,$$

461 where F_{inc} is the cumulative distribution function of the incubation period (this was assumed
462 to be known; the exact incubation period distribution we used is given under “Parameter
463 estimation” below).

464

465 *Mechanistic model*

466 Under the mechanistic model [12], infectors who develop symptoms progress through
467 independent latent (E), presymptomatic infectious (P) and symptomatic infectious (I) stages.
468 The duration of each stage was assumed to be gamma distributed, and infectiousness was
469 assumed to be constant during each stage. Under these assumptions, an explicit expression
470 can be derived for the expected infectiousness, $\beta(\tau | \tau_{inc})$, of a host (who develops
471 symptoms) at time since infection τ , conditional on their incubation period τ_{inc} . Details of the
472 mechanistic approach, including the formula for $\beta(\tau | \tau_{inc})$, are provided in the
473 Supplementary Material.

474

475 When we fitted this model to the household transmission data, three model parameters were
476 estimated in addition to β_0 . These parameters correspond to:

- 477 • The ratio between the mean latent (E) period and the mean incubation (combined E
478 and P) period (where the latter was assumed known).
- 479 • The mean symptomatic infectious (I) period.
- 480 • The ratio between the transmission rates when potential infectors are in the P and I
481 stages.

482

483 Likelihood function

484 The models were fitted to the household data using data augmentation Markov chain Monte
485 Carlo (MCMC; see further details below). For a household (of size n) in which n_I household
486 members are infected during the study (of whom n_S develop symptoms and n_A remain
487 asymptomatic throughout infection) and $n_U = n - n_I$ are uninfected, augmented data consist
488 of the entire sequence of infection times of individuals in the household ($t_1 < \dots < t_{n_I}$), as
489 well as the exact symptom onset times ($t_{s,j}$) of each host, j , who develops symptoms.

490

491 When deriving an expression for the likelihood contribution from the household, given these
492 augmented data, we made several simplifying assumptions:

- 493 • The virus was introduced once into each household (i.e., no subsequent infections
494 from the community occurred following the infection of the primary case).
- 495 • No co-primary cases.
- 496 • Potential bias towards more recent infection of the primary host when community
497 prevalence was increasing, or less recent when prevalence was decreasing [8, 26, 38],
498 was neglected.

499

500 We denote the expected infectiousness of household member j , at time τ since infection, by
501 $\beta_j(\tau)$. For the mechanistic model in which transmission and symptoms are not independent,
502 this infectiousness is conditional on the duration of the incubation period, $t_{s,j} - t_j$, if host j
503 developed symptoms. The total (instantaneous) force of infection exerted at time t on each
504 susceptible household member is then

506
$$\lambda(t) = \sum_{j=1}^{n_I} \beta_j(t - t_j),$$

505 where $\beta_j(t - t_j) = 0$ for $t \leq t_j$, and the cumulative force of infection is

507
$$\Lambda(t) = \int_{-\infty}^t \lambda(s) ds.$$

508

509 For $k = 2, \dots, n_I$, conditional on the sequence of infection times up to time t_k , the probability
510 that host k becomes infected at time t_k is given by

511
$$\lambda(t_k) \exp(-\Lambda(t_k)),$$

512 where $\exp(-\Lambda(t_k))$ represents the probability that host k avoided infection up to time t [36,
513 37].

514

515 For $k = n_I + 1, \dots, n$, conditional on the entire sequence of infection times, t_1, \dots, t_{n_I} , the

516 probability that host k remains uninfected is given by $\exp(-\Lambda(\infty))$. In the case of

517 independent transmission and symptoms, we have

518
$$\exp(-\Lambda(\infty)) = \exp(-\beta_0(n_S + \alpha_A n_A)/n),$$

519 whereas for the mechanistic model, $\exp(-\Lambda(\infty))$ instead depends on the incubation periods

520 of those hosts who develop symptoms (see the Supplementary Material).

521

522 The likelihood contribution from the household, $L(\theta)$, where θ is the vector of unknown

523 model parameters, given the augmented data, can therefore be written as

524
$$L(\theta) = \prod_{k=1}^n L_{k,1}(\theta) L_{k,2}(\theta).$$

525 Here, $L_{k,1}(\theta)$ is the contribution to the likelihood from the transmission, or absence of

526 transmission, to host k , i.e.,

527
$$L_{k,1}(\theta) = \begin{cases} 1, & \text{for } k = 1; \\ \lambda(t_k) \exp(-\Lambda(t_k)), & \text{for } k = 2, \dots, n_I; \\ \exp(-\Lambda(\infty)), & \text{for } k = n_I + 1, \dots, n. \end{cases}$$

528 $L_{k,2}(\theta)$ is the contribution from the incubation period of host k (where applicable), i.e.,

529
$$L_{k,2}(\theta) = \begin{cases} f_{inc}(t_{s,k} - t_k), & \text{if host } k \text{ becomes infected and develops symptoms;} \\ 1, & \text{otherwise;} \end{cases}$$

530 where f_{inc} is the probability density function of the incubation period (this was assumed to be

531 known; the exact incubation period distribution we used is given below).

532

533 Each household was assumed to be independent, so that the overall likelihood was given by

534 the product of the contributions from each household.

535

536 Parameter estimation

537 *Incubation period*

538 For the independent transmission and symptoms model, we assumed a lognormal incubation
539 period distribution with mean 5.8 days and standard deviation 3.1 days [31]. For the
540 mechanistic model, we assumed a gamma distributed incubation period with the same mean
541 and standard deviation; this was for mathematical convenience, since the incubation period
542 was decomposed into the sum of independent gamma distributed latent and presymptomatic
543 infectious periods. Results for the independent transmission and symptoms model using a
544 gamma distributed incubation period are shown in Figure S8, and we account for uncertainty
545 in the exact parameters of the incubation period distribution in Figure S9.

546

547 *Parameter fitting procedure*

548 Unknown model parameters were estimated using data augmentation MCMC. The observed
549 data comprised information about whether or not individuals were ever infected and/or
550 displayed symptoms, symptom onset dates, and for some individuals an upper bound on the
551 infection time (corresponding to the date of a positive PCR test). These data were augmented
552 with precise times of infection and symptom onset (where applicable) for each host. No prior
553 assumptions were made about the order of transmissions within the household.

554

555 Below, we outline the parameter fitting procedure that we used for the independent
556 transmission and symptoms model. The procedure used for the mechanistic model was
557 similar and is described in the Supplementary Material.

558

559 Lognormal priors were assumed for fitted model parameters (these parameters were the mean
560 and standard deviation of the generation time distribution, in addition to the overall

561 infectiousness, β_0). The priors for the mean and standard deviation of the generation time
562 distribution had medians of 5 days and 2 days, respectively (these choices were informed by
563 previous estimates of the SARS-CoV-2 generation time distribution [8, 13, 14]), and were
564 chosen to ensure a prior probability of only 0.025 that these parameters exceeded very high
565 values of 10 days and 7 days, respectively. The exact priors we used are detailed in Table S2.

566

567 We denote the vector of model parameters by θ , and the augmented data by

$$568 \quad \mathbf{t} = (\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(M)}),$$

569 where $\mathbf{t}^{(m)}$ represents the augmented data from household $m = 1, \dots, M$, and M is the total
570 number of households. We write the (overall) likelihood as

$$571 \quad L(\theta; \mathbf{t}) = \prod_{m=1}^M L^{(m)}(\theta; \mathbf{t}^{(m)}),$$

572 where the likelihood contribution, $L^{(m)}(\theta; \mathbf{t}^{(m)})$, from each household, m , was computed as
573 described in the previous section, and denote the prior density of θ by $\pi(\theta)$.

574

575 In each step of the chain, we carried out (in turn) one of the following:

- 576 1. Propose new values for the vector of model parameters, θ , using independent
577 normal proposal distributions for each parameter (around the corresponding
578 parameter values in the previous step of the chain). Accept the proposed
579 parameters, θ_{prop} , with probability

$$580 \quad \min\left(\frac{L(\theta_{prop}; \mathbf{t})\pi(\theta_{prop})}{L(\theta_{old}; \mathbf{t})\pi(\theta_{old})}, 1\right),$$

581 where θ_{old} denotes the vector of parameter values from the previous step of
582 the chain, and where the augmented data, \mathbf{t} , remain unchanged in this step.

- 583 2. Propose new values for the precise symptom onset times of each symptomatic
584 infected host, using independent uniform proposal distributions (within the

585 day of symptom of onset for each host). For each household, m , accept the
586 proposed augmented data, $\mathbf{t}_{prop}^{(m)}$, from that household with probability

587
$$\min\left(\frac{L^{(m)}(\theta; \mathbf{t}_{prop}^{(m)})}{L^{(m)}(\theta; \mathbf{t}_{old}^{(m)}), 1\right),$$

588 where $\mathbf{t}_{old}^{(m)}$ denotes the corresponding augmented data from the previous step
589 of the chain, and where the model parameters, θ , remain unchanged in this
590 step (i.e., proposed times are accepted/rejected independently for each
591 household, according to the likelihood contribution from that household).

592 3. Propose new values for the infection time of one randomly chosen
593 symptomatic infected host in each household (in households where there was
594 at least one), using independent normal proposal distributions (around the
595 equivalent times in the previous step of the chain). For each household, m ,
596 accept the proposed augmented data, $\mathbf{t}_{prop}^{(m)}$, from that household with
597 probability

598
$$\min\left(\frac{L^{(m)}(\theta; \mathbf{t}_{prop}^{(m)})}{L^{(m)}(\theta; \mathbf{t}_{old}^{(m)}), 1\right).$$

599 4. Propose new values for the infection time of one randomly chosen
600 asymptomatic infected host in each household (in households where there was
601 at least one), using independent normal proposal distributions (around the
602 equivalent times in the previous step of the chain). For each household, m ,
603 accept the proposed augmented data, $\mathbf{t}_{prop}^{(m)}$, from that household with
604 probability

605
$$\min\left(\frac{L^{(m)}(\theta; \mathbf{t}_{prop}^{(m)})}{L^{(m)}(\theta; \mathbf{t}_{old}^{(m)}), 1\right).$$

606

607 The chain was run for 10,000,000 iterations; the first 2,000,000 iterations were discarded as
608 burn-in. Posteriors were obtained by recording only every 100 iterations of the chain.

609

610

611 **Governance statement:**

612 The household study was approved by the PHE Research Ethics and Governance Group as
613 part of the portfolio of PHE's enhanced surveillance activities in response to the pandemic.

614

615 **Acknowledgments:**

616 Thanks to Pauline Waight, who managed the data for the household study, and to the PHE
617 staff who collected the data and tested the PCR and serum samples. Thanks also to Rob
618 Challen, Julia Gog, Matt Keeling and other members of the Juniper Consortium
619 (www.maths.org/juniper/) for helpful comments about this research.

620

621 **Competing interests:**

622 AE received a research grant from Taisho Pharmaceutical Co., Ltd. All the other authors
623 declare no competing interests.

624

625 **References:**

626 [1] Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. OUP
627 Oxford, 1992.

628 [2] Diekmann O, Heesterbeek JAP. *Mathematical Epidemiology of Infectious Diseases:
629 Model Building, Analysis and Interpretation*. John Wiley & Sons, 2000.

- 630 [3] Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth
631 rates and reproductive numbers. *Proc R Soc B Biol Sci* 2007; 274: 599–604.
- 632 [4] Svensson Å. A note on generation times in epidemic models. *Math Biosci* 2007; 208:
633 300–311.
- 634 [5] Griffin J, Casey M, Collins Á, et al. Rapid review of available evidence on the serial
635 interval and generation time of COVID-19. *BMJ Open* 2020; 10: e040263.
- 636 [6] Fraser C. Estimating Individual and Household Reproduction Numbers in an Emerging
637 Epidemic. *PLOS ONE* 2007; 2: e758.
- 638 [7] Ashcroft P, Lehtinen S, Angst DC, et al. Quantifying the impact of quarantine duration on
639 COVID-19 transmission. *eLife* 2021; 10: e63704.
- 640 [8] Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests
641 epidemic control with digital contact tracing. *Science* 2020; 368: eabb6936.
- 642 [9] Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the
643 effective reproductive number, Rt. *PLOS Comput Biol* 2020; 16: e1008409.
- 644 [10] Abbott S, Hellewell J, Thompson RN, et al. Estimating the time-varying reproduction
645 number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open*
646 *Res* 2020; 5: 112.
- 647 [11] Thompson RN, Hollingsworth TD, Isham V, et al. Key questions for modelling COVID-
648 19 exit strategies. *Proc R Soc B Biol Sci* 2020; 287: 20201405.
- 649 [12] Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID-19
650 symptom onset highlights the importance of continued contact tracing. *eLife* 2021; 10:
651 e65534.
- 652 [13] Ferretti L, Ledda A, Wymant C, et al. The timing of COVID-19 transmission. *medRxiv*
653 2020; 2020.09.04.20188516.
- 654 [14] Ganyani T, Kremer C, Chen D, et al. Estimating the generation interval for coronavirus
655 disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*
656 2020; 25: 2000257.
- 657 [15] Sun K, Wang W, Gao L, et al. Transmission heterogeneities, kinetics, and controllability
658 of SARS-CoV-2. *Science* 2021; 371: eabe2424.
- 659 [16] Hu S, Wang W, Wang Y, et al. Infectivity, susceptibility, and risk factors associated with
660 SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *Nat*
661 *Commun* 2021; 12: 1533.
- 662 [17] Cheng H-Y, Jian S-W, Liu D-P, et al. Contact tracing assessment of COVID-19
663 transmission dynamics in Taiwan and risk at different exposure periods before and
664 after symptom onset. *JAMA Intern Med* 2020; 180: 1156.
- 665 [18] Xia W, Liao J, Li C, et al. Transmission of corona virus disease 2019 during the
666 incubation period may lead to a quarantine loophole. *medRxiv* 2020;
667 2020.03.06.20031955.

- 668 [19] Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission
669 dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive
670 and modelling study. *Lancet Infect Dis* 2020; 20: 793–802.
- 671 [20] He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility
672 of COVID-19. *Nat Med* 2020; 26: 672–675.
- 673 [21] Ali ST, Wang L, Lau EHY, et al. Serial interval of SARS-CoV-2 was shortened over time
674 by nonpharmaceutical interventions. *Science* 2020; 369: 1106–1109.
- 675 [22] Bushman M, Worby C, Chang H-H, et al. Transmission of SARS-CoV-2 before and
676 after symptom onset: impact of nonpharmaceutical interventions in China. *Eur J*
677 *Epidemiol* 2021; 36: 429–439.
- 678 [23] Challen R, Brooks-Pollock E, Tsaneva-Atanasova K, et al. Meta-analysis of the SARS-
679 CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19
680 reproduction number. *medRxiv* 2020; 2020.11.17.20231548.
- 681 [24] Public Health England. "The First Few Hundred (FF100)" Enhanced Case and Contact
682 Protocol v12. 2020.
- 683 [25] Boddington NL, Charlett A, Elgohari S, et al. COVID-19 in Great Britain:
684 epidemiological and clinical characteristics of the first few hundred (FF100) cases: a
685 descriptive case series and case control analysis. *medRxiv* 2020;
686 2020.05.18.20086157.
- 687 [26] Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval,
688 infectiousness profile and generation time. *J R Soc Interface* 2021; 18: 20200756.
- 689 [27] Cauchemez S, Donnelly CA, Reed C, et al. Household Transmission of 2009 Pandemic
690 Influenza A (H1N1) Virus in the United States. *N Engl J Med* 2009; 361: 2619–2627.
- 691 [28] Ashcroft P, Huisman JS, Lehtinen S, et al. COVID-19 infectivity profile correction. *Swiss*
692 *Med Wkly* 2020; 150: w20336.
- 693 [29] Wells CR, Townsend JP, Pandey A, et al. Optimal COVID-19 quarantine and testing
694 strategies. *Nat Commun* 2021; 12: 356.
- 695 [30] Manfredi P, D'Onofrio A (eds). *Modeling the Interplay Between Human Behavior and*
696 *the Spread of Infectious Diseases*. New York: Springer-Verlag, 2013.
- 697 [31] McAloon C, Collins Á, Hunt K, et al. Incubation period of COVID-19: a rapid systematic
698 review and meta-analysis of observational research. *BMJ Open* 2020; 10: e039652.
- 699 [32] Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission
700 potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living
701 systematic review and meta-analysis. *PLOS Med* 2020; 17: e1003346.
- 702 [33] Deng Y, You C, Liu Y, et al. Estimation of incubation period and generation time based
703 on observed length-biased epidemic cohort with censoring for COVID-19 outbreak in
704 China. *Biometrics*; 1–13.
- 705 [34] Knight J, Mishra S. Estimating effective reproduction number using generation time
706 versus serial interval, with application to covid-19 in the Greater Toronto Area,
707 Canada. *Infect Dis Model* 2020; 5: 889–896.

- 708 [35] Bacallado S, Zhao Q, Ju N. Letter to the editor: Generation interval for COVID-19
709 based on symptom onset data. *Eurosurveillance* 2020; 25: 2001381.
- 710 [36] Cauchemez S, Carrat F, Viboud C, et al. A Bayesian MCMC approach to study
711 transmission of influenza: application to household longitudinal data. *Stat Med* 2004;
712 23: 3469–3487.
- 713 [37] Ferguson NM, Cummings DAT, Cauchemez S, et al. Strategies for containing an
714 emerging influenza pandemic in Southeast Asia. *Nature* 2005; 437: 209–214.
- 715 [38] Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. *J*
716 *R Soc Interface* 2019; 16: 20180670.
- 717 [39] Knock ES, Whittles LK, Lees JA, et al. The 2020 SARS-CoV-2 epidemic in England:
718 key epidemiological drivers and impact of interventions. *medRxiv* 2021;
719 2021.01.11.21249564.
- 720 [40] Pouwels KB, House T, Pritchard E, et al. Community prevalence of SARS-CoV-2 in
721 England from April to November, 2020: results from the ONS Coronavirus Infection
722 Survey. *Lancet Public Health* 2021; 6: e30–e38.
- 723 [41] Park SW, Bolker BM, Funk S, et al. Roles of generation-interval distributions in shaping
724 relative epidemic strength, speed, and control of new SARS-CoV-2 variants. *medRxiv*
725 2021; 2021.05.03.21256545.
- 726 [42] Public Health England. *SARS-CoV-2 variants of concern and variants under*
727 *investigation in England*. Technical briefing 7, 2021.
- 728 [43] Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of
729 SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; eabg3055.
- 730 [44] Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in
731 England: Insights from linking epidemiological and genetic data. *medRxiv* 2021;
732 2020.12.30.20249034.
- 733 [45] Kissler SM, Fauver JR, Mack C, et al. Densely sampled viral trajectories suggest longer
734 duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2.
735 *medRxiv* 2021; 2021.02.16.21251535.
- 736 [46] Cevik M, Kuppalli K, Kindrachuk J, et al. Virology, transmission, and pathogenesis of
737 SARS-CoV-2. *BMJ* 2020; 371: m3862.
- 738 [47] Hart WS, Maini PK, Yates CA, et al. A theoretical framework for transitioning from
739 patient-level to population-scale epidemiological dynamics: influenza A as a case
740 study. *J R Soc Interface* 2020; 17: 20200230.
- 741 [48] Ke R, Zitzmann C, Ribeiro RM, et al. Kinetics of SARS-CoV-2 infection in the human
742 upper and lower respiratory tracts and their relationship with infectiousness. *medRxiv*
743 2020; 2020.09.25.20201772.
- 744 [49] Marc A, Kerioui M, Blanquart F, et al. Quantifying the relationship between SARS-CoV-
745 2 viral load and infectiousness. *medRxiv* 2021; 2021.05.07.21256341.
- 746 [50] Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-
747 CoV-2 infection course. *Science* 2021; eabi5273.

- 748 [51] Endo A, Uchida M, Kucharski AJ, et al. Fine-scale family structure shapes influenza
749 transmission risk in households: Insights from primary schools in Matsumoto city,
750 2014/15. *PLOS Comput Biol* 2019; 15: e1007589.
- 751 [52] Cauchemez S, Ferguson NM, Fox A, et al. Determinants of influenza transmission in
752 South East Asia: insights from a household cohort study in Vietnam. *PLOS Pathog*
753 2014; 10: e1004310.
- 754 [53] Fraser C, Riley S, Anderson RM, et al. Factors that make an infectious disease
755 outbreak controllable. *Proc Natl Acad Sci* 2004; 101: 6146–6151.
- 756