

1           **High infectiousness immediately before COVID-19 symptom onset**  
2                           **highlights the importance of contact tracing**

3                           W.S. Hart<sup>1\*</sup>, P.K. Maini<sup>1</sup>, R.N. Thompson<sup>1,2</sup>.

4   **Affiliations:**

5   <sup>1</sup>Mathematical Institute, University of Oxford.

6   <sup>2</sup>Christ Church, University of Oxford.

7   \*Correspondence to: [william.hart@keble.ox.ac.uk](mailto:william.hart@keble.ox.ac.uk).

8  
9   **Abstract:**

10   Understanding changes in infectiousness during COVID-19 infections is critical to assess the  
11   effectiveness of public health measures such as contact tracing. Data from known source-  
12   recipient pairs can be used to estimate the average infectiousness profile of infected  
13   individuals, and to evaluate the proportion of presymptomatic transmissions. Here, we infer  
14   the infectiousness profile of COVID-19 infections using a mechanistic approach, and show  
15   that this method provides an improved fit to data from source-recipient pairs compared to  
16   previous studies. Our results indicate a higher proportion of presymptomatic transmissions  
17   than previously thought, with many transmissions occurring shortly before symptom onset.  
18   High infectiousness immediately prior to symptom onset highlights the importance of contact  
19   tracing, even if contacts from a short time window before symptom onset alone are traced.

20

21

22   NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 **Main text:**

2 The precise proportion of SARS-CoV-2 transmissions arising from non-symptomatic (either  
3 presymptomatic or asymptomatic) infectors remains uncertain (1, 2). Statistical models can  
4 be used to assess the relative contributions of presymptomatic and symptomatic transmission  
5 using data from known source-recipient transmission pairs (3–6). The distributions of three  
6 important epidemiological time periods – the generation time (i.e., the difference between the  
7 infection times of the source and recipient) (3, 4, 7, 8), the time from onset of symptoms to  
8 transmission (TOST) (4, 9, 10), and the serial interval (i.e., the difference between the  
9 symptom onset times of the source and recipient) (4, 11) – can also be inferred. The  
10 generation time and TOST distributions also indicate the average infectiousness profile of a  
11 host at each time since infection and time since symptom onset, respectively (9, 12). This is  
12 important for assessing the effectiveness of public health measures such as contact tracing (3)  
13 and for deciding the duration of quarantine/isolation periods (13). Estimates of the generation  
14 time distribution of SARS-CoV-2 obtained from transmission pair data have typically  
15 involved an assumption that the infectiousness of a host is independent of their symptom  
16 status (3, 7, 8, 14, 15). However, such an assumption may both be unjustified (15, 16) and  
17 lead to a poor fit to data (4).

18

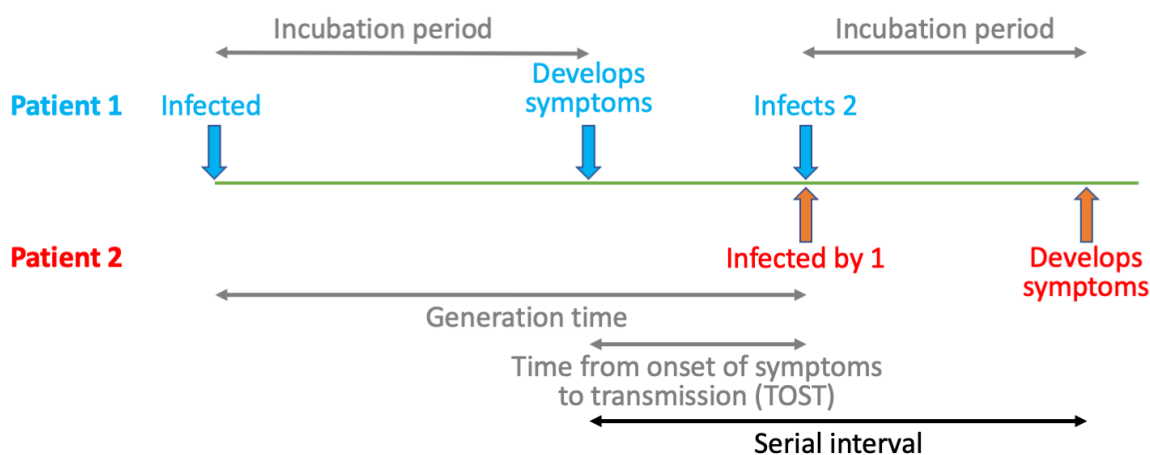
19 Here, we develop a mechanistic approach for inference from transmission pair data. Our  
20 method provides an improved fit to data from SARS-CoV-2 transmission pairs compared to  
21 previously used approaches, namely: (i) a model assuming that transmission and symptoms  
22 are independent (3, 7, 8, 14), and (ii) a previous statistical method in which this assumption is  
23 relaxed (4). We use the fitted models to infer the distributions of epidemiological time  
24 intervals and the contribution of presymptomatic infectious individuals to transmission.  
25 According to our best-fitting model, we find that the predicted proportion of presymptomatic

1 transmissions is higher than estimated using standard approaches, with a substantial  
2 proportion of transmissions occurring in a short time window prior to symptom onset.  
3 Finally, we consider the implications of our results for contact tracing.

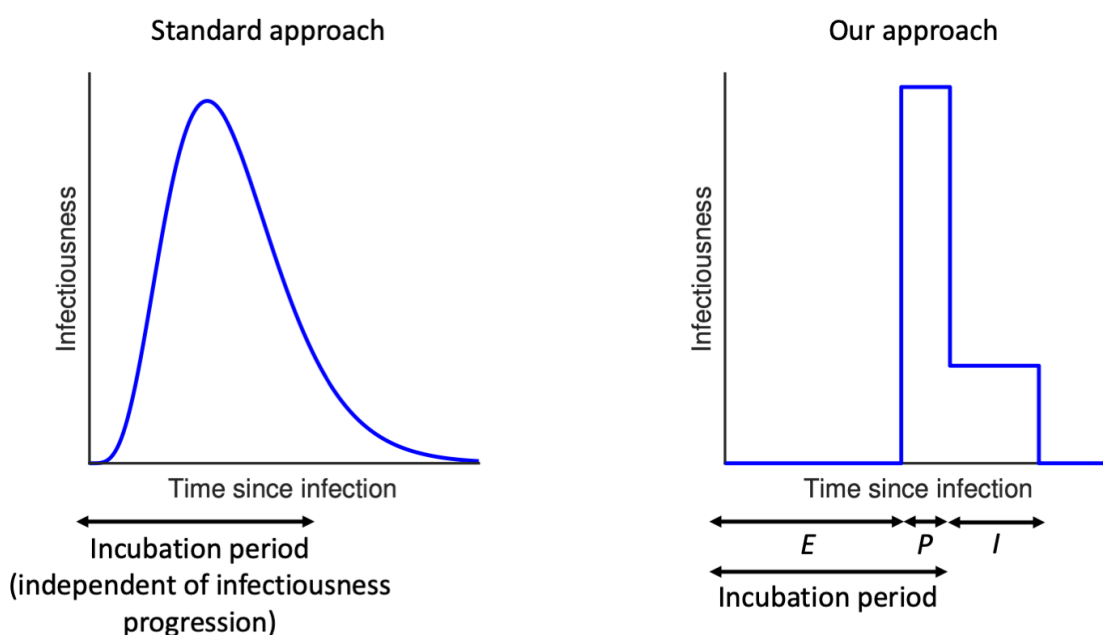
4

5 Transmission pair data (Fig. 1A) generally comprise symptom onset dates for known source-  
6 recipient pairs. These data may be supplemented with partial information about infection  
7 times, consisting of a range of possible exposure dates for the source and/or recipient (3).  
8 While the serial interval for each pair can be calculated directly from the data (with some  
9 uncertainty, given the unknown precise times of symptom appearance on the onset dates  
10 (*I7*)), other time intervals, including the generation time and TOST (which is negative for  
11 presymptomatic transmissions), are unobserved. In almost all previous approaches that have  
12 been used to estimate the generation time distribution of SARS-CoV-2 from transmission  
13 pair data (Fig. 1B, left panel), the infectiousness of the source at a given time since infection  
14 is assumed to be independent of their incubation period (3, 7, 8, *I4*). In contrast, in our  
15 mechanistic approach (Fig. 1B, right panel), which is based on compartmental modelling,  
16 each infected host passes through three stages of infection – latent (*E*), presymptomatic  
17 infectious (*P*), and symptomatic infectious (*I*). Infectiousness is assumed to be constant  
18 during each stage but may vary between presymptomatic and symptomatic infectious hosts.

## A. Observed and unobserved time intervals in transmission pair data



## B. Models of infectiousness



1

2 **Fig. 1. Schematic illustrating time intervals in data from source-recipient pairs and approaches for**

3 **inference from transmission pair data.** A. Important epidemiological time periods in transmission pair data,

4 with time intervals that are not observed directly in grey. B. Assumptions made about the relationship between

5 infectiousness and symptoms within individuals. In standard approaches (left panel), the infectiousness of a host

6 at a given time since infection is independent of their incubation period. In our approach (right panel), we

7 assume that individuals are not infectious during the latent ( $E$ ) period, and that infectiousness may vary between

8 the presymptomatic infectious ( $P$ ) and symptomatic infectious ( $I$ ) periods, for example due to changing

9 behaviour in response to symptoms (*18*).

10

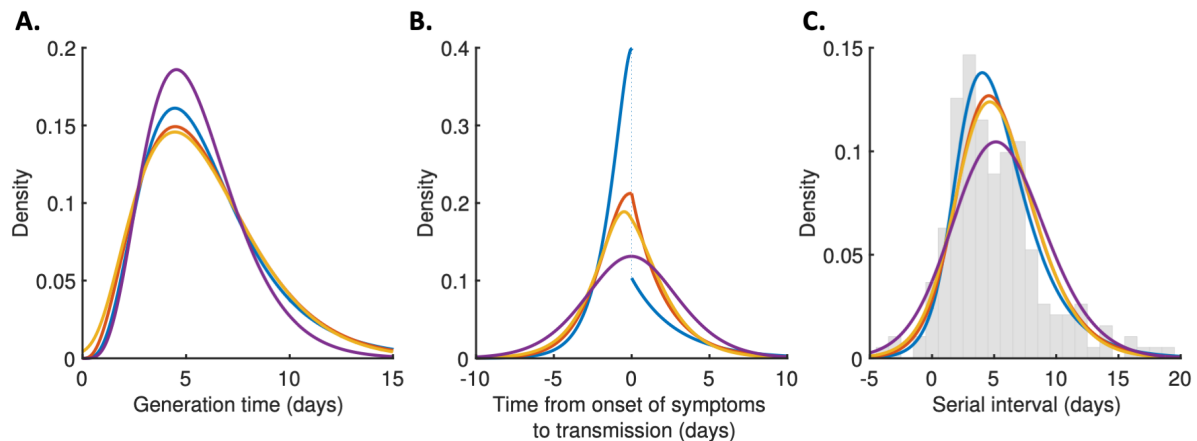
11 We considered four different models of infectiousness:

- 1        1. Our mechanistic approach, with the relative infectiousness levels for presymptomatic
- 2            ( $P$ ) and symptomatic ( $I$ ) infectious hosts estimated from the transmission pair data
- 3            (the “variable infectiousness model”).
- 4        2. Our mechanistic approach, with identical infectiousness levels for presymptomatic
- 5            ( $P$ ) and symptomatic ( $I$ ) infectious hosts (the “constant infectiousness model”).
- 6        3. The best-fitting model from (4) (the “Ferretti model”).
- 7        4. The standard approach (3, 8) in which infectiousness is independent of symptom
- 8            status (the “independent transmission and symptoms model”).

9    Following (4), we fitted each model to data from 191 SARS-CoV-2 transmission pairs  
10    obtained by combining five different datasets (3, 9, 19–21). To account for uncertainty in the  
11    precise times of symptom appearance within the day of onset for the source and recipient  
12    (22), we used data augmentation Markov chain Monte Carlo (MCMC). The best fit to the  
13    data was obtained using our mechanistic approach in which infectiousness varies between  
14    presymptomatic infectious and symptomatic hosts (the variable infectiousness model;  $\Delta\text{AIC}$   
15    = 0). The constant infectiousness model gave the next best fit ( $\Delta\text{AIC} = 4.5$ ), followed by the  
16    Ferretti model ( $\Delta\text{AIC} = 8.3$ ). Finally, the standard assumption of independent transmission  
17    and symptoms led to a significantly worse fit compared to the other models ( $\Delta\text{AIC} = 44.6$ ).

18  
19    For each model, we calculated the predicted distributions of the generation time (Fig. 2A),  
20    TOST (Fig. 2B) and serial interval (Fig. 2C) under the set of model parameters that gave the  
21    best fit to the data. The empirical serial interval distribution is also plotted in Fig. 2C, for  
22    visual confirmation of the goodness of fit of the different models. We found that the  
23    variability in the generation time between individuals was lower for the independent  
24    transmission and symptoms model compared to the other three models (Fig. 2A). On the  
25    other hand, the TOST distribution was most concentrated around the time of symptom onset

1 for the best-fitting variable infectiousness model, and least concentrated for the independent  
2 transmission and symptoms model (Fig. 2B). In our best-fitting model, infectiousness was  
3 found to decrease immediately after onset, likely due to behavioural factors that reduce the  
4 transmission risk following symptom onset (18).



5

6 **Fig. 2. Distributions of epidemiological time intervals estimated by fitting different models to data from**

7 **191 SARS-CoV-2 transmission pairs.** A. Generation time, indicating the relative expected infectiousness of a

8 host at each time since infection. B. Time from onset of symptoms to transmission (TOST), indicating the

9 relative expected infectiousness of a host at each time since symptom onset. C. Serial interval, indicating the

10 periods between sources and recipients developing symptoms. In panel C, the empirical serial interval

11 distribution from the transmission pair data is shown as grey bars. In addition, discretised versions of the serial

12 interval distributions, calculated using the method in (23), are shown in the Supplementary Materials (Fig. S1).

13 In all panels, lines represent: variable infectiousness model (blue), constant infectiousness model (red), Ferretti

14 model (orange), and independent transmission and symptoms model (purple).

15

16 Using the posterior distributions of model parameters that were obtained when we fitted the

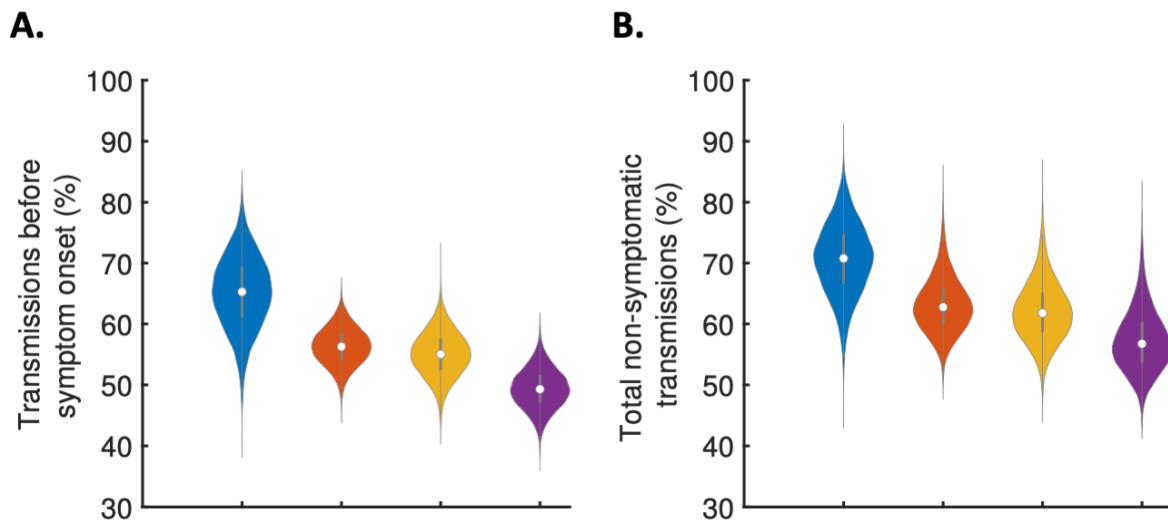
17 models to data, we calculated the posterior distribution of the proportion of transmissions

18 occurring prior to symptom onset for each model (Fig. 3A). The median (95% CI)

19 presymptomatic proportion was 0.65 (0.53-0.76), 0.56 (0.50-0.62), 0.55 (0.48-0.62), and 0.49

20 (0.43-0.56) under the variable infectiousness model, constant infectiousness model, Ferretti

1 model, and independent transmission and symptoms model, respectively. Our central  
2 estimate of 65% of transmissions occurring prior to symptom onset using our best-fitting  
3 model is higher than estimated in previous studies that have fitted statistical models of  
4 infectiousness using transmission pair data (3, 4, 9) – for example, estimates of 37% and 55%  
5 were obtained under an assumption of independent transmission and symptoms in (3).



6  
7 **Fig. 3. The contribution of non-symptomatic infectious individuals to transmission.** A. Violin plots  
8 indicating posterior distributions for the proportion of transmissions occurring prior to symptom onset for  
9 individuals who develop symptoms (i.e., neglecting transmissions from individuals who remain asymptomatic  
10 throughout infection) for the different models. B. Posterior distributions of the total proportion of non-  
11 symptomatic transmissions, accounting for transmissions from asymptomatic infectious individuals, for the  
12 different models. In both panels, violins represent: variable infectiousness model (blue), constant infectiousness  
13 model (red), Ferretti model (orange), and independent transmission and symptoms model (purple).

14  
15 The estimates in Fig. 3A describe the proportion of transmissions that occur prior to symptom  
16 onset, and therefore apply only to individuals who go on to develop symptoms. However,  
17 these estimates can be combined with the results of a previous study (1) in which the extent  
18 of asymptomatic transmission (i.e., transmissions from individuals who never display  
19 symptoms) was characterised (Fig. S2), to obtain estimates for the total proportion of non-

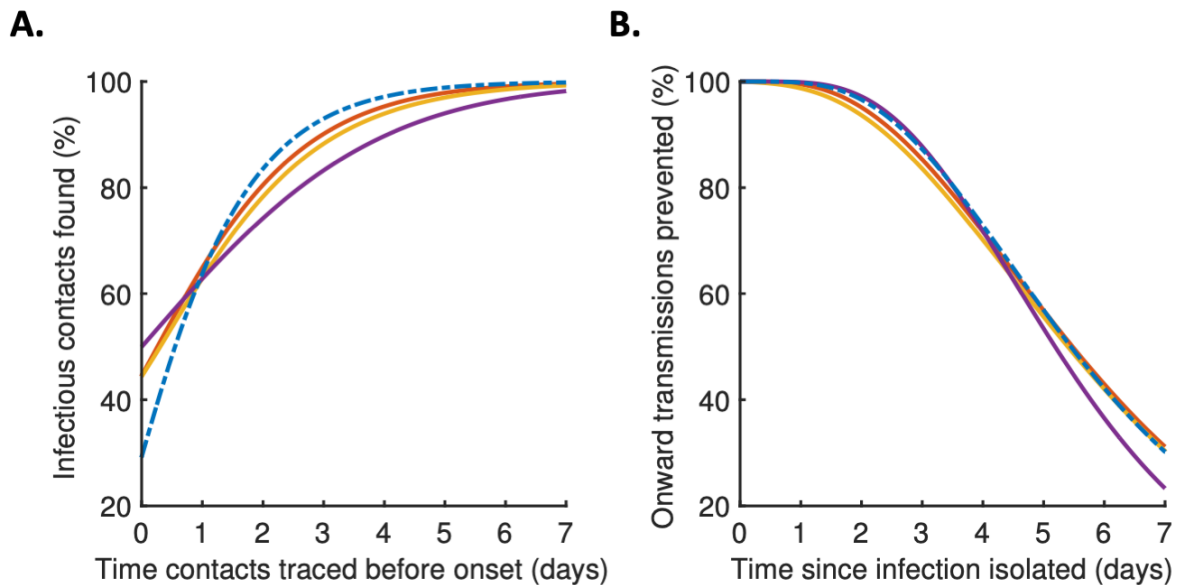
1 symptomatic (i.e., either presymptomatic or asymptomatic) transmissions in an entire  
2 population of hosts for the different models (Fig. 3B). Again, the non-symptomatic  
3 proportion was highest for the variable infectiousness model, and lowest for the independent  
4 transmission and symptoms model.

5

6 Finally, we considered the implications of these results for contact tracing (Fig. 4). In Fig.  
7 4A, we show the proportion of infectious contacts of hosts who go on to develop symptoms  
8 that are identified if contacts are traced up to different times before the symptom onset time  
9 of the index case (i.e., for different contact elicitation windows). Long duration contact  
10 elicitation windows are impractical and place significant strain on contact tracing systems,  
11 leading to contact elicitation windows of two days being used in countries such as the UK  
12 (24) and USA (25). In the best-fitting variable infectiousness model, 84% of infectious  
13 contacts are estimated to be identified when tracing up to 2 days before symptom onset (Fig.  
14 4A, blue dashed). This is due to the clustering of transmission events around the symptom  
15 onset time (*cf.* Fig. 2B) and compares to a lower estimate of only 74% if the standard  
16 assumption of independence between transmission and symptoms is made (Fig. 4A, purple).  
17 We also explored the effect of the timing of isolation of infected individuals identified  
18 through contact tracing, and estimated the reduction in onwards transmissions from infected  
19 contacts if contact tracing is conducted quickly (Fig. 4B). Compared to the best-fitting  
20 variable infectiousness model, the standard independent transmission and symptoms model  
21 estimated a similar reduction in transmission when a host was isolated within 4 days of  
22 exposure, but under-predicted the efficacy of isolation later in infection. We assumed that  
23 contact identification and isolation is perfectly effective in this analysis, but we also explored  
24 the sensitivity of our results to this assumption in the Supplementary Materials. In each case



1 that we considered, under the best-fitting model a two-day contact elicitation window is more  
2 effective than estimated using existing approaches (Fig. S3).



3  
4 **Fig. 4. Implications for contact tracing.** A. Effect of the contact elicitation window: the proportion of  
5 infectious contacts found for different times up to which contacts are traced before the symptom onset time of  
6 the index case. B. Effect of the timing of isolation: the proportion of onward transmissions prevented through  
7 isolation, for different time periods between exposure and isolation. In both panels, lines represent: variable  
8 infectiousness model (blue dashed), constant infectiousness model (red), Ferretti model (orange), and  
9 independent transmission and symptoms model (purple).

10

11 As we have shown, our mechanistic approach provides an improved fit to transmission pair  
12 data, and predicts a higher proportion of transmissions occurring in a short time window  
13 before symptom onset, compared to previous approaches. These conclusions are robust to the  
14 exact incubation period distribution that we assumed (26) when fitting the different models to  
15 transmission pair data (Fig. S4). Our best-fitting model outperforms a model predicated on a  
16 critical assumption – that infectiousness is independent of symptom status – which underlies  
17 most previous studies in which the generation time distribution of COVID-19 is estimated (3,  
18 7, 8, 14). That assumption neglects potential relationships between symptoms and viral

1 shedding, as well as behavioural changes in response to symptom onset (18). Some  
2 alternative assumptions have also been considered when fitting to transmission pair data,  
3 such as the possibility that infectiousness depends only on the time since symptom onset,  
4 independent of the time of infection (4, 27). If the serial interval is always positive, which is  
5 not always the case for COVID-19 (11), this is equivalent to assuming that the serial interval  
6 and generation time distributions are identical (15, 23, 28). In a previous study (4), another  
7 model (the Ferretti model) was developed in which a host's infectiousness could depend on  
8 both the time since infection and the time since symptom onset, and was found to outperform  
9 models in which their infectiousness depends on either one of these two times alone.  
10 However, as we have demonstrated, our mechanistic approach provides an improved fit to  
11 data compared to the statistical Ferretti model. In addition, our method has the advantage of  
12 being useful for parameterising population-scale compartmental epidemic forecasting  
13 models, since the time periods in our approach correspond naturally to compartments (29).

14

15 In summary, using a mechanistic approach to infer key epidemiological quantities from  
16 transmission pair data indicates that a higher proportion of SARS-CoV-2 transmissions occur  
17 prior to symptoms than previously thought. Furthermore, a significant proportion of  
18 transmissions arise shortly before symptom onset, indicating that contact tracing is beneficial  
19 even if the contact elicitation window is short. Continued use and refinement of contact  
20 tracing programmes in countries worldwide is therefore of clear public health importance.

21

## 1 **Materials and Methods:**

### 2 Notation and general details

3 Here, we outline the notation used in this section when describing the different models that  
4 we considered. For a given transmission pair, we label the source as 1 and the recipient 2, and  
5 define:

6  $t_{ik} = (\text{time of infection of host } k), \quad k = 1,2,$

7  $t_{sk} = (\text{time of symptom onset of host } k), \quad k = 1,2,$

8  $\tau_{inc,k} = (\text{incubation period of host } k), \quad k = 1,2,$

9  $\tau_{gen} = (\text{generation time}),$

10  $x_{tost} = (\text{time from symptom onset of 1 to transmission to 2 (TOST)}),$

11  $x_{ser} = (\text{serial interval}).$

12 Here,  $t$  is used to denote calendar times,  $\tau$  for time intervals relative to the time of infection,  
13 and  $x$  for time intervals relative to the time of symptom onset. We denote the probability  
14 density functions of the incubation period, generation time, TOST and serial interval as  $f_{inc}$ ,  
15  $f_{gen}$ ,  $f_{tost}$  and  $f_{ser}$ , respectively, and use a capital  $F$  for the corresponding cumulative  
16 distribution functions.

17

18 In addition, we denote the expected infectiousness of a host at time since infection  $\tau$  as  $\beta(\tau)$ ,  
19 and the expected infectiousness at time since symptom onset  $x$  as  $b(x)$ . Note that

20 
$$\beta(\tau) = R_0 f_{gen}(\tau),$$

21 
$$b(x) = R_0 f_{tost}(x),$$

22 where  $R_0$  is the basic reproduction number (for hosts that develop symptoms at some stage  
23 during infection). We also let  $\beta(\tau | \tau_{inc})$  and  $b(x | \tau_{inc})$  be the expected infectiousness at  
24 time  $\tau$  since infection and at time  $x$  since onset, respectively, conditional on an incubation

1 period of  $\tau_{inc}$  (note that  $\beta(\tau | \tau_{inc}) = b(\tau - \tau_{inc} | \tau_{inc})$  and  $b(x | \tau_{inc}) =$   
2  $\beta(x + \tau_{inc} | \tau_{inc})$ ).

3

4 We considered several different models for infectiousness (details of individual models are  
5 given below). In each model, the conditional infectiousness,  $\beta(\tau | \tau_{inc})$ , or equivalently,  
6  $b(x | \tau_{inc})$ , is specified. The distributions of the generation time and TOST can be recovered  
7 from this conditional infectiousness by averaging over the incubation period distribution  
8 (which is assumed to be known):

9 
$$\beta(\tau) = R_0 f_{gen}(\tau) = \int_0^{\infty} \beta(\tau | \tau_{inc}) f_{inc}(\tau_{inc}) d\tau_{inc},$$

10 
$$b(x) = R_0 f_{tost}(x) = \int_0^{\infty} b(x | \tau_{inc}) f_{inc}(\tau_{inc}) d\tau_{inc}.$$

11 Alternative (equivalent) expressions for the generation time and TOST distributions are  
12 available for some of the models considered (these are detailed in the “Models of  
13 infectiousness” subsection below).

14

15 To obtain an expression for the serial interval distribution, we note that

16 
$$x_{ser} = x_{tost} + \tau_{inc,2}.$$

17 We assume throughout that  $x_{tost}$  and  $\tau_{inc,2}$  are independent, so that the serial interval  
18 distribution is given by the convolution

19 
$$f_{ser}(x_{ser}) = \int_0^{\infty} f_{tost}(x_{ser} - \tau_{inc}) f_{inc}(\tau_{inc}) d\tau_{inc}.$$

20 The proportion of presymptomatic transmissions (out of all transmissions generated by  
21 individuals who develop symptoms) can be calculated as

22 
$$q_P = \int_{-\infty}^0 f_{tost}(x_{tost}) dx_{tost},$$

23 although simpler equivalent expressions for individual models are also detailed later.

1

## 2 Data

3 Following (4), we considered COVID-19 transmission pair data from five different studies  
4 (3, 9, 19–21), totalling 191 source-recipient pairs. In all 191 transmission pairs, both the  
5 source and the recipient developed symptoms, and the symptom onset date of each host was  
6 recorded. In four of the five studies (3, 9, 19, 20), intervals of exposure were available for  
7 either the source or recipient (or both), whereas in the other (21), only symptom onset dates  
8 were recorded.

9

## 10 Incubation period

11 In the main text, the incubation period was assumed to follow a Gamma distribution with  
12 shape parameter 5.807 and scale parameter 0.948 (26). This corresponds to a mean  
13 incubation period of 5.5 days and a standard deviation of 2.3 days. However, to demonstrate  
14 that our main conclusions are robust to the exact incubation period distribution used, we also  
15 repeated our analyses using an alternative, more dispersed, Gamma distributed incubation  
16 period with a mean of 5.3 days and a standard deviation of 3.2 days (30) (Fig. S4).

17

## 18 Models of infectiousness

### 19 *Independent transmission and symptoms model*

20 In this model, the infectiousness of each host at a given time since infection is assumed to be  
21 independent of their incubation period, so that

$$22 \quad \beta(\tau \mid \tau_{inc}) = \beta(\tau) = R_0 f_{gen}(\tau),$$

23 where the generation time distribution,  $f_{gen}$ , is prescribed. We assumed (3, 8) that

$$24 \quad \tau_{gen} \sim \text{Gamma}(a, b).$$

1 The shape parameter ( $a$ ) and scale parameter ( $b$ ) were estimated when we fitted the model to  
2 transmission pair data.

3

4 The TOST distribution for this model is given by

$$5 \quad f_{tost}(x_{tost}) = \int_0^{\infty} f_{gen}(x_{tost} + \tau_{inc}) f_{inc}(\tau_{inc}) d\tau_{inc},$$

6 whilst the proportion of presymptomatic transmissions is

$$7 \quad q_P = \int_0^{\infty} f_{gen}(\tau)(1 - F_{inc}(\tau)) d\tau.$$

8 Derivations of these expressions are given in the Supplementary Text.

9

10 *Ferretti model*

11 Ferretti et al. (4) proposed a model in which the conditional infectiousness was specified as  
12 the re-scaled skew-logistic distribution,

$$13 \quad b(x | \tau_{inc}) = \begin{cases} \frac{CR_0 e^{-(xm_{inc}/\tau_{inc}-\mu)/\sigma}}{(1 + e^{-(xm_{inc}/\tau_{inc}-\mu)/\sigma})^{\alpha+1}}, & -\tau_{inc} \leq x < 0, \\ \frac{CR_0 e^{-(x-\mu)/\sigma}}{(1 + e^{-(x-\mu)/\sigma})^{\alpha+1}}, & x \geq 0. \end{cases}$$

14 Here,  $m_{inc}$  is the mean incubation period,  $\mu$ ,  $\sigma$  and  $\alpha$  are estimated parameters, and we set

$$15 \quad C = \frac{\alpha}{\sigma(1 - (1 + e^{(m_{inc}+\mu)/\sigma})^{-\alpha})}$$

16 in order to ensure the correct scaling for the infectiousness (see the Supplementary Text).

17

18 The proportion of presymptomatic transmissions is

$$19 \quad q_P = \frac{(1 + e^{\mu/\sigma})^{-\alpha} - (1 + e^{(m_{inc}+\mu)/\sigma})^{-\alpha}}{1 - (1 + e^{(m_{inc}+\mu)/\sigma})^{-\alpha}}.$$

20 A derivation of this expression is given in the Supplementary Text.

21

1 *Our mechanistic model*

2 In our mechanistic approach, we divided each infection into three stages: latent ( $E$ ),  
3 presymptomatic infectious ( $P$ ), and symptomatic infectious ( $I$ ). The stage durations were  
4 assumed to be independent, and infectiousness was assumed to be constant over the duration  
5 of each stage. We denote the stage durations by  $y_{E/P/I}$ , their density and cumulative  
6 distribution functions by  $f_{E/P/I}$  and  $F_{E/P/I}$ , and the infectiousness of hosts in the  $P$  and  $I$   
7 stages by  $\beta_{P/I}$ , respectively. We also define the ratio

$$8 \quad \alpha = \beta_P / \beta_I.$$

9 Note that in this model, the basic reproduction number is

$$10 \quad R_0 = \beta_P m_P + \beta_I m_I,$$

11 where  $m_{P/I}$  are the respective mean durations of the  $P$  and  $I$  stages.

12

13 We further assumed that the durations of each stage followed Gamma distributions, with

$$14 \quad y_E \sim \text{Gamma}\left(k_E, \frac{1}{k_{inc}\gamma}\right),$$

$$15 \quad y_P \sim \text{Gamma}\left(k_P, \frac{1}{k_{inc}\gamma}\right),$$

$$16 \quad y_I \sim \text{Gamma}\left(k_I, \frac{1}{k_I\mu}\right),$$

17 where

$$18 \quad k_{inc} = k_E + k_P.$$

19 In particular, the scale parameters of  $y_E$  and  $y_P$  were both assumed to be equal to  $1/(k_{inc}\gamma)$ ,  
20 in order to ensure a Gamma distributed incubation period,

$$21 \quad \tau_{inc} = y_E + y_P \sim \text{Gamma}\left(k_{inc}, \frac{1}{k_{inc}\gamma}\right).$$

22 We fixed  $k_{inc} = 5.807$  and  $\gamma = 1/(5.807 \times 0.948)$ , in order to obtain the specified  
23 incubation period distribution (see “Incubation period” subsection above). When we fitted the

1 model to data, we assumed that  $k_I = 1$ , so that the symptomatic infectious period followed an  
 2 exponential distribution. The parameters  $k_E$  and  $\mu$  were estimated in the fitting procedure.  
 3 We considered two versions of the model: one in which we assumed  $\alpha = 1$  (the constant  
 4 infectiousness model), and one in which  $\alpha$  was also estimated (the variable infectiousness  
 5 model).

6  
 7 For this model, the infectiousness of a host at time  $x$  since symptom onset, conditional on an  
 8 incubation period of  $\tau_{inc}$ , can be calculated to be

$$9 \quad b(x | \tau_{inc}) = \begin{cases} \alpha C R_0 (1 - F_{Beta}(-x/\tau_{inc}; k_P, k_E)), & -\tau_{inc} \leq x < 0, \\ C R_0 (1 - F_I(x)), & x \geq 0, \end{cases}$$

10 where  $F_{Beta}(s; a, b)$  is the cumulative distribution function of a Beta distributed random  
 11 variable with parameters  $a$  and  $b$ , and

$$12 \quad C = \frac{\beta_I}{R_0} = \frac{k_{inc} \gamma \mu}{\alpha k_P \mu + k_{inc} \gamma}.$$

13 The TOST distribution is given by

$$14 \quad f_{tost}(x_{tost}) = \begin{cases} \alpha C (1 - F_P(-x_{tost})), & x_{tost} < 0, \\ C (1 - F_I(x_{tost})), & x_{tost} \geq 0. \end{cases}$$

15 The generation time can be written as

$$16 \quad \tau_{gen} = y_E + y^*,$$

17 where  $y^*$  is the time between the start of the  $P$  stage and the transmission occurring, and  
 18 therefore the generation time distribution is given by the convolution

$$19 \quad f_{gen}(\tau_{gen}) = \int_0^{\tau_{gen}} f^*(\tau_{gen} - y_E) f_E(y_E) dy_E,$$

20 where the density,  $f^*$ , of  $y^*$  satisfies

$$21 \quad f^*(y^*) = C \left( \alpha (1 - F_P(y^*)) + \int_0^{y^*} (1 - F_I(y^* - y_P)) f_P(y_P) dy_P \right).$$

22 The proportion of presymptomatic transmissions is



1 
$$q_P = \frac{\beta_P m_P}{R_0} = \frac{\alpha k_P \mu}{\alpha k_P \mu + k_{inc} \gamma}.$$

2 Derivations of these formulae are given in the Supplementary Text.

3

#### 4 Likelihood and model fitting

5 For a single transmission pair (labelled  $n$ ), suppose that the times of infection for the source  
6 and recipient are known to lie in the intervals  $[t_{i1,L}, t_{i1,R}]$  and  $[t_{i2,L}, t_{i2,R}]$ , respectively (where  
7 these intervals may be infinite), and that their symptom onset times,  $t_{s1}$  and  $t_{s2}$ , are known  
8 exactly. In this case, the likelihood of the parameters,  $\theta$ , of the model of infectiousness under  
9 consideration (when only that transmission pair is observed) is given by

10 
$$L^{(n)}(\theta) = \frac{1}{R_0} \int_{t_{i2,L}}^{t_{i2,R}} \int_{t_{i1,L}}^{t_{i1,R}} b(t_{i2} - t_{s1} \mid t_{s1} - t_{i1}, \theta) f_{inc}(t_{s1} - t_{i1}) f_{inc}(t_{s2} - t_{i2}) dt_{i1} dt_{i2},$$

11 where the dependence of the conditional expected infectiousness,  $b(x \mid \tau_{inc}, \theta)$ , on the model  
12 parameters,  $\theta$ , is indicated explicitly. A derivation of this expression is given in the  
13 Supplementary Text. Assuming that each transmission pair in our dataset was independent,  
14 the overall likelihood was therefore given by the product of the contributions,  $L^{(n)}(\theta)$ , from  
15 each individual transmission pair, i.e.,

16 
$$L(\theta) = \prod_{n=1}^N L^{(n)}(\theta),$$

17 where  $N$  is the total number of transmission pairs.

18

19 In order to account for uncertainty in the exact symptom onset times within the day of onset,  
20 we fitted the models to the data using data augmentation MCMC. In particular, in alternating  
21 steps of the chain, we updated either the vector of model parameters,  $\theta$ , or the symptom onset  
22 times of each source and recipient. The chain was run for 2.5 million steps, of which the first

1 500,000 were discarded as burn-in. Posterior distributions of parameters were obtained by  
2 recording only every 100 iterations of the chain (assuming a uniform prior distribution for  
3 each model parameter), whilst the parameter values that maximised the likelihood were taken  
4 as point estimates. Full details of the model fitting procedure are given in the Supplementary  
5 Text.

### 7 Distributions of the presymptomatic and total non-symptomatic proportion of transmissions

8 Expressions for the proportion of transmissions,  $q_p$ , generated prior to symptom onset, are  
9 given for the individual models above. Once asymptomatic cases are accounted for, the  
10 overall non-symptomatic proportion of transmissions can be written as

$$11 \quad \frac{p_A r_A + (1 - p_A) q_p}{p_A r_A + (1 - p_A)},$$

12 where  $p_A$  is the proportion of infected individuals who remain asymptomatic, and  $r_A$  is the  
13 ratio between the average number of secondary cases generated by an asymptomatic host and  
14 the number generated by a host who at some point develops symptoms. A derivation of this  
15 expression is given in the Supplementary Text.

16  
17 For each model, we used the posterior parameter distributions that were obtained when we  
18 fitted the model to data to obtain a sample from the posterior distribution of  $q_p$ . In order to  
19 estimate the total proportion of non-symptomatic transmissions, we assumed the distributions

$$20 \quad p_A \sim \text{Beta}(85, 186),$$

$$21 \quad r_A \sim \text{Lognormal}(-1.04, 0.65^2),$$

22 which are consistent with estimates in (1). These distributions are shown in Fig. S2. We then  
23 combined samples from the assumed distributions of  $p_A$  and  $r_A$  with the sample that we  
24 generated from the posterior distribution of  $q_p$  to obtain a distribution for the total proportion  
25 of non-symptomatic transmissions.

1

## 2 Contact tracing

3 First, we considered the proportion of the infectious contacts of a symptomatic index case  
4 that will be found, if contacts are traced up to  $d_1$  days before the time of symptom onset (of  
5 the index host). In this case, assuming that it is possible to trace a fraction  $\varepsilon_1$  of the host's  
6 total contacts between times  $-d_1$  and  $\infty$  since symptom onset, then the proportion of  
7 infectious contacts found is equal to

$$8 \quad \varepsilon_1(1 - F_{tost}(-d_1)).$$

9 Note that we assumed that the TOST distribution for a detected index case does not differ  
10 from that of a random infected individual.

11

12 We then considered the proportion of transmissions that can be prevented, if an infected  
13 individual (identified through contact tracing) is isolated  $d_2$  days after exposure. Assuming  
14 that a proportion  $\varepsilon_2$  of infectious contacts that would otherwise occur are prevented during  
15 the isolation period, the overall proportion of onward infections prevented through isolation  
16 is

$$17 \quad \varepsilon_2(1 - F_{gen}(d_2)).$$

18

19 In the main text, we assumed that  $\varepsilon_1 = \varepsilon_2 = 1$  (i.e., contact identification and isolation are  
20 both 100% effective), although values of  $\varepsilon_1$  and  $\varepsilon_2$  below 1 are considered in the  
21 Supplementary Materials (Fig. S3).

22

## 1   **References:**

- 2   1.    D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, A. M.  
3       Ipekci, G. Salanti, N. Low, Occurrence and transmission potential of asymptomatic  
4       and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-  
5       analysis. *PLoS Med.* **17**, e1003346 (2020).
- 6   2.    M. Casey, J. Griffin, C. G. McAloon, A. W. Byrne, J. M. Madden, D. McEvoy, A. B.  
7       Collins, K. Hunt, A. Barber, F. Butler, E. A. Lane, K. O. Brien, P. Wall, K. A. Walsh,  
8       S. J. More, Pre-symptomatic transmission of SARS-CoV-2 infection: a secondary  
9       analysis using published data. *medRxiv* (2020), doi:10.1101/2020.05.08.20094870.
- 10  3.    L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker,  
11       D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic  
12       control with digital contact tracing. *Science.* **368**, eabb6936 (2020).
- 13  4.    L. Ferretti, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. Abeler-Dörner, M. Kendall,  
14       A. Nurtay, H.-Y. Cheng, T.-C. Ng, H.-H. Lin, R. Hinch, J. Masel, A. M. Kilpatrick, C.  
15       Fraser, The timing of COVID-19 transmission. *medRxiv* (2020),  
16       doi:10.1101/2020.09.04.20188516.
- 17  5.    W. Zhang, Estimating the presymptomatic transmission of COVID19 using incubation  
18       period and serial interval data. *medRxiv* (2020), doi:10.1101/2020.04.02.20051318.
- 19  6.    Y. Liu, Centre for Mathematical Modelling of Infectious Diseases nCoV Working  
20       Group, S. Funk, S. Flasche, The contribution of pre-symptomatic infection to the  
21       transmission dynamics of COVID-2019. *Wellcome Open Res.* **5**, 58 (2020).
- 22  7.    Y. Deng, C. You, Y. Liu, J. Qin, X. H. Zhou, Estimation of incubation period and  
23       generation time based on observed length-biased epidemic cohort with censoring for  
24       COVID-19 outbreak in China. *Biometrics*, 1–13 (2020).

- 1 8. T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, N. Hens,  
2 Estimating the generation interval for coronavirus disease (COVID-19) based on  
3 symptom onset data, March 2020. *Eurosurveillance*. **25**, 2000257 (2020).
- 4 9. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y.  
5 Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu,  
6 L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral  
7 shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
- 8 10. P. Ashcroft, J. S. Huisman, S. Lehtinen, J. A. Bouman, C. L. Althaus, R. R. Regoes, S.  
9 Bonhoeffer, COVID-19 infectivity profile correction. *Swiss Med. Wkly.* **150**, w20336  
10 (2020).
- 11 11. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, Serial interval of  
12 COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* **26**, 1341–  
13 1343 (2020).
- 14 12. C. Fraser, Estimating individual and household reproduction numbers in an emerging  
15 epidemic. *PLoS One.* **2**, e758 (2007).
- 16 13. P. Ashcroft, S. Lehtinen, S. Bonhoeffer, Quantifying the impact of quarantine duration  
17 on COVID-19 transmission. *medRxiv* (2020), doi:10.1101/2020.09.24.20201061.
- 18 14. J. Knight, S. Mishra, Estimating effective reproduction number using generation time  
19 versus serial interval, with application to covid-19 in the Greater Toronto Area,  
20 Canada. *Infect. Dis. Model.* **5**, 889–896 (2020).
- 21 15. S. Lehtinen, P. Ashcroft, S. Bonhoeffer, On the relationship between serial interval,  
22 infectiousness profile and generation time. *medRxiv* (2020),  
23 doi:10.1101/2020.09.18.20197210.
- 24 16. S. Bacallado, Q. Zhao, N. Ju, Letter to the editor: Generation interval for COVID-19

- 1 based on symptom onset data. *Eurosurveillance*. **25**, 2001381 (2020).
- 2 17. R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P.  
3 A. Demarsh, E. Dahlgvist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, A.  
4 Cori, Improved inference of time-varying reproduction numbers during infectious  
5 disease outbreaks. *Epidemics*. **29**, 100356 (2019).
- 6 18. P. Manfredi, A. D’Onofrio, Eds., *Modeling the interplay between human behavior and*  
7 *the spread of infectious diseases* (Springer, New York, 2013).
- 8 19. W. Xia, J. Liao, C. Li, Y. Li, X. Qian, X. Sun, H. Xu, G. Mahai, X. Zhao, L. Shi, J.  
9 Liu, L. Yu, M. Wang, Q. Wang, A. Namat, Y. Li, J. Qu, Q. Liu, X. Lin, S. Cao, S.  
10 Huan, J. Xiao, F. Ruan, H. Wang, Q. Xu, X. Ding, X. Fang, F. Qiu, J. Ma, Y. Zhang,  
11 A. Wang, Y. Xing, S. Xu, Transmission of corona virus disease 2019 during the  
12 incubation period may lead to a quarantine loophole. *medRxiv* (2020),  
13 doi:10.1101/2020.03.06.20031955.
- 14 20. H. Y. Cheng, S. W. Jian, D. P. Liu, T. C. Ng, W. T. Huang, H. H. Lin, Contact Tracing  
15 Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different  
16 Exposure Periods before and after Symptom Onset. *JAMA Intern. Med.* **180**, 1156–  
17 1163 (2020).
- 18 21. J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L.  
19 Yi, X. Chen, Q. Wu, Y. Liang, X. Wang, J. Yang, K. Sun, I. M. Longini, M. E.  
20 Halloran, P. Wu, B. J. Cowling, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, H.  
21 Yu, Evolving epidemiology and transmission dynamics of coronavirus disease 2019  
22 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.*  
23 **20**, 793–802 (2020).
- 24 22. R. N. Thompson, Novel coronavirus outbreak in Wuhan, China, 2020: intense

- 1 surveillance is vital for preventing sustained transmission in new locations. *J. Clin.*  
2 *Med.* **9**, 498 (2020).
- 3 23. A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to  
4 estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**,  
5 1505–1512 (2013).
- 6 24. NHS test and trace: how it works, (available at [https://www.gov.uk/guidance/nhs-test-](https://www.gov.uk/guidance/nhs-test-and-trace-how-it-works)  
7 [and-trace-how-it-works](https://www.gov.uk/guidance/nhs-test-and-trace-how-it-works)).
- 8 25. Investigating a COVID-19 Case, (available at [https://www.cdc.gov/coronavirus/2019-](https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/investigating-covid-19-case.html)  
9 [ncov/php/contact-tracing/contact-tracing-plan/investigating-covid-19-case.html](https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/investigating-covid-19-case.html)).
- 10 26. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman,  
11 N. G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (CoVID-  
12 19) from publicly reported confirmed cases: estimation and application. *Ann. Intern.*  
13 *Med.* **172**, 577–582 (2020).
- 14 27. E. L. Davis, T. C. D. Lucas, A. Borlase, T. M. Pollington, S. Abbott, D. Ayabina, T.  
15 Crellen, J. Hellewell, L. Pi, G. F. Medley, T. D. Hollingsworth, P. Klepac, An  
16 imperfect tool: COVID-19 “test & trace” success relies on minimising the impact of  
17 false negatives and continuation of physical distancing. *medRxiv* (2020),  
18 doi:10.1101/2020.06.09.20124008.
- 19 28. T. Britton, G. S. Tomba, Estimation in emerging epidemics: Biases and remedies. *J. R.*  
20 *Soc. Interface.* **16**, 20180670 (2019).
- 21 29. W. S. Hart, P. K. Maini, C. A. Yates, R. N. Thompson, A theoretical framework for  
22 transitioning from patient-level to population-scale epidemiological dynamics:  
23 influenza A as a case study. *J. R. Soc. Interface.* **17**, 20200230 (2020).
- 24 30. N. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. Akhmetzhanov, S. Jung, B. Yuan,

1 R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics  
2 of 2019 novel coronavirus infections with right truncation: a statistical analysis of  
3 publicly available case data. *J. Clin. Med.* **9**, 538 (2020).

4

5 **Acknowledgments:** Thanks to members of the Wolfson Centre for Mathematical Biology at  
6 the University of Oxford for useful discussions about this work.

7 **Funding:** WSH was funded by an EPSRC Excellence Award for his doctoral studies. RNT  
8 was funded by a Junior Research Fellowship from Christ Church, Oxford. The funders had  
9 no role in study design, data collection and analysis, decision to publish, or preparation of the  
10 manuscript.

11 **Author contributions:** RNT and WSH conceived the research; All authors designed the  
12 study; WSH carried out the research; WSH drafted the manuscript; RNT and PKM  
13 supervised the research; All authors revised the manuscript and gave final approval for  
14 publication.

15 **Competing interests:** Authors declare no competing interests.

16 **Data and materials availability:** The transmission pair data used in our analyses are  
17 available in the Supplementary Materials. Code for reproducing our results is available at  
18 <https://github.com/will-s-hart/COVID-19-Infectiousness-Profile>.