

Transmission of SARS-CoV-2 before and after symptom onset: impact of nonpharmaceutical interventions in China

Mary Bushman^{1*}, Colin Worby², Hsiao-Han Chang³, Moritz Kraemer⁴, William P. Hanage¹

¹Harvard T.H. Chan School of Public Health, Boston, MA, USA

²Broad Institute, Cambridge, MA, USA

³National Tsing Hua University, Hsinchu City, Taiwan

⁴University of Oxford, Oxford, UK

*Correspondence to: mbushman@hsph.harvard.edu

Abstract

Nonpharmaceutical interventions, such as contact tracing and quarantine, are currently the primary means of controlling the spread of SARS-CoV-2; however, it remains uncertain which interventions are most effective at reducing transmission at the population level. Using serial interval data from before and after the rollout of nonpharmaceutical interventions in China, we estimate that the relative frequency of presymptomatic transmission increased from 34% before the rollout to 71% afterward. The shift toward earlier transmission indicates a disproportionate reduction in transmission post-symptom onset. We estimate that, following the rollout of nonpharmaceutical interventions, transmission post-symptom onset was reduced by 82% whereas presymptomatic transmission decreased by only 16%. These findings suggest that interventions which limit opportunities for transmission in the later stages of infection, such as contact tracing and isolation, may have been particularly effective at reducing transmission of SARS-CoV-2.

Keywords: COVID-19, SARS-CoV-2, presymptomatic transmission, nonpharmaceutical interventions

Introduction

In January 2020, in Wuhan, China, what began as a cluster of viral pneumonia cases rapidly spiraled into an epidemic of a new disease, COVID-19, caused by a novel coronavirus, designated SARS-CoV-2. As cases mounted in Wuhan and began cropping up elsewhere, China introduced a comprehensive set of measures, termed nonpharmaceutical interventions (NPIs), to contain the virus. On January 23, a lockdown was enacted in Wuhan, which shut down public transit and travel out of the city. Other cities throughout Hubei province (of which Wuhan is capital) announced similar lockdowns over the next few days [1]. The rest of China was subject to social distancing measures: mass transit and public gatherings were severely curtailed, and the New Year holiday (*Chunyun*) was extended, which kept most schools, workplaces, and businesses closed [1-3]. In addition, numerous measures were implemented to rapidly identify and isolate suspected cases. These included temperature checks at borders and travel hubs, quarantine of new arrivals, isolation of both confirmed and suspected cases, and contact tracing with quarantine and medical observation.

China's robust public health response was decidedly effective in controlling the spread of SARS-CoV-2 [4-7]. As of December 1, 2020, China had a cumulative incidence of 63 cases per million residents and cumulative mortality of 3 deaths per million, compared to 7937 cases and 186 deaths per million persons globally [8]. Given the magnitude and intensity of the response

in China, which would be difficult to replicate in many settings, it would be useful to know which interventions were most effective in limiting the spread of the virus.

5 Evaluating the effectiveness of specific NPIs has been a challenge throughout the COVID-19 pandemic because multiple interventions are typically used concurrently. Several studies have estimated the combined impact of multiple NPIs [5-7, 9]; others have estimated the effect of specific control measures using statistical approaches [10-20] or modeling [21-25]. The majority of studies focus on the effectiveness of social distancing interventions, while relatively few explicitly consider the impact of quarantine or isolation measures. However, in China, the impact of isolation has been indirectly observed in a shift toward earlier transmission, reflected in shorter serial intervals [26]. Here, we use serial interval data and incidence data from China to show that, in the first few weeks after NPIs were implemented, presymptomatic transmission decreased slightly but transmission post-symptom onset declined dramatically. These results suggest that “post-onset” interventions, such as case isolation, may have been more effective than other control measures at limiting transmission of SARS-CoV-2.

Results

20 We compiled published data, including symptom onset dates, for 873 infector-infectee pairs from China (see Methods) [27-30]. Among these case pairs, the majority of transmission events occurred outside Hubei, with 84% of secondary cases acquired in other provinces (Table S1). Symptom onset dates ranged from January 7 to February 29, 2020, a period that spans the rollout of nonpharmaceutical interventions in China. The mean serial interval, defined as the time between onset of symptoms in infector and infectee, was 4.64 days over the entire period, similar to published estimates [27, 28, 31, 32], but this distribution shifted markedly over time. A linear regression of serial intervals vs. symptom onset dates of primary cases reveals a significant decrease in the length of serial intervals ($p = 4 \times 10^{-34}$, Figure 1a), an observation also made by Ali et al. [26].

30 We then divided case pairs into two time periods using the date of symptom onset in each primary case. Those with onset before January 23 - marking the lockdown of Wuhan and the start of a national rollout of nonpharmaceutical interventions – were denoted pre-lockdown ($n = 207$) and the rest denoted post-lockdown ($n = 666$). Pre- and post-lockdown serial intervals were significantly different between the two time periods ($p = 4 \times 10^{-19}$, Figure 1b), with a mean of 7.57 days (standard deviation 5.13 days) before the lockdown and a mean of 3.73 days (standard deviation 4.93 days) afterward.

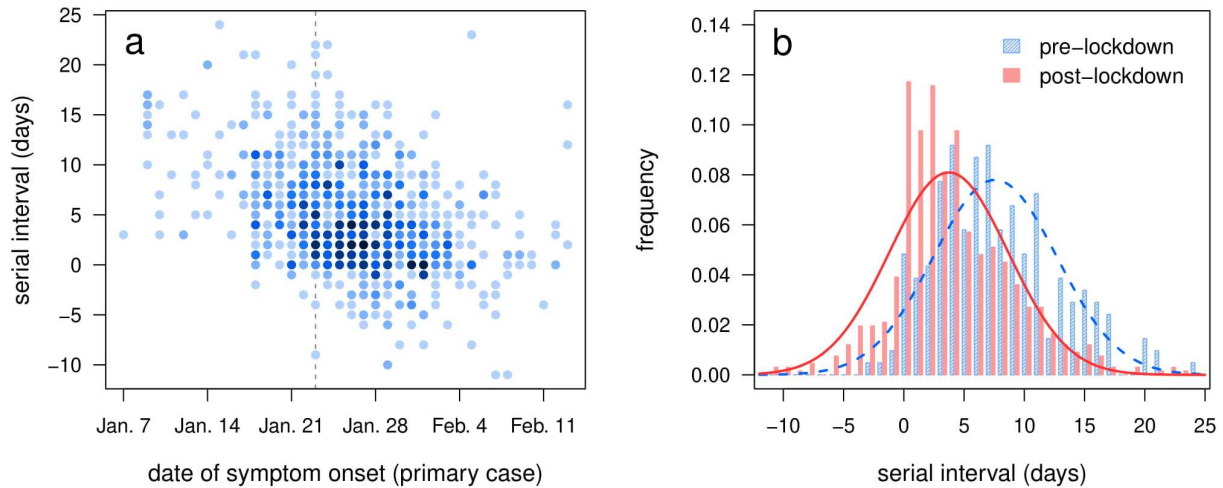


Fig. 1 Serial intervals of case pairs in China. (a) Serial interval vs. date of symptom onset in the primary case of each pair. Shading reflects the number of overlapping points, with darker colors indicating higher numbers. Dashed vertical line indicates Jan 23, the date of the initial lockdown in Wuhan. (b) Serial interval histograms and best-fit normal distributions for primary cases with symptom onset before Jan 23 (pre-lockdown; blue, hatched bars and dashed line) and on/after Jan 23 (post-lockdown; red, solid bars and solid line).

Next, we used a Markov chain Monte Carlo (MCMC) approach to estimate the distribution of the generation interval for each time period by fitting to serial interval data. We replicated this analysis using three different priors for the incubation period distribution (Table S2), based on work by Lauer et al. [33], Zhang et al. [30]; and Backer et al. [34]; we also used two different models for the generation interval distribution. Under the incubation-independent model, all generation intervals are drawn from a single gamma distribution; under the incubation-dependent model, the generation interval is still gamma-distributed but the probability distribution for each case is horizontally “stretched” in proportion to the incubation period, which means that cases with longer incubation periods will tend to have longer generation intervals. We used a modified deviance information criterion [35] to assess model fit (S1 Text) and found that the best model for both time periods was the incubation-independent model – which, somewhat surprisingly, suggests generation intervals do not increase with incubation period - paired with the prior based on Lauer et al. (Table S3).

As expected, the fitted generation interval distributions had similar means but smaller variances than the serial interval distributions [36]. Under the best model, pre-lockdown generation intervals had a mean of 7.50 days (standard deviation 3.95 days) while post-lockdown generation intervals had a mean of 3.90 days (standard deviation 3.15 days). Under the remaining models, mean generation intervals ranged from 7.49 to 7.64 days for the pre-lockdown period and from 3.84 to 4.04 days for the post-lockdown period (Table S4).

Using the inferred generation interval distributions, we estimated the relative frequency of presymptomatic transmission under each model by calculating the probability of the generation interval being shorter than the incubation period. Under the best model, the frequency of presymptomatic transmission was estimated to be 34.4% in the pre-lockdown period (95%

credible interval 28.3% - 41.3%) and 71.0% in the post-lockdown period (95% CI 67.6% - 74.2%) under the best model (Figure 2). Across all models, the estimated frequency of presymptomatic transmission ranged from 30.7% to 47.0% for the pre-lockdown period, with 95% CIs collectively extending from 24.0% to 53.7% (Table S5). For the post-lockdown period, estimates ranged from 68.1% to 80.6%, with 95% CIs extending from 64.5% to 83.5%. We note that the estimates for the pre-lockdown period are lower than many published estimates of the frequency of presymptomatic transmission [27, 29, 37-40], which raises the possibility that many estimates might be inflated by the effects of nonpharmaceutical interventions.

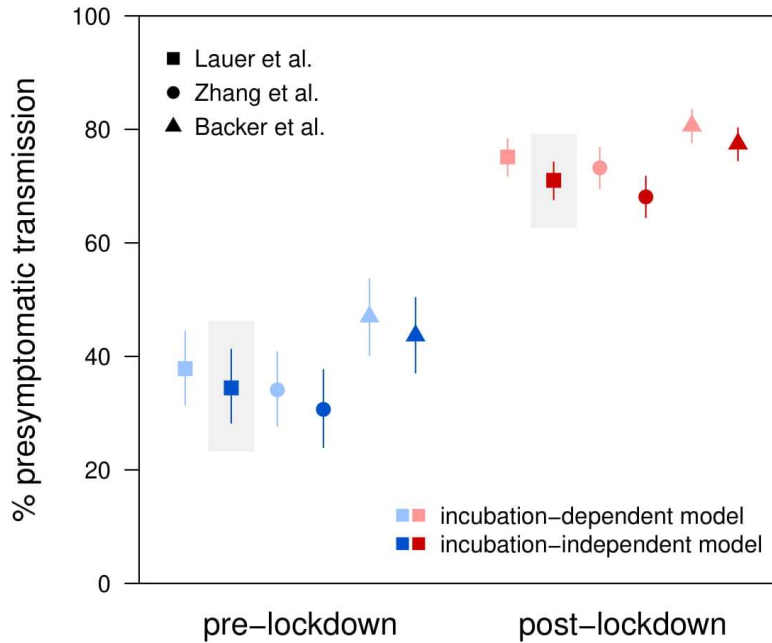


Fig. 2 Estimated relative frequency of presymptomatic transmission before and after the first lockdown on January 23. Points and lines show posterior means and 95% credible intervals, respectively. Colors denote time periods (blue, pre-lockdown; red, post-lockdown), shades denote generation interval distribution models (light, incubation-dependent model; dark, incubation-independent model), and symbols denote sources for incubation period prior distributions (squares, Lauer et al. [33]; circles, Zhang et al. [30]; triangles, Backer et al. [34]). Estimates from the best model are highlighted in gray.

The shift toward presymptomatic transmission following the rollout of NPIs suggests a disproportionate reduction in transmission post-symptom onset. We therefore estimated the change in transmission during each phase of the infection (before and after symptom onset) following the implementation of control measures. To do so, it was first necessary to estimate the total reduction in transmission of SARS-CoV-2. For this, we used province-level incidence data for all of China, combining the numbers for all provinces except Hubei because the transmission events represented in the case-pair data mostly occurred outside Hubei. We calculated the daily case reproduction number (R_t), shown in Figure 3, using a version of the Wallinga-Teunis method [41] which was modified to allow for time-varying serial intervals, as changes in the serial interval distribution have been shown to affect inference of reproduction numbers [26].

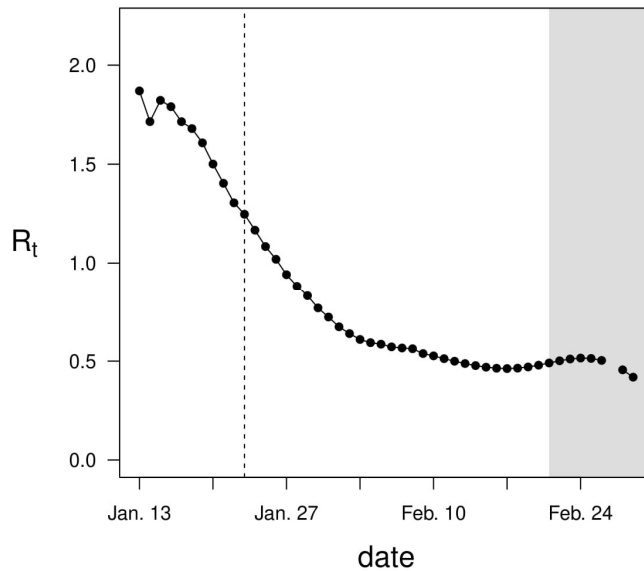


Fig. 3 Daily reproduction numbers estimated using incidence data from all Chinese provinces except Hubei. Dashed vertical line, start of initial lockdown in Wuhan on Jan. 23; gray shading, region in which R_t is likely to be underestimated due to right-truncation.

We used the estimated R_t values to calculate the mean reproduction numbers for the pre- and post-lockdown periods; estimates past February 20 were not used due to the potential for underestimation of R_t as a result of right-truncation (Figure 3) [42]. The mean reproduction numbers were estimated to be 1.64 for the pre-lockdown period and 0.666 for the post-lockdown period, corresponding to a 59.4% reduction in overall transmission. Treating this net change as a weighted average of the changes in presymptomatic transmission and transmission post-symptom onset, we calculated the change in the absolute frequency of transmission during each phase of the infection. Under the best model, we estimate that presymptomatic transmission decreased by 15.5% (95% CI: -30.6% to +2.56%) while transmission post-symptom onset decreased by 82.0% (95% CI: -84.5% to -79.0%; Figure 4). Across all models, estimates of the change in presymptomatic transmission ranged from -29.9% to -8.62%, with 95% CIs extending from -39.4% to +16.3% (Table S6). Estimates of the change in transmission post-symptom onset ranged from -85.1% to -81.3%, with 95% CIs extending from -87.9% to -78.3%.

One hypothesis to explain the disproportionate reduction in transmission post-symptom onset is that “post-onset” interventions, such as case isolation, were highly effective at preventing transmission in the later stages of infection; however, several alternative explanations must be considered. One possibility is that travel restrictions reduced the frequency of case pairs with different infection locations (primary and secondary cases infected in different cities), which might be expected to have longer-than-average serial intervals. Although the proportion of case pairs infected in different cities decreased after the lockdown ($p=1\times 10^{-4}$; Table S7), a two-way ANOVA found no significant effect of location matching on serial intervals ($p>0.05$).

Another possibility is that social distancing increased time spent at home, with more frequent exposure leading to earlier transmission between family members or household contacts – similar to the relationship between force of infection and age of first infection. The proportion of transmission events occurring between family members increased following the lockdown

($p=3\times 10^{-5}$), although there was no change in the proportion of transmission events taking place within households ($p=0.09$; Table S7). However, neither familial relationship nor household contact had a significant effect on serial intervals, nor did either factor have a significant interaction with time period ($p>0.05$).

5

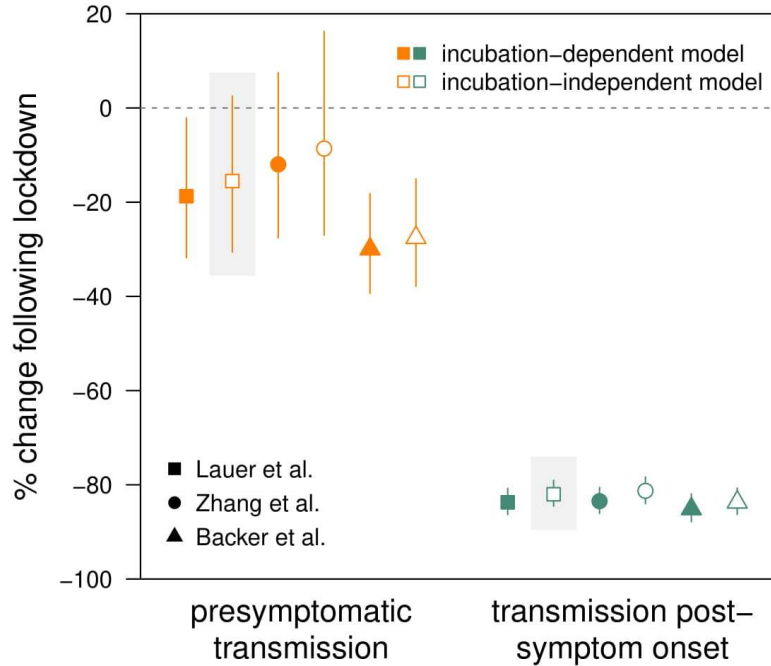


Fig. 4 Estimated percent change in absolute frequency of presymptomatic transmission and transmission post-symptom onset. Points and lines show posterior means and 95% credible intervals, respectively. Orange, presymptomatic transmission; green, transmission post-symptom onset; closed symbols, incubation-dependent model; open symbols, incubation-independent model. Symbol shapes denote sources for incubation period prior distributions (squares, Lauer et al. [33]; circles, Zhang et al. [30]; triangles, Backer et al. [34]). Estimates from the best model are highlighted in gray.

10

Discussion

15

Following the implementation of nonpharmaceutical interventions in China, the transmission of SARS-CoV-2 changed in two ways. Compared to the period preceding the lockdown of Wuhan on January 23, 2020, the post-lockdown period was characterized by a significant reduction in the reproduction number (R_t), indicating decreased transmission, and a decrease in the length of generation intervals, reflecting earlier transmission. Specifically, we found that overall transmission of SARS-CoV-2 declined by 59.4% after the lockdown, while presymptomatic transmission expanded from 34% to 71% of all transmission events. The shift toward earlier transmission implies a disproportionate reduction in transmission following symptom onset; we estimate that presymptomatic transmission decreased by roughly 16% after the lockdown, whereas transmission post-symptom onset decreased by approximately 82%.

20

25

Several factors might contribute to the observed shift toward earlier transmission: a reduction in the proportion of primary cases imported from other cities, eliminating a potential delay in transmission; increased time at home, resulting in more frequent exposure to infectious household contacts; and/or interventions that disproportionately reduce transmission in the later stages of infection. Our results do not support either of the first two possibilities: whether a linked pair of cases were infected in the same city was not a significant predictor of the serial

30

interval, nor was the existence of a shared household or familial relationship. The third possibility is indirectly supported by observed correlations between isolation delays and serial intervals, which suggest that case isolation is capable of shifting the serial interval distribution [25, 26, 43].

Our findings therefore suggest that “post-onset” interventions, such as case isolation, may have been responsible for the dramatic reduction in transmission post-symptom onset that followed the implementation of NPIs in China. If so, this would indicate that isolation and associated interventions, such as contact tracing, were among the largest contributors to the rapid reduction in transmission of SARS-CoV-2, while the comparatively modest reduction in presymptomatic transmission would indicate a relatively small impact of social distancing measures. However, it is possible that the effects of social distancing were offset by increased household transmission, an effect that might be expected to dissipate once most household transmission chains go extinct.

A key limitation of this work is the lack of information regarding transmission from asymptomatic infections, which may comprise 40-45% of all SARS-CoV-2 infections [44]. Because the serial interval is defined by symptom onset dates, alternative methods are required to examine the impact of NPIs on asymptomatic transmission. For instance, data on exposure windows could be used to estimate the time of infection for primary cases, similar to the approach used for estimating incubation periods [30, 33, 34, 45, 46]; in conjunction with exposure or symptom onset dates for secondary cases, an approach similar to the one employed here could then be used to infer generation intervals for transmission pairs with asymptomatic primary cases.

In summary, we find that the implementation of nonpharmaceutical interventions in China was followed not only by a rapid decrease in the rate of SARS-CoV-2 transmission, but a significant shift in the timing of viral transmission, with more transmission occurring in the presymptomatic (incubation) period. The leading hypothesis to explain these observations is that interventions, particularly case isolation, were highly effective in limiting transmission in the later stages of infection, while other measures, such as social distancing, had a more limited impact on transmission in the earlier stages. These findings suggest that rapid case detection and isolation, if rigorously implemented, may be a highly effective strategy for interrupting transmission of SARS-CoV-2.

Methods

Data sources

We obtained data on serial intervals for linked pairs of cases reported in China in January and February 2020. We combined data that were previously collected and published by the following sources: Xu et al. with 679 case pairs, compiled from provincial and urban health commission reports [27]; Du et al. with 468 case pairs, compiled from provincial health agency reports [28]; He et al. with 41 case pairs, compiled from government and media reports [29]; and Zhang et al. with 35 case pairs, compiled from health agency and media reports [30]. We cross-checked all datasets and eliminated duplicate case pairs, which we identified as those with matching sex, age, and symptom onset date for both cases. This left a total of 873 unique case pairs for use in parameter estimation. Among these, a majority of transmission events took place outside Hubei province (home to the city of Wuhan), with 84.2% of secondary cases infected outside Hubei (Table S1).

Analysis of serial intervals

Since dates of transmission were unknown, we dated each serial interval using the date of symptom onset in the primary case. We used simple linear regression (serial interval versus date) to test the hypothesis that serial intervals declined over time. We then divided serial intervals into two time periods based on the date of symptom onset in the primary case of each pair. Case pairs were designated “pre-lockdown” if the primary case developed symptoms before the lockdown of Wuhan on Jan 23 (i.e. symptom onset occurred on or before Jan 22) and as “post-lockdown” otherwise. The terms pre-lockdown and post-lockdown refer here to the time periods preceding and following a specific event – the lockdown of Wuhan – rather than periods characterized by the presence or absence of lockdown measures; true lockdowns were not widely implemented beyond Hubei province. We used a two-sided Student’s t-test to test the hypothesis that the mean serial intervals differed between the pre- and post-lockdown periods.

Generation interval distributions

Consider a linked pair of cases, with the secondary case arising by transmission from the primary case. We denote the incubation periods of the primary and secondary cases by θ_1 and θ_2 , respectively. We use δ to refer the serial interval and τ to denote the generation interval. The relationship between these quantities is given by $\tau = \delta + \theta_1 - \theta_2$ (Figure 5).

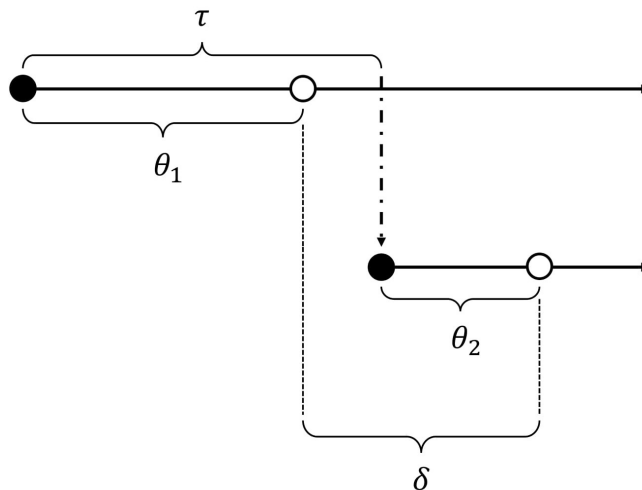


Fig. 5 Diagram showing incubation period (θ), serial interval (δ), and generation interval (τ) for a linked pair of cases. Closed circle, time of infection; open circle, time of symptom onset; dot-dash arrow, transmission from primary case to secondary case.

We assumed that the generation interval τ followed a gamma distribution f_τ with unknown parameters α and β . Due to uncertainty regarding the relationship between the incubation period and the generation interval, we used two different models for the generation interval distribution. In the incubation-independent model, we assumed that a single generation interval distribution, with shape parameter α and rate parameter β , applied to all individuals, regardless of incubation period. In the “incubation-dependent” model, we assumed that individuals with longer incubation periods would tend to have longer generation intervals; specifically, for a primary case with incubation period θ_1 , we assumed that the generation interval followed a gamma distribution with shape parameter α and rate parameter β/θ_1 . This is equivalent to defining a new random variable $X = \tau/\theta_1$ where X follows a gamma distribution with shape α

and rate β . This formulation causes the generation interval distribution to be horizontally “stretched out” in proportion to θ_1 – for instance, it results in the expected generation interval being a fixed multiple of θ_1 , rather than a fixed length of time.

5 *Parameter estimation*

We used a Markov chain Monte Carlo (MCMC) algorithm to estimate the parameters of the pre- and post-lockdown generation interval distributions by fitting to serial interval data from each time period. Because an observed serial interval depends on the incubation period of each infection as well as the generation interval, we also estimated the unobserved incubation periods θ_1 and θ_2 for each pair of cases (data augmentation). The model assumed that the incubation periods were drawn from a prior f_θ ; we used three different priors drawn from the literature (see below). The generation interval was assumed to be drawn from a gamma distribution f_τ with unknown parameters α and β , with minimally informative priors f_α and f_β , respectively (Table S2). We can express the joint posterior density of the unknown parameters as follows:

$$\begin{aligned}
 f_{\alpha, \beta | \bar{\delta}}(\alpha, \beta | \bar{\delta}) &\propto \prod_{i=1}^N f_{\delta | \alpha, \beta}(\delta_i | \alpha, \beta) f_\alpha(\alpha) f_\beta(\beta) \\
 &= \prod_{i=1}^N \left(\int_0^\infty \int_0^\infty f_{\delta | \alpha, \beta}(\delta_i | \alpha, \beta, \theta_1, \theta_2) f_\alpha(\alpha) f_\beta(\beta) f_\theta(\theta_1) f_\theta(\theta_2) d\theta_1 d\theta_2 \right) \\
 &= \prod_{i=1}^N \left(\int_0^\infty \int_0^\infty f_{\tau | \alpha, \beta}(\delta_i + \theta_1 - \theta_2 | \alpha, \beta) f_\alpha(\alpha) f_\beta(\beta) f_\theta(\theta_1) f_\theta(\theta_2) d\theta_1 d\theta_2 \right)
 \end{aligned}$$

We estimated the unknown quantities using a Metropolis-Hastings algorithm, with each iteration taking place in two parts. The parameters α and β were updated first, with proposed values α' and β' being accepted with probability $\min\left(1, \frac{\prod_{i=1}^N f_{\tau | \alpha, \beta}(\delta_i + \theta_{1(i)} - \theta_{2(i)} | \alpha', \beta') f_\alpha(\alpha') f_\beta(\beta')}{\prod_{i=1}^N f_{\tau | \alpha, \beta}(\delta_i + \theta_{1(i)} - \theta_{2(i)} | \alpha, \beta) f_\alpha(\alpha) f_\beta(\beta)}\right)$. The incubation periods θ_1 and θ_2 for each case pair were then updated, with proposed values $\theta_{1(i)}'$ and $\theta_{2(i)}'$ being accepted with probability $\min\left(1, \frac{f_{\tau | \alpha, \beta}(\delta_i + \theta_{1(i)}' - \theta_{2(i)}' | \alpha, \beta) f_\theta(\theta_{1(i)}') f_\theta(\theta_{2(i)}')}{f_{\tau | \alpha, \beta}(\delta_i + \theta_{1(i)} - \theta_{2(i)} | \alpha, \beta) f_\theta(\theta_{1(i)}) f_\theta(\theta_{2(i)})}\right)$.

We ran the algorithm for 250,000 iterations, discarded the first 50,000 and thinned the remainder by keeping every 10th iteration. The resulting set of 20,000 observations was used to approximate the joint posterior distribution of the parameters of interest. The aggregated posterior distributions of the incubation periods θ_1 and θ_2 for each model are illustrated in Figures S1-S4, while posterior means and 95% credible intervals for the generation interval parameters are reported in Tables S3-S4. We report the mean and standard deviation of the generation interval distribution alongside the parameters α and β , since these latter quantities are difficult to interpret. For the incubation-independent model, the mean and variance are simply α/β and α/β^2 , respectively. For the incubation-dependent model, the mean generation interval is given by $E[X]E[\theta]$, where $X = \tau/\theta$ follows a gamma distribution with shape α and rate β . The variance is given by $(\text{Var}[X] + E[X]^2)(\text{Var}[\theta] + E[\theta]^2) - (E[X]^2)(E[\theta]^2)$. We calculated the mean and variance of the generation interval for each iteration in the converged and thinned

Markov chain in order to approximate posterior distributions, which we used to obtain posterior means and 95% credible intervals.

Incubation period distributions

Since the incubation period distribution affects the inferred generation interval distribution, we replicated our analysis using three different priors (f_θ) for the incubation period. The distributions and their sources are as follows: a lognormal distribution with mean 5.52 days and standard deviation 2.41 days, based on Lauer et al. [33]; a lognormal distribution with mean 5.21 days and standard deviation 2.59 days, based on Zhang et al. [30]; and a Weibull distribution with mean 6.49 days and standard deviation 2.35 days, based on Backer et al. [34]. The parameters for these distributions can be found in Table S2.

Model fit

We assessed model fit using a modified deviance information criterion (DIC) for data-augmented models (S1 Text) [35]. Broadly speaking, DIC is a generalization of the Akaike information criterion (AIC); it penalizes model complexity as well as poor model fit (low likelihood of observing the data under the specified model). As with AIC, the “best” model is the one with the lowest DIC value.

Pre-symptomatic transmission

We next used the joint posterior distribution of the generation interval parameters to approximate the distribution of the percentage of transmission expected to take place prior to symptom onset, which we denote φ . For the incubation-independent model, given α and β , the proportion of transmission expected to occur before symptom onset is given by $\varphi = \int_0^\infty \int_0^u f_\theta(u) f_\tau(v) dv du$, where f_τ is a gamma distribution with shape α and rate β . For the incubation-dependent model, the expected proportion of transmission occurring before symptom onset is given by $\varphi = \int_0^\theta f_\tau(u) du$, where f_τ is a gamma distribution with shape α and rate β/θ . We calculated φ for each iteration in the converged and thinned Markov chain in order to approximate the posterior distribution of φ , which we use to obtain posterior means and 95% credible intervals.

Method to estimate reproduction numbers (R_t) with time-varying serial intervals

We introduce a modified version of the Wallinga-Teunis method of estimating the reproduction number at time t (denoted R_t). In the basic method described by Wallinga & Teunis (2004), the probability that case i was infected by case j is given by

$$p_{ij} = \frac{w(t_i - t_j)}{\sum_{k \neq i} w(t_i - t_k)}$$

where t_i is the time at which case i developed symptoms, and w the serial interval distribution. The reproduction number for case j is therefore given by

$$R_j = \sum_{i \neq j} p_{ij}$$

and the reproduction number at time t (R_t) is equal to the reproduction number for any case with symptom onset at time t .

We now present a modification of this approach, which allows for time-varying serial interval distributions. Rather than a fixed serial interval distribution w , let w_t be the distribution of serial intervals for primary cases with symptom onset in the time window $[t - z, t + z]$. Then the probability case i was infected by case j is given by

$$p_{ij} = \frac{w_{t_j}(t_i - t_j)}{\sum_{k \neq i} w_{t_k}(t_i - t_k)}$$

and the reproduction number for case j is similarly given by $R_j = \sum_{i \neq j} p_{ij}$, which we can rewrite as follows:

$$R_j = \sum_{x=t_{\min}}^{t_{\max}} \frac{(n_x - \delta_{xt_j})w_{t_j}(x - t_j)}{\sum_{y=t_{\min}}^{t_{\max}} (n_y - \delta_{xy})w_y(x - y)}$$

where t_{\min} and t_{\max} are the first and last dates of symptom onset, n_t is the number of cases with symptom onset at time t , and δ_{ij} is the Kronecker delta function, with

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

such that case j is subtracted from the set of potential infectees with symptom onset at time t_j . With R_t assumed to be equal to the reproduction number for any case with symptom onset at time t , we can modify the above to get the final expression for R_t :

$$R_t = \sum_{x=t_{\min}}^{t_{\max}} \frac{(n_x - \delta_{xt})w_t(x - t)}{\sum_{y=t_{\min}}^{t_{\max}} (n_y - \delta_{xy})w_y(x - y)}$$

Estimation of time-varying serial interval distributions

We used the serial interval data described above to estimate the time-varying serial interval distributions; we assumed serial intervals for primary cases with symptom onset at time t followed a normal distribution with mean and standard deviation calculated using serial intervals for primary cases with symptom onset in the window $[t - z, t + z]$ with $z = 3$ (corresponding to a 7-day moving window). For time windows extending beyond the range of primary case symptom onset dates, the nearest 7-day window falling within this range was used to estimate the serial interval distribution.

Estimation of reproduction number (R_t) using case incidence data

The incidence data based on reported case pairs are less than ideal for estimation of R_t due to small daily case numbers (median of 9 cases per day) as well as potential sampling biases. We therefore compared case pair incidence data to the incidence data for all of China minus Hubei province. Incidence curves for the two datasets were similar, although the non-Hubei curve was shifted to the right (Figure S5), presumably due to the delay between symptom onset and case

reporting (nationwide data did not include symptom onset dates). After shifting the non-Hubei incidence data back by 7 days to align the peaks of the two curves, we find good agreement between the case pair data and non-Hubei data (Figure S6).

5 We estimated R_t using both the case pair incidence data and the time-shifted non-Hubei incidence data, and find good agreement between the two sets of estimates (Figure S7). However, because the non-Hubei data feature larger case numbers, and presumably smaller sampling error, we use the R_t estimates based on the these data for subsequent analysis.

10 *Estimating mean reproduction numbers before and after lockdown*

We used R_t estimates from before and after the lockdown (before Jan. 23 and on/after Jan. 23, respectively) to estimate the mean reproduction numbers for the pre- and post-lockdown periods. In order to reduce the effects of sampling error, we only used R_t estimates for days with at least 10 cases. In addition, because the Wallinga-Teunis method uses future case incidence to estimate the reproduction number, it will tend to underestimate R_t toward the end of a time series due to right-truncation. We therefore calculated the probability of a secondary case developing symptoms by t_{\max} if the primary case developed symptoms at time t , given the serial interval distribution at time t (Figure S8). We discarded R_t estimates beyond the time point at which this probability dropped below 90%.

15 *Estimating reductions in presymptomatic and (post)symptomatic transmission*

Let R_{pre} and R_{post} denote the mean reproduction numbers for the prelockdown and postlockdown periods, respectively. Similarly, let φ_{pre} and φ_{post} denote the relative frequency of presymptomatic transmission for each time period. Then the change in the absolute frequency of presymptomatic transmission (from prelockdown to postlockdown) is given by

$$25 \quad r_{\text{presym}} = \frac{\varphi_{\text{post}}R_{\text{post}} - \varphi_{\text{pre}}R_{\text{pre}}}{\varphi_{\text{pre}}R_{\text{pre}}}$$

30 and the change in the absolute frequency of (post)symptomatic transmission is given by

$$35 \quad r_{\text{postsym}} = \frac{(1 - \varphi_{\text{post}})R_{\text{post}} - (1 - \varphi_{\text{pre}})R_{\text{pre}}}{(1 - \varphi_{\text{pre}})R_{\text{pre}}}$$

Posterior distributions for the changes in presymptomatic and (post)symptomatic transmission were obtained using the posterior distributions for the relative frequency of presymptomatic transmission for each time period.

40 *Software*

All of the analysis for this study was conducted in R (version 3.6.1).

Declarations

45 *Funding:* This work was supported by the National Institute of Allergy and Infectious Diseases (R01AI128344).

Competing interests: The authors declare that no competing interests exist.

5 *Ethical approval:* Because the data utilized for the current project were not collected specifically for the present study and because no one on the study team had access to identifying data, the study was not considered human subjects research and a Harvard Longwood Campus Institutional Review Board application was not required.

Consent to participate: Not applicable

10 *Consent for publication:* Not applicable

Availability of data and material: The serial interval and incidence data are freely available through Open Science Framework (DOI: 10.17605/OSF.IO/EZMKQ).

15 *Code availability:* The code files and model output are freely available through Open Science Framework (DOI: 10.17605/OSF.IO/EZMKQ).

20 *Author contributions:* All authors contributed to study conception and design. MB, CJW and HHC developed the methodology. MB analyzed the data and drafted the manuscript. All authors revised the manuscript and approved the final version.

20 **References**

1. Zhou L, Wu Z, Li Z, et al. 100 Days of COVID-19 Prevention and Control in China. Clin Infect Dis. 2020. doi:10.1093/cid/ciaa725
- 25 2. Zanin M, Xiao C, Liang T, et al. The public health response to the COVID-19 outbreak in mainland China: a narrative review. J Thorac Dis. 2020;12(8):4434-49. doi:10.21037/jtd-20-2363
3. Xu TL, Ao MY, Zhou X, et al. China's practice to prevent and control COVID-19 in the context of large population movement. Infect Dis Poverty. 2020;9(1):115. doi:10.1186/s40249-020-00716-0
- 30 4. Lai S, Ruktanonchai NW, Zhou L, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature. 2020;585(7825):410-3. doi:10.1038/s41586-020-2293-x
5. Tian H, Liu Y, Li Y, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science. 2020;368(6491):638-42. doi:10.1126/science.abb6105
- 35 6. Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. Lancet. 2020;395(10233):1382-93. doi:10.1016/s0140-6736(20)30746-7
7. Kraemer MUG, Yang C-H, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020;368(6490):493-7. doi:10.1126/science.abb4218
- 40 8. World Health Organization. COVID-19 weekly epidemiological update, 1 December 2020: World Health Organization2020.
9. Gatto M, Bertuzzo E, Mari L, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proceedings of the National Academy of Sciences. 2020;117(19):10484-91. doi:10.1073/pnas.2004978117
- 45 10. Soucy J-PR, Sturrock SL, Berry I, et al. Estimating effects of physical distancing on the COVID-19 pandemic using an urban mobility index. medRxiv. 2020:2020.04.05.20054288. doi:10.1101/2020.04.05.20054288

11. Dehning J, Zierenberg J, Spitzner FP, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*. 2020;369(6500):eabb9789. doi:10.1126/science.abb9789
12. Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*. 2020;584(7820):262-7. doi:10.1038/s41586-020-2404-8
13. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. 2020;584(7820):257-61. doi:10.1038/s41586-020-2405-7
14. Banholzer N, van Weenen E, Kratzwald B, et al. Impact of non-pharmaceutical interventions on documented cases of COVID-19. *medRxiv*. 2020:2020.04.16.20062141. doi:10.1101/2020.04.16.20062141
15. Wang Y, Tian H, Zhang L, et al. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ Global Health*. 2020;5(5):e002794. doi:10.1136/bmjgh-2020-002794
16. VoPham T, Weaver MD, Hart JE, Ton M, White E, Newcomb PA. Effect of social distancing on COVID-19 incidence and mortality in the US. *medRxiv*. 2020:2020.06.10.20127589. doi:10.1101/2020.06.10.20127589
17. Matrajt L, Leung T. Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease. *Emerg Infect Dis*. 2020;26(8):1740-8. doi:10.3201/eid2608.201093
18. Islam N, Sharp SJ, Chowell G, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *Bmj*. 2020;370:m2743. doi:10.1136/bmj.m2743
19. Anderson SC, Edwards AM, Yerlanov M, et al. Estimating the impact of COVID-19 control measures using a Bayesian model of physical distancing. *medRxiv*. 2020:2020.04.17.20070086. doi:10.1101/2020.04.17.20070086
20. Haug N, Geyrhofer L, Londei A, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*. 2020. doi:10.1038/s41562-020-01009-0
21. Teslya A, Pham TM, Godijk NG, Kretzschmar ME, Bootsma MCJ, Rozhnova G. Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. *PLoS Med*. 2020;17(7):e1003166. doi:10.1371/journal.pmed.1003166
22. Prem K, Liu Y, Russell TW, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*. 2020;5(5):e261-e70. doi:10.1016/S2468-2667(20)30073-6
23. Davies NG, Klepac P, Liu Y, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. 2020;26(8):1205-11. doi:10.1038/s41591-020-0962-9
24. Koo JR, Cook AR, Park M, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis*. 2020;20(6):678-88. doi:10.1016/S1473-3099(20)30162-6
25. Sun K, Wang W, Gao L, et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*. 2020:eabe2424. doi:10.1126/science.abe2424
26. Ali ST, Wang L, Lau EHY, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science*. 2020;369(6507):1106-9. doi:10.1126/science.abc9004
27. Xu XK, Liu XF, Wu Y, et al. Reconstruction of Transmission Pairs for novel Coronavirus Disease 2019 (COVID-19) in mainland China: Estimation of Super-spreading Events, Serial Interval, and Hazard of Infection. *Clin Infect Dis*. 2020. doi:10.1093/cid/ciaa790
28. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis*. 2020;26(6):1341-3. doi:10.3201/eid2606.200357

29. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med*. 2020;26(5):672-5. doi:10.1038/s41591-020-0869-5
30. Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis*. 2020;20(7):793-802. doi:10.1016/S1473-3099(20)30230-9
- 5 31. Mettler SK, Kim J, Maathuis MH. Clinical onset serial interval and diagnostic serial interval of SARS-CoV-2/COVID-19 in South Korea. *medRxiv*. 2020:2020.05.05.20070946. doi:10.1101/2020.05.05.20070946
32. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis*. 2020;93:284-6. doi:10.1016/j.ijid.2020.02.060
- 10 33. Lauer SA, Grantz KH, Bi Q, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020;172(9):577-82. doi:10.7326/M20-0504
34. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill*. 2020;25(5). doi:10.2807/1560-7917.ES.2020.25.5.2000062
- 15 35. Celeux G, Forbes F, Robert CP, Titterton DM. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006;1(4):651-73. doi:10.1214/06-ba122
36. Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. *J R Soc Interface*. 2019;16(150):20180670. doi:10.1098/rsif.2018.0670
- 20 37. Tindale LC, Stockdale JE, Coombe M, et al. Evidence for transmission of COVID-19 prior to symptom onset. *Elife*. 2020;9. doi:10.7554/eLife.57149
38. Ren X, Li Y, Yang X, et al. Evidence for pre-symptomatic transmission of coronavirus disease 2019 (COVID-19) in China. *Influenza Other Respir Viruses*. 2020. doi:10.1111/irv.12787
- 25 39. Ganyani T, Kremer C, Chen D, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill*. 2020;25(17). doi:10.2807/1560-7917.ES.2020.25.17.2000257
40. Hu S, Wang W, Wang Y, et al. Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *medRxiv*. 2020. doi:10.1101/2020.07.23.20160317
- 30 41. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. 2004;160(6):509-16. doi:10.1093/aje/kwh255
- 35 42. Gostic KM, McGough L, Baskerville E, et al. Practical considerations for measuring the effective reproductive number, Rt. *medRxiv*. 2020. doi:10.1101/2020.06.18.20134858
43. Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis*. 2020;20(8):911-9. doi:10.1016/S1473-3099(20)30287-5
- 40 44. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med*. 2020;173(5):362-7. doi:10.7326/M20-3012
- 45 45. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine*. 2020;382(13):1199-207. doi:10.1056/NEJMoa2001316
46. Tan WYT, Wong LY, Leo YS, Toh M. Does incubation period of COVID-19 vary with age? A study of epidemiologically linked cases in Singapore. *Epidemiol Infect*. 2020;148:e197. doi:10.1017/S0950268820001995