

Supplementary Appendix

Index

Supplementary File S1: Outbreaker model specification	2
Posterior distribution	2
Ward likelihood	3
Supplementary File S2: Supplementary methods	9
Definitions of onset and association	9
Ascertainment probability	9
Symptom onset date imputation	11
Proportion of imported cases	11
Parameterisation of the importations	12
Outbreaker model run specifications	14
Secondary case distribution	14
Supplementary Figure S1. Treemap of all PANGO lineages represented in the Wave 1 and Wave 2 datasets	17
Supplementary Figure S2. Network of all transmission chains in Wave 1 and 2	18
Supplementary Figure S3. Sensitivity analysis of Wave 1	19
Supplementary Figure S4. Sensitivity analysis of wave 2	20
Supplementary Figure S5. Schematic representation of the outbreaker ward model	21
Supplementary Figure S6. Integrating over transmission pathways to a given ward	22
Supplementary Figure S6. Trace of the log-posterior values and in exploratory runs	23
Supplementary Table S1. Notation of outbreaker2 model	24
Supplementary Table S2. Percentage of each onset and association with infectors or onward secondary infections in the outbreaker model	25
Supplementary File S3: Supplementary Authors	26
Supplementary File S4: Additional References	47

Supplementary File S1: Outbreaker model specification

Posterior distribution

The joint posterior distribution of the augmented data and parameters given the data can then be defined as:

$$P(A, \theta|D) \propto P(\theta) \prod_i \Omega_i^1 \Omega_i^2 \Omega_i^3 \Omega_i^4 \Omega_i^5 \Omega_i^6$$

where the likelihood terms model the incubation period (Ω_i^1), generation time (Ω_i^2), mutation events (Ω_i^3), case reporting process (Ω_i^4), contact process (Ω_i^5) and ward infection process (Ω_i^6). Given the absence of contact data, the contact term Ω_i^5 is fixed at a value of 1.

The prior distributions of the parameters are assumed independent such that:

$$P(\theta) = p(\mu)p(\pi)p(\sigma)p(\tau)$$

The very high identity among SARS-COV-2 genetic sequences tends to link all the observations in the data and this results in posterior values for these parameters—for instance, π converges to very high values indicating that the vast majority of the cases were present in our data—that are not compatible with clinical and epidemiological sources of information. To account for this (i) we assigned sensible fixed values to the ward parameters (σ and τ) and the ascertainment parameter (π) (**Supplementary Table S1**) and explored the variability around these values in one way sensitivity analysis; (ii) we modified the original outbreaker genetic likelihood by modelling the SNP distance between direct transmission pairs with a Poisson distribution; this places less likelihood weight on the genetic sequence data than the original formulation and allows a more evenly weighted

integration with ward data. By modelling mutation events as a feature of infection, assuming no within-host diversity and using a per-infectious-generation, per-base mutation rate μ and number of comparable nucleotides $l(s_i, s_{\alpha_i})$, the genetic likelihood Ω_i^3 of the SNP distance $d(s_i, s_{\alpha_i})$ between the samples of a case i and its most recently sampled ancestor α_i separated by κ_i generations is given by:

$$d(s_i, s_{\alpha_i}) \sim \text{Poisson}(\mu \times \kappa_i \times l(s_i, s_{\alpha_i}))$$

The mutation rate (μ) represents a probability and was assigned a Beta distributed prior (**Supplementary Table S1**).

Ward likelihood

Though the infectious process is unobserved, the probability of infection between cases can be estimated from the ward data, more specifically from the pair of wards they are registered in on the most likely inferred day of infection (**Supplementary Figure S5**). We assume that infection is more likely to occur between individuals on the same ward, and model this assumption explicitly by defining the parameter σ as the probability of a primary case infecting a secondary case registered on the same ward on the day of infection, conditional on the primary case having caused an infection. We refer to such events as “within-ward” infections.

The probability of a primary case infecting a secondary case registered on a different ward on the day of infection is therefore $1 - \sigma$, which we refer to as “between-ward” infections. This accounts for uncertainty in using ward data to estimate the infectious process, for example if a patient visits another ward during their stay in the hospital or if viral particles

are transferred between wards via contaminated equipment or ventilation. However, $1 - \sigma$ defines the marginal probability of transmission pairs being registered on *any* two different wards, while the ward likelihood must consider the probability of infection occurring between the two specific wards where the transmission pair is based.

To calculate this probability, the likelihood of infection moving from one ward to another must be defined. Formally, given N^w different wards in the hospital, we define a transition matrix $Q_{k,l}$ ($N^w \times N^w$) describing the probability of a case registered in ward k infecting a case registered in ward l , conditional on a between-ward infection. The rows of this matrix therefore sum to 1, under the assumption that all outgoing infections occur in one of the other wards, and $Q_{k,k} = 0$. given that we lack information on the proximity between wards or the number of staff shared between them, we made the assumption that infection occurs with equal probability to any other ward such that $Q_{k,l} = 1/(N^w - 1)$.

Given the probability of within-ward infection σ and the matrix of transition probabilities $Q_{k,l}$, we define a matrix X as a function of σ as follows:

$$X(\sigma) = \begin{bmatrix} \sigma & (1 - \sigma)Q_{1,2} & \dots & (1 - \sigma)Q_{1,N^w} & (1 - \sigma)Q_{2,1} & \sigma & (1 - \sigma)Q_{2,3} & (1 - \sigma)Q_{2,N^w} & \dots & (1 - \sigma)Q_{3,2} & \sigma & \dots & (1 - \sigma)Q_{N^w,1} & (1 - \sigma)Q_{N^w,2} & \dots & \sigma \end{bmatrix}$$

where individual elements $X(\sigma)[k, l]$ describe the probability of a case in ward k infecting a case in ward l , conditional on the primary case having caused an infection). In other words, this matrix describes the expected proportion of infections originating in ward k that end

up in ward l . Using the ward designation of 0 to indicate a case not being registered in the hospital, we make the assumption that infection can only occur between cases in the hospital by defining $X(\sigma)[k, l] = 0$ if $k = 0$ or $l = 0$. This completes the ward likelihood for direct transmission pairs (**Supplementary Figure S5**), which describes the probability of a primary case on ward k infecting a secondary case on ward l on the day of infection of the secondary case, and is given by $X(\sigma)[k, l]$.

To keep the dimensionality of the ward model fixed, we analytically integrate over unobserved ward states and calculate the probability of the most recent sampled ancestor α_i , registered on a given ward k , causing infection in case i , registered on a given ward l , after κ_i generations of infection. This approach requires knowledge of the registered ward of α_i on the day of infection of the first unobserved case between α_i and i (**Supplementary Figure S5**). However, the default *Outbreaker* model does not impute infection times of unobserved cases and as such this ward cannot be defined. We therefore impute an additional set of augmented data T_i^{anc} , called the ancestral infection time and defined as the infection time of the earliest unobserved case in an unobserved transmission chain separating α_i and i . T_i^{anc} is thus only defined when κ_i is greater than 1. A complete integration of T_i^{anc} into the outbreaker likelihood would require calculating $P(T_i^{anc})$ as part of the generation time likelihood Ω_i^2 . However, this calculation would introduce variable model dimensionality that could only be explored using reversible-jump MCMC^{1,2}. To avoid this complexity, we make the simplifying assumption that $P(T_i^{anc}) = 1$ for $T_{\alpha_i}^{inf} < T_i^{anc} < T_i^{inf}$ and $P(T_i^{anc}) = 0$ otherwise, and keep the original generation time likelihood Ω_i^2 .

Under these assumptions, the ward likelihood can be calculated by integrating over all potential transmission routes between the registered ward of α_i on day T_i^{anc} and the registered ward of i on day T_i^{inf} . For each unobserved case there exist two unobserved ward states, one on the day of infection and one on the day of onward infection (**Supplementary Figure S5**). As cases tend to remain on a single ward for a given period of time, these ward states cannot be considered independent. We therefore condition the later ward state on the earlier ward state by defining a parameter τ as the daily probability of a case being registered on the same ward it was in on the day of infection. If the case is found on a different ward, the specific probability of being transferred from a given ward k to a different ward l is calculated using the same transition probability $Q_{k,l}$ defined above. We therefore assume that ward transfers of unobserved cases follow the same transition probabilities as infections occurring between wards, though this assumption could be easily relaxed if the necessary data to inform these estimates were available. We then define a matrix Y as a function of τ :

$$Y(\tau) = \begin{bmatrix} \tau (1 - \tau)Q_{1,2} & \dots & (1 - \tau)Q_{1,N^w} & (1 - \tau)Q_{2,1} & \tau (1 - \tau)Q_{2,3} & (1 - \tau)Q_{2,N^w} & \dots & (1 - \tau)Q_{3,2} & \tau & \dots & (1 - \tau)Q_{N^w,1} & (1 - \tau)Q_{N^w,2} & \dots & \tau \end{bmatrix}$$

where individual elements $Y(\tau)[k, l]$ indicate the probability of an unobserved case being transferred from a given ward k on the day of infection to a given ward l on the day of onward infection.

Given this parameterization it is possible to integrate over the unobserved transmission pathways linking a given ward k to a given ward l over κ_i generations of infection. We do so

by first defining all potential transmission routes between any two wards in a *single* unobserved generation of infection (**Supplementary Figure S6**). Fast integration over these transmission routes that accounts for all combinations of ward transfers and infection events can be achieved by matrix multiplication of $Y(\tau)$ and $X(\sigma)$ such that the elements of $Y(\tau) \times X(\sigma)$ describe the probability of an unobserved case that was infected on ward k causing infection on ward l after one unobserved generation of infection.

The probability of each route (**Supplementary Figure S6**) is conditional on the primary case becoming infected on that given ward in the previous generation of infection. The matrix product $Y(\tau) \times X(\sigma)$ can therefore be multiplied by itself in a recursive manner to integrate over transmission pathways for greater numbers of unobserved cases, where each multiplication by $Y(\tau) \times X(\sigma)$ represents the movement of infection in one unobserved generation.

This recursive approach can be initialised as the ward states of the most recent sampled ancestor α_i are observed. The probability of the first unobserved case becoming infected on any given ward i , given its infector α_i is registered in the observed ward k on the day of onward infection, is simply given by $X(\sigma)[k, l]$. We therefore define a matrix Z as a function of σ , τ and κ :

$$Z(\sigma, \tau, \kappa_i) = X(\sigma) \times (Y(\tau) \times X(\sigma))^{\kappa_i - 1}$$

where \times indicates the matrix multiplication operator and $^{\wedge}$ the exponentiation operator.

The first term calculates the probability of movement of infection from the sampled ancestor α_i to the first unobserved case, while the exponentiated term accounts for

successive unobserved generations of infection. Individual elements $Z(\sigma, \tau, \kappa_i)[k, l]$ therefore describe the probability of infection originating in a given ward k and ending up on ward l after κ_i generations of infection.

Given these probabilities, the ward likelihood Ω_i^6 can then be defined for a given case i as:

$$\begin{aligned}\Omega_i^6 &= \{p\left(W_{i,T_i^{inf}}|\alpha_i, T_i^{inf}, W_{\alpha_i, T_i^{inf}}, \kappa_i, \sigma, \tau\right) \quad \kappa_i \\ &= 1 p\left(W_{i,T_i^{inf}}|\alpha_i, T_i^{inf}, T_i^{anc}, W_{\alpha_i, T_i^{anc}}, \kappa_i, \sigma, \tau\right) \quad \kappa_i > 1\end{aligned}$$

and calculated as:

$$\Omega_i^6 = \{Z(\sigma, \tau, \kappa_i) \left[W_{\alpha_i, T_i^{inf}}; W_{i, T_i^{inf}} \right] \quad \kappa_i = 1 \quad Z(\sigma, \tau, \kappa_i) \left[W_{\alpha_i, T_i^{anc}}; W_{i, T_i^{inf}} \right] \quad \kappa_i > 1\}$$

The ward likelihood therefore describes the probability of infection originating in the ward that α_i is registered in on the day of onward infection (T_i^{inf} if $\kappa_i = 1$ or T_i^{anc} if $\kappa_i > 1$), and ending up in the ward that i is registered in on its day of infection (T_i^{inf}) after κ_i generations of infection.

Supplementary File S2: Supplementary methods

Definitions of onset and association

All COVID positive cases in patients were classified in 5 groups as in ⁴:

- (1) Community-Onset Community-Associated: Positive specimen date up to 14 days before, or within 2 days after, hospital admission, with no discharge from hospital in 14 days before specimen date.
- (2) Community-Onset Suspected Healthcare-Associated: Positive specimen date up to 14 days before, or within 2 days after, hospital admission, with discharge from hospital in 14 days before specimen date.
- (3) Hospital-Onset Suspected Healthcare-Associated: Positive specimen dated 8-14 days after hospital admission; or specimen date 3-14 days after admission, with discharge from hospital in 14 days before specimen date.
- (4) Hospital-Onset Intermediate Healthcare-Associated: Positive specimen date 3-7 days after hospital admission, with no discharge from hospital in 14 days before specimen date.
- (5) Hospital-Onset Healthcare-Associated: Positive specimen date to 15 or more days after hospital admission.

Ascertainment probability (π)

The ascertainment parameter, π , was estimated by combining (i) the proportion of known cases which have high quality sequence, (ii) the proportion of these cases where their location is known and (iii)

the proportion of cases which are undetected. While the first two of these proportions were known, the third was more uncertain.

Wave 1 estimate

2,288 patient (1,184) and staff (1,104) cases tested SARS-CoV-2 positive at Sheffield Teaching Hospitals NHS Foundation Trust. 886/1184 (74.8%) patients and 842/1104 (76.3%) staff cases were sequenced to a genome coverage of greater than 90%. 220 cases were excluded because they were outpatients (106), non-clinical staff (37), external staff (47) or community based staff (30). 206 staff cases were excluded because their location of work was unknown. After exclusion, 86.4% (1,302/1,508) of staff and inpatient sequences were included. Combining the proportion of sequenced staff and inpatients known cases and included cases resulted in an ascertainment parameter of 65.4% ($0.755 \times 0.864 \times 100$). Assuming 80% of cases were detected, the reporting estimate for the model would be 52.3%.

Wave 2 estimate

2,021 patient (1183) and staff (838) cases tested positive during this time period. 669/1183 (56.6%) patient cases and 651/838 (77.7%) staff cases were sequenced to a genome coverage of greater than 90%. 332 cases were excluded because they were household contacts (141), non-clinical staff (74), community based staff (28) or outpatients (89). 109 staff cases were excluded because their location of work was unknown. After exclusion 89.0% (879/988) of sequenced staff and inpatient cases were included. Applying the same rationale as above results in a reporting parameter of 46.5%.

Based on these estimates, a sensitivity analysis was carried out varying the parameter from 20% to 60% for both waves.

Symptom onset date imputation

For 30.1% of wave 1 and 22.0% of wave 2 individuals, the symptom onset was not available. To account for these missing data, we estimated the distributions of delay (in days) between symptom onset and testing for inpatients and staff for whom this delay was known. For inpatients, we further stratified these distributions by six infection categories: Community Onset–Community Associated, Community Onset–Suspected Healthcare Associated, Hospital Onset–Healthcare Associated, Hospital Onset–Indeterminate Healthcare Associated, Hospital Onset–Suspected Healthcare Associated and undefined³. For staff, most were classified into a ‘not admitted’ category, with those who required admission stratified into the six categories listed above. All negative delays were removed from these 13 distributions. Next, for each individual for whom a date of onset was not recorded, we randomly sampled a delay between onset and testing from the respective known distributions associated with each infection category and individual type (staff or inpatient). If a distribution was unable to be defined due to absence of data, we instead sampled from the undefined patient or staff category. Finally, we calculated the imputed onset date of individuals with a sampled delay between symptom onset to testing by subtracting the sample delay from the recorded date of testing.

Proportion of imported cases

The precise proportion of imported cases was unknown and this uncertainty was captured as part of the final global sensitivity analysis. The proportion of imported patient cases was estimated from the admission date, symptom onset date and the incubation period. As we

were unable to accurately estimate the proportion of imports for staff the sampling distributions outlined below include a staff import proportion ranging from 0% to 100%. The rationale for the Wave 1 and Wave 2 import proportions are outlined below.

Wave 1 import proportion

The estimated import proportion for patients in Wave 1 was 80% (~624/780; 95% confidence interval: 76% - 83%). Combining the inpatient and staff estimate (261/522, 50%) gives an estimated import percentage of ~68% (885/1302). Considering both extreme scenarios of no staff being imports and all staff being imports gives a range in the import percentage of ~48% to ~88%. To account for this uncertainty in the final results the import proportion for Wave 1 is varied 500 times, sampling from a normal distribution with mean 0.7 and standard deviation 0.075.

Wave 2 import proportion

Combining the patient (63%, ~365/580; 95% confidence interval: 58% - 67%) and staff (150/299, 50%) estimate gives an estimated import percentage of ~59% (515/879). Considering both extreme scenarios as above gives a range of ~42% to ~76%. The import proportion for wave 2 is sampled from the normal distribution with mean 0.6 and standard deviation 0.06.

Parameterisation of the importations

The *Outbreaker2* model requires the number of imports (i.e. community-associated infections) to be known in advance. Here, we relaxed this assumption by integrating over the distribution of the number of imports in independent MCMC chains, for each of which the

number of imports was randomly drawn from a binomial distribution with N trials and a probability drawn from a Beta distribution (see proportion of imported cases section below). We define N as the total number of cases in the dataset, and the mean of the Beta distribution (p) as the probability across all cases that they are community-acquired. We first calculated, for each patient, the probability that they acquired the infection in the community as the probability that their incubation period was longer than their delay between admission and symptom onset. We then calculated the average patient-specific probability of a community-acquired infection. As admission data is unavailable for staff, we set bounds for the fraction of staff who acquired infection in the community as between 0 and 1. We then calculated the fraction of community-acquired infections across the whole dataset as the weighted sum of the patient-specific and the staff-specific probabilities as described in the proportion of imported cases section above.

To assign the identity of the imports, we calculated the mean individual likelihood $\Omega_i = \Omega_i^1 \Omega_i^2 \Omega_i^3 \Omega_i^4 \Omega_i^5 \Omega_i^6$ of case i being infected by another individual observed in the dataset in a pre-run of the MCMC (1100 iterations with a thinning frequency of 1/50 and a burn-in of 500 iterations). For each MCMC chain in the main analysis, the randomly drawn number of imports were randomly assigned identities on the basis of the inverse of the individual likelihood Ω_i .

Outbreaker model run specifications

Outbreaker was run in parallel integrating across uncertainty in the symptom onset data and the number and identity of imports. Specifically, we used a global sensitivity analysis in our base case analysis via an ensemble approach in which we generated a total of 10,000 posterior transmission networks. These 10,000 samples were generated from a sample of 100 imputed symptom onset datasets. In addition we drew five samples of the number and identity of imports and ran an MCMC across all 500 combinations. We then sampled 20 posterior networks from each of the converged chains of these 500 MCMC runs. Each parallel MCMC chain was run for 1100 iterations with a thinning frequency of 1/50 and a burn-in of 100 iterations. We ran the analysis using a timescale of quarter-day increments consistent with the ward location data being available to the nearest 6-hour interval.

Secondary case distribution

Our simulations calculate a posterior sample of the distribution of the number of observed secondary cases, X , by evaluating, for each case, the total number of connected cases (either direct transmission or with intermediate unsampled individuals). To calculate the true number of secondary cases (Z), including both observed and unobserved, we define the conditional probability,

$$\begin{aligned} P(Z = j | X = i) &= P(Y = j - i) && \text{for } j > i \\ P(Z = j | X = i) &= 1 - \sum_{j>i} (1 - \pi)^{j-i} && \text{for } j = i \\ P(Z = j | X = i) &= 0, && \text{else} \end{aligned}$$

We then calculate the unconditional probability distribution of true cases using the expression:

$$P(Z = j) = \alpha \sum_{i \leq j} P(Z = j | X = i)P(X = i)$$

where α is the normalising constant to constrain $\sum_{j \geq 0} P(Z = j) = 1$.

We define Y as the number of secondary cases that both them and all their descendants are unsampled. Specifically,

$$P(Y = k) = \left(\sum_{c \geq 1} P(C = c)(1 - \pi)^c \right)^k \quad \text{for } k > 0$$

$$P(Y = 0) = 1 - P(Y \geq 1)$$

where π is the reporting probability and $P(C)$ is the probability distribution of the number of cases in a cluster of descendants (including itself). Specifically, $P(C)$ can be defined in terms of the number of true secondary cases, Z , as:

$$P(C = 1) = P(Z = 0),$$

$$P(C = 2) = P(Z = 1),$$

$$P(C = 3) = P(Z = 2) + P(Z = 1)^2, \text{ etc.}$$

Assuming $P(Z = k)(1 - \pi)^j \approx 0$ when $j > 2, k > 1$, we can write,

$$P(Y = i) = ((1 - \pi)P(Z = 0) + (1 - \pi)^2 P(Z = 1))^i \text{ for } i > 0$$

$$P(Y = 0) = 1 - P(Y \geq 1)$$

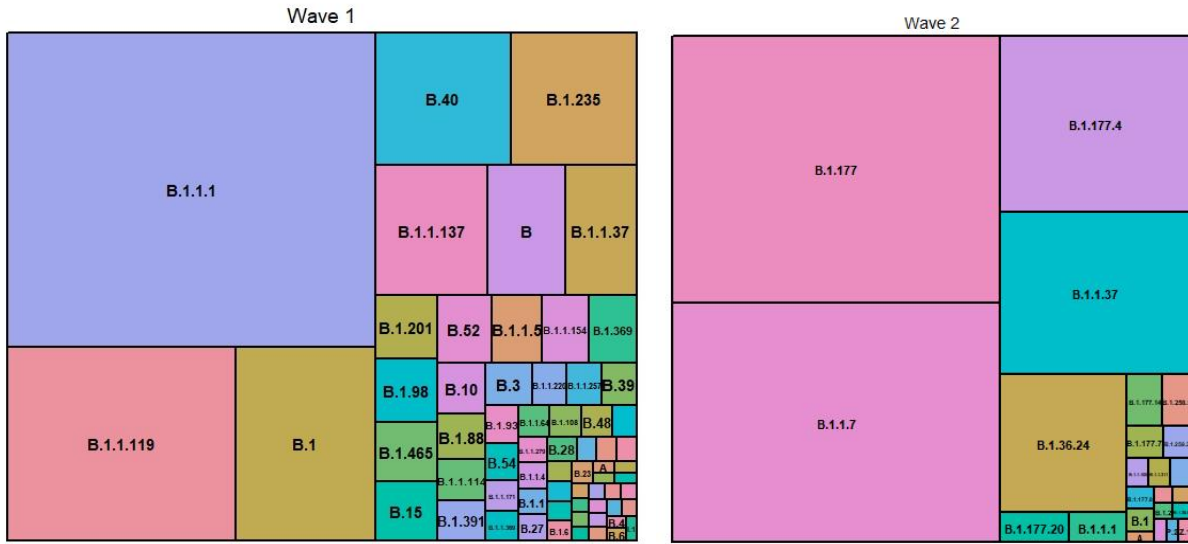
Now, we define

$$P(Y = k | p, q) = ((1 - \pi)p + (1 - \pi)^2 q)^k$$

$$P(Y = 0 | p, q) = 1 - P(Y \geq 1 | p, q)$$

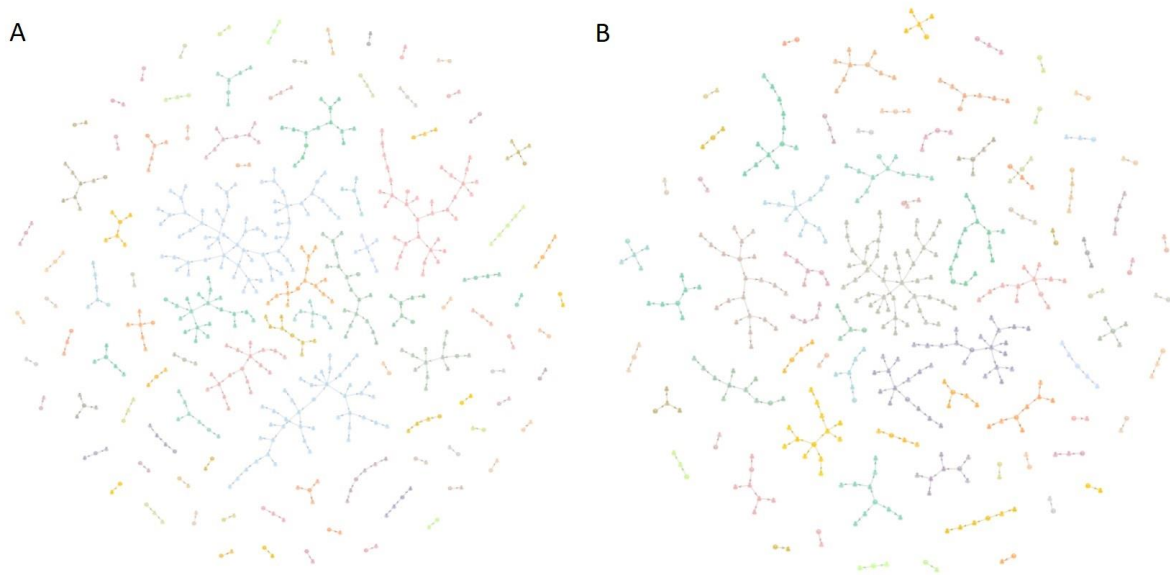
and we generate samples from the two distributions $P(X)$ and $P(Y | p, q)$ to generate a sample of Z and, from this sample, we estimate $P(Z | p, q)$. We then calculate the unique p and q that satisfies $P(Y = k | p, q) = ((1 - \pi)P(Z = 0) + (1 - \pi)^2P(Z = 1))^k$.

We therefore calculate a secondary case distribution, $P(Z_h = i)$ for each sampled transmission tree, h , in the posterior distribution.



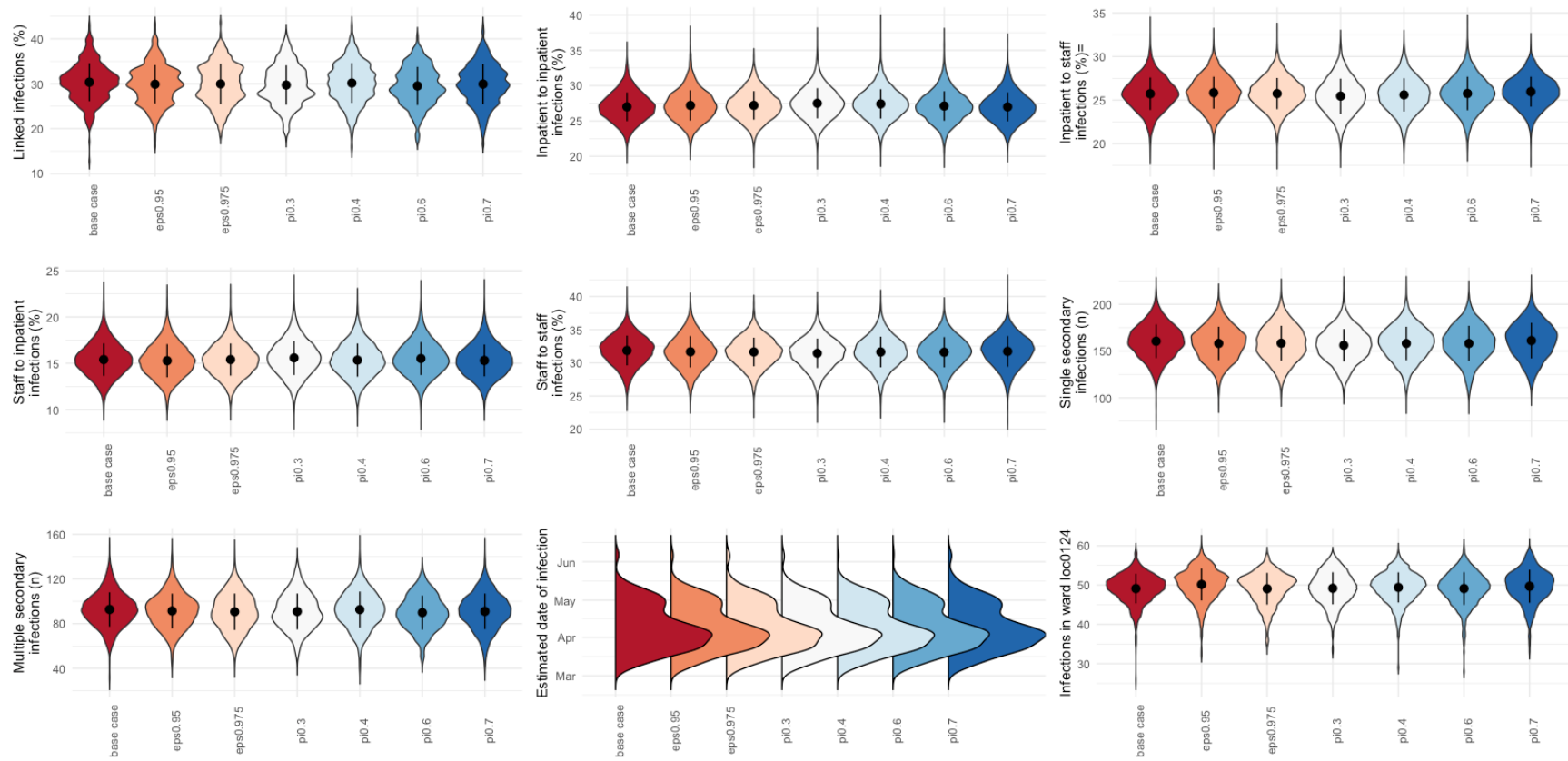
Supplementary Figure S1. Treemap of all PANGO lineages represented in the Wave 1 and Wave 2 datasets.

The size of each rectangle is proportional to the number of each lineage in the dataset.



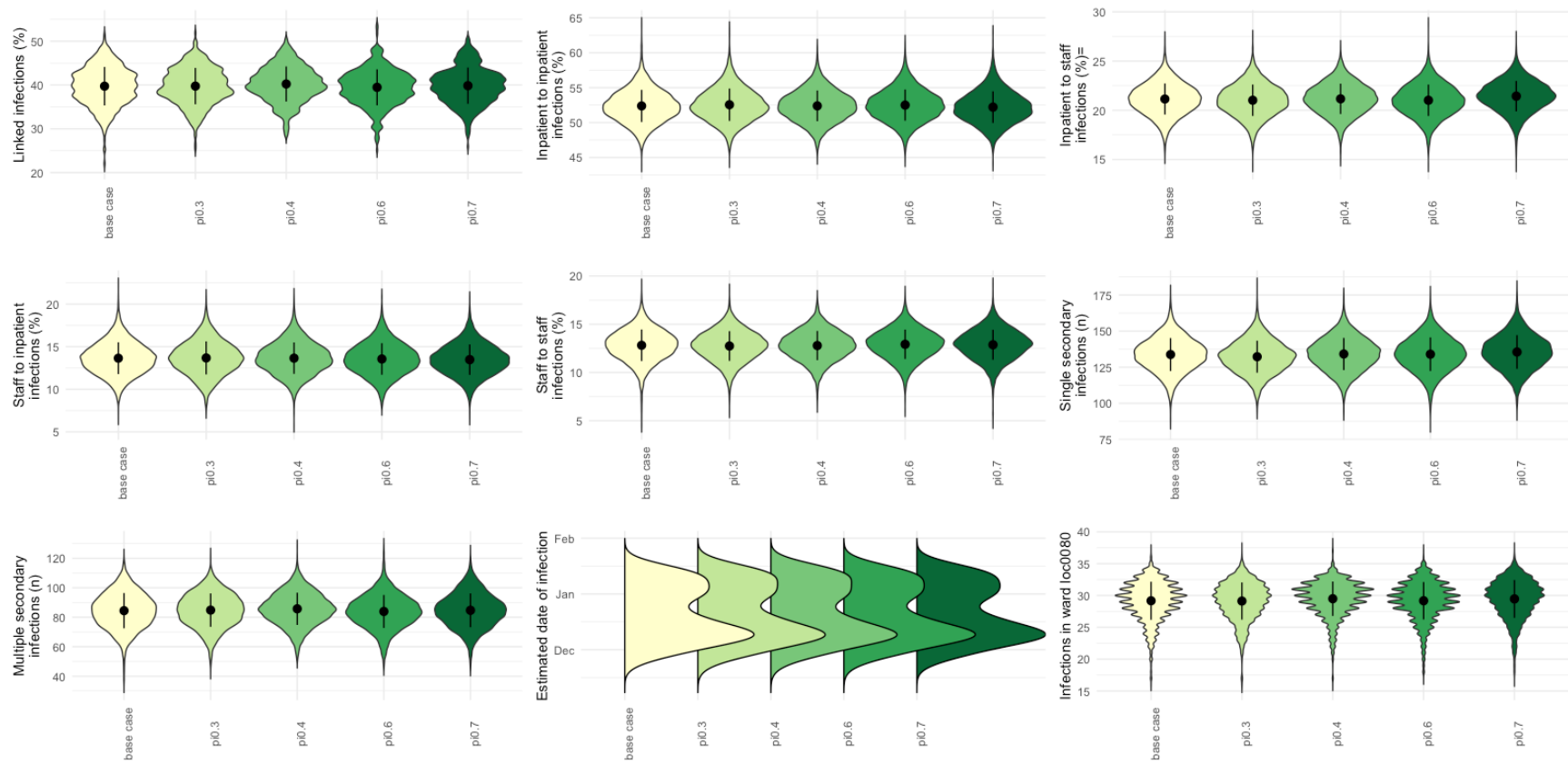
Supplementary Figure S2. Network of all transmission chains in Wave 1 and 2

(A) All transmission chains from a representative network in Wave 1. (B) All transmission chains from a representative network in Wave 2. Each colour represents a different transmission chain.



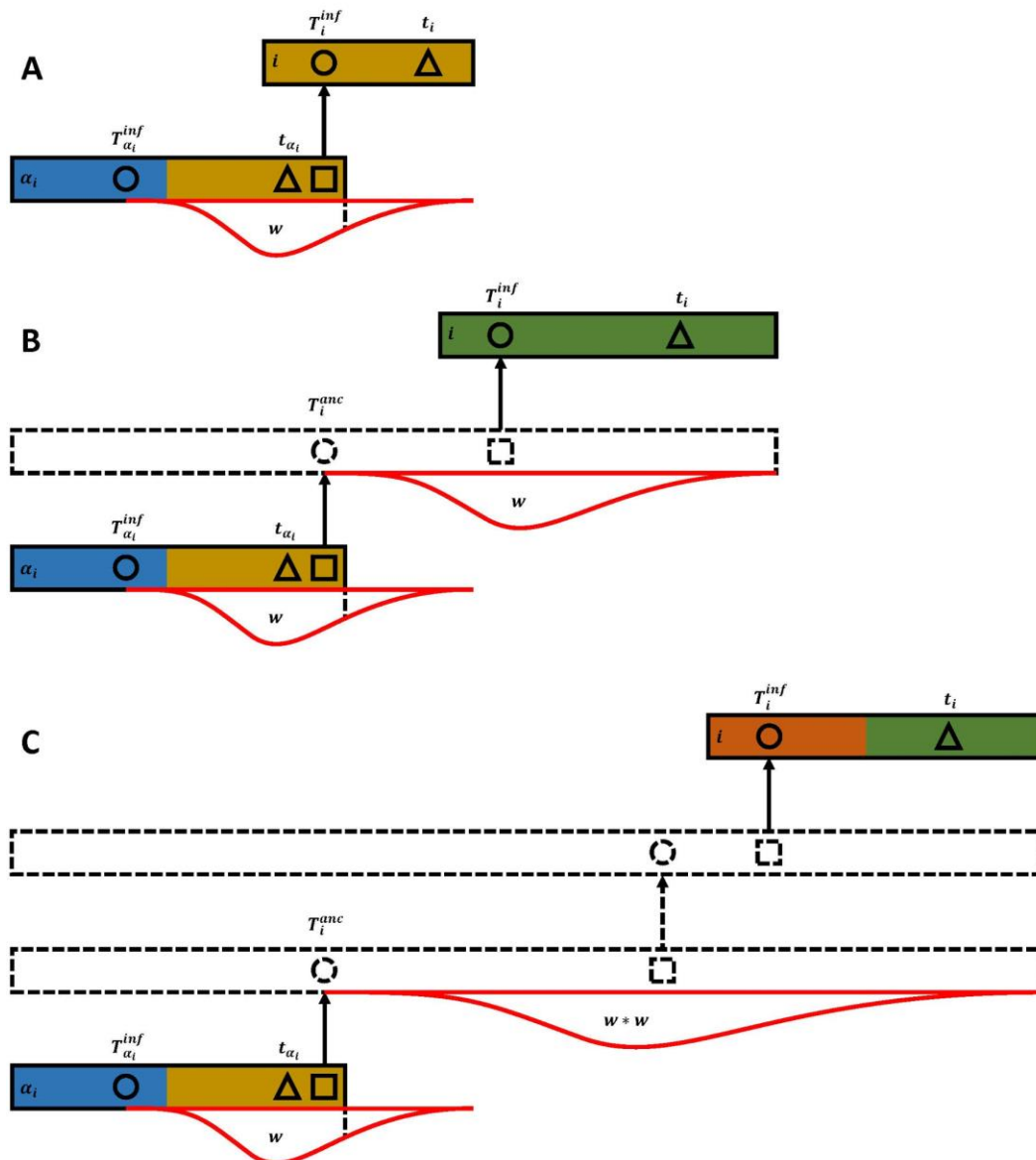
Supplementary Figure S3. Sensitivity analysis of Wave 1.

eps (σ) and pi (π) refer to the probability of α_i and i being registered on the same ward on the day of infection and the proportion of cases sampled in the outbreak, respectively (as in the notation Supplementary Table S1).



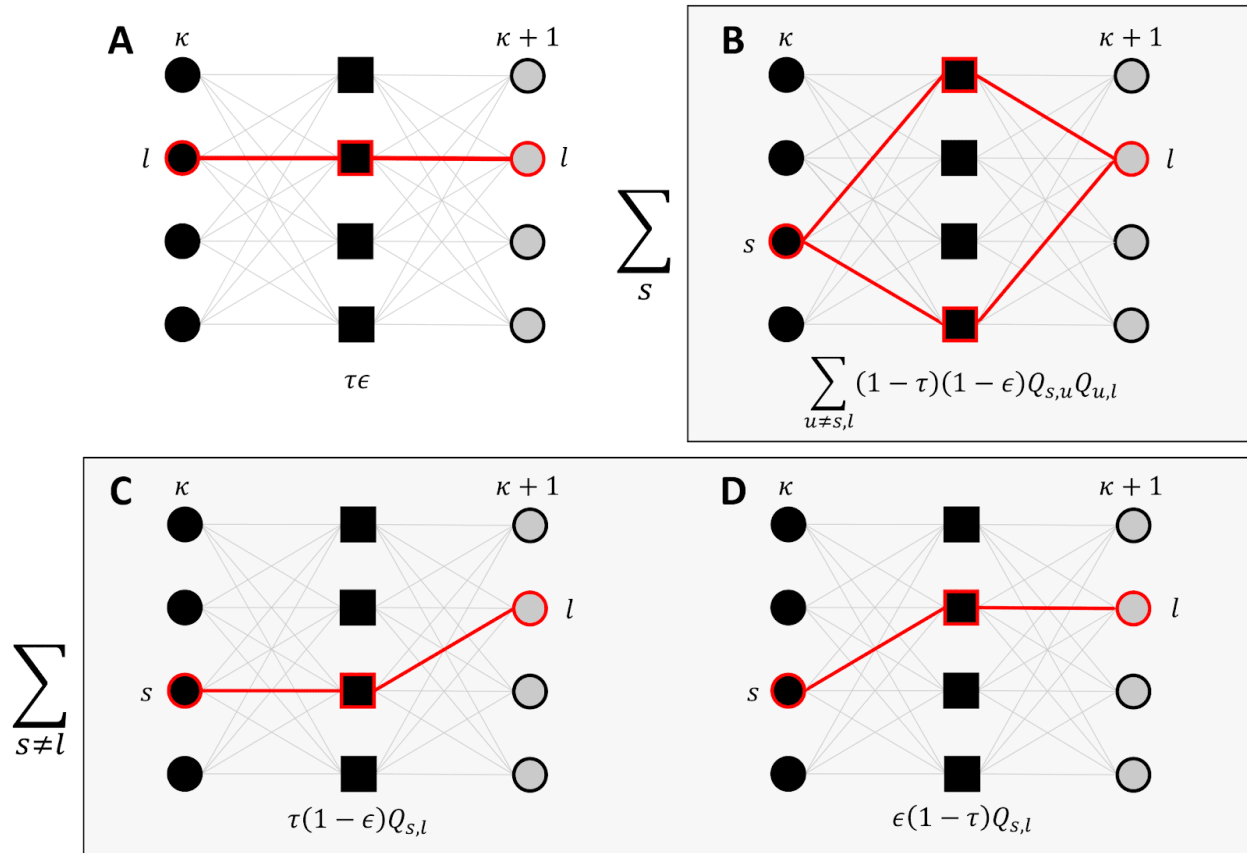
Supplementary Figure S4. Sensitivity analysis of wave 2.

π (π) refers to the proportion of cases sampled in the outbreak as in the notation Supplementary table S1.



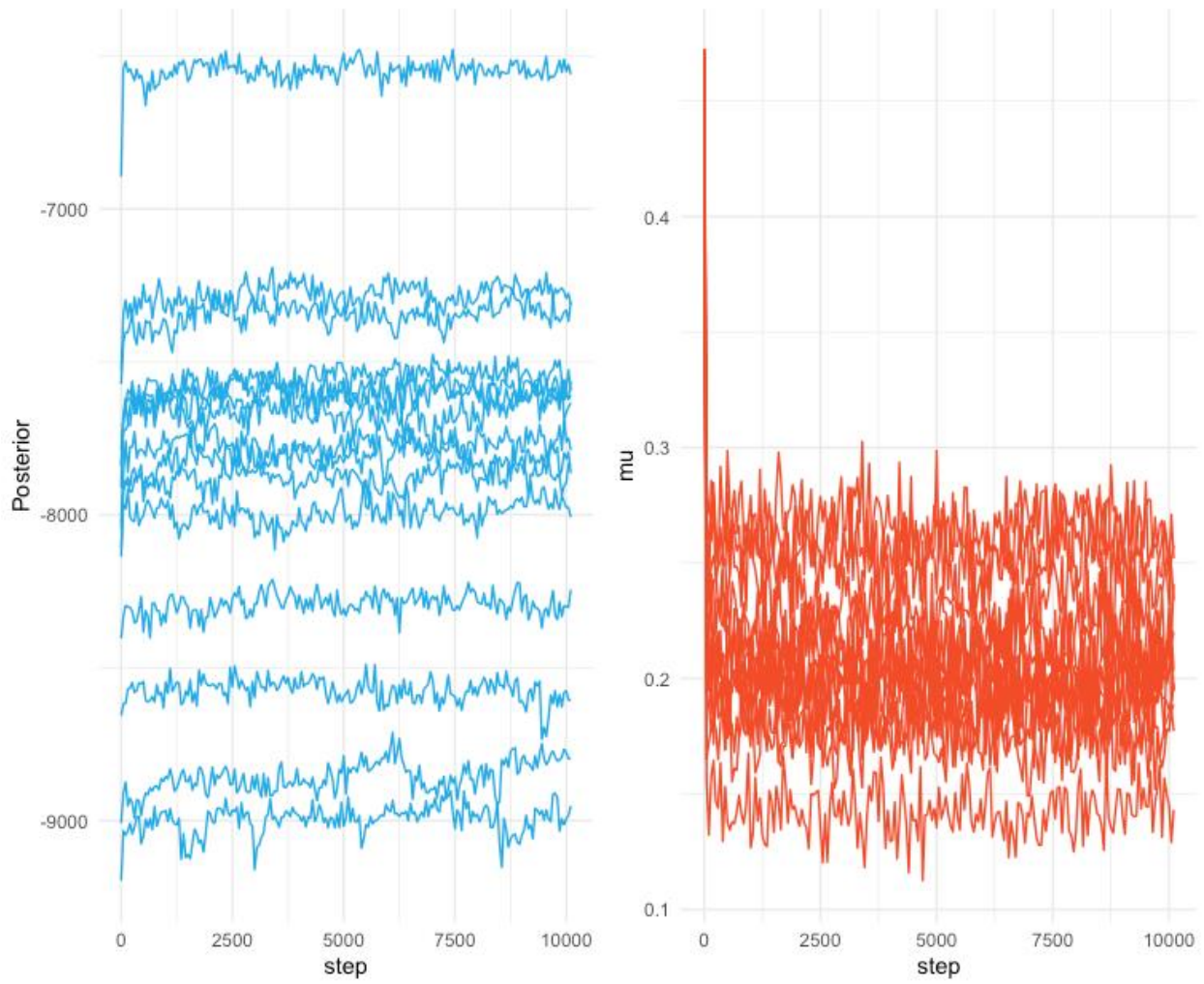
Supplementary Figure S5. Schematic representation of the outbreak ward model

Rows represent individual cases, colours indicate the ward a case is registered in, dotted outlines indicate unobserved cases. Circles represent infection times, triangles dates of symptom onset, squares dates of onward infection. Red curved lines indicate the generation time distribution. (A) A case i and its infector are both sampled. The wards of the infectee and infector are directly observed and the probability of infection between these wards on the day of infection can be directly calculated. (B) A case i and its most recent sampled ancestor α_i are separated by an unobserved case. The ancestral infection time T_i^{anc} is imputed and the potential ward states of the unobserved case on the day of infection and day of onward infection are integrated over. (C) A case i and its most recent sampled ancestor α_i are separated by two unobserved cases. The ancestral infection time T_i^{anc} is imputed and the ward states of both unobserved cases integrated over.



Supplementary Figure S6. Integrating over transmission pathways to a given ward l .

Each row represents a ward. Black circles represent the ward the primary case is infected on. Black squares represent the ward of the primary case on the day of onward infection. Grey circles represent the ward the secondary case is infected on. A case remains on the ward it was infected on with probability τ . An infection occurs within a ward with probability σ . All probabilities described in this figure are conditional on the primary case becoming infected on that given ward. A) The primary case is infected on ward l , remains on the same ward and infection occurs within the ward. B) The primary case is infected on any ward s , moves to an intermediate ward that is not s or l , and infects a case on ward l . The probability of this scenario must integrate over all potential initial wards s . C) The primary case is infected on a ward s that is not l , remains on ward s and causes infection in ward l . D) The primary case is infected on a ward s that is not l , moves to ward l and causes infection within ward l . The probability of scenarios C) and D) must integrate over all potential initial wards that are not l . Fast integration over the transmission pathways given in panels A-D can be achieved by the matrix multiplication $Y(\tau) \times X(\sigma)$.



Supplementary Figure S6. Trace of the log-posterior values and μ in exploratory runs

Supplementary Table S1. Notation of *outbreaker2* model

Symbol	Type	Description	Value (when applicable)
i	Data	Index of cases	
N	Data	Number of cases in the sample	
s_i	Data	Sequence of case i	
t_i	Data	Collection date of s_i	
$W_{i,t}$	Data	Registered ward of case i on day t	
w	Function	Generation time distribution	Discrete Gamma (shape = 4.83, scale = 1.05) ^{5,6}
f	Function	Incubation period distribution	Discrete Gamma (shape = 5.89, scale=1.07) ^{5,6}
$d(s_i, s_j)$	Function	Number of mutations between s_i and s_j	
$l(s_i, s_j)$	Function	Number of comparable nucleotide positions between s_i and s_j	
α_i	Augmented data	Index of the most recent sampled ancestor of case i	
κ_i	Augmented data	Number of generations between α_i and i	
T_i^{inf}	Augmented data	Date of infection of i	
T_i^{anc}	Augmented data	Date of infection of the earliest unobserved case in an unobserved transmission chain separating α_i and i	
μ	Parameter	Mutation rate of SARS-CoV-2, per site per generation of infection	Beta ($\alpha = 89.9, \beta = 809.1$) ⁷
π	Parameter	Proportion of cases sampled in the outbreak	0.5
σ	Parameter	Probability of α_i and i being registered on the same ward on the day of infection	0.99
τ	Parameter	Daily probability of an unobserved case being registered on the same ward it was in on the day of infection	

Supplementary Table S2. Percentage of each onset and association with infectors or onward secondary infections in the outbreaker model

	Assigned infector (hospital acquired infections) % (95% Credible interval)		Secondary infections % (95% Credible interval)	
	Wave 1	Wave 2	Wave 1	Wave 2
Community onset community associated	0	0	3.9 (2.1-6.0)	3.5 (1.7-5.9)
Community onset suspected healthcare associated	35.0 (24.1-45.6)	42.2 (31.3-52.2)	18.1 (10.1-26.6)	24.0 (14.9-32.8)
Hospital onset indeterminate hospital acquired	32.1 (21.3-41.0)	52.3 (41.6-61.8)	27.0 (18.9-35.1)	39.1 (29.5-47.7)
Hospital onset suspected hospital acquired	71.3 (56.8-82.4)	83.6 (69.3-93.2)	59.6 (45.9-72.1)	52.4 (42.7-61.8)
Hospital onset hospital acquired	69.8 (55.0-80.0)	77.4 (61.0-89.0)	45.2 (35.0-55.0)	51.2 (41.0-60.0)

Supplementary File S3: Supplementary Authors

Sheffield COVID-19 Genomics Group

Alison Cope, Nasar Ali, Rasha Raghei, Joe Heffer, Nikki Smith, Max Whiteley, Manoj Pohare, Samantha E. Hansford, Luke R. Green, Dennis Wang, Michael Anckorn, Adrienn Angyal, Rebecca Brown, Hailey Hornsby, Mehmet Yavuz, Danielle C. Groves, Paul J. Parsons, Rachel M. Tucker, Magdalena B. Dabrowska, Thomas Saville, Jose Schutter and Matthew D. Wyles

CMMID COVID-19 Working Group

Mihaly Koltai, Kerry LM Wong, Joel Hellewell, Emilie Finch, Matthew Quaife, James D Munday, Timothy W Russell, Kaja Abbas, Oliver Brady, Nicholas G. Davies, Alicia Rosello, Gwenan M Knight, Simon R Procter, William Waites, Nikos I Bosse, Stefan Flasche, Paul Mee, Rachel Lowe, Kathleen O'Reilly, Carl AB Pearson, Christopher I Jarvis, W John Edmunds, Fiona Yueqian Sun, Amy Gimma, Samuel Clifford, Rosanna C Barnard, Rachael Pung, Kiesha Prem, Sam Abbott, Sebastian Funk, Akira Endo, Sophie R Meakin, Yang Liu, Adam J Kucharski, Graham Medley, Ciara V McCarthy, Yalda Jafari, Mark Jit, Damien C Tully, Hamish P Gibbs, Lloyd A C Chapman, Billy J Quilty, Frank G Sandmann, David Hodgson and Rosalind M Eggo.

The COVID-19 Genomics UK (COG-UK) consortium

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Samuel C Robson PhD ¹³.

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Prof Nicholas J Loman PhD ⁴¹ and Dr Thomas R Connor PhD ^{10, 69}.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Tanya Golubchik PhD ⁵.

Funding acquisition, Metadata curation, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Rocio T Martinez Nunez PhD ⁴².

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:

Dr Catherine Ludden PhD ⁸⁸.

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:

Dr Sally Corden PhD ⁶⁹.

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:

Ian Johnston ⁹⁹ and Dr David Bonsall PhD ⁵.

Funding acquisition, Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:

Prof Colin P Smith PhD ⁸⁷ and Dr Ali R Awan PhD ²⁸.

Funding acquisition, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Giselda Bucca PhD ⁸⁷.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:

Dr M. Estee Torok FRCP ^{22, 101}.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:

Dr Kordo Saeed MD/ FRCPath ^{81, 110} and Dr Jacqui A Prieto PhD ^{83, 109}.

Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:

Dr David K Jackson PhD ⁹⁹.

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Dr William L Hamilton PhD ²².

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Visualisation:

Dr Luke B Snell MSc/ MBBS ¹¹.

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:

Dr Catherine Moore ⁶⁹.

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:

Dr Ewan M Harrison PhD ^{99, 88}.

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:

Dr Sonia Goncalves PhD ⁹⁹ and Dr Leigh M Jackson Ph.D. ⁹¹.

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:

Prof Ian G Goodfellow PhD ²⁴, Dr Derek J Fairley PhD ^{3, 72}, Prof Matthew W Loose PhD ¹⁸ and Joanne Watkins MSc ⁶⁹.

Leadership and supervision, Metadata curation, Samples and logistics, and Software and analysis tools:

Rich Livett MSc ⁹⁹.

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:

Dr Samuel Moses MD ^{25, 106}.

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:

Dr Roberto Amato PhD ⁹⁹, Dr Sam Nicholls PhD ⁴¹ and Dr Matthew Bull PhD ⁶⁹.

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:

Prof Darren L Smith PhD ^{37, 58, 105}.

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Jeff Barrett PhD ⁹⁹ and Prof David M Aanensen PhD ^{14, 114}.

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:

Dr Martin D Curran PhD ⁶⁵, Dr Surendra Parmar PhD ⁶⁵, Dr Dinesh Aggarwal MRCP ^{95, 99, 64} and Dr James G Shepherd MBChB/MRCP ⁴⁸.

Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:

Dr Matthew D Parker PhD ⁹³.

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:

Dr Sharon Glaysher PhD ⁶¹.

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Matthew Bashton PhD ^{37, 58}, Dr Anthony P Underwood PhD ^{14, 114}, Dr Nicole Pacchiarini PhD ⁶⁹ and Dr Katie F Loveson PhD ⁷⁷.

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dr Alessandro M Carabelli PhD ⁸⁸.

Funding acquisition, Leadership and supervision, and Metadata curation:

Dr Kate E Templeton PhD ^{53, 90}.

Funding acquisition, Leadership and supervision, and Project administration:

Dr Cordelia F Langford PhD ⁹⁹, John Sillitoe BEng ⁹⁹, Dr Thushan I de Silva PhD ⁹³ and Dr Dennis Wang PhD ⁹³.

Funding acquisition, Leadership and supervision, and Sequencing and analysis:

Prof Dominic Kwiatkowski ^{99, 107}, Prof Andrew Rambaut DPhil ⁹⁰, Dr Justin O'Grady PhD ^{70, 89} and Dr Simon Cottrell PhD ⁶⁹.

Leadership and supervision, Metadata curation, and Sequencing and analysis:

Prof Matthew T.G. Holden PhD ⁶⁸ and Prof Emma C Thomson PhD/FRCP ⁴⁸.

Leadership and supervision, Project administration, and Samples and logistics:

Dr Husam Osman PhD ^{64, 36}, Dr Monique Andersson PhD ⁵⁹, Prof Anoop J Chauhan ⁶¹ and Dr Mohammed O Hassan-Ibrahim PhD/FRCPath ⁶.

Leadership and supervision, Project administration, and Sequencing and analysis:

Dr Mara Lawniczak ⁹⁹.

Leadership and supervision, Samples and logistics, and Sequencing and analysis:

Prof Ravi Kumar Gupta PhD ^{88, 113}, Dr Alex Alderton PhD ⁹⁹, Dr Meera Chand ⁶⁶, Dr Chrystala Constantinidou PhD ⁹⁴, Dr Meera Unnikrishnan PhD ⁹⁴, Prof Alistair C Darby PhD ⁹², Prof Julian A Hiscox PhD ⁹² and Prof Steve Paterson PhD ⁹².

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:

Dr Inigo Martincorena ⁹⁹, Prof David L Robertson PhD ⁴⁸, Dr Erik M Volz PhD ³⁹, Dr Andrew J Page PhD ⁷ and Prof Oliver G Pybus DPhil ²³.

Leadership and supervision, Sequencing and analysis, and Visualisation:

Dr Andrew R Bassett PhD ⁹⁹.

Metadata curation, Project administration, and Samples and logistics:

Dr Cristina V Ariani PhD ⁹⁹, Dr Michael H Spencer Chapman MBBS ^{99, 88}, Dr Kathy K Li MBChB/FRCPath ⁴⁸, Dr Rajiv N Shah BMBS/MRCP/MSc ⁴⁸, Dr Natasha G Jesudason MBChB MRCP FRCPath ⁴⁸ and Dr Yusri Taha MD/PhD ⁵⁰.

Metadata curation, Project administration, and Sequencing and analysis:

Martin P McHugh MSc ⁵³, Dr Rebecca Dewar PhD ⁵³.

Metadata curation, Samples and logistics, and Sequencing and analysis:

Dr Aminu S Jahun PhD ²⁴, Dr Claire McMurray PhD ⁴¹, Ms Sarojini Pandey MSc ⁸⁴, Dr James P McKenna PhD ³, Dr Andrew Nelson PhD ^{58, 105}, Dr Gregory R Young PhD ^{37, 58}, Dr Clare M McCann PhD ^{58, 105} and Mr Scott Elliott ⁶¹.

Metadata curation, Samples and logistics, and Visualisation:

Ms Hannah Lowe MSc ²⁵.

Metadata curation, Sequencing and analysis, and Software and analysis tools:

Dr Ben Temperton Ph.D. ⁹¹, Dr Sunando Roy PhD ⁸², Dr Anna Price PhD ¹⁰, Dr Sara Rey PhD ⁶⁹ and Mr Matthew Wyles ⁹³.

Metadata curation, Sequencing and analysis, and Visualisation:

Stefan Rooke MSc ⁹⁰ and Dr Sharif Shaaban PhD ⁶⁸.

Project administration, Samples and logistics, Sequencing and analysis:

Dr Mariateresa de Cesare PhD ⁹⁸.

Project administration, Samples and logistics, and Software and analysis tools:

Laura Letchford BSc ⁹⁹.

Project administration, Samples and logistics, and Visualisation:

Miss Siona Silveira MSc ⁸¹, Dr Emanuela Pelosi FRCPATH ⁸¹ and Dr Eleri Wilson-Davies MD/FRCPATH ⁸¹.

Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Dr Myra Hosmillo PhD ²⁴.

Sequencing and analysis, Software and analysis tools, and Visualisation:

Áine O'Toole MSc ⁹⁰, Dr Andrew R Hesketh PhD ⁸⁷, Mr Richard Stark MSc ⁹⁴, Dr Louis du Plessis PhD ²³, Dr Chris Ruis PhD ⁸⁸, Dr Helen Adams PhD ⁴ and Dr Yann Bourgeois PhD ⁷⁶.

Funding acquisition, and Leadership and supervision:

Dr Stephen L Michell PhD ⁹¹, Prof Dimitris Grammatopoulos PhD/FRCPATH ^{84, 112}, Dr Jonathan Edgeworth PhD/FRCPATH ¹², Prof Judith Breuer MD ^{30, 82}, Prof John A Todd PhD ⁹⁸ and Dr Christophe Fraser PhD ⁵.

Funding acquisition, and Project administration:

Dr David Buck PhD ⁹⁸ and Michaela John BSc ⁹.

Leadership and supervision, and Metadata curation:

Dr Gemma L Kay PhD ⁷⁰.

Leadership and supervision, and Project administration:

Steve Palmer⁹⁹, Prof Sharon J Peacock^{88, 64} and David Heyburn⁶⁹.

Leadership and supervision, and Samples and logistics:

Danni Weldon BSc⁹⁹, Dr Esther Robinson PhD^{64, 36}, Prof Alan McNally PhD^{41, 86}, Dr Peter Muir PhD⁶⁴, Dr Ian B Vipond PhD⁶⁴, Dr John BoYes MBChB²⁹, Dr Venkat Sivaprakasam PhD⁴⁶, Dr Tranpriti Saluja FRCPATH/MD⁷⁵, Dr Samir Dervisevic FRCPATH⁵⁴ and Dr Emma J Meader FRCPATH⁵⁴.

Leadership and supervision, and Sequencing and analysis:

Dr Naomi R Park PhD⁹⁹, Karen Oliver BSc⁹⁹, Dr Aaron R Jeffries Ph.D.⁹¹, Dr Sascha Ott PhD⁹⁴, Dr Ana da Silva Filipe PhD⁴⁸, Dr David A Simpson PhD⁷² and Dr Chris Williams MB BS⁶⁹.

Leadership and supervision, and Visualisation:

Dr Jane AH Masoli MBChB^{73, 91}.

Metadata curation, and Samples and logistics:

Dr Bridget A Knight PhD.^{73, 91}, Dr Christopher R Jones Ph.D.^{73, 91}, Mr Cherian Koshy MSc CSci FIBMS¹, Miss Amy Ash BSc¹, Dr Anna Casey PhD⁷¹, Dr Andrew Bosworth PhD^{64, 36}, Dr Liz Ratcliffe PhD⁷¹, Dr Li Xu-McCrae PhD³⁶, Miss Hannah M Pymont MSc⁶⁴, Ms Stephanie Hutchings⁶⁴, Dr Lisa Berry PhD⁸⁴, Ms Katie Jones MSc⁸⁴, Dr Fenella Halstead PhD⁴⁶, Mr Thomas Davis MSc²¹, Dr Christopher Holmes PhD¹⁶, Prof Miren Iturriza-Gomara PhD⁹², Dr Anita O Lucaci PhD⁹², Dr Paul Anthony Randell MBBCh^{38, 104}, Dr Alison Cox PhD^{38, 104}, Pinglawathee Madona^{38, 104}, Dr Kathryn Ann Harris PhD³⁰, Dr Julianne Rose Brown PhD³⁰, Dr Tabitha W Mahungu FRCPATH⁷⁴, Dr Dianne Irish-Tavares FRCPATH⁷⁴, Dr Tanzina Haque

FRCPath PhD ⁷⁴, Dr Jennifer Hart MRCP ⁷⁴, Mr Eric Witele MSc ⁷⁴, Mrs Melisa Louise Fenton DipHE ⁷⁵, Mr Steven Liggett ⁷⁹, Dr Clive Graham MD ⁵⁶, Ms Emma Swindells Bsc ⁵⁷, Ms Jennifer Collins BSc ⁵⁰, Mr Gary Eltringham BSc ⁵⁰, Ms Sharon Campbell MSc ¹⁷, Dr Patrick C McClure PhD ⁹⁷, Dr Gemma Clark PhD ¹⁵, Dr Tim J Sloan PhD ⁶⁰, Mr Carl Jones ¹⁵ and Dr Jessica Lynch PhD MBChB ^{2, 111}.

Metadata curation, and Sequencing and analysis:

Dr Ben Warne MRCP ⁸, Steven Leonard PhD ⁹⁹, Jillian Durham BSc ⁹⁹, Dr Thomas Williams MD ⁹⁰, Dr Sam T Haldenby PhD ⁹², Dr Nathaniel Storey PhD ³⁰, Dr Nabil-Fareed Alikhan PhD ⁷⁰, Dr Nadine Holmes PhD ¹⁸, Dr Christopher Moore PhD ¹⁸, Mr Matthew Carlile BSc ¹⁸, Malorie Perry MSc ⁶⁹, Dr Noel Craine DPhil ⁶⁹, Prof Ronan A Lyons MD ⁸⁰, Miss Angela H Beckett MSc ¹³, Salman Goudarzi PhD ⁷⁷, Christopher Fearn MRes ⁷⁷, Kate Cook ⁷⁷, Hannah Dent BSc ⁷⁷ and Hannah Paul MRes ⁷⁷.

Metadata curation, and Software and analysis tools:

Robert Davies ⁹⁹.

Project administration, and Samples and logistics:

Beth Blane BSc ⁸⁸, Sophia T Girgis MSc ⁸⁸, Dr Mathew A Beale PhD ⁹⁹, Katherine L Bellis ^{99, 88}, Matthew J Dorman ⁹⁹, Eleanor Drury ⁹⁹, Leanne Kane ⁹⁹, Sally Kay ⁹⁹, Dr Samantha McGuigan ⁹⁹, Dr Rachel Nelson PhD ⁹⁹, Liam Prestwood ⁹⁹, Dr Shavanthi Rajatileka PhD ⁹⁹, Dr Rahul Batra MD ¹², Dr Rachel J Williams PhD ⁸², Dr Mark Kristiansen PhD ⁸², Dr Angie Green PhD ⁹⁸, Miss

Anita Justice MSc ⁵⁹, Dr Adhyana I.K Mahanama MD ^{81, 102} and Dr Buddhini Samaraweera MD ^{81, 102}.

Project administration, and Sequencing and analysis:

Dr Nazreen F Hadjirin PhD ⁸⁸ and Dr Joshua Quick PhD ⁴¹.

Project administration, and Software and analysis tools:

Mr Radoslaw Poplawski BSc ⁴¹.

Samples and logistics, and Sequencing and analysis:

Leanne M Kermack MSc ⁸⁸, Nicola Reynolds PhD ⁷, Grant Hall BS ²⁴, Yasmin Chaudhry BSc ²⁴, Malte L Pinckert MPhil ²⁴, Dr Iliana Georgana PhD ²⁴, Dr Robin J Moll PhD ⁹⁹, Dr Alicia Thornton ⁶⁶, Dr Richard Myers ⁶⁶, Dr Joanne Stockton PhD ⁴¹, Miss Charlotte A Williams BSc ⁸², Dr Wen C Yew PhD ⁵⁸, Alexander J Trotter MRes ⁷⁰, Miss Amy Trebes MSc ⁹⁸, Mr George MacIntyre-Cockett BSc ⁹⁸, Alec Birchley MSc ⁶⁹, Alexander Adams BSc ⁶⁹, Amy Plimmer ⁶⁹, Bree Gatica-Wilcox MPhil ⁶⁹, Dr Caoimhe McKerr PhD ⁶⁹, Ember Hilvers MA ⁶⁹, Hannah Jones ⁶⁹, Dr Hibo Asad PhD ⁶⁹, Jason Coombes BSc ⁶⁹, Johnathan M Evans MSc ⁶⁹, Laia Fina ⁶⁹, Lauren Gilbert A-Levels ⁶⁹, Lee Graham BSc ⁶⁹, Michelle Cronin ⁶⁹, Sara Kumziene-SummerhaYes MSc ⁶⁹, Sarah Taylor ⁶⁹, Sophie Jones MSc ⁶⁹, Miss Danielle C Groves BA ⁹³, Mrs Peijun Zhang MSc ⁹³, Miss Marta Gallis MSc ⁹³ and Miss Stavroula F Louka MSc ⁹³.

Samples and logistics, and Software and analysis tools:

Dr Igor Starinskij Msc MRCP ⁴⁸.

Sequencing and analysis, and Software and analysis tools:

Dr Chris J Illingworth PhD ⁴⁷, Dr Chris Jackson PhD ⁴⁷, Ms Marina Gourtovaia MSc ⁹⁹, Gerry Tonkin-Hill ⁹⁹, Kevin Lewis ⁹⁹, Dr Jaime M Tovar-Corona PhD ⁹⁹, Dr Keith James PhD ⁹⁹, Dr Laura Baxter PhD ⁹⁴, Dr Mohammad T. Alam PhD ⁹⁴, Dr Richard J Orton PhD ⁴⁸, Dr Joseph Hughes PhD ⁴⁸, Dr Sreenu Vattipally PhD ⁴⁸, Dr Manon Ragonnet-Cronin PhD ³⁹, Dr Fabricia F. Nascimento PhD ³⁹, Mr David Jorgensen MSc ³⁹, Ms Olivia Boyd MSc ³⁹, Ms Lily Geidelberg MSc ³⁹, Dr Alex E Zarebski PhD ²³, Dr Jayna Raghvani PhD ²³, Dr Moritz UG Kraemer DPhil ²³, Joel Southgate MSc ^{10, 69}, Dr Benjamin B Lindsey MRCP ⁹³ and Mr Timothy M Freeman MPhil ⁹³.

Software and analysis tools, and Visualisation:

Jon-Paul Keatley ⁹⁹, Dr Joshua B Singer PhD ⁴⁸, Leonardo de Oliveira Martins PhD ⁷⁰, Dr Corin A Yeats PhD ¹⁴, Dr Khalil Abudahab PhD ^{14, 114}, Mr Ben EW Taylor MEng ^{14, 114} and Mirko Menegazzo ¹⁴.

Leadership and supervision:

Prof John Danesh ⁹⁹, Wendy Hogsden MSc ⁴⁶, Dr Sahar Eldirdiri MBBS MSc FRCPATH ²¹, Mrs Anita Kenyon MSc ²¹, Dr Jenifer Mason MBBS ⁴³, Mr Trevor I Robinson MSc ⁴³, Prof Alison Holmes MD ^{38, 103}, Dr James Price PhD ^{38, 103}, Prof John A Hartley PhD ⁸², Dr Tanya Curran PhD ³, Dr Alison E Mather PhD ⁷⁰, Dr Giri Shankar ⁶⁹, Dr Rachel Jones ⁶⁹, Dr Robin Howe ⁶⁹ and Dr Sian Morgan FRCPATH ⁹.

Metadata curation:

Dr Elizabeth Wastenge MD ⁵³, Dr Michael R Chapman PhD ^{34, 88, 99}, Mr Siddharth Mookerjee MPH ^{38, 103}, Dr Rachael Stanley PhD ⁵⁴, Mrs Wendy Smith ¹⁵, Prof Timothy Peto PhD ⁵⁹, Dr David Eyre PhD ⁵⁹, Dr Derrick Crook ⁵⁹, Dr Gabrielle Vernet MBBS ³³, Dr Christine Kitchen PhD ¹⁰, Huw Gulliver ¹⁰, Dr Ian Merrick PhD ¹⁰, Prof Martyn Guest PhD ¹⁰, Robert Munn BSc ¹⁰, Dr Declan T Bradley ^{63, 72} and Dr Tim Wyatt ⁶³.

Project administration:

Dr Charlotte Beaver ⁹⁹, Luke Foulser ⁹⁹, Sophie Palmer ⁸⁸, Carol M Churcher ⁸⁸, Ellena Brooks MA ⁸⁸, Kim S Smith ⁸⁸, Dr Katerina Galai PhD ⁸⁸, Georgina M McManus BSc ⁸⁸, Dr Frances Bolt PhD ^{38, 103}, Dr Francesc Coll PhD ¹⁹, Lizzie Meadows MA ⁷⁰, Dr Stephen W Attwood PhD ²³, Dr Alisha Davies ⁶⁹, Elen De Lacy MSc ⁶⁹, Fatima Downing ⁶⁹, Sue Edwards ⁶⁹, Dr Garry P Scarlett PhD ⁷⁶, Mrs Sarah Jeremiah MSc ⁸³ and Dr Nikki Smith PhD ⁹³.

Samples and logistics:

Danielle Leek Bsc ⁸⁸, Sushmita Sridhar BS ^{88, 99}, Sally Forrest BSc ⁸⁸, Claire Cormie ⁸⁸, Harmeet K Gill PhD ⁸⁸, Joana Dias MSc ⁸⁸, Ellen E Higginson PhD ⁸⁸, Mailis Maes MPhil ⁸⁸, Jamie Young BSc ⁸⁸, Michelle Wantoch PhD ⁷, Sanger Covid Team (www.sanger.ac.uk/covid-team) ⁹⁹, Dorota Jamrozy ⁹⁹, Stephanie Lo ⁹⁹, Dr Minal Patel PhD ⁹⁹, Verity Hill ⁹⁰, Ms Claire M Bewshea MSc ⁹¹, Prof Sian Ellard FRCPath ^{73, 91}, Dr Cressida Auckland FRCPath ⁷³, Dr Ian Harrison ⁶⁶, Dr Chloe Bishop ⁶⁶, Dr Vicki Chalker ⁶⁶, Dr Alex Richter PhD ⁸⁵, Dr Andrew Beggs PhD ⁸⁵, Dr Angus Best PhD ⁸⁶, Dr Benita Percival PhD ⁸⁶, Dr Jeremy Mirza PhD ⁸⁶, Dr Oliver Megram PhD ⁸⁶, Dr Megan Mayhew PhD ⁸⁶, Dr Liam Crawford PhD ⁸⁶, Dr Fiona Ashcroft PhD ⁸⁶, Dr Emma Moles-

Garcia PhD ⁸⁶, Dr Nicola Cumley PhD ⁸⁶, Mr Richard Hopes ⁶⁴, Dr Patawee Asamaphan PhD ⁴⁸, Mr Marc O Niebel MSc ⁴⁸, Prof Rory N Gunson PhD FRCPath ¹⁰⁰, Dr Amanda Bradley PhD ⁵², Dr Alasdair Maclean PhD ⁵², Dr Guy Mollett MBChB ⁵², Dr Rachel Blacow MBChB ⁵², Mr Paul Bird MSc ¹⁶, Mr Thomas Helmer ¹⁶, Miss Karlie Fallon ¹⁶, Dr Julian Tang ¹⁶, Dr Antony D Hale MBBS ⁴⁹, Dr Louissa R Macfarlane-Smith PhD ⁴⁹, Katherine L Harper MBiol ⁴⁹, Miss Holli Carden MSc ⁴⁹, Dr Nicholas W Machin MSc ^{45, 64}, Ms Kathryn A Jackson MSc ⁹², Dr Shazaad S Y Ahmad MSc ^{45, 64}, Dr Ryan P George PhD ⁴⁵, Dr Lance Turtle PhD MRCP ⁹², Mrs Elaine O'Toole BSc ⁴³, Mrs Joanne Watts BSc ⁴³, Mrs Cassie Breen BSc ⁴³, Mrs Angela Cowell MSc ⁴³, Ms Adela Alcolea-Medina ^{32, 96}, Ms Themoula Charalampous MSc ^{12, 42}, Amita Patel ¹¹, Dr Lisa J Levett PhD ³⁵, Dr Judith Heaney PhD ³⁵, Dr Aileen Rowan PhD ³⁹, Prof Graham P Taylor DSc ³⁹, Dr Divya Shah PhD ³⁰, Miss Laura Atkinson MSc ³⁰, Mr Jack CD Lee MSc ³⁰, Mr Adam P Westhorpe BSc ⁸², Dr Riaz Jannoo PhD ⁸², Dr Helen L Lowe PhD ⁸², Miss Angeliki Karamani MSc ⁸², Miss Leah Ensell BSc ⁸², Mrs Wendy Chatterton MSc ³⁵, Miss Monika Pusok MSc ³⁵, Mrs Ashok Dadrah MSc ⁷⁵, Miss Amanda Symmonds MSc ⁷⁵, Dr Graciela Sluga MD/MSc ⁴⁴, Dr Zoltan Molnar PhD ⁷², Mr Paul Baker MD ⁷⁹, Prof Stephen Bonner ⁷⁹, Ms Sarah Essex ⁷⁹, Dr Edward Barton MD ⁵⁶, Ms Debra Padgett BSc ⁵⁶, Ms Garren Scott BSc ⁵⁶, Ms Jane Greenaway MSc ⁵⁷, Dr Brendan AI Payne MD ⁵⁰, Dr Shirelle Burton-Fanning MD ⁵⁰, Dr Sheila Waugh MD ⁵⁰, Dr Veena Raviprakash MD ¹⁷, Ms Nicola Sheriff BSc ¹⁷, Ms Victoria Blakey BSc ¹⁷, Ms Lesley-Anne Williams BSc ¹⁷, Dr Jonathan Moore MD ²⁷, Ms Susanne Stonehouse BSc ²⁷, Dr Louise Smith ⁵⁵, Dr Rose K Davidson PhD ⁸⁹, Dr Luke Bedford ²⁶, Dr Lindsay Coupland PhD ⁵⁴, Ms Victoria Wright BSc ¹⁸, Dr Joseph G Chappell PhD ⁹⁷, Dr Theocharis Tsoleridis PhD ⁹⁷, Prof Jonathan Ball PhD ⁹⁷, Mrs Manjinder Khakh ¹⁵, Dr Vicki M Fleming PhD ¹⁵, Dr Michelle M Lister PhD ¹⁵, Dr Hannah C Howson-Wells PhD ¹⁵, Dr Louise Berry ¹⁵, Dr Tim Boswell ¹⁵, Dr Amelia Joseph ¹⁵,

Dr Iona Willingham ¹⁵, Dr Nichola Duckworth ⁶⁰, Dr Sarah Walsh ⁶⁰, Dr Emma Wise PhD ^{2, 111}, Dr Nathan Moore PhD ^{2, 111}, Miss Matilde Mori BSc ^{2, 108, 111}, Dr Nick Cortes MRCP FRCPath ^{2, 111}, Dr Stephen Kidd PhD ^{2, 111}, Dr Rebecca Williams BMBS ³³, Laura Gifford MSc ⁶⁹, Miss Kelly Bicknell ⁶¹, Dr Sarah Wyllie ⁶¹, Miss Allyson Lloyd ⁶¹, Mr Robert Impey MSc ⁶¹, Ms Cassandra S Malone MSc ⁶, Mr Benjamin J Cogger BSc ⁶, Nick Levene MSc ⁶², Lynn Monaghan ⁶², Dr Alexander J Keeley MRCP ⁹³, Dr David G Partridge FRCP FRCPath ^{78, 93}, Dr Mohammad Raza ^{78, 93}, Dr Cariad Evans ^{78, 93} and Dr Kate Johnson ^{78, 93}.

Sequencing and analysis:

Emma Betteridge BSc ⁹⁹, Ben W Farr BSc ⁹⁹, Scott Goodwin MSc ⁹⁹, Dr Michael A Quail PhD ⁹⁹, Carol Scott ⁹⁹, Lesley Shirley MSc ⁹⁹, Scott AJ Thurston BSc ⁹⁹, Diana Rajan MSc ⁹⁹, Dr Iraad F Bronner PhD ⁹⁹, Louise Aigrain PhD ⁹⁹, Dr Nicholas M Redshaw PhD ⁹⁹, Dr Stefanie V Lensing PhD ⁹⁹, Shane McCarthy ⁹⁹, Alex Makunin ⁹⁹, Dr Carlos E Balcazar PhD ⁹⁰, Dr Michael D Gallagher PhD ⁹⁰, Dr Kathleen A Williamson PhD ⁹⁰, Thomas D Stanton BSc ⁹⁰, Ms Michelle L Michelsen BSc ⁹¹, Ms Joanna Warwick-Dugdale BSc ⁹¹, Dr Robin Manley Ph.D. ⁹¹, Ms Audrey Farbos MSc ⁹¹, Dr James W Harrison Ph.D. ⁹¹, Dr Christine M Sambles Ph.D. ⁹¹, Dr David J Studholme Ph.D. ⁹¹, Dr Angie Lackenby ⁶⁶, Dr Tamyo Mbisa ⁶⁶, Dr Steven Platt ⁶⁶, Mr Shahjahan Miah ⁶⁶, Dr David Bibby ⁶⁶, Dr Carmen Manso ⁶⁶, Dr Jonathan Hubb ⁶⁶, Dr Gavin Dabrera ⁶⁶, Dr Mary Ramsay ⁶⁶, Dr Daniel Bradshaw ⁶⁶, Dr Ulf Schaefer ⁶⁶, Dr Natalie Groves ⁶⁶, Dr Eileen Gallagher ⁶⁶, Dr David Lee ⁶⁶, Dr David Williams ⁶⁶, Dr Nicholas Ellaby ⁶⁶, Hassan Hartman ⁶⁶, Nikos Manesis ⁶⁶, Vineet Patel ⁶⁶, Juan Ledesma ⁶⁷, Ms Katherine A Twohig ⁶⁷, Dr Elias Allara ^{64, 88}, Ms Clare Pearson ^{64, 88}, Mr Jeffrey K. J. Cheng MSc ⁹⁴, Dr Hannah E. Bridgewater PhD ⁹⁴, Ms Lucy R. Frost BSc ⁹⁴, Ms Grace Taylor-Joyce BSc ⁹⁴, Dr Paul E Brown PhD ⁹⁴, Dr Lily Tong PhD

⁴⁸, Ms Alice Broos BSc ⁴⁸, Mr Daniel Mair BSc ⁴⁸, Mrs Jenna Nichols BSc ⁴⁸, Dr Stephen N Carmichael PhD ⁴⁸, Dr Katherine L Smollett PhD ⁴⁰, Dr Kyriaki Nomikou PhD ⁴⁸, Dr Elihu Aranday-Cortes PhD/DVM ⁴⁸, Ms Natasha Johnson BSc ⁴⁸, Dr Seema Nickbakhsh PhD ^{48, 68}, Dr Edith E Vamos PhD ⁹², Dr Margaret Hughes PhD ⁹², Dr Lucille Rainbow PhD ⁹², Mr Richard Eccles MSc ⁹², Ms Charlotte Nelson MSc ⁹², Dr Mark Whitehead PhD ⁹², Dr Richard Gregory PhD ⁹², Mr Matthew Gemmell MSc ⁹², Ms Claudia Wierzbicki BSc ⁹², Ms Hermione J Webster BSc ⁹², Ms Chloe L Fisher MSc ²⁸, Mr Adrian W Signell BSc ²⁰, Dr Gilberto Betancor PhD ²⁰, Mr Harry D Wilson BSc ²⁰, Dr Gaia Nebbia PhD FRCPath ¹², Dr Flavia Flaviani PhD ³¹, Mr Alberto C Cerda MSc ⁹⁶, Ms Tammy V Merrill MSc ⁹⁶, Rebekah E Wilson MSc ⁹⁶, Mr Marius Cotic MSc ⁸², Miss Nadua Bayzid BSc ⁸², Dr Thomas Thompson PhD ⁷², Dr Erwan Acheson PhD ⁷², Prof Steven Rushton PhD ⁵¹, Prof Sarah O'Brien PhD ⁵¹, David J Baker BEng ⁷⁰, Steven Rudder ⁷⁰, Alp Aydin MSci ⁷⁰, Dr Fei Sang PhD ¹⁸, Dr Johnny Debebe PhD ¹⁸, Dr Sarah Francois PhD ²³, Dr Tetyana I Vasylyeva DPhil ²³, Dr Marina Escalera Zamudio PhD ²³, Mr Bernardo Gutierrez MSc ²³, Dr Angela Marchbank BSc ¹⁰, Joshua Maksimovic FD ⁹, Karla Spellman FD ⁹, Kathryn McCluggage MSc ⁹, Dr Mari Morgan PhD ⁶⁹, Robert Beer BSc ⁹, Safiah Afifi BSc ⁹, Trudy Workman HNC ¹⁰, William Fuller BSc ¹⁰, Catherine Bresner BSc ¹⁰, Dr Adrienn Angyal PhD ⁹³, Dr Luke R Green PhD ⁹³, Dr Paul J Parsons PhD ⁹³, Miss Rachel M Tucker MSc ⁹³, Dr Rebecca Brown PhD ⁹³ and Mr Max Whiteley PhD ⁹³

Software and analysis tools:

James Bonfield BSc ⁹⁹, Dr Christoph Puethé ⁹⁹, Mr Andrew Whitwham BSc ⁹⁹, Jennifer Liddle ⁹⁹, Dr Will Rowe PhD ⁴¹, Dr Igor Siveroni PhD ³⁹, Dr Thanh Le-Viet PhD ⁷⁰ and Amy Gaskin MSc

Visualisation:

Dr Rob Johnson PhD ³⁹.

1 Barking, Havering and Redbridge University Hospitals NHS Trust, **2** Basingstoke Hospital, **3** Belfast Health & Social Care Trust, **4** Betsi Cadwaladr University Health Board, **5** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **6** Brighton and Sussex University Hospitals NHS Trust, **7** Cambridge Stem Cell Institute, University of Cambridge, **8** Cambridge University Hospitals NHS Foundation Trust, **9** Cardiff and Vale University Health Board, **10** Cardiff University, **11** Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, **12** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **13** Centre for Enzyme Innovation, University of Portsmouth (PORT), **14** Centre for Genomic Pathogen Surveillance, University of Oxford, **15** Clinical Microbiology Department, Queens Medical Centre, **16** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **17** County Durham and Darlington NHS Foundation Trust, **18** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **19** Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, **20** Department of Infectious Diseases, King's College London, **21** Department of Microbiology, Kettering General Hospital, **22** Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, **23** Department of Zoology, University of Oxford, **24** Division of Virology, Department of Pathology, University of Cambridge, **25** East Kent Hospitals University NHS Foundation Trust, **26** East Suffolk and North Essex NHS Foundation Trust, **27** Gateshead Health NHS Foundation Trust, **28** Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, **29** Gloucestershire Hospitals NHS Foundation Trust, **30** Great Ormond Street Hospital for Children NHS Foundation Trust, **31** Guy's and St. Thomas' BRC, **32** Guy's and St. Thomas' Hospitals, **33** Hampshire Hospitals NHS Foundation Trust, **34** Health Data Research UK Cambridge, **35** Health Services

Laboratories, **36** Heartlands Hospital, Birmingham, **37** Hub for Biotechnology in the Built Environment, Northumbria University, **38** Imperial College Hospitals NHS Trust, **39** Imperial College London, **40** Institute of Biodiversity, Animal Health & Comparative Medicine, **41** Institute of Microbiology and Infection, University of Birmingham, **42** King's College London, **43** Liverpool Clinical Laboratories, **44** Maidstone and Tunbridge Wells NHS Trust, **45** Manchester University NHS Foundation Trust, **46** Microbiology Department, Wye Valley NHS Trust, Hereford, **47** MRC Biostatistics Unit, University of Cambridge, **48** MRC-University of Glasgow Centre for Virus Research, **49** National Infection Service, PHE and Leeds Teaching Hospitals Trust, **50** Newcastle Hospitals NHS Foundation Trust, **51** Newcastle University, **52** NHS Greater Glasgow and Clyde, **53** NHS Lothian, **54** Norfolk and Norwich University Hospital, **55** Norfolk County Council, **56** North Cumbria Integrated Care NHS Foundation Trust, **57** North Tees and Hartlepool NHS Foundation Trust, **58** Northumbria University, **59** Oxford University Hospitals NHS Foundation Trust, **60** PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, **61** Portsmouth Hospitals University NHS Trust, **62** Princess Alexandra Hospital Microbiology Dept., **63** Public Health Agency, **64** Public Health England, **65** Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, **66** Public Health England, Colindale, **67** Public Health England, Colindale, **68** Public Health Scotland, **69** Public Health Wales NHS Trust, **70** Quadram Institute Bioscience, **71** Queen Elizabeth Hospital, **72** Queen's University Belfast, **73** Royal Devon and Exeter NHS Foundation Trust, **74** Royal Free NHS Trust, **75** Sandwell and West Birmingham NHS Trust, **76** School of Biological Sciences, University of Portsmouth (PORT), **77** School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), **78** Sheffield Teaching Hospitals, **79** South Tees Hospitals NHS Foundation Trust, **80** Swansea University, **81** University Hospitals Southampton NHS Foundation Trust, **82** University College London, **83** University Hospital Southampton NHS Foundation Trust, **84** University Hospitals Coventry and Warwickshire, **85** University of Birmingham, **86** University of Birmingham Turnkey Laboratory, **87** University of Brighton, **88** University of Cambridge, **89** University of East Anglia, **90** University of Edinburgh, **91** University of Exeter, **92**

University of Liverpool, **93** University of Sheffield, **94** University of Warwick, **95** University of Cambridge, **96** Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, **97** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **98** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **99** Wellcome Sanger Institute, **100** West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, **101** Department of Medicine, University of Cambridge, **102** Ministry of Health, Sri Lanka, **103** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **104** North West London Pathology, **105** NU-OMICS, Northumbria University, **106** University of Kent, **107** University of Oxford, **108** University of Southampton, **109** University of Southampton School of Health Sciences, **110** University of Southampton School of Medicine, **111** University of Surrey, **112** Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, **113** Wellcome Africa Health Research Institute Durban and **114** Wellcome Genome Campus.

Supplementary File S4: Additional References

- W1 Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995; **82**: 711–32.
- W2 Cauchemez S, Temime L, Guillemot D, *et al.* Investigating Heterogeneity in Pneumococcal Transmission. *Journal of the American Statistical Association*. 2006; **101**: 946–58.
- W3 Meredith LW, Hamilton WL, Warne B, *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* 2020; **20**: 1263–71.
- W4 Contribution of nosocomial infections to the first wave in the United Kingdom. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961210/S1056_Contribution_of_nosocomial_infections_to_the_first_wave.pdf (accessed June 28, 2021).
- W5 Challen R, Brooks-Pollock E, Tsaneva-Atanasova K, Danon L. Meta-analysis of the SARS-CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction number. DOI:10.1101/2020.11.17.20231548.
- W6 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**: 577–82.
- W7 Boni MF, Lemey P, Jiang X, *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 2020; **5**: 1408–17.