THE ROYAL SOCIETY
PUBLISHING

A Bayesian approach to identifying the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2

Jessica R. E. Bridgen[1], Joseph M. Lewis[2,3,4], Stacy Todd[2], Miriam Taegtmeyer[2,4], Jonathan M. Read[1] and Chris P. Jewell[5]

[1]Centre for Health Informatics, Computing, and Statistics, Lancaster Medical School, Lancaster University, Lancaster, UK
[2]Tropical and Infectious Diseases Unit, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK
[3]Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, UK
[4]Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK
[5]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

JREB, 0000-0001-5497-2700; JML, 0000-0002-3837-5188; ST, 0000-0003-3215-9625; MT, 0000-0002-5377-2536; JMR, 0000-0002-9697-0962; CPJ, 0000-0002-7902-2178

Nosocomial infections threaten patient safety, and were widely reported during the COVID-19 pandemic. Effective hospital infection control requires a detailed understanding of the role of different transmission pathways, yet these are poorly quantified. Using patient and staff data from a large UK hospital, we demonstrate a method to infer unobserved epidemiological event times efficiently and disentangle the infectious pressure dynamics by ward. A stochastic individual-level, continuous-time state-transition model was constructed to model transmission of SARS-CoV-2, incorporating a dynamic staff–patient contact network as time-varying parameters. A Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm was used to estimate transmission rate parameters associated with each possible source of infection, and the unobserved infection and recovery times. We found that the total infectious pressure exerted on an individual in a ward varied over time, as did the primary source of transmission. There was marked heterogeneity between wards; each ward experienced unique infectious pressure over time. Hospital infection control should consider the role of between-ward movement of staff as a key infectious source of nosocomial infection for SARS-CoV-2. With further development, this method could be implemented routinely for real-time monitoring of nosocomial transmission and to evaluate interventions.

1. Introduction

Healthcare-associated infections are a significant burden to health systems worldwide, and are associated with increased morbidity and mortality [1]. Transmission of SARS-CoV-2 in healthcare settings was widely reported during the first wave of the COVID-19 pandemic [2–5]. Hospital patients are considered to be especially vulnerable to severe COVID-19 and healthcare workers have been shown to be at an increased risk of infection [6–8]. Identifying nosocomial transmission routes of SARS-CoV-2 is therefore critical to patient and staff safety. Universal point-of-care testing of patients and staff was not available during the initial stages of the pandemic in the UK to rapidly identify and isolate individuals. A triage system was recommended in hospitals to isolate and test patients with suspected COVID-19 [9]. Although hospitals employed a range of infection prevention control (IPC) protocols to minimize transmission, high levels of hospital infections were reported [2,10]. Healthcare-associated infections, or nosocomial infections, are defined as infections that occur either as a direct result of healthcare interventions or from being in contact with a healthcare setting [11].
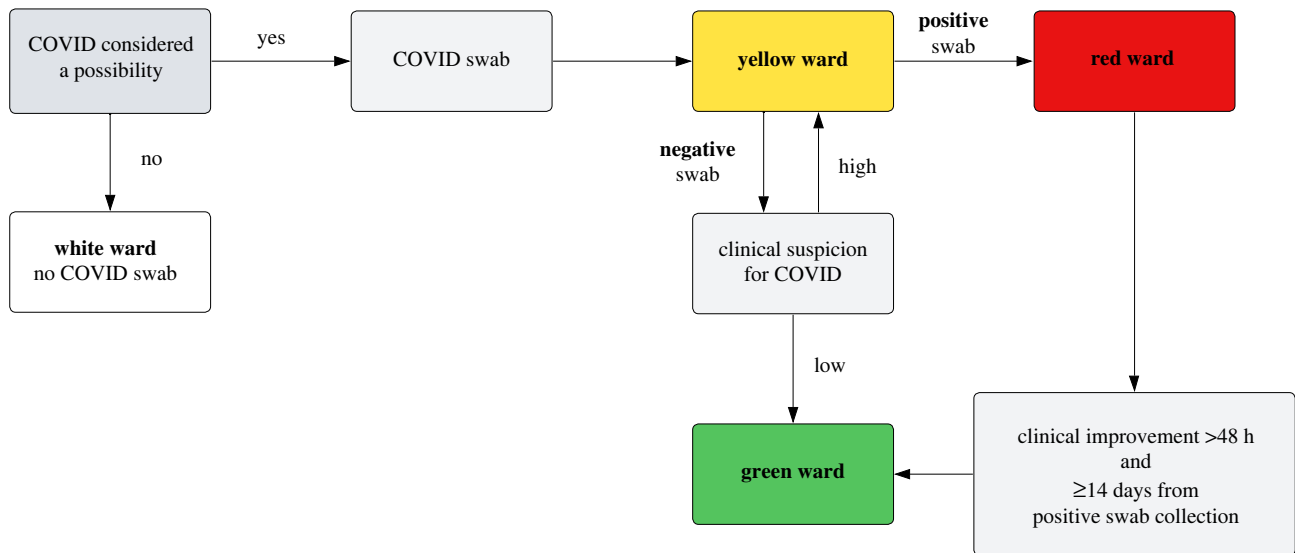
Figure 1. Patient pathway from initial admission to ward allocation, by suspected and confirmed SARS-CoV-2 infection status, during the study period.

To determine nosocomial transmission routes of a respiratory pathogen, the network of transmission opportunities within the hospital, hereafter called the contact network, must be considered. The routine running of a hospital in the UK, comprising wards, bays and side rooms, can give rise to additional structural and network transmission opportunities. Partitioning of patients into wards may be by diagnosis, severity of illness, or by infection status during an outbreak. Patients may have direct contact with staff, visitors and patients in the same ward, but also indirect contact with other patients and staff through shared equipment and objects, relevant to fomite transmission, airflow, relevant to airborne transmission, and staff acting as vectors for infectious diseases. Modelling studies have been conducted to investigate nosocomial transmission routes of numerous pathogens, most commonly *Methicillin-resistant Staphylococcus aureus* (MRSA) and *Vancomycin-resistant Enterococcus* (VRE) [12]. While transmission from patients to patients and healthcare workers to patients has been identified, there remains a lack of consensus as to the primary routes of nosocomial infection [7,13,14].

One of the challlenges of identifying transmission routes in a hospital is distinguishing between hospital-acquired and community-acquired infections. Knight *et al.* report large uncertainty in the classification of nosocomial and community-acquired infections of SARS-CoV-2 during the first part of 2020 [15]. Routinely collected hospital data tend to record a timestamp of when a patient tested positive for an infection. For COVID-19, this provides a marker as to when a patient was infectious. However, the time that a patient became infected is unobserved, and the time a patient recovered from infection and ceased to be infectious is also often unobserved. This makes the inference for nosocomial models much more difficult and computationally intensive, since we need to explore all possible configurations of event times consistent with the epidemiological landscape (such as contact networks and physical proximity within the hospital).

Generally, surveillance data only capture a partially observed epidemic process, and therefore requires a form of data augmentation to estimate unobserved epidemiological event times. O'Neill & Roberts initially introduced a Bayesian data augmentation approach to inference of general stochastic epidemic models, where unobserved event times are treated as parameters to be estimated [16]. This has proved to be a popular method for conducting inference on partially observed epidemic models [17,18]. A drawback of this method is that repeatedly calculating the likelihood can become extremely costly computationally and its use is therefore limited by population size. As an alternative method, McKinley *et al.* demonstrate using approximate likelihood ratios to conduct inference on partially observed epidemics, though this relies on repeatedly simulating the epidemic [19]. Previous studies which model nosocomial transmission have used data augmentation to handle unobserved infection times [20,21]. However, these studies do not include a time-varying contact network parameter to model interactions between staff and patients.

Here, we use routinely collected data from a large acute-care hospital in the UK to quantify the temporal and network dynamics of nosocomial transmission of SARS-CoV-2. We demonstrate a Bayesian approach to conducting inference with a time-varying covariate, a fine-scale patient–staff contact network, to estimate unobserved epidemiological event times and provide insight into within-hospital transmission dynamics.

2. Covariate data

Patient testing data from a large acute-care hospital in the UK were used in conjunction with staff rota data (staff shift times and ward assignments) to identify routes of nosocomial transmission of SARS-CoV-2 during the first wave of the COVID-19 pandemic. A patient pathway was developed with a colour-coded ward system to separate patients based on their SARS-CoV-2 infection status; figure 1. Universal testing for SARS-CoV-2 was not available for patient admissions or staff during our study period. Patient diagnosis was therefore based on clinical suspicion of COVID-19 and a confirmatory test. Some wards (e.g. the isolation ward) could be assigned different colour areas within a single ward based on the availability of side rooms (enclosed patient rooms), which might be treated differently to communal areas. Ward colours could also change over time in view of capacity demands. Personal protective equipment (PPE) guidance for staff differed by ward colour. Our study period covered a four-week time span from 12 April to 10 May 2020. During this time, 3816 staff worked at least one shift at the hospital, consisting of 2948 healthcare
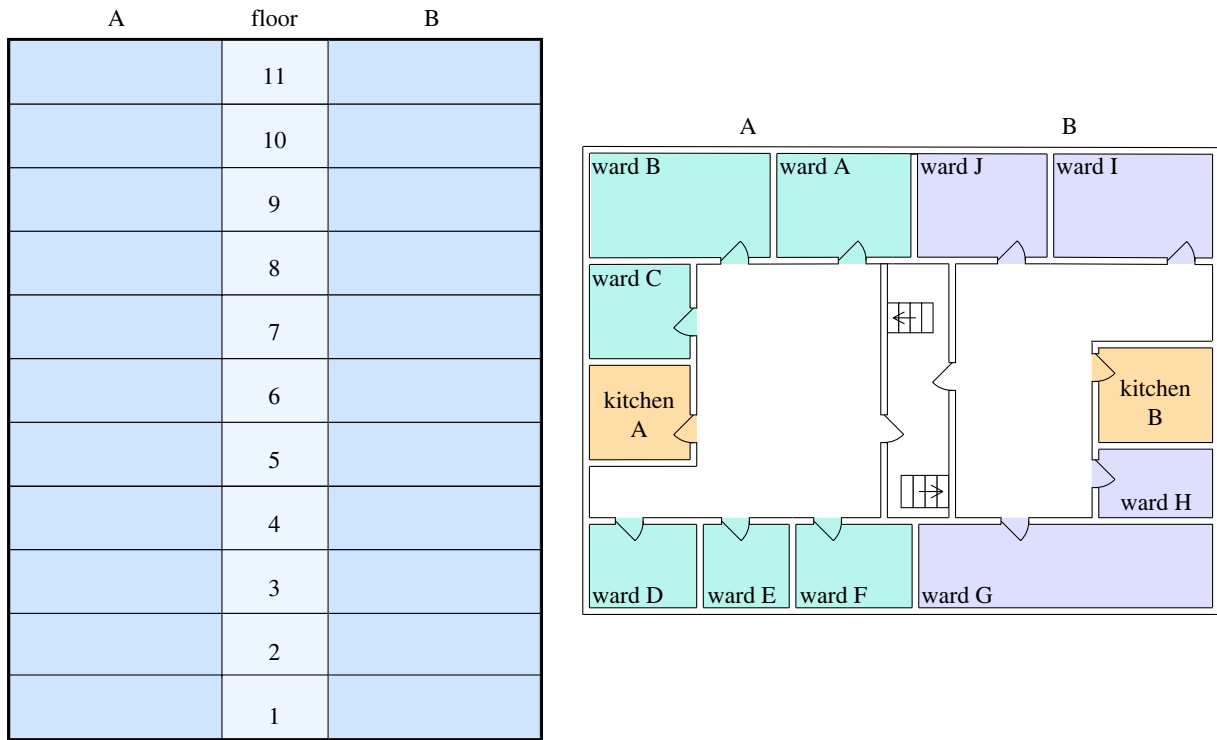
Figure 2. A schematic illustration of the hospital layout. Each floor of the hospital has two kitchens, one on side A and one on side B. The spatial adjacency tensor W, as defined previously, would consider all wards on the same floor and side of the hospital as connected. The number and size of wards on each floor of the hospital vary.

staff, 232 medical doctors and 636 ancillary staff. There were $N = 2981$ patients admitted (including day attenders) in our study period to $p = 55$ wards.

Anonymized patient location, staff work patterns and SARS-CoV-2 testing results were extracted from routine trust electronic records. Patient data were linked to SARS-CoV-2 testing results through anonymous hashed identifiers. The study received Health Research Authority (HRA) approval via the Integrated Research Application System (IRAS) and trust approval as a research project, accessing only routinely collected anonymized data (IRAS ID 288257).

2.1. Hospital network

Patient movements (ward transfers, admission and discharges) and staff shift times were recorded continuously in time. However, the recorded times of these movements were not likely to be exact in practice, and the volume of changes to the patient and staff structure leads to prohibitively large data structures in RAM. Hence, patient and staff movements were aggregated to 1 h time intervals to form the basis of our discrete-time contact network.

The hospital contact network is represented in three distinct ways. Firstly, a weighted ward connectivity tensor, C, of shape $[p \times p \times T]$, where element c_{qrt} is the connectivity between ward q and r at time t, in units of number of staff working across multiple wards. This primarily consists of doctors who are assigned to a group of wards per shift. Similarly, we define a spatial adjacency tensor of shape $[p \times p \times T]$ denoted W, where w_{qrt} is the number of staff using the kitchen *if* the wards share a kitchen, and 0 otherwise. For example, in figure 2, wards A and B share kitchen A, so w_{abt} equals the total number of staff allocated to side A (wards A–F) on that floor of the hospital. Lastly, a membership tensor of shape $[N \times p \times T]$ denoted M, where $m_{iqt} = 1$ if individual i is a member of ward q at time t and zero otherwise. Visualizations of the contact network can be found in the electronic supplementary material.

We introduce the dot subscript notation to indicate a slice of a tensor. For example, $C_{q \cdot t}$ refers to a slice of tensor C along the second axis, which can be thought of as a vector representing the connectivity between ward q and all hospital wards ($\forall r$) at time t.

3. Modelling

In this section, we first describe the model structure in terms of the data generating process, before outlining our approach to Bayesian inference with the appropriate likelihood function, posterior distribution and Markov chain Monte Carlo (MCMC) algorithm.

3.1. Model structure

Transmission of SARS-CoV-2 is represented here by a stochastic individual-level, continuous-time state-transition model. At any point in time, patients are considered to belong to one of four mutually exclusive epidemiological states: *susceptible to infection (S)*, *infected but not infectious (E)*, *infected and infectious (I)* or *recovered/removed from the population (R)*. We denote the sets of individuals in each state at time t as $S(t), E(t), I(t), R(t)$, respectively, and for convenience write $X(t) = \{S(t), E(t), I(t), R(t)\}$. Any particular patient, i is assumed to progress between the states according to the transitions [SE], [EI] and [IR] (where [PQ] denotes a transition from state P to state Q) with transition rates $\lambda_i^{SE}(t)$, $\lambda_i^{EI}(t)$ and $\lambda_i^{IR}(t)$, respectively, as defined below.

At time t, the rate at which patient $i = 1, \ldots, N$ becomes infected, i.e. $\lambda_i^{\mathrm{SE}}(t)$, is assumed to be a function of the time-evolving infectious landscape. Within the hospital, we assume that i experiences infection pressure from four separate sources: other patients in the same ward; patients in other wards connected by staff being assigned to more than one ward; the spatial structure of wards connected by adjoining kitchens; constant 'background' infection, representing sources of infection not explicitly modelled by the hospital structure. Within-hospital routes of infection are represented by our patient–staff contact networks (C, W, M). Within the continuous-time model, changes to the patient–ward and staff–ward structures (either via patient movements or staff–ward allocation) are assumed to occur at discrete intervals of 1 hour, aggregating the precisely time-stamped patient movement and staff shift data to the hour. Continuous time was used to model the epidemic process to account for the fluidity of events in the hospital through time, thus avoiding the necessary act of choosing a time step. In addition, this leads to more efficient sampling from the posterior distribution of censored event times, avoiding large swings in posterior density that otherwise occur in discrete-time systems.

The rate at which an individual transitions from state S to E can be described as a time-dependent infectious pressure. We assume that infections occur at points of a right-continuous-time inhomogeneous Poisson process at a rate equal to the sum of infectious pressure on susceptibles immediately before that time point. The infectious pressure is density dependent and is defined at the individual level. We account for the number of infected on each ward at time t, the patient–staff contact network at time t and a background infection rate. An individual can either be present in the hospital H ($i \in H$) and admitted to ward q ($i \in q$), or in the community ($i \notin H$). The infectious pressure on a susceptible individual i ($i \in S(t)$) at time t can be defined as

$$\lambda_i^{\mathrm{SE}}(t) = \begin{cases} \beta_1 I_q(t) + (\beta_2 \boldsymbol{C}_{q \cdot t} + \beta_3 \boldsymbol{W}_{q \cdot t})^{\mathsf{T}} \cdot \boldsymbol{I}(t) + \beta_4 & \text{if } i \in S(t), i \in q, i \in H \\ \beta_5 & \text{if } i \in S(t), i \notin H \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

where β_1 denotes the transmission rate for within-ward mixing and $I_q(t)$ represents the number of infected individuals on ward q at time t. β_2 and β_3 are transmission rates for between-ward mixing, with $\boldsymbol{C}_{q \cdot t}$ denoting the connectivity between ward q and all other wards at time t, and $\boldsymbol{W}_{q \cdot t}$ describing the connectivity between ward q and all other wards by ward proximity at time t. $\boldsymbol{I}(t)$ denotes a vector whose q_{th} entry is equal to $I_q(t)$ and thus represents the number of infected individuals on each ward at time t. β_4 is a background hospital transmission rate. To allow for individuals to be infected before and between hospital admissions, we define a constant infectious pressure β_5 that is exerted on to susceptibles in the community.

Similarly, we can define an individual's [EI] transition rate as follows:

$$\lambda_i^{\mathrm{EI}}(t) = \begin{cases} \alpha & \text{if } i \in E(t) \\ 0 & \text{otherwise,} \end{cases} \quad (3.2)$$

where $i \in E(t)$ represents an individual i residing in the E state at time t.

Likewise, an individual's [IR] transition rate is defined as

$$\lambda_i^{\mathrm{IR}}(t) = \begin{cases} \gamma & \text{if } i \in I(t) \\ 0 & \text{otherwise,} \end{cases} \quad (3.3)$$

where $i \in I(t)$ represents an individual i in the I state at time t.

Thus the [EI] and [IR] transition rates are assumed to be constant across individuals and time, and for identifiability reasons, we fix $\alpha = 1/4 \, \mathrm{day}^{-1}$ and $\gamma = 1/5 \, \mathrm{day}^{-1}$, respectively [22].

The epidemic process is assumed to be Markovian. The data generating process is outlined in algorithm 1.

The epidemic process is assumed to be Markovian. The data generating process is outlined in Algorithm 1.

Algorithm 1: Gillespie's Direct algorithm

Input: initial state $X(0) = \{S(0), E(0), I(0), R(0)\}$, wallclock time $t = 0$, contact matrix change
time $s = 0$, transition rate functions $\lambda^{\mathrm{se}}(t), \lambda^{\mathrm{ei}}(t), \lambda^{\mathrm{ir}}(t)$

while $\sum_{i=1}^{N} [\lambda_i^{\mathrm{se}}(t) + \lambda_i^{\mathrm{ei}}(t) + \lambda_i^{\mathrm{ir}}(t)] > 0$ **do** // transition rates are greater than 0

$\quad t^* \sim$ Exponential $\left(\sum_{i=1}^{N} [\lambda_i^{\mathrm{se}}(t) + \lambda_i^{\mathrm{ei}}(t) + \lambda_i^{\mathrm{ir}}(t)] \right)$ // draw time to next transmission event

$\quad$ **if** $t + t^* > s$ **then**
$\quad\quad$ Set $t := s$
$\quad\quad$ Set $s := s + \frac{1}{24}$ // advance s by 1 hour
$\quad\quad$ Set $k := 3N + 1$
$\quad$ **else**
$\quad\quad k \sim$ Discrete $\left(\frac{\vec{h}(t)}{\sum_{i=1}^{3N} h_i(t)} \right)$ // Choose event

$\quad\quad$ where $\vec{h}(t) = (\lambda_1^{\mathrm{se}}(t), \ldots, \lambda_N^{\mathrm{se}}(t), \lambda_1^{\mathrm{ei}}(t), \ldots, \lambda_N^{\mathrm{ei}}(t), \lambda_1^{\mathrm{ir}}(t), \ldots, \lambda_N^{\mathrm{ir}}(t))^T$

$\quad\quad$ Let $j = \lfloor k/N \rfloor$, $i = (k \bmod N)$ // enumerate transition and individual
$\quad\quad$ **switch** j **do**
$\quad\quad\quad$ **case** 0 **do** // S->E
$\quad\quad\quad\quad S(t + t^*) = S(t) - \{i\}$
$\quad\quad\quad\quad E(t + t^*) = E(t) + \{i\}$
$\quad\quad\quad$ **case** 1 **do** // E->I
$\quad\quad\quad\quad E(t + t^*) = E(t) - \{i\}$
$\quad\quad\quad\quad I(t + t^*) = I(t) + \{i\}$
$\quad\quad\quad$ **case** 2 **do** // I->R
$\quad\quad\quad\quad I(t + t^*) = I(t) - \{i\}$
$\quad\quad\quad\quad R(t + t^*) = R(t) + \{i\}$
$\quad\quad$ **end**
$\quad\quad t := t + t^*$
$\quad$ **end**
$\quad$ Record t, k
end

Result: The ordered lists **t, k**

3.2. Bayesian inference

A Metropolis–Hastings MCMC algorithm is used to estimate the transmission rate parameters $\theta = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ and the unobserved [SE] and [IR] state transition times for each infection event according to the method of Jewell *et al.* [18]. Each infection

transition event is associated with an observed [EI] transition time, where an individual's [EI] transition time is assumed to be 2 days before the collection of their first positive swab. A sensitivity analysis was conducted to assess the robustness of our findings given different delays between a patients first positive test and their [EI] transition time. The inclusion of a time-varying covariate, the contact networks, in the continuous-time model adds to the complexity of computing the likelihood. We define the likelihood of observing $(\mathbf{t}, \mathbf{k})$ transitions, in terms of an ordered event list, where an event is considered to be a transition event (transitioning between states: $S \to E$, $E \to I$ or $I \to R$) or a hospital network update, as follows:

$$f(\mathbf{t}, \mathbf{k}|X(0), \theta, C, W, H, t_0 = 0) = \prod_{l=1}^{|\mathbf{t}|}\left[h_{k_l}(t_l)e^{-\sum_{i=1}^{3N} h_i(t_l)(t_l - t_{l-1})}\right], \tag{3.4}$$

where $\mathbf{t}$ and $\mathbf{k}$ are the (time-ordered) lists of event times and indices as in algorithm 1, $X(0)$ denotes the initial conditions, θ the parameters as above, and covariate data C, W and H. $\boldsymbol{h}(t_k)$ is a vector of length $3N + 1$ of hazard rates for the three transitions, with $N = 2981$ individuals and a covariate marker to indicate a hospital network update. That is, $\boldsymbol{h}(t) = (\lambda_1^{SE}(t), \ldots, \lambda_N^{SE}(t), \lambda_1^{EI}(t), \ldots, \lambda_N^{EI}(t), \lambda_1^{IR}(t), \ldots, \lambda_N^{IR}(t), 1)^\top$ similar to algorithm 1. Thus $h_{k_l}(t_l)$ denotes the hazard rate for the k_lth transition at time t_l. Note that the trailing 1 corresponds to a contact matrix change (a hospital network update), such that for a timepoint at which the contact matrices change the likelihood term represents the probability of surviving an epidemiological state transition in the preceding time interval.

The following prior distributions were chosen for all transmission parameters $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ denoted by $f(\theta)$. The prior distributions were chosen to favour small beta values which were still bound by zero.

$$\beta_i \overset{iid}{\sim} \text{Gamma}(1.1, 1000). \tag{3.5}$$

The joint conditional posterior can therefore be defined as

$$\begin{aligned}\pi(\theta, \mathbf{t}_z, \mathbf{k}_z|X(0), \mathbf{t}_{-z}, \mathbf{k}_{-z}, C, W, H, t_0 = 0) \\ \propto f(\mathbf{t}, \mathbf{k}|X(0), \theta, C, W, H, t_0 = 0)f(\theta),\end{aligned} \tag{3.6}$$

where $z = ([SE], [IR])$ and $\mathbf{t}_z$ and $\mathbf{t}_{-z}$ denote partitioning of elements of the lists $\mathbf{t}$ into unobserved and observed events, respectively (and similarly for $\mathbf{k}$). As the likelihood is intractable to integration, a Metropolis-within-Gibbs MCMC algorithm is used to sample from the joint posterior. Full details of the MCMC algorithm used can be found in the electronic supplementary material. To evaluate chain convergence, the Gelman and Rubin potential scale reduction statistic is calculated for three chains with differing starting values drawn from the prior distribution [23]. The posterior predictive distribution is analysed visually to assess how well our modelled estimates fit the observed data.

3.3. Code implementation

The model and MCMC were implemented in Python v. 3.9 using TensorFlow and TensorFlow Probability for graphics processing unit (GPU) acceleration [24,25]. The model code is available at: https://zenodo.org/doi/10.5281/zenodo.10567172.

3.4. Infection hazard attributable fraction

Having computed the joint posterior distribution, we are able to investigate how within-ward and between-ward dynamics contribute to nosocomial transmission. We calculate the attributable fraction, denoted *AF*, per infection for each transmission type: within-ward transmission β_1; between-ward transmission rates β_2 and β_3; background hospital transmission rate β_4; community transmission rate β_5. For example, the attributable fraction for individual i infected from between-ward transmission on ward q can be defined as follows:

$$P(AF_{i_{bw}}) = \frac{(\beta_2 \boldsymbol{C}_{q\cdot t} + \beta_3 \boldsymbol{W}_{q\cdot t})^\top \cdot \boldsymbol{I}(\widetilde{t})}{\lambda_i^{SE}(\widetilde{t})}, \quad \widetilde{t} = \mathbf{t}_{\text{idx}(\mathbf{k}=i)}, \tag{3.7}$$

where $\text{idx}(\mathbf{k} = i)$ denotes the index of the element of $\mathbf{k}$ which is equal to i (i.e. corresponding to the [SE] event for individual i).

The attributable fractions are aggregated for each set of sampled parameter estimates to compute the mean attributable fraction and 95% credible intervals (95%CrI) for each transmission type per infection. Similarly, using the sampled parameter estimates, we can identify which wards and associated ward colours nosocomial infections take place in.

4. Results

The results presented are of SARS-CoV-2 infections recorded in a UK hospital one month into the first wave of the pandemic. In a population size of 2981 patients from 12 April to 10 May 2020, we have identified 131 infection events, [EI] transitions, which are a combination of community-acquired and nosocomially acquired infections. Historic testing data, 3 days prior to our study start date, were used to identify the 58 patients which are considered to be initially infectious, residing in the I state, in our model.
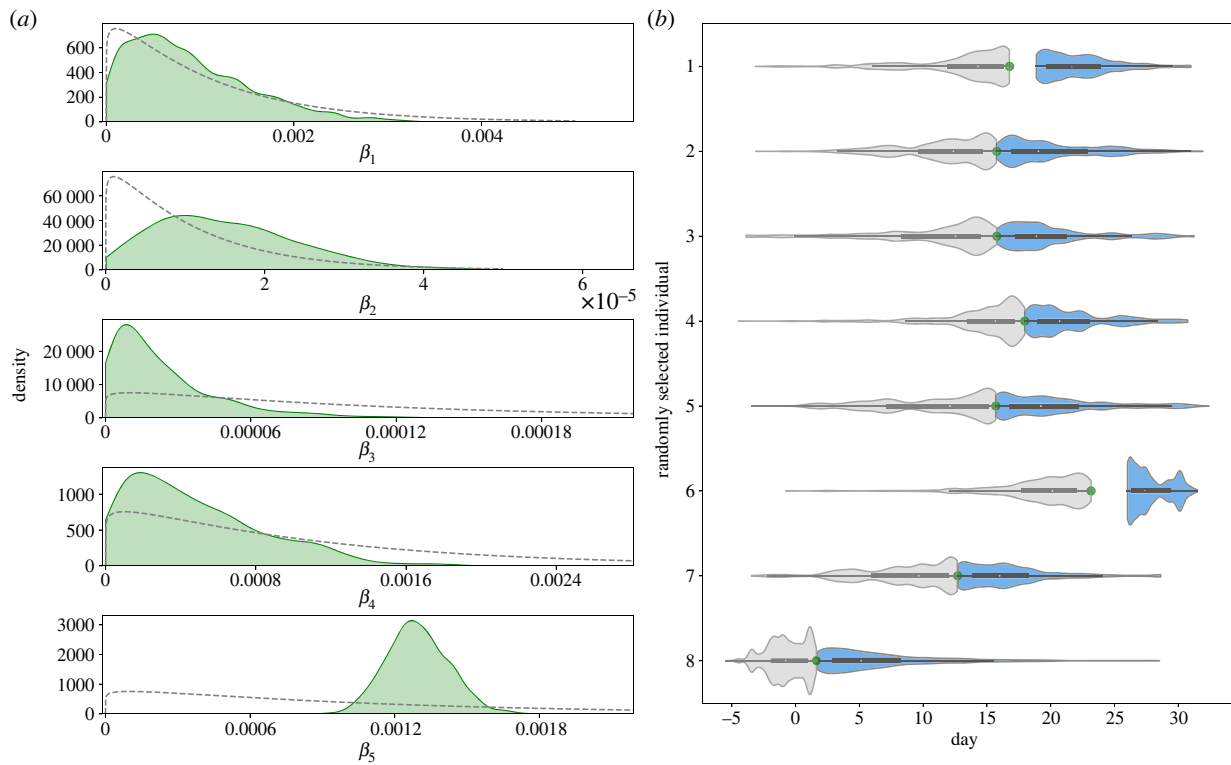
Figure 3. (a) Kernel density estimates for each transmission rate parameter. Dashed lines represent the associated prior distribution. (b) Density estimates of [SE] and [IR] transition times for a random sample of eight observed [EI] events. Distributions for the associated [SE] and [IR] transition times are shown in grey and blue, respectively. The observed [EI] transition time for each randomly selected individual is represented as a green circle positioned between the [SE] and [IR] distributions.
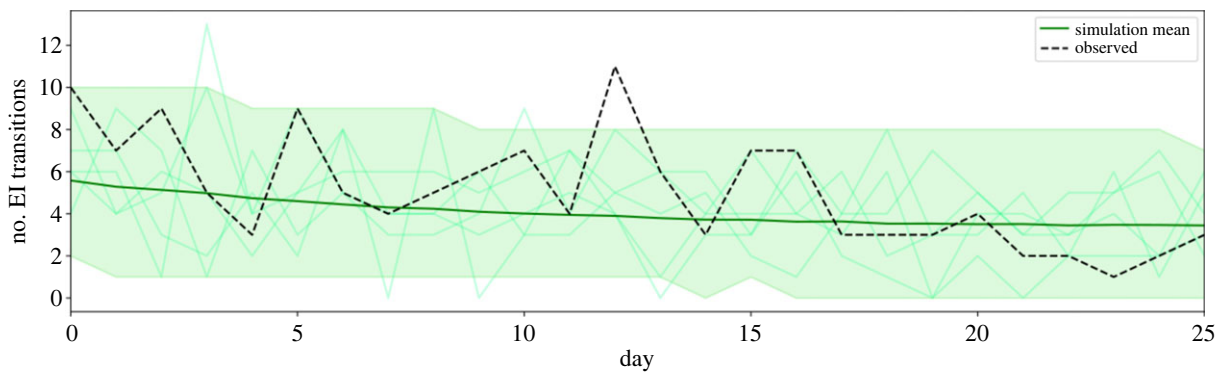
Figure 4. Posterior predictive formed from 10 000 stochastic simulations over the joint posterior. Mean simulated number of [EI] transitions in dark green with 95% credible interval as the shaded area. Five individual simulations are displayed in the faint green lines. The observed number of [EI] transitions per day is shown by the dashed line.

4.1. Parameter estimation

To estimate our parameters of interest ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, t_z$), we ran a Metropolis–Hastings MCMC algorithm for 11 000 iterations removing the first 1000 samples as burn-in. Convergence across all parameters is seen which is confirmed by the potential scale reduction statistic computed for three independent chains; see electronic supplementary material, table S1. The effective sample size was also computed for each transmission rate parameter for three independent chains; see electronic supplementary material. In figure 3, we present the kernel density estimates for each of the transmission rate parameters and for a random sample of transition event times. A constraint when drawing event times, is that an individual must have been admitted to hospital before their [IR] transition time. As the observed [EI] transition times are set to 2 days prior to a patient's first positive hospital test, there may be a period of time pre-admission which is unexplored by the MCMC algorithm, as seen with sample 1 and 6 of figure 3b.

The posterior predictive distribution is used to evaluate how well our model fits the observed data. An epidemic is simulated forward using the data generating model (algorithm 1) for each of the 10 000 sets of parameter estimates sampled by the MCMC algorithm. We compare the number of [EI] transitions occurring each day from the simulations with the observed data; figure 4. With the exception of one peak of [EI] transitions on day 12, the observed data sit comfortably within the 95%CrI of the aggregated simulation data. Figure 4 shows that in a given simulation, the peaks and troughs of the number of [EI] transitions per day may similarly sit outside of the credible interval.
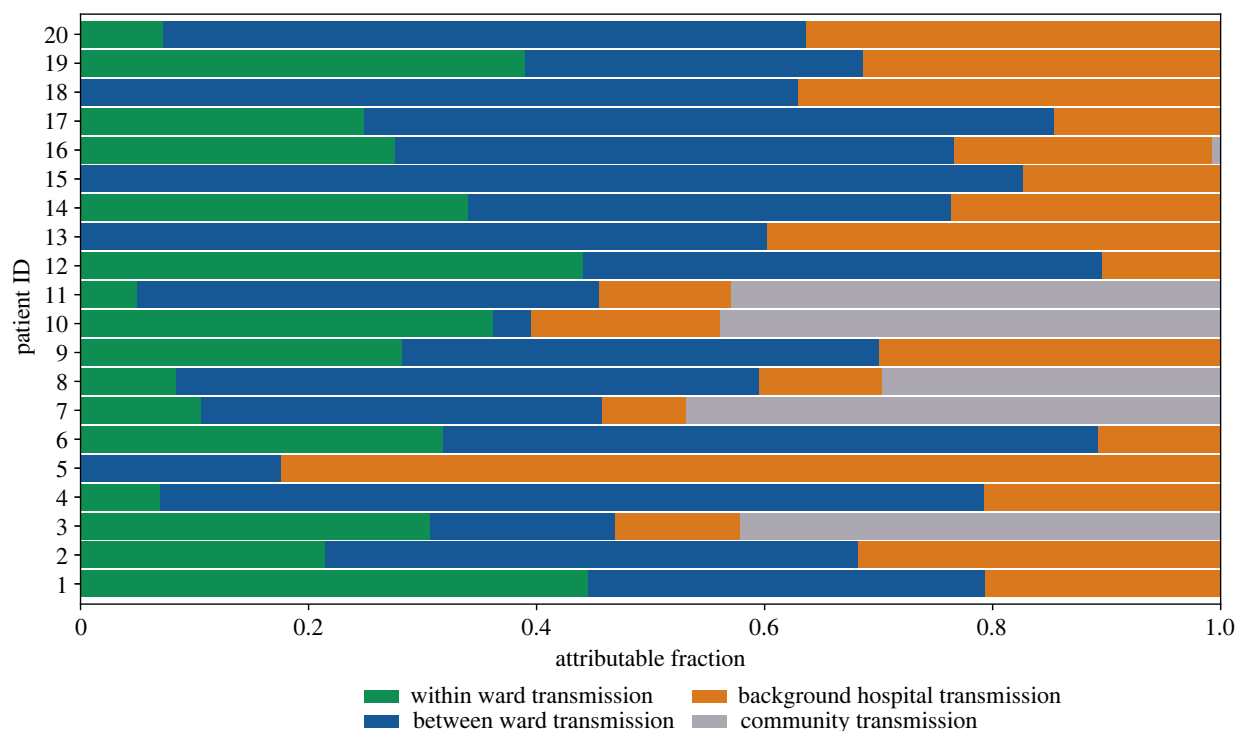
Figure 5. Mean attributable fraction for each transmission type per infection event for 10 000 posterior samples. Infection events are displayed if they have a mean community transmission attributable fraction less than 0.5.

Table 1. Percentage of total nosocomial infections by designated ward colour.

ward colour	percentage of nosocomial infections (95%CrI)
green	40.04 (33.33–60.00)
green-red-yellow	13.59 (6.67–25.00)
white	10.29 (5.56–23.53)
yellow	10.00 (5.56–20.00)
white-yellow	9.20 (5.56–20.00)
red	5.68 (5.00–9.09)
green-red	5.61 (5.00–9.09)
red-yellow	5.61 (5.26–9.09)

4.2. Nosocomial transmission routes

One of the parameters of interest is the associated [SE] transition time for each infection. Estimating this enables us to identify which infections were most likely to be contracted nosocomially rather than in the community. Moreover, for each infection, we analyse how the different types of transmission contribute to the infectious pressure exerted on an individual directly before their infection event. Hospital-acquired infections are likely for 15.3% (20/131) of detected infections. These infections have a mean community attributable fraction of less than 0.5, with the majority (15/20) having a mean community attributable fraction of less than 0.1; figure 5. Between-ward transmission is the highest mean attributable fraction for 13 of the 20 infections identified as nosocomial. Infection events of patient 4 and patient 15 have the highest mean attributable fractions for between-ward transmission of 0.72 (95%CrI 0.12–0.99) and 0.83 (95%CrI 0.29–1.00); figure 5. Four nosocomial infections have a mean within-ward attributable fraction of zero, indicating that these infections occurred when there were no infectious individuals admitted to the individual's ward.

To assess the plausibility of our model output, we further explored the hospital data associated with the 15 individuals with the lowest community attributable fraction. One individual appeared to have been tested on the day they were admitted to hospital; on further inspection, this individual had been discharged from a 15-day hospital spell 2 days prior to re-admission which was captured within our study period. The other 14 individuals identified had been admitted to hospital for an average of 13.3 days (95%CrI 6.62–22.76) before their positive test sample was collected.

Using our modelling framework, we are able to identify the wards in which nosocomial infections occurred. Nosocomial infections are identified based on whether an individual's [SE] transition time is during a hospital admission spell. For each set of sampled parameters, the percentage distribution of nosocomial infections by ward is calculated and aggregated for the 10 000 samples. We
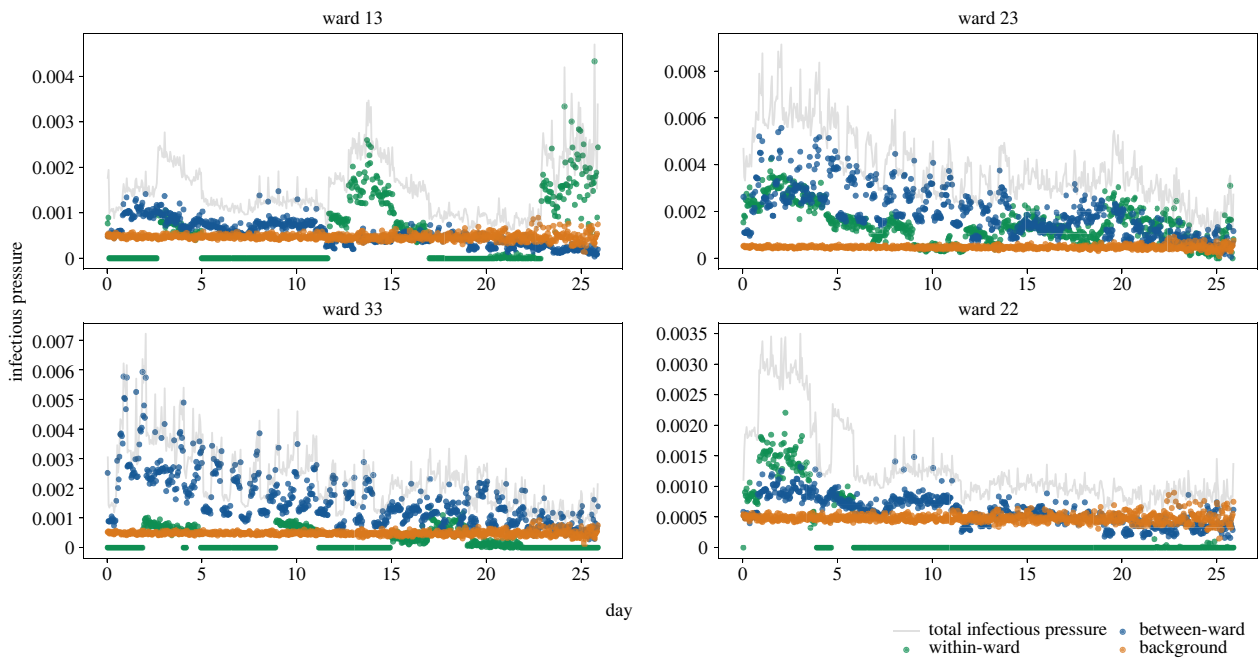
Figure 6. Mean infectious pressure for an individual on the stated ward at each [SE] transition, calculated for 10 000 posterior samples. The wards displayed are those with the highest mean number of nosocomial infections.

find that four wards account for the locations of 63.1% of nosocomially acquired infections, with wards 13 and 23 accounting for 21.1% (95%CrI 11.76–30.00) and 17.5% (95%CrI 10.00–25.00), respectively. Similarly, we compute the percentage distribution of nosocomial infections by ward type; table 1. Nosocomial infections are recorded most frequently in green wards (40.0%, 95%CrI 33.33–60.00) and wards which are assigned three colours of green, red and yellow (13.6%, 95%CrI 6.67–25.00).

The modelling suggests that the infectious pressure exerted on an individual in a ward changes considerably over time, this may be explained by the dynamic contact network. The within-ward infectious pressure will increase with the number of infectious individuals on the ward. Similarly, if there is an increase of infectious individuals on wards considered to be connected, the between-ward infectious pressure will increase. Interestingly, the source of infectious pressure exerted on individuals in the four wards that recorded the highest number of nosocomial infections was also dynamic; figure 6. For the first 23 days of our study period, each of these four wards were classified as green wards with the exception of ward 23 which was a mixed ward, either classified as green-red-yellow or green-red ward. We find infectious pressure dynamics varied by ward and ward colour; figure 6. An individual on ward 33 would have experienced an infectious pressure driven by between-ward dynamics whereas individuals on wards 13 and 22 would have experienced clear peaks in within-ward infectious pressure.

We conducted a sensitivity analysis of the delay between an individual's [EI] transition time and an individual's first positive test in hospital to assess the robustness of our results; electronic supplementary material, figures S3–S5. Delays of 1 day and 3 days were considered in addition to the 2 days used for our primary analysis. A by-product of altering the [EI] transition times is that the initial conditions of our model change, with a different number of individuals initially residing in the I state; electronic supplementary material, table S3. The number of nosocomial infections identified therefore varied. However, between-ward dynamics were found to be the key driver of infectious pressure in each analysis, and the same four wards were identified as housing the most nosocomial infections.

5. Discussion

This study provides a statistical framework for conducting inference on hospital outbreak dynamics to quantify the relative contribution of transmission routes for nosocomial infections. We have demonstrated that transmission parameters and unobserved event times can be inferred by incorporating a discrete time-varying patient–staff contact network and testing data into a continuous-time stochastic epidemic model. By estimating these parameters, we are able to disentangle the different components of infectious pressure and identify routes of nosocomial transmission of SARS-CoV-2 during the first wave of the pandemic. While computing the likelihood for these models can be computationally intensive, we found that with GPU acceleration (1 × NVIDIA V100 32GB), 11 000 iterations of our MCMC took approximately 61 min.

We estimate that 15.3% of identified SARS-CoV-2 infections in patients identified in hospital were nosocomially acquired, other studies which examined hospital infections during the first wave of the pandemic have reported similar findings [2,6,13,26]. Nonetheless, we expect that this is an under-representation of the full extent of within-hospital transmission. Asymptomatic infections are not captured in this study as universal testing of admissions had not yet been introduced. Similarly, if a patient had been discharged while unknowingly infected, their infection would be unrecorded. A simple method which is used to identify nosocomial infections is to consider the time interval between admission and symptom onset. However, this can lead to an infected patient with multiple hospital spells in quick succession being misclassified as a community-acquired infection [15]. The modelling

approach presented here enabled us to identify an individual who was discharged from a long hospital stay and re-admitted 2 days later, as having a likely nosocomially acquired infection.

We found that the total infectious pressure exerted on an individual in a ward changes over time, as does the primary source of transmission. When comparing wards which housed nosocomial infections, it was clear that infectious pressure dynamics varied greatly by ward. Moreover, these dynamics varied across wards which were designated with the same ward colour. For most nosocomial infections, the most likely source of infection was captured by between-ward dynamics, suggesting that the patient pathway implemented was successful in separating susceptible patients from infectious patients. This finding is supported by several other studies which found that healthcare workers were a likely source of nosocomial infection [14,27,28]. Evans *et al.* found that indirect transmission from infected patients was the most likely route for nosocomial transmission, where indirect transmission may happen through healthcare workers acting as vectors for transmission or through fomite transmission [7]. Without staff infection data it is difficult to pinpoint the exact cause of between-ward transmission. Between-ward transmission in our model is driven by the staff–patient contact network, which indicates that the hospital contact network is a key route for nosocomial transmission.

Nosocomial infections were most likely to be contracted within four wards. These wards tended to be wards designated as green on the patient pathway. Patients would be placed in green wards after receiving a negative test result for SARS-CoV-2 and if they were also of low clinical suspicion of having COVID-19, or if they were being stepped down from a red ward. Separating patients who were not suspected of having COVID-19 on admission appears to have been effective, with fewer nosocomial infections occurring on white wards than green wards. However, undetected asymptomatic transmission may be more likely to occur on white wards as patients were not tested on admission.

There are several limitations to our approach. Firstly, in order to conduct the inference, we fix an individual's [EI] transition time to be 2 days prior to their first positive test. This allows for a delay between patients becoming infectious and being tested [29]. The effect of this assumption was tested in our sensitivity analysis (electronic supplementary material), and although we note that the initial conditions, β_1, β_2 and β_3, do change in response to changing the delay, the overall conclusion of which wards presented the highest risk is robust. A feasible innovation to our inference algorithm would be to estimate this interval, along with formal inference on the initial conditions, as part of the MCMC. We do not consider reinfections owing to the short length of our study period. Additionally, we assume that individuals progress from exposed to infectious and infectious to recovered at constant rates. As our study period is during the initial phase of the pandemic, vaccination status and virus variants were not considered. The data augmentation methodology used could be developed further to better cope with unknown disease status on admission and occult infections at the end of the time window. Furthermore, the force of infection exerted on the community could be based on prevalence estimates. Community prevalence was not well known for our study period, with estimates often based on hospital admission data, as SARS-CoV-2 tests were not readily available to the public. Nevertheless, this should be considered for future studies. We also did not explore different model structures, using the standard SEIR structure often used for COVID-19 models. Computationally, calculating the likelihood in continuous time is expensive, it would be worthwhile to investigate the potential gains and losses of a fully discrete model.

Our findings allow us to draw some important conclusions regarding effectiveness of infection prevention and control measures in this hospital early in the pandemic. At this time, universal SARS-CoV-2 testing was not logistically possible, and a lack of isolation facilities meant that patients had to be cohorted based on COVID-19 risk. Firstly, we conclude that stratifying patient risk of SARS-CoV-2 infection based on clinical assessment (primarily presence of respiratory symptoms) was successful, in that few nosocomial infections probably occurred in white (low risk) and yellow (possible COVID-19) wards. Secondly, most nosocomial infections probably occurred in green wards, and were primarily driven by between-ward transmission and hence the staff–patient contact network. These findings can inform future planning for outbreaks of respiratory pathogens and provide support for strategies to reduce staff–patient transmission, such as asymptomatic staff testing and lateral flow testing of patients on admission, which were implemented in the NHS later in the pandemic [30].

We have presented a Bayesian approach to quantifying routes of nosocomial transmission of SARS-CoV-2 which could be applied to other respiratory infections and extended to outbreaks of bacterial infections. To our knowledge, inference methods such as the ones presented here have not been used for real-time IPC monitoring owing to the computational complexity and impractical runtimes. We have demonstrated an efficient method to infer epidemiological event times and nosocomial transmission routes while accounting for the intricacies of the hospital network. This could be used to retrospectively evaluate and simulate interventions. Additionally, with further development and the appropriate infrastructure in place, we believe that methods such as these could be implemented at a similar scale during a prolonged hospital outbreak to alert IPC teams to potential hotspots of transmission and assess effectiveness of interventions.

References

1. World Health Organization. 2011 *Report on the burden of endemic health care-associated infection worldwide*. Technical report.
2. Read JM *et al.* 2021 Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave. *Lancet* **398**, 1037–1038. (doi:10.1016/S0140-6736(21)01786-4)
3. Abbas M, Robalo Nunes T, Martischang R, Zingg W, Iten A, Pittet D, Harbarth S. 2021 Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers. *Antimicrob. Resist. Infect. Control* **10**, 7. (doi:10.1186/s13756-020-00875-7)
4. Wu Z, McGoogan JM. 2020 Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **23**, 1239–1242. (doi:10.1001/jama.2020.2648)
5. Suárez-García I, Martínez de Aramayona López MJ, Sáez Vicente A, Lobo Abascal P. 2020 SARS-CoV-2 infection among healthcare workers in a hospital in Madrid, Spain. *J. Hosp. Infect.* **106**, 357–363. (doi:10.1016/j.jhin.2020.07.020)
6. Rickman HM *et al.* 2021 Nosocomial transmission of coronavirus disease 2019: a retrospective study of 66 hospital-acquired cases in a London teaching hospital. *Clin. Infect. Dis.* **72**, 690–693. (doi:10.1093/cid/ciaa816)
7. Evans S, Stimson J, Pople D, Bhattacharya A, Hope R, White PJ, Robotham JV. 2022 Quantifying the contribution of pathways of nosocomial acquisition of COVID-19 in English hospitals. *Int. J. Epidemiol.* **51**, 393–403. (doi:10.1093/ije/dyab241)
8. Ward H *et al.* 2021 SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat. Commun.* **12**, 905. (doi:10.1038/s41467-021-21237-w)
9. European Centre for Disease Prevention and Control. 2020 *Infection prevention and control and preparedness for COVID-19 in healthcare settings - second update*. See https://www.ecdc.europa.eu/sites/default/files/documents/Infection-prevention-control-for-the-care-of-patients-with-2019-nCoV-healthcare-settings_update-31-March-2020.pdf.
10. GOV UK. 2020 *COVID-19: infection prevention and control (IPC)*. January. See https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control.
11. GOV UK. 2012 *Healthcare associated infection (HCAI): operational guidance and standards*. July. See https://www.gov.uk/government/publications/healthcare-associated-infection-hcai-operational-guidance-and-standards (accessed 21 May 2023).
12. van Kleef E, Robotham JV, Jit M, Deeny SR, Edmunds WJ. 2013 Modelling the transmission of healthcare associated infections: a systematic review. *BMC Infect. Dis.* **13**, 294. (doi:10.1186/1471-2334-13-294)
13. Bhattacharya A *et al.* 2021 Healthcare-associated COVID-19 in England: a national data linkage study. *J. Infect.* **83**, 565–572. (doi:10.1016/j.jinf.2021.08.039)
14. Aghdassi SJS *et al.* 2022 Risk factors for nosocomial SARS-CoV-2 infections in patients: results from a retrospective matched case–control study in a tertiary care university center. *Antimicrob. Resist. Infect. Control* **11**, 1–10. (doi:10.1186/s13756-022-01056-4)
15. Knight GM *et al.* 2022 The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020. *BMC Infect. Dis.* **22**, 556. (doi:10.1186/s12879-022-07490-4)
16. O'Neill PD, Roberts GO. 1999 Bayesian inference for partially observed stochastic epidemics. *J. R. Stat. Soc. A Stat. Soc.* **162**, 121–129. (doi:10.1111/1467-985X.00125)
17. Streftaris G, Gibson GJ. 2004 Bayesian inference for stochastic epidemics in closed populations. *Stat. Modell.* **4**, 63–75. (doi:10.1191/1471082X04st065oa)
18. Jewell CP, Kypraios T, Neal P, Roberts GO. 2009 Bayesian analysis for emerging infectious diseases. *Bayesian Anal.* **4**, 465–496. (doi:10.1214/09-BA417)
19. McKinley T, Cook AR, Deardon R. 2009 Inference in epidemic models without likelihoods. *Int. J. Biostat.* **5**, 24. (doi:10.2202/1557-4679.1171)
20. Forrester ML, Pettitt AN, Gibson GJ. 2007 Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostatistics* **8**, 383–401. (doi:10.1093/biostatistics/kxl017)
21. Kypraios T, O'Neill PD, Huang SS, Rifas-Shiman SL, Cooper BS. 2010 Assessing the role of undetected colonization and isolation precautions in reducing methicillin-resistant *staphylococcus aureus* transmission in intensive care units. *BMC Infect. Dis.* **10**, 29. (doi:10.1186/1471-2334-10-29)
22. GOV UK. 2022 *COVID-19: epidemiology, virology and clinical features*. See https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features.
23. Brooks SP, Gelman A. 1998 General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**, 434–455. (doi:10.1080/10618600.1998.10474787)
24. Abadi M *et al.* 2015 TensorFlow: large-scale machine learning on heterogeneous systems. Software available from https://www.tensorflow.org/.
25. Dillon JV *et al.* 2017 TensorFlow distributions. November. *arXiv*. (https://arxiv.org/abs/1711.10604)
26. Carter B *et al.* 2020 Nosocomial COVID-19 infection: examining the risk of mortality. the COPE-Nosocomial study (COVID in older PEople). *J. Hosp. Infect.* **106**, 376–384. (doi:10.1016/j.jhin.2020.07.013)
27. Borges V *et al.* 2021 Nosocomial outbreak of SARS-CoV-2 in a 'Non-COVID-19' hospital ward: virus genome sequencing as a key tool to understand cryptic transmission. *Viruses* **13**, 604. (doi:10.3390/v13040604)
28. Thoma R, Kohler P, Haller S, Maenner J, Schlegel M, Flury D. 2022 Ward-level risk factors associated with nosocomial coronavirus disease 2019 (COVID-19) outbreaks: a matched case-control study. *Antimicrob. Steward. Healthc. Epidemiol.* **2**, e49. (doi:10.1017/ash.2022.11)
29. European Centre for Disease Prevention and Control. 2023 Transmission of COVID-19. See https://www.ecdc.europa.eu/en/infectious-disease-topics/z-disease-list/covid-19/facts/transmission-covid-19.
30. NHS England. 2021 *Standard operating procedure for rollout of lateral flow devices for asymptomatic staff testing for COVID-19 in primary care*. Technical report.
31. Bridgen JRE, Lewis JM, Todd S, Taegtmeyer M, Read JM, Jewell CP. 2024 A Bayesian approach to identifying the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2. Zenodo.
32. Bridgen JRE, Lewis JM, Todd S, Taegtmeyer M, Read JM, Jewell CP. 2024 A Bayesian approach to identifying the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2. Figshare. (doi:10.6084/m9.figshare.c.7075564)