**Title 1:** Estimating nosocomial transmission of micro-organisms in hospital settings using patient records and culture data

**Title 2:** Hospital traffic and surveillance determine nosocomial transmission and detection of pathogenic bacteria in hospital settings in New York City

**Authors:** Jaime Cascante Vega1, Rami Yaari1, Tal Robin1, Lingsheng Wen2, Jason Zucker2, Anne-Catrin Uhlemann2, Sen Pei1,\*, Jeffrey Shaman1,3,\*

**Affiliations:**

1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA.

2. Division of Infectious Diseases, Department of Medicine, Columbia University, College of Physicians and Surgeons, New York, New York, USA.

3. Columbia Climate School, Columbia University, New York, NY, USA.

\* Corresponding authors: Sen Pei, Jeffrey Shaman

\*Corresponding authors. Email: sp3449@cumc.columbia.edu, jls106@cumc.columbia.edu

**One Sentence Summary:** We use electronic health records and clinical culture data to inform an agent based model and estimated transmission of detection of pathogenic bacterial in a major New York City hospital system. We found hospital traffic and micro-orgnaism surveillance determine nosocomial transmission and detection of pathogenic bacteria.

**Abstract (244/250 words)**

Pathogenic bacteria are a major threat to patient health in hospitals. Here we leverage electronic health records from a major New York City hospital system collected during 2020-2021 to support inference of nosocomial transmission for eight micro-organisms. We develop an agent-based model informed by patient hospitalization records to simulate importation from the community, nosocomial transmission, and patient spontaneous decolonization of bacteria. The model is coupled with a Bayesian inference algorithm to estimate the likelihood of detection upon testing and nosocomial transmission rates. We evaluate parameter identifiability for this model-inference system and find that it is able to discriminate nosocomial transmission and effective sensitivity. We apply the framework to estimate both quantities for eight prevalent bacterial pathogens: *Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, Staphylococcus aureus* (both sensitive, MSSA, and resistant, MRSA, phenotypes), *Staphylococcus epidermidis*, *Enterococcus faecium* and *Enterococcus faecalis*. We find that nosocomial transmission for *E. coli* is negligible, and *MSSA* has a lower nosocomial transmission rate than *MRSA.* While bacterial pathogens have different levels of importation rates, nosocomial transmission rates were similar among organisms, except *E. coli*. We also find that estimated likelihoods of detection are similar for all pathogens. This work highlights how fine-scale patient data can support inference of the epidemiological properties of micro-organisms and how hospital traffic, patient contact and surveillance determine epidemiological features. Evaluation of the surveillance and transmission potential for different pathogens could ultimately support the development of in-hospital control measures, as well as the design of surveillance strategies.

**Main Text:**

**Introduction**

Antimicrobial resistance micro-organisms (AMRO) are a major threat to human health worldwide and have emerged as one of the leading public health threats of the 21st century *(1)*. An estimated 4.95 million deaths were associated with bacterial AMRO in 2019 globally, and mortality caused by AMRO is projected to reach 10 million by 2050 *(2)*. Hospital-acquired (nosocomial) infections by bacterial pathogens, including those resistant to antibiotics, are a major contributor to mortality, length of stay in hospital and health-care-associated costs *(2)*.

Understanding the burden, spread and detection of AMR pathogens, and micro-organisms in general, within hospital settings, is critical for effective control planning and design of testing/culture protocols. Quantification of the epidemiological characteristics of these pathogens remains challenging due to limited observation of micro-organisms carriage, difficulty assessing interventions in real-world hospital settings and incomplete understanding of underlying data generation processes *(3–5)*.

To circumvent these difficulties, mathematical models have been applied to study pathogen transmission in hospital settings, to quantify and understand the relative roles of different routes of transmission *(4)*, and to characterize the hospital conditions and settings that sustain transmission of both resistant and sensitive strains *(6)*. In the context of AMRO, theory has been used to understand the emergence of resistance and its interplay with community-acquired infections *(7, 8)*, to evaluate antibiotic treatment protocols *(9)*, to assess control measures to reduce nosocomial transmission *(6, 10–13)*. More recently, models have been used in conjunction with empirical observations to assess the role of competition among different strains between hosts and the role of within-host microbiome pathogen interactions *(14)*.

Most existing modeling studies focus on general theoretical frameworks of AMROs *(6–9, 14–18)*, providing insights on possible mechanisms behind transmission, co-existence and other processes in the hospital settings. Investigations using models combined with inference methods *(20)*—model-inference systems—usually have focused on a single pathogen of interest such as (MRSA) *(10, 13, 21, 22)*, *P. aeuruginosa (4), C. difficile (23),* and *Vancomycin-resistant enterococci* (VRE) *(24)* using carriage detection *(10, 13, 22, 24)*, intensive care unit (ICU) clinical culture data *(4, 22)* or sequencing data *(23)*. Studies on the epidemiological characteristics of communities of co-circulating organisms in a single hospital network supported by real-world data are absent.

Understanding the transmission dynamics of healthcare-associated (HA) infections is challenging. The epidemiology and transmission HA infections differ from community-acquired (CA) infections in a number of important ways. First, hospital networks are open systems with significant variations in admission rate ranging from 23% to 52% (See Figure 1A). Variations are explained by heterogeneous patient traffic at the building and ward scale. In contrast, communities are relatively closed systems with individuals moving in or out at much lower rates. Second, as many bacterial species exist as commensals on human hosts, detection of colonization is difficult and differential by body site. Indeed, patients with infections of the bloodstream or lower respiratory tract are more likely to be detected in hospitals than those colonized at other sites *(7, 25)*. Third, facilities within a single hospital system (e.g., infusion, pediatric, emergency wards, surgical, among others) may differ substantially in their control and surveillance of micro-organisms, as well as in the hospital traffic features, a complication compounded by patient transfers between wards and hospital buildings (see Supplementary Material Figure S1 for patient transfer matrces at both spatial scales) *(26, 27)*. Fourth, a lack of observational data for communities of micro-organisms spreading among patients and the environment in the same hospital system often imposes further challenges for comparing epidemiological features. Lastly, the emergence of strains with different genetic backgrounds that were once confined to hospital circulation such as CA-MRSA clone-USA300 in the United States with enhanced features compared to HA-MRSA makes epidemiology in communities and healthcare settings different *(8, 25)*.

In this study, we use an agent-based model (ABM) and patient clinical culture data for eight prevalent pathogenic bacterial species collected in a major New York City (NYC) hospital system to simulate nosocomial transmission and in-hospital colonization prevalence. The ABM is informed by real-world patient movement in the hospital system and incorporates importation of micro-organisms from the community, nosocomial transmission, decolonization via host clearance and patient transfer across hospital wards. Further, the ABM is coupled with a Bayesian inference algorithm to enable assimilation of clinical testing records. To account for the heterogeneity of testing frequency and micro-organism prevalence among facilities, we use the clinical testing records to design a patient-level observational model that mimics the detection of micro-organisms in hospitals. We fix the importation rates by surveying the literature and estimate nosocomial transmission and the likelihood of detection given carriage upon testing. We estimate these two epidemiological features for eight pathogenic bacteria that cause substantial mortality associated with AMR worldwide *(2)*. The ABM captures multiple sources of heterogeneity (e.g., patient length of stay, contact patterns, individual observational model, etc.) that otherwise cannot be represented by compartmental models. To simulate and quantify the transmission and detection of different micro-organisms using the same model-inference framework, we simplify the underlying biological processes that are unique to each microparasitic infection. Our aim is, in consequence, to understand the general similarities and quantify the epidemiological properties of communities of circulating bacteria among patients in hospitals. We show that while bacterial pathogens have different levels of importation rates, nosocomial transmission rates were similar. We also found the likelihood of detection is similar for all pathogens.

**Results**

### *Empirical patterns of hospital traffic and heterogeneity of nosocomial infections burden*

In Figure 1A, we present weekly and monthly incidence for *E. coli* (total positives cultures, n=2890), *K. pneumoniae* (n=1139), *P. aeruginosa* (n=809), MSSA (n=773), MRSA (n=486), *S. epidermidis* (n=694), *E. faecalis* (n=596) and *E. faecium (n=*263*)*. We estimated multiple positive clinical culture results during a patient visit and deduplicated them in the observational model. During the study period, the daily number of hospitalized patients fluctuated between 1,000 and 2,500 with most of the variation explained by differences in the day of the week (Figure 1B). Daily numbers of new admissions ranged between 100 and 1,000, including outpatients. During the first COVID-19 wave in New York City, the numbers of in-patients and admissions were generally lower; after June 2020, patient traffic was higher and relatively stationary. Ward size, defined as average ward occupancy per day during the study period, was heterogeneous (SM Figure S2A) with the majority of wards experiencing an occupancy below 10 patients. However, a few wards (e.g., emergency rooms) admitted more than 100 patients each day. Average ward size was 9 and 20 when excluding wards with size equal to 1.

Hospital-level variation in patient movement, admission and discharge rates, is explained mostly by the heterogeneity at the building scale within the hospital network with rates ranging from 0-50% (see building traffic for Allen, Harkness Pavilion and Milstein hospitals in SM Figure S1) to 80-150% in buildings mostly admitting and discharging outpatients (see Presbyterian hospital and Rest in SM Figure S3). This heterogeneity at the building scale is in turn explained by the variation of ward traffic affecting each building (see SM Figure S4). Outfluxes of patients at the ward facility scale follow admission patterns with a few wards per building admitting the majority of patients per week and manifesting stable patient traffic (see Figure 1C and SM Figure S4). To visualize patient traffic within each ward, we investigated temporal occupancy. Figure 1C shows the weekly number of patients relative to average occupancy for each ward and highlights the 5 most and least populated wards. The least populated wards were empty most of the time with irregular occupancy during the study period, whereas hospitalizations in the most populated wards were relatively stable. The relationship between newly admitted and hospitalized patients (Figures 1D and 1E) was linear and depended on ward size (see SM Figure S5). Clinical culture numbers were also heterogeneous across wards. Most cultures were sampled from a small subset of wards, as shown by the weekly number of cultures collected in each (see Figure 1D). The distribution of time in hospital obtained from the patient hospitalization records varied was 3.85 days across the entire hospital system, and 3.17, 4.18, 6.22, 5.05, 1.25 and 1.85 days in Allen Hospital, Harkness Pavilion, Milstein Hospital, Mschony, Presbyterian Hospital and a fictitious unit 'Rest' with all the other wards, respectively (SM Figure S6).

### *Model framework and setting: micro-organism prevalence and effective sensitivity*

We assumed community-acquired carriage was at a steady state and parametrized the probability of importing a micro-organism from the community. Studies for the pathogens considered in this study setting, northern Manhattan, are limited. We therefore searched the literature for studies in other settings and used these to inform the importation rate, a proxy for community prevalence, for each micro-organism. Our search terms included 'prevalence', 'colonization', and 'carriage rate', and we used reviews for some of the microparasites that report pooled estimates across different geographical locations. In the Supplementary Material section *Prevalence estimates* we present these different values, sources, and a brief description of the study, with the geographical location and population of interest. We report the values used for this analysis in Table 1. In the analyses, we fixed importation rates, , of micro-organisms using the estimated community prevalence and designed an individual-level observational model to quantify the likelihood of detection upon testing given an effective sensitivity, .

### *Identifiability: synthetic simulations and inference*

To investigate the identifiability of the model-inference system, we explored inferences on simulated trajectories with known parameters. To determine whether the model-inference system is capable of recovering parameters for a variety of parameter combinations, we fixed the importation rate to 25% and 50% and varied the effective sensitivity and the nosocomial transmission rate . The parameter estimates of the inference consistently captured the true parameter values for both detection rates and nosocomial transmission rates (Figure 2A and 2B). The model-inference system was able to explore different regions of the prior range (limits of each axis) and captured the true parameter values. Inferences with were more biased but resulted in a broader distribution than inferences with . Estimates were substantially biased in only 2 instances, see scenarios 2 and 4 in Figure 2. Visual inspection confirms that the marginal posterior estimates asymptotically converge to the true parameter values in the inference algorithm (see SM Figure S7A and S7B). We also investigated how statistical uncertainty and Monte Carlo error impact the inference (see *Making sense of bias* in Methods). In particular, goodness-of-fit can be predicted by the oddness of a stochastic realization (SM Figure S8).

We ran model simulations using parameters drawn from posterior distributions and compared them with simulations generated using the true parameters (SM Figure S9A.1 and S9B.1, for respectively). We find that, qualitatively, even biased inferences (scenarios 2 and 4 in Figure 2) reproduce well both the true ensemble simulations and the assimilated data. We quantified the calibration to study if observed uncertainty is reproduced by the model (see *Calibration of inference* in Methods) for both the inferred posterior and truth ensemble simulations (SM Figure S9A.2 and S9B.2, for respectively). Calibration of the inference and true ensemble simulations were similar, the credible intervals were slightly broad (above the diagonal), but better for the true ensemble simulations, as expected. Lower values of weekly incident simulated data (low and low ) resulted in less calibrated ensembles.

Lastly, we explored the ability model-inference system to estimate solutions in different regions of the parameter range. We used parameter combination pairs that in free simulation produce aggregated infections at the hospital level (aggregation of buildings time series shown in Figure 1A) that match observations for each pathogenic bacteria. We found the model-inference system capable of discriminating among the scenarios. This analysis covers a broad range of prevalence, implying 'low' to 'high' importation rates, ; 4% and 16% for MRSA and *K. pneumoniae* to 50% and 70% for *E. faecium* and *E. coli* respectively.

### *Inference using real data*

We applied the model-inference system to estimate epidemiological properties for eight different microbial bacterial pathogens. In Figure 3 we present the posterior joint estimates of the effective sensitivity, , and nosocomial transmission rate, . For most species (except MRSA) the system localizes the posterior to the same region of the prior range (limits of each axis). Subplot insets zoom to the solution region to highlight the level of sensitivity to We find *E. coli* has the lowestestimated mean nosocomial transmission rate across tested bacteria and the highest mean effective sensitivity (18.53%, 18.63%, and 17.31% for community prevalence of 55%, 63% and 70% respectively, see Table 2 for confidence intervals). *E. coli* is followed by *MSSA* with mean nosocomial transmission rates of 0.141, 0.121, and 0.0994 and effective sensitivity of 1.63%, 1.65%, and 1.58% for prevalence of 25%, 29% and 35%, respectively (Table 2 for 95% CI). For the rest of the micro-organisms, except MRSA, we find that mean nosocomial transmission rate estimates are consistently between 0.17 and 0.19, and effective sensitivity between 0.5% (*E. faecium)* and 2.5% (*K. pneumoniae*). Effective sensitivity is similar for all pathogens, except *E. coli* (SM Figure S10B). Mean nosocomial transmission estimates are consistently between 0.15 and 0.2, except for *E. coli* and MSSA (SM Figure S10A). The nosocomial transmission rate estimates for MSSA are the second lowest with mean values between 0.099 and 0.14. For MRSA*,* the model-inference system finds one of two solutions depending on the value of . For equal to 5% and 10%, the system estimates high equal to 0.17 and 0.18, respectively, and low , 1.96 and 1.38, respectively (see Table 2 for 95% CI); for the lowest community prevalence of 3.9%, the estimated nosocomial transmission rate is 0.0017 and the estimated sensitivity is 17.15%.

To assess goodness-of-fit, we simulated each pathogen using posterior parameter estimates. We find that simulated detected nosocomial infections span the observed numbers at the hospital and building level for all micro-organisms (Figure 4A and SM Figure S11 for each spatial scale, respectively). We produced reliability plots to examine whether model simulated uncertainty is appropriately calibrated to observations (Methods). We find that the coverage of simulated CIs is just below expected values (under the diagonal line), indicating slightly narrow uncertainty (Figure 4B). The model simulations allow estimation of the relative contributions of importation and nosocomial transmission to the overall burden of different bacterial pathogens as well as the prevalence in the hospital.

**Discussion**

**Discussion:**

1. Opening paragraph: Restate the overall approach, restate findings, discuss a few important features of the findings and their implications.

2. How nosocomial transmission findings and effective sensitivitty match prior findings.

3. Further discussion of how the network contact patterns guide exposure risk.

4. Discussion of \rho, what it means, why is it low, math bit in the supplement.

5. Limitations of the approach.

6. What models like this tell us and what else needs to be done.

The epidemiological properties of microorganisms present in hospitals are difficult to quantify due to short lenght of stay in hospital, sparse surveillance data and imcomplete understanding of the mechanisms behind transmission. Estimating these epidemiological properties can help understand the mechanism behind the risk of contagion, such as nosocomial transmission, and help support improved AMR control in healthcare systems. Similarly, measuring detection could guide quantitative design of surveillance in clinics, such as clinical culture allocation and understand tradeoffs between screening and diagnosis. In this paper, we used an agent-based model informed by patient hospitalization records and patient level clinical culture testing data from a large hospital system in New York City to study eight co-circulating bacterial pathogens. The fine-grained patient level data enriched patient movement across the hospital system. To simualte transsmission of individual bacterial pathogens we parametrized the risk of acquiring bacteria proportional to a nosocomial transmission rate , and the fraction of carriers an individual was contacted during the day. We designed an individual-level observational model to represent the likelihood of detection upon testing given the effective sensitivity . We coupled this model with an inference algorithm to quantify these epidemiological features for these organisms and found limited variability in these estimated properties, suggesting that network contact patterns, admission, discharges and transfers at ward facility commonly guide exposure and transmission risk.

*WHY FOCUS ON THIS DETAIL?* We assumed the decolonization rate was constant for all micro-organisms. However, it is known for example that *S. pneumoniae* bacterial clearances are strain dependent *(28)*. Host clearance rates for the bacteria studied in this work are reported between 5 months to 2 years *(10, 24)*. In SM section *The ordinary differential equation* we show the basic reproductive number is non-sensitive to clearance rate given the fast replacement of patients in the hospital network. This plus the stochastic nature of the ABM and the desegregation of the transmission rate by building a time-varying contact network will make the inference robust to changes in the bacterial decolonization period (Methods).

The record of observed nosocomial infection (Figure 1A) is a stochastic realization among an ensemble of possible outcomes of an underlying stochastic process. In order to evaluate the ability of the system to infer key epidemiological parameters, we studied how the oddness of a stochastic realization impacts the estimate. We demonstrated the method is able to recover the parameters reliably across different levels of importation rates, nosocomial transmission rates and detection levels.

We found that while *E. coli* had the highest likelihood of detection, it has the lowest nosocomial transmission rate, which is almost negligible, suggesting that most nosocomial infections can be associated with infections with host commensal strains that are not transmitted, as was reported empirically *(38)*. We found nosocomial transmission rates of *S. aureus* phenotypes were different, suggesting a difference in the fitness of the two strains. We found MSSAhas a lower nosocomial transmission rate than MRSA. The enormous known diversity of *S. aureus* hampers our understanding and discussion of the possible source of the difference. From differences in fitness and dynamical interactions of CA-MRSA and HA-MRSA *(8, 25)* to different levels of resistance in different strains *(39)* and within-host dynamics *(5, 14, 39, 40)*. As well as between-species interactions *(41)*. Further research that represents in the same framework dynamics of MSSA and MRSA in hospital settings could improve understanding of the sources of the estimated differences.

Transmission is a fundamental property that governs epidemiological dynamics and an important step in the life cycle of bacterial pathogens *(29)*. It is also one of the most challenging processes to understand and quantify. Estimating this property could improve understanding of the mechanisms facilitating contagion risk and ultimately support improved control in healthcare systems. Our model reperesents both direct and indirect modes of transmission in the same parameter . Short lengths of stay and heterogeneous admissions/discharges and hospitalizations across the hospital system, points to transmission possibly dominated by indirect modes of transmission, that do not depend on host-to-host contact, especially environmental or fomite as well as HCW-mediated transmission *(4, 30–35)*. In fact, the Center for Disease Control and Prevention (CDC) healthcare infection control webpage *(36, 37)* highlights the environment, surfaces and devices, as a common reservoir of germs in healthcare places. Explicit inclusion of such reservoirs in future models will help to understand and disentangle the contribution of environmental transmission to the risk of exposure and subsequent consequences to infection and resistance emergence.

Nosocomial infection data are typically used to approximate and infer transmission; however, collection of these observations is rarely systematized. Surveillance of micro-organism infection and colonization in clinics is usually the product of patient screenings carried out at the discretion of clinicians and diagnostic of patients with infection who present symptoms. Additionally, a clinical culture involves swabbing a patient at a specific body site, which compounded by bacterial differential niches across the body makes detection of carriage with a single clinical culture uncertain. To overcome this our observational model was designed to quantify in one parameter, , individual-level probabilities of testing, absorbing imperfect sensitivity of the clinical cultures, probabilities of testing depending on the patient status and imperfect observation across body sites. All these factors contribute to the uncertainty in detecting carriage. In the SM section *Understanding the effective sensitivity* guided by a theoretical transmission model, we show that results from the interaction of 4 main mechanisms: **i)** the biology of the infection, **ii)** properties of the clinical culture, that also produce heterogeneities across body sites, **iii)** discharge rates and **iv)** hospital surveillance settings. Factors are intrinsic to the hospital dynamics or the clinical culture except hospital surveillance settings, screening and diagnostic, and by treating patients' infections the biology of host-pathogen interaction. Our mean estimates of except for *E. coli*, were 'low' ranging from 0.5% to 2.5%. We show that there is a range of non-linear parameter combinations of the factors described above that could result in the estimated values of (see SM Figure S12 and SM section *Understanding the effective sensitivity* ). Research that clarifies the biology of the host-pathogen interaction in the hospital, specifically understanding the mechanism behind the transition from asymptomatic carriage to symptomatic infections as well as data availability about surveillance in the clinics, will permit further understanding of the estimated values. Ultimately, models that embed in their parameters probabilities of testing across a range of states of the host and model those states, such as asymptomatic carriage and symptomatic infection, will also help to have a quantitative framework to design better surveillance strategies.

Other limitations of the current study include the absence of information on the demographics, age profiles and co-morbidities of patients in the hospital network that are known to have differential susceptibility to bacterial acquisition and infection *(1)*(cite).

The current work exploits the availability of individual-level patient records to estimate the transmission properties and detection of pathogenic bacteria in hospital settings. Individual-level data and models can be employed in future research to understand the impact of individual-level interventions on disease control, such as the disinfection of shared medical devices to diminish the risk of contagion (CDC recommendations). Additionally, the ABM presented here could be used to test counterfactual scenarios of individual-level interventions that replace infections caused by resistant strains with infections caused by sensitive strains to quantify the impact of resistance at the hospital level and its broader implications to resistance emergence *(1)*.

**Materials and Methods**

### *Overview*

To estimate key epidemiological characteristics of bacterial pathogens, we developed an ABM to simulate the dynamics of these organisms in hospital settings. The model was informed by patient hospitalization and clinical culture data from electronic healthcare records collected between February 1 2020 and February 28 2021. We coupled the ABM with a Bayesian inference algorithm and, using simulated nosocomial infection data, validated the ability of this ABM inference system to identify likelihood of detection and nosocomial transmission rates. The process model, as is an ABM, tracks the state of patients at daily time scales. Patients were either susceptible to colonization or carriers with a micro-organism (See *The transmission model* in Methods). The observational model designed at the individual level allows us to map from the carriers to detected individuals via a simulated clinical culture (See *The individual observational model* in Methods). We parametrize the patient observational model with a likelihood of detection given carriage upon testing given the effective sensitivity, . Data assimilation is conducted at the building level therefore we consistently aggregate simulated patient detections at this scale (See *Hospital-level observational model* in Methods). We studied the performance of the model-inference system to estimate the parameter using synthetic data and investigate source of bias (See *Identifiability: Synthetic simulations and inference* and *Identifiability: Making sense of the bias*). We then assimilated the clinical culture data and estimated the effective sensitivity and nosocomial transmission rates for eight co-circulating bacterial pathogens (See *Inference using real-world data*). Using these estimated parameters and the ABM, we were able to reproduce the time series of positive cases for five major buildings in the hospital network and the aggregation of all the other wards in a fictitious unit 'Rest'. The estimated effective sensitivity and nosocomial transmission rates were compared for the eight pathogens.

### *Data*

Data for this study derives from 3 hospitals of a Northern Manhattan Hospital System, including a quaternary care center, pediatric hospital, and community hospital. The hospitals contain 221 wards of different types including emergency, infusion, cardiology, pediatrics, etc. Hospitalization and clinical culture data were collected during the study period from February 1 2020 to February 28 2021. The hospitalization data include admission, discharge, and transfer of patients within the hospital system. The dates and wards in which each patient stayed during hospitalization were used to construct a time-evolving contact network. Clinical culture records to confirm nosocomial infection were available for the eight most prevalent organisms in the hospital system: *E. coli*, *K. pneumoniae,* *P. aeruginosa*, methicillin-susceptible *S. aureus* (MSSA), MRSA, *S. epidermidis*, *E. faecalis* and *E. faecium*. The date and patient associated with each test were used to inform the observation model (see Methods below). We plotted the number of tests across the 3 hospitals and it's buildings (SM Figure S13) and in the wards in each hospital (SM Figure S14), which show a similar order in cultures across the hospital but substantial heterogeneity inside the hospital at the ward scale.

### *The transmission model*

We use a process ABM to simulate transmission in the study hospital system *(10)*. For the ABM, the patient-to-patient daily contact networks were constructed using hospitalization records. Two patients staying in the same ward on the same day were connected in the contact network as they are expected to have close contact, and care from the same hospital workers who might facilitate transmission of microbial pathogens and share the same environment. Due to patient movement, admissions, discharges, and transfers, the contact network is time-varying and was updated daily. We assumed each patient is either susceptible (S) or colonized (C). Patients in contact on a given day (those who shared the same ward) can be converted to carriers proportional to the number of patients in the ward carrying a particular micro-organism. We model the force of infection in ward , , following the law of mass action, as described below. We assumed colonized patients can spontaneously decolonize and become susceptible after a decolonization period days *(12)*. Colonization in hospitals is attributed to two mechanisms: importation from the community and nosocomial transmission. Specifically, we defined these two processes as follows:

1. **Importation from the community:** Using hospital admission records, we assumed a newly admitted patient is colonized with a particular pathogen with an importation probability . The number of patients admitted from the community is shown in Figure 2B. We treated all admissions independently but keep track of patient states on consecutive days and assumed their colonization status did not change.
2. **Nosocomial contact transmission:** We defined a force of infection for each ward as . The force of infection is proportional to the number of individuals carrying the pathogen in a given ward on day , denoted by . We define a frequency-dependent transmission rate per ward as , where is the average ward daily occupancy (See *Empirical patterns of hospital traffic and heterogeneity of nosocomial infections burden* in Results). The force of infection per ward is computed as:

Thus, the force of infection per ward, or probability of colonization for susceptible individuals staying in ward at time , is given by , and the transition equations governing the change of state for a patient, , residing in ward during time are shown below by, where is the decolonization rate of patient .

### *The individual observational model*

Colonized patients may develop clinical infections due to the invasion of typically sterile body sites such as blood. At other body sites (i.e. respiratory and urinary tract) the presence of pathogens may indicate both infection or colonization, depending on additional clinical variables *(1, 6)*. Other individuals carrying pathogens may be discovered through routine screening or cultures ordered discretionarily by clinicians for patients without clinical manifestations. In the process ABM, we did not explicitly distinguish between colonized and clinically infected patients. Instead, we apply an observational to encapsulate detection of carriers. This observational model represents the detection probability for clinical cultures taken from individuals in the hospital network during the study period; it captures the heterogeneous observation of micro-organisms across wards. Weekly number of cultures across hospital wards and weekly number of admitted patients have a strong positive correlation (cultures are proportional to admissions) (Figure 1D-E).

A key challenge in colonization detection is that cultures collected from one body site may miss colonization of other sites even though the patient carries the bacterium. For instance, blood cultures of a patient colonized by *E. coli* in the urinary tract will likely test negative, and this patient may not necessarily present symptoms. This compounded with differential detection of screening and diagnosis makes the parametrization challenging. To define our observational model, we therefore estimated the ‘effective sensitivity’ of detecting colonization upon testing defined by the culture data, i.e. the probability of identifying carriage given that the patient is carrying a particular micro-organism. This effective sensitivity not only represents culture test sensitivity, in the strict sense, but also the likelihood a specimen will be taken from a colonized site (and capture bacteria) given that colonization or infection exists anywhere on a patient as well as probability of detection depending on the state of the patient (colonization, and infection). The defined observational model indicates the likelihood that a colonized patient is detected given the effective sensitivity or .

### *Hospital-level observational model*

Due to substantial testing heterogeneity, observations of positive cultures were dominated by a few wards with a disproportionately large number of tests. To facilitate use of more granular observations at the sub-hospital level while avoiding excessive noise, we used the building aggregation to sum the number of nosocomial infection data and perform inference. SM Figure S12 shows the number of clinical cultures per building, we merge buildings composed mostly by outpatients and with few clinical cultures into a fictitious unit presented as 'Rest'. The number of transfers between wards is shown in SM Figure S1. Wards within each identified building have more frequent within-building transfers than cross-cluster transfers.

### *Inference*

Inference for dynamic and latent variables and parameters is often treated as a filtering problem, in which the state space is sequentially estimated as observations become available *(42, 43)*. This framework has been applied to many infectious diseases, including influenza, dengue, malaria, cholera, Ebola, enterovirus D68, and COVID-19 *(44–53)*, typically using compartmental models that represent transmission dynamics with a set of ordinary differential equations (ODEs). However, there are fewer instances in which filters have been used in conjunction with ABMs, possibly due to the considerably higher dimension of the state space for these models 17. For this ABM system, we solve the inverse problem without solving the filtering problem and therefore only estimate parameters, i.e., infer the effective sensitivity and the nosocomial transmission rate , without re-adjusting the state space.

We use the ensemble adjustment Kalman filter (EAKF *(42)*), which assumes both the prior and the observations are normally distributed. Compared to other data assimilation techniques, the EAKF is amenable for use with high-dimensional models such as numerical weather models (cite). We assimilated weekly observations of nosocomial infection detections for micro-organisms using an iterated filtering framework (IF). Iterated filtering for dynamical systems was proposed in the context of epidemiology using particle filters or sequential Monte Carlo *(54, 55)*, but has also been proposed in the context of inverse problems using Kalman filters *(56, 57)*. We base most of our implementation in the algorithms proposed in the epidemiological context but grab some ideas from the algorithms using Kalman filters. The implementation of the IF-EAKF in Python was made using the packages NumPy and SciPy *(58, 59)*. Further details and hyperparameters for this implementation are available in the Supplementary Material section Simulation-based inference framework.

### *Identifiability: Synthetic simulations and inference*

To verify that the simulation-based inference system is able to accurately estimate the two key epidemiological parameters, we first investigated the identifiability of the system using simulated data (i.e., the ability of the framework to infer the parameters when they are known). We generate synthetic observations of incident micro-organism colonization using the ABM and assigned effective sensitivity and nosocomial transmission rates . The synthetic observations were then assimilated into the full model-inference system to assess the system ability to accurately estimate the parameters.

*Identifiability: Making sense of the bias*

In the synthetic inferences the true parameters **P** are known and the simulator allows repetition of the experiment. We define the oddness of a stochastic realization as the average Continuous Ranked Probability Score (CRPS) over data assimilation times between an ensemble and a single stochastic realization. Additionally, the parameters of the simulator are known to impact the spread of the simulated data (statistical uncertainty) *(50, 60, 61)*. In consequence, we define the ensemble spread using the CRPS of ensemble simulations and its mean. Ensemble spread is a common measure in weather forecasting but is usually computed as the average deviation of an ensemble of trajectories from the mean. The CRPS is defined as shown below, F is the cumulative density distribution computed from the ensembles and y is a trajectory *(62, 63)*. is the indicator function where if or otherwise. The CRPS is in fact a standard proper score to evaluate epidemic forecasts *(62)*.

In synthetic tests, inferences about the mean trajectory could be made. The resulting inferred parameter density provides a benchmark of the method given the noise-less signal possible. Quality of inferred parameter densities conducted on single stochastic realizations can be weighted by the goodness-of-fit of the benchmark . We measure the goodness-of-fit of a posterior density given **P** as proposed in *(57)*. Specifically, the inferred posterior can be parametrized with a vector of means , and a covariance matrix and the goodness-of-fit assessed with the L2 norm scaled by , as shown below.

We first studied how the oddness of the observation explain the goodness-of-fit of the inferred parameters (SM Figure S7A). In this controlled experiment the true parameters and the ensemble spread are known. Monte Carlo errors *(64)* are controlled by normalizing by and removing statistical uncertainty . In SM Figure S7B we show the controlled goodness-of-fit of against the normalized oddness .

### *Inference using real-world data*

We next used actual weekly observations for the six buildings to infer the epidemiological parameters for all eight bacterial pathogens. The model-inference framework was applied to each pathogen separately. To account for uncertainty in the reported estimates for community prevalence for the pathogens (See Result section *Model framework and setting: micro-organism prevalence and effective sensitivity* and Table 1) we conducted inferences on 3 values for the importation rate . We, therefore, studied the sensitivity of the simulation-based inference to this parameter. The values are related to high, middle and low importation rates based on the literature review, also providing a picture of the uncertainty obtained for the estimates for and not only captured by the uncertainty obtained from the simulation-based inference.

### *Calibration of inferences*

We measure the consistency of the modeled nosocomial detection and the observed ones statistically. We produced reliability plots *(65)*. Specifically, we computed the percentage of observations falling within a given credible interval (CI) of the quantity obtained from model simulations. For a perfectly calibrated simulation, X% of observed values should fall within the X% CI generated by model simulations, producing a diagonal line (y=x) in the reliability plot.

**List of Supplementary Materials**

### *Supplementary Materials and Methods*

* Prevalence estimates.
* The ordinary differential equation. (done)
* Understanding the effective sensitivity . (done)
* Simulation-based inference framework. (done)

### *Supplementary Material Figures*

* Figure S1. Transfer matrices at the ward and building levels. (done)
* Figure S2. Ward size distribution. (done)
* Figure S3. Hospital traffic at the building level. (done)
* Figure S4. Hospital traffic at the ward level. (done)
* Figure S5. Relationship between hospitalizations-admissions. (done)
* Figure S6. Length of stay distributions at hospital and building levels. (done)
* Figure S7. Convergence plots of inferences for synthetic data. (done)
* Figure S8. Goodness-of-fit of posterior inferences on synthetic data, Monte Carlo and statistical uncertainty analysis. (done)
* Figure S9. Hospital-level simulation of synthetic tests and calibration. (done)
* Figure S10. Marginal posterior parameter estimates for each pathogen. (done)
* Figure S11. Simulation with posterior parameter estimates at the building level and calibration for each pathogen.
* Figure S12. Understanding the effective sensitivity . (done)
* Figure S13. Total clinical cultures in each hospital and its buildings.
* Figure S14. Clinical cultures per ward.

**Acknowledgments**

This study was supported by funding from the Centers for Disease Control and Prevention U01CK000592 and 75D30122C14289. We thank Matteo Perini, the Shaman lab and the CDC Modeling Infectious Diseases in Healthcare Network (MInD – Healthcare) for comments and discussions.

**Author Contributions:**

Conceptualization: J.C.V., A.C.U., S.P. and J.S.

Data curation: J.Z. and L.W.

Data processing: J.C.V. and R.Y.

Methodology: J.C.V., T.R., S.P. and J.S.

Investigation: J.C.V., R.Y., A.C.U., J.Z., S.P. and J.S.

Visualization: J.C.V.

Funding acquisition: S.P. and J.S.

Project administration: J.C.V., S.P. and J.S.

Supervision: S.P. and J.S.

Writing - original draft: J.C.V.

Writing - review & editing: All authors revised and reviewed the manuscript.

**Competing Interest Statement:** J.S. and Columbia University disclose partial ownership of SK Analytics. J.S. discloses consulting for BNI. ACU discloses research support from Merck. All other authors declare no competing interests.

**Data and materials availability:** All data, code, and materials used in the analysis must be available in some form to any researcher for purposes of reproducing or extending the analysis. Include a note explaining any restrictions on materials, such as materials transfer agreements (MTAs). Note accession numbers to any data relating to the paper and deposited in a public database; include a brief description of the data set or model with the number. If all data are in the paper and supplementary materials, include the sentence “All data are available in the main text or the supplementary materials.”

**Keywords:** epidemiology, microorganisms, antimicrobial resistance, Bayesian inference, individual-based model, agent-based model, inverse problems

**Figures and Tables:**

Main Text

Figures 1 to 6

Figure 1. Data: empirical colonization of microbial organisms, hospital admissions and number of clinical cultures.

Figure 2. Identifiability, parameter estimates of simulated data.

Figure 3. Parameter estimates of pathogenic bacteria using clincal culture data.

Figure 4. Hospital level fit and reliability plots.

Tables 1 to 2

             Table 1. Community prevalance from empirical studies.

             Table 2. Parameter estimates of pathogenic bacteria.

**A picture containing text, line, screenshot, plot

Description automatically generated**

**Figure 1. Data: empirical colonization of microbial organisms, hospital admissions and number of clinical cultures.** **A)** Nosocomial infection data for pathogenic bacteria studied by building; faded dots are the weekly incident detections, and solid lines are monthly incident detections. Buildings are color-coded: Milstein Hospital, Allen Hospital, Presbyterian Hospital, Harkness Pavilion, Mschony, and a fictitious unit Rest. From left to right and upper to lower plots: *E. coli*, *K. pneumoniae*, *P. aeruginosa*, *MSSA*, *MRSA*, *S. epidermis,* *E. faecalis* and *E. faecium*. **B)** Numbers of hospitalized (green), admitted patients (salmon) and discharged patients (blue) during the study period at daily resolution (weekly variability is evident). **C)** Numbers of in-hospital patients normalized by ward size (average occupancy per day during the study period); red lines show the 5 most populated wards, blue lines the 5 least populated, and the remaining wards are shown in gray in the background. We remove wards with size equal to 1 for better visualization. **D)** Heatmap showing the number of weekly cultures in each ward during the study period. **E)** Heatmap showing the number of patients admitted weekly to each ward during the study period.

**A picture containing text, screenshot, line, number

Description automatically generated**

**Figure 2. Identifiability, parameter estimates of simulated data.** Joint posterior estimates for **A)** importation rate =25% and **B)** =50%. The posterior estimate is highlighted with a density plot (darker means more probable), and posterior ensemble members are shown as purple dots. In each subplot the true value used for simulating the stochastic trajectory to infer is highlighted in the title of each scenario, with a yellow cross and the intersection of the two black dashed lines. x-axis shows the effective sensitivity (%) and y-axis the nosocomial transmission rate . Note that in both **A)** and **B)** the prior range is the limits of each axis, increments from left to right and from upper to lower plots.

**A picture containing screenshot

Description automatically generated**

**Figure 3. Parameter estimates of pathogenic bacteria.** Joint estimates for the effective sensitivity (%) (x-axis) and the nosocomial transmission rate (y-axis). Importation rate values for sensitvity analysis are color-coded with high, medium and low prevalences, and indicated in each subplot. Color-coded dashed lines show the mean estimates, and posterior is shown with both a density plot (darker indicates more probable) and with the posterior ensemble members plotted as dots.

**A collage of graphs

Description automatically generated with low confidence**

**Figure 4. Hospital level fit and calibration:** **A)** Ensemble simulation of modeled nosocomial data with estimated parameters, solid lines show the mean and ribbons the 95% CI. Importation rate is highlighted in the legend and color-coded for high, medium and low prevalences in all plots. Weekly nosocomial infection data is displayed with red dots, this aggreagation was not used for data assimilation. **B)** Reliability plot (Methods). Hospital-level fit with 4 different confidence intervals (25%, 50%, 75%, 97.5%). Importation rate is color-coded. The black dotted line is the reference perfect calibration.

**Table 1.** Human prevalence range from the literature, we present the 3 values used in the inferences (SI *Prevalence estimates)*.

|  |  |  |
| --- | --- | --- |
|  | Importation rate / Community prevalence (%) | Reference(s) |
| *E. coli* | 70.0 63.0 55.0 |  |
| *K. pneumoniae* | 35.0 23.0 15.0 |  |
| *P. aeruginosa* | 25.0 18.8 11.6 |  |
| MSSA | 35.0 29.0 25.0 |  |
| MRSA | 10.0 5.0 3.9 |  |
| *S. epidermidis* | 90.0 75.0 58.0 |  |
| *E. faecalis* | 55.0 47.6 36.8 |  |
| *E. faecium* | 50.0 40.6 36.8 |  |

**Table 2.** **Parameter estimates of pathogenic bacteria**. Posterior estimates for the eight bacterial pathogens; the value for the importation rate γ is shown in the second column. Nosocomial transmission rates β is presented as rate per day in the third column and likelihoods of detection upon testing in the fourth column.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Importation rate, (%)** | **Nosocomial transmission rate**  **Mean (95% CI)** | **Effective sensitivity (%)**  **Mean (95% CI)** |
| *E. coli* | 55 | 2.54E-03 (1.72E-03, 3.58E-03) | 18.53 (17.30, 19.43) |
|  | 63 | 8.66E-04 (3.78E-04, 1.58E-03) | 18.63 (17.85, 19.23) |
|  | 70 | 9.62E-04 (4.06E-04, 1.73E-03) | 17.31 (16.33, 18.02) |
| *K. pneumoniae* | 15 | 0.187 (0.180, 0.194) | 2.51 (2.36, 2.65) |
|  | 23 | 0.189 (0.180, 0.195) | 2.14 (1.98, 2.29) |
|  | 35 | 0.181 (0.165, 0.193) | 2.05 (1.92, 2.17) |
| *P. aeruginosa* | 11.6 | 0.189 (0.183, 0.194) | 1.93 (1.78, 2.07) |
|  | 18.8 | 0.190 (0.182, 0.195) | 1.61 (1.47, 1.75) |
|  | 25 | 0.187 (0.178, 0.194) | 1.55 (1.43, 1.67) |
| *MSSA* | 25 | 0.141 (0.122, 0.164) | 1.63 (1.47, 1.79) |
|  | 29 | 0.121 (0.103, 0.140) | 1.65 (1.49, 1.81) |
|  | 35 | 9.94E-02 (8.43E-02, 0.116) | 1.58 (1.40, 1.73) |
| *MRSA* | 10 | 0.170 (0.136, 0.184) | 1.38 (1.21, 1.60) |
|  | 3.9 | 1.70E-02 (9.16E-03, 2.74E-02) | 17.15 (12.48, 18.96) |
|  | 5 | 0.182 (0.168, 0.190) | 1.96 (1.72, 2.19) |
| *S. epidermidis* | 58 | 0.171 (0.117, 0.190) | 1.24 (1.12, 1.35) |
|  | 75 | 0.174 (0.134, 0.190) | 1.21 (1.10, 1.33) |
|  | 90 | 0.173 (0.107, 0.189) | 1.23 (1.12, 1.36) |
| *E. faecalis* | 36 | 0.167 (0.121, 0.186) | 1.05 (0.96, 1.17) |
|  | 47.6 | 0.149 (9.69E-02, 0.182) | 1.07 (0.96, 1.19) |
|  | 55.0 | 0.159 (0.101, 0.185) | 1.09 (0.96, 1.21) |
| *E. faecium* | 36.8 | 0.182 (0.163, 0.192) | 0.58 (0.54, 0.63) |
|  | 40.6 | 0.179 (0.147, 0.191) | 0.57 (0.54, 0.65) |
|  | 50.0 | 0.178 (0.123, 0.193) | 0.58 (0.54, 0.66) |

**References**

1. M. E. A. de Kraker, M. Lipsitch, Burden of Antimicrobial Resistance: Compared to What? *Epidemiologic Reviews* **43**, 53–64 (2022).

2. C. J. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. Robles Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, S. C. Johnson, A. J. Browne, M. G. Chipeta, F. Fell, S. Hackett, G. Haines-Woodhouse, B. H. Kashef Hamadani, E. A. P. Kumaran, B. McManigal, R. Agarwal, S. Akech, S. Albertson, J. Amuasi, J. Andrews, A. Aravkin, E. Ashley, F. Bailey, S. Baker, B. Basnyat, A. Bekker, R. Bender, A. Bethou, J. Bielicki, S. Boonkasidecha, J. Bukosia, C. Carvalheiro, C. Castañeda-Orjuela, V. Chansamouth, S. Chaurasia, S. Chiurchiù, F. Chowdhury, A. J. Cook, B. Cooper, T. R. Cressey, E. Criollo-Mora, M. Cunningham, S. Darboe, N. P. J. Day, M. De Luca, K. Dokova, A. Dramowski, S. J. Dunachie, T. Eckmanns, D. Eibach, A. Emami, N. Feasey, N. Fisher-Pearson, K. Forrest, D. Garrett, P. Gastmeier, A. Z. Giref, R. C. Greer, V. Gupta, S. Haller, A. Haselbeck, S. I. Hay, M. Holm, S. Hopkins, K. C. Iregbu, J. Jacobs, D. Jarovsky, F. Javanmardi, M. Khorana, N. Kissoon, E. Kobeissi, T. Kostyanev, F. Krapp, R. Krumkamp, A. Kumar, H. H. Kyu, C. Lim, D. Limmathurotsakul, M. J. Loftus, M. Lunn, J. Ma, N. Mturi, T. Munera-Huertas, P. Musicha, M. M. Mussi-Pinhata, T. Nakamura, R. Nanavati, S. Nangia, P. Newton, C. Ngoun, A. Novotney, D. Nwakanma, C. W. Obiero, A. Olivas-Martinez, P. Olliaro, E. Ooko, E. Ortiz-Brizuela, A. Y. Peleg, C. Perrone, N. Plakkal, A. Ponce-de-Leon, M. Raad, T. Ramdin, A. Riddell, T. Roberts, J. V. Robotham, A. Roca, K. E. Rudd, N. Russell, J. Schnall, J. A. G. Scott, M. Shivamallappa, J. Sifuentes-Osornio, N. Steenkeste, A. J. Stewardson, T. Stoeva, N. Tasak, A. Thaiprakong, G. Thwaites, C. Turner, P. Turner, H. R. van Doorn, S. Velaphi, A. Vongpradith, H. Vu, T. Walsh, S. Waner, T. Wangrangsimakul, T. Wozniak, P. Zheng, B. Sartorius, A. D. Lopez, A. Stergachis, C. Moore, C. Dolecek, M. Naghavi, Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).

3. S. Lehtinen, Co-colonisation and coexistence. *Nat Ecol Evol* **3**, 334–335 (2019).

4. T. M. Pham, M. Kretzschmar, X. Bertrand, M. Bootsma, on behalf of COMBACTE-MAGNET Consortium, R. D. Kouyos, Ed. Tracking Pseudomonas aeruginosa transmissions due to environmental contamination after discharge in ICUs using mathematical models. *PLoS Comput Biol* **15**, e1006697 (2019).

5. N. G. Davies, S. Flasche, M. Jit, K. E. Atkins, Within-host dynamics shape antibiotic resistance in commensal bacteria. *Nat Ecol Evol* **3**, 440–449 (2019).

6. M. Lipsitch, C. T. Bergstrom, B. R. Levin, The epidemiology of antibiotic resistance in hospitals: Paradoxes and prescriptions. *Proc Natl Acad Sci USA* **97**, 1938 (2000).

7. C. T. Bergstrom, M. Lo, M. Lipsitch, Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proc Natl Acad Sci U S A* **101**, 13285 (2004).

8. R. Kouyos, E. Klein, B. Grenfell, B. R. Levin, Ed. Hospital-Community Interactions Foster Coexistence between Methicillin-Resistant Strains of Staphylococcus aureus. *PLoS Pathog* **9**, e1003134 (2013).

9. S. Bonhoeffer, M. Lipsitch, B. R. Levin, Evaluating treatment protocols to prevent antibiotic resistance. *Proc Natl Acad Sci USA* **94**, 12106 (1997).

10. S. Pei, F. Morone, F. Liljeros, H. Makse, J. L. Shaman, Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. *eLife* **7**, e40977 (2018).

11. P. Paul, R. B. Slayton, A. J. Kallen, M. S. Walters, J. A. Jernigan, Modeling Regional Transmission and Containment of a Healthcare-associated Multidrug-resistant Organism. *Clin Infect Dis* **70**, 388–394 (2020).

12. B. S. Cooper, G. F. Medley, S. P. Stone, C. C. Kibbler, B. D. Cookson, J. A. Roberts, G. Duckworth, R. Lai, S. Ebrahim, Methicillin-resistant Staphylococcus aureus in hospitals and the community: Stealth dynamics and control catastrophes. *Proceedings of the National Academy of Sciences* **101**, 10223–10228 (2004).

13. S. Pei, F. Liljeros, J. Shaman, Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2111190118 (2021).

14. D. R. Smith, L. Temime, L. Opatowski, Microbiome-pathogen interactions drive epidemiological dynamics of antibiotic resistance: A modeling study applied to nosocomial pathogen control. *eLife* **10**, e68764 (2021).

15. M. Lipsitch, M. H. Samore, Antimicrobial Use and Antimicrobial Resistance: A Population Perspective. *Emerg. Infect. Dis.* **8**, 347–354 (2002).

16. D. J. Austin, K. G. Kristinsson, R. M. Anderson, The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance. *Proceedings of the National Academy of Sciences* **96**, 1152–1156 (1999).

17. S. E. Drohan, S. A. Levin, B. T. Grenfell, R. Laxminarayan, Incentivizing hospital infection control. *Proc Natl Acad Sci USA* **116**, 6221–6225 (2019).

18. S. Lehtinen, F. Blanquart, M. Lipsitch, C. Fraser, Mechanisms that maintain coexistence of antibiotic sensitivity and resistance also promote high frequencies of multidrug resistance. .

19. G. M. Knight, N. G. Davies, C. Colijn, F. Coll, T. Donker, D. R. Gifford, R. E. Glover, M. Jit, E. Klemm, S. Lehtinen, J. A. Lindsay, M. Lipsitch, M. J. Llewelyn, A. L. P. Mateus, J. V. Robotham, M. Sharland, D. Stekel, L. Yakob, K. E. Atkins, Mathematical modelling for antibiotic resistance control policy: do we know enough? *BMC Infect Dis* **19**, 1011 (2019).

20. K. Cranmer, J. Brehmer, G. Louppe, The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055–30062 (2020).

21. A. S. de Vos, S. J. de Vlas, J. A. Lindsay, M. E. E. Kretzschmar, G. M. Knight, Understanding MRSA clonal competition within a UK hospital; the possible importance of density dependence. *Epidemics* **37**, 100511 (2021).

22. M. Forrester, A. Pettitt, G. Gibson, Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostatistics* **8**, 383–401 (2007).

23. D. W. Eyre, M. Laager, A. S. Walker, B. S. Cooper, D. J. Wilson, on behalf of the CDC Modeling Infectious Diseases in Healthcare Program (MInD-Healthcare), R. D. Kouyos, Ed. Probabilistic transmission models incorporating sequencing data for healthcare-associated Clostridioides difficile outperform heuristic rules and identify strain-specific differences in transmission. *PLoS Comput Biol* **17**, e1008417 (2021).

24. B. S. Cooper, G. F. Medley, S. J. Bradley, G. M. Scott, An Augmented Data Method for the Analysis of Nosocomial Infection Data. *American Journal of Epidemiology* **168**, 548–557 (2008).

25. A.-C. Uhlemann, M. Otto, F. D. Lowy, F. R. DeLeo, Evolution of community- and healthcare-associated methicillin-resistant Staphylococcus aureus. *Infection, Genetics and Evolution* **21**, 563–574 (2014).

26. D. L. Smith, S. A. Levin, R. Laxminarayan, Strategic interactions in multi-institutional epidemics of antibiotic resistance. *Proc Natl Acad Sci U S A* **102**, 3153 (2005).

27. the EuSCAPE Working Group, the ESGEM Study Group, S. David, S. Reuter, S. R. Harris, C. Glasner, T. Feltwell, S. Argimon, K. Abudahab, R. Goater, T. Giani, G. Errico, M. Aspbury, S. Sjunnebo, E. J. Feil, G. M. Rossolini, D. M. Aanensen, H. Grundmann, Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. *Nat Microbiol* **4**, 1919–1929 (2019).

28. S. Lehtinen, F. Blanquart, N. J. Croucher, P. Turner, M. Lipsitch, C. Fraser, Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1075–1080 (2017).

29. R. M. Anderson, R. M. May, Population biology of infectious diseases: Part I. *Nature* **280**, 361–367 (1979).

30. D. J. Weber, D. Anderson, W. A. Rutala, The role of the surface environment in healthcare-associated infections: *Current Opinion in Infectious Diseases* **26**, 338–344 (2013).

31. E. Tajeddin, M. Rashidan, M. Razaghi, S. S. S. Javadi, S. J. Sherafat, M. Alebouyeh, M. R. Sarbazi, N. Mansouri, M. R. Zali, The role of the intensive care unit environment and health-care workers in the transmission of bacteria associated with hospital acquired infections. *Journal of Infection and Public Health* **9**, 13–23 (2016).

32. R. W. Loftus, J. R. Brown, M. D. Koff, S. Reddy, S. O. Heard, H. M. Patel, P. G. Fernandez, M. L. Beach, H. L. Corwin, J. T. Jensen, D. Kispert, B. Huysman, T. M. Dodds, K. L. Ruoff, M. P. Yeager, Multiple Reservoirs Contribute to Intraoperative Bacterial Transmission. *Anesthesia & Analgesia* **114**, 1236–1248 (2012).

33. L. Chaoui, R. Mhand, F. Mellouki, N. Rhallabi, Contamination of the Surfaces of a Health Care Environment by Multidrug-Resistant (MDR) Bacteria. *International Journal of Microbiology* **2019**, 1–7 (2019).

34. S. Reuter, A. Sigge, H. Wiedeck, M. Trautmann, Analysis of transmission pathways of Pseudomonas aeruginosa between patients and tap water outlets\*: *Critical Care Medicine* **30**, 2222–2228 (2002).

35. D. Lu, A. Aleta, Y. Moreno, Assessing the Risk of Spatial Spreading of Diseases in Hospitals. *Front. Phys.* **10**, 882314 (2022).

36. Center for Disease Control and Prevention, Learn About Infection Control in Health Care (available at https://www.cdc.gov/infectioncontrol/projectfirstline/healthcare.html).

37. Center for Disease Control and Prevention, Learn Where Germs Live in Health Care (available at https://www.cdc.gov/infectioncontrol/projectfirstline/healthcare/where-germs-live.html).

38. A. D. Harris, M. Kotetishvili, S. Shurland, J. A. Johnson, J. G. Morris, L. L. Nemoy, J. K. Johnson, How important is patient-to-patient transmission in extended-spectrum β-lactamase Escherichia coli acquisition. *American Journal of Infection Control* **35**, 97–101 (2007).

39. Q. Leclerc, A. Clements, H. Dunn, J. Hatcher, J. A. Lindsay, L. Grandjean, G. M. Knight, *Quantifying patient- and hospital-level antimicrobial resistance dynamics in* Staphylococcus aureus *from routinely collected data* (Epidemiology, 2023; http://medrxiv.org/lookup/doi/10.1101/2023.02.15.23285946).

40. B. J. Arnold, I.-T. Huang, W. P. Hanage, Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* **20**, 206–218 (2022).

41. M. E. Schoen, M. A. Jahne, J. Garland, L. Ramirez, A. J. Lopatkin, K. A. Hamilton, Quantitative Microbial Risk Assessment of Antimicrobial Resistant and Susceptible *Staphylococcus aureus* in Reclaimed Wastewaters. *Environ. Sci. Technol.* **55**, 15246–15255 (2021).

42. J. L. Anderson, An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.* **129**, 2884–2903 (2001).

43. M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50**, 174–188 (2002).

44. R. Subramanian, Q. He, M. Pascual, Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *PNAS* **118** (2021), doi:10.1073/pnas.2019716118.

45. V. Romeo-Aznar, L. Picinini Freitas, O. Gonçalves Cruz, A. A. King, M. Pascual, Fine-scale heterogeneity in population density predicts wave dynamics in dengue epidemics. *Nat Commun* **13**, 996 (2022).

46. M. Santos-Vega, P. P. Martinez, K. G. Vaishnav, V. Kohli, V. Desai, M. J. Bouma, M. Pascual, The neglected role of relative humidity in the interannual variability of urban malaria in Indian cities. *Nat Commun* **13**, 533 (2022).

47. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).

48. S. Pei, J. Shaman, S. Y. Del Valle, Ed. Aggregating forecasts of multiple respiratory pathogens supports more accurate forecasting of influenza-like illness. *PLoS Comput Biol* **16**, e1008301 (2020).

49. S. Pei, X. Teng, P. Lewis, J. Shaman, Optimizing respiratory virus surveillance networks using uncertainty propagation. *Nat Commun* **12**, 222 (2021).

50. S. Pei, M. A. Cane, J. Shaman, V. E. Pitzer, Ed. Predictability in process-based ensemble forecast of influenza. *PLoS Comput Biol* **15**, e1006783 (2019).

51. S. W. Park, M. Pons-Salort, K. Messacar, C. Cook, L. Meyers, J. Farrar, B. T. Grenfell, Epidemiological dynamics of enterovirus D68 in the United States and implications for acute flaccid myelitis. *Sci. Transl. Med.* **13**, eabd2400 (2021).

52. W. Yang, S. Kandula, M. Huynh, S. K. Greene, G. V. Wye, W. Li, H. T. Chan, E. McGibbon, A. Yeung, D. Olson, A. Fine, J. Shaman, Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases* **21**, 203–212 (2021).

53. I. Ukawuba, J. Shaman, M. Meier-Schellersheim, Ed. Inference and dynamic simulation of malaria using a simple climate-driven entomological model of malaria transmission. *PLoS Comput Biol* **18**, e1010161 (2022).

54. E. L. Ionides, C. Breto, A. A. King, Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103**, 18438–18443 (2006).

55. E. L. Ionides, D. Nguyen, Y. Atchadé, S. Stoev, A. A. King, Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc Natl Acad Sci USA* **112**, 719–724 (2015).

56. D. Sanz-Alonso, A. M. Stuart, A. Taeb, Inverse Problems and Data Assimilation (2023) (available at http://arxiv.org/abs/1810.06191).

57. D. Z. Huang, T. Schneider, A. M. Stuart, Iterated Kalman Methodology For Inverse Problems (2022) (available at http://arxiv.org/abs/2102.01580).

58. C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy. *Nature* **585**, 357–362 (2020).

59. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).

60. D. Alonso, A. J. McKane, M. Pascual, Stochastic amplification in epidemics. *J. R. Soc. Interface.* **4**, 575–582 (2007).

61. B. M. Bolker, B. T. Grenfell, Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12648–12653 (1996).

62. J. Bracher, E. L. Ray, T. Gneiting, N. G. Reich, V. E. Pitzer, Ed. Evaluating epidemic forecasts in an interval format. *PLoS Comput Biol* **17**, e1008618 (2021).

63. T. Gneiting, F. Balabdaoui, A. E. Raftery, Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **69**, 243–268 (2007).

64. E. L. Ionides, C. Breto, J. Park, R. A. Smith, A. A. King, Monte Carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface.* **14**, 20170126 (2017).

65. The International Research Institute for Climate and Society, Description of the IRI Climate Forecast Verification Scores (available at https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1006783&type=printable).