Specific Aims

In recent decades a number of antimicrobial resistant (AR) pathogens have emerged in healthcare facilities. The prevalence of these drug-resistant agents has imposed a substantial morbidity and mortality burden in the United States and complicated the containment of healthcare associated infections (HAIs). To better prevent and control antimicrobial resistant HAIs, it is necessary to quantify and evaluate potential intervention strategies against HAIs. However, testing and optimizing interventions in real-world healthcare settings is impractical, or even impossible. Mathematical models depicting the transmission dynamics of AR pathogens within and among healthcare facilities provide an alternate, structured *in silica* environment for exploring, quantifying and optimizing prevention and response strategies. These models, together with healthcare data, can be used to analyze HAI transmission dynamics, predict outbreaks of AR pathogens, and identify effective intervention bundles that more efficiently reduce HAI incidence. While never a perfect representation of real-world HAI dynamics, data-driven model simulations can support the development of evidence-based policy planning for actual HAI outbreaks.

In this project, we propose to develop and use a hierarchy of model structures to explore AR transmission dynamics and intervention effects at a variety of scales. We will use hospitalization records and diagnostic data for multiple AR pathogens from four major hospitals in New York City to conduct a series of studies. Our specific aims for this project are:

Aim 1: Model the spread of AR pathogens in healthcare systems at individual and facility levels. We will develop agent-based and metapopulation models to simulate the spread of AR pathogens among individual patients and healthcare facilities, respectively. These models, constructed using real-world patient-to-patient contact networks and inter-facility patient movement, will incorporate a number of transmission mechanisms, including nosocomial contact transmission, environmental contamination and interaction with the community.

Aim 2: Infer key epidemiological parameters and asymptomatic carriage probabilities using sparse observations. In this capacity we propose to develop and apply novel computational methods to estimate model parameters and variables in both the agent-based and metapopulation model constructs. Having calibrated the models using historical diagnostic records, we will estimate incidence attributable to nosocomial contact transmission, environmental contamination and importation, as well as prevalence of asymptomatic carriage in hospital. In addition, we will develop an efficient individual-level inference framework for work with high-dimensional agent-based models and sparse observations of infection/colonization. The colonization probability of an individual will be then inferred to identify high-risk patients and potential AR spreaders.

Aim 3: Predict outbreaks of HAI pathogens in healthcare facilities. We will use the metapopulation model developed in Aim 1 in conjunction with the inference methods employed in Aim 2 to design a framework for forecasting the progression of HAI spread within healthcare systems. Probabilistic predictions for onset time, near-term incidence and the estimated number of asymptomatic carriers will be generated using information as it would be available in real time. The forecasting framework will be validated using model-simulated outbreaks as well as historical records of multiple AR pathogens. Wards/clinics subject to a high forecast probability of AR invasion can thus be identified in order to facilitate targeted, preemptive control.

Aim 4: Evaluate and optimize intervention combinations given logistical constraints. A suite of intervention strategies such as targeted screening, patient isolation, barrier precautions, and environment decolonization will be tested. Specifically, identification of asymptomatic carriers in Aim 2 can be used to inform a targeted screening strategy, whose effectiveness will be compared with traditional approaches based on length of stay and contact tracing. These strategies will be evaluated both in isolation and in combination, and their effectiveness against multiple AR pathogens will be assessed and compared. Further, we will develop a framework to optimize intervention bundles of multiple strategies to achieve maximum HAI reductions.

Aim 5: Work with the CDC and the network to test interventions using different modeling approaches. We propose to collaborate within the network to perform inter-comparisons and syntheses of findings across modeling groups. We will quantify and communicate disagreement in intervention outcomes obtained using distinct model constructs as a means of specifying uncertainty. Reference outbreak scenarios will be defined in consultation with the CDC and participating groups, and a set of intervention combinations will be tested using different modeling approaches. Standardized communication and visualization methods that disseminate model projections and intervention effectiveness findings to public health officials will be developed to assist decision-making and HAI outbreak control efforts.

A. Background and Significance

Antimicrobial resistant (AR) pathogens remain a major cause of healthcare associated infections (HAIs) in the United States[1-4]. Due to a lack of effective treatment and high mortality rates, the prevalence of these existing and emerging drug-resistant agents continues to impose a heavy burden on U.S. healthcare systems. A recent study of AR pathogen-associated infections in healthcare settings reported that 4515 hospitals had at least one HAI between 2011 and 2014[3]. Most HAIs occur in general hospitals, though they are prevalent in other healthcare settings as well. In 2019, five pathogens were highlighted as urgent threats by the CDC[2]: carbapenem-resistant *Acinetobacter*[5-8], *Candida auris* (*C. auris*)[9-10], *Clostridioides difficile* (*C. diff*)[11], carbapenem-resistant *Enterobacteriaceae* (CRE)[12-14], and drug-resistant *Neisseria gonorrhoeae*[15,16]. Other serious threats include multidrug-resistant (MDR) *Pseudomonas aeruginosa* (*P. aeruginosa*)[17-19], vancomycin-resistant *enterococci* (VRE)[20-22], ESBL-producing *Enterobacteriaceae*[23,24], methicillin-resistant *Staphylococcus aureus* (MRSA)[25-27], and drug-resistant *Streptococcus pneumoniae*[28,29].

To better control existing AR pathogen-associated HAIs and prepare for the possible emergence of a novel AR organism, better, more targeted identification and intervention strategies need to be developed. Due to the difficulty testing interventions in real-world hospital settings, mathematical models are the principal option for initially vetting and quantifying the effectiveness of interventions against HAIs. A number of models with different levels of complexity have been used to simulate HAI outbreaks and evaluate intervention strategies against single or multiple AR pathogens[30-34]. These models range from simple compartmental models[35-39] to metapopulation models connecting multiple healthcare facilities[40-43] to individual-level agent-based models that incorporate detailed person-to-person interactions during hospitalization[44-50]. As intervention findings are highly dependent on the processes represented in these models, it is important to use a well-specified model structure that realistically simulates transmission dynamics.

As demonstrated in clinical studies, multiple processes can contribute to the emergence and transmission of HAIs: 1) **transmission due to person-to-person contact**, 2) **environmental contamination of the healthcare facility**, and 3) **importation from outside the healthcare facility**. In hospital settings, the relative importance of these processes is not well established. In some healthcare facilities, healthcare worker-mediated contact remains an important route of nosocomial transmission of AR pathogens[2]. At the same time, more recent evidence indicates that environmental contamination contributes to hospital infection with MRSA and VRE[51], and *Candida* species have been recovered on high-touch surfaces such as floors and sink drains in multiple hospitals, including one experiencing a *C. auris* outbreak[52]. In addition, with the increasing prevalence of certain AR pathogens in the general population, cases originating from the broader community may enter healthcare systems and initiate HAI outbreaks in hospitals[53,54]. For example, a substantial proportion of MRSA cases in Sweden have been imported from abroad due to travel and healthcare contacts in foreign countries[55,56]. Given such entwined transmission dynamics, the above three processes should all be properly represented in mathematical models so that different outbreak scenarios with distinct primary modes of transmission can be flexibly simulated.

HAI control efforts could be improved with more accurate estimation of the burden of disease and the primary modes of transmission in healthcare settings. Modeling can help identify such features through inference of colonization rates and key epidemiological parameters such as nosocomial transmissibility, environmental contamination levels and importation rates from the community. For instance, an augmented data method has been developed to estimate transmission and importation rates in an agent-based model using screening outcomes from a single ward[57]. This approach was later generalized and applied to inter-facility models with time-varying contact patterns[58,59]. More recently, we developed an efficient inference framework for use with large-scale agent-based models and applied this system to hospitalization records of three quarters of a million patients in 66 Swedish hospitals[44]. Using MRSA surveillance data collected over 6 years, we estimated that the majority of reported cases were imported, which suggests better screening at admission should be prioritized. Such situational information is critical for the design of cost-effective interventions.

Even with a good estimation of the burden of disease, controlling HAIs will likely remain difficult. Most AR pathogens, including MRSA[60,61], CRE[62-64], *C. diff*[65-70] and *P. aeruginosa*[71,72], can colonize patients without symptoms for long periods of time. In the United States, around 5% of patients in hospital carry MRSA[73]. Most colonized patients are asymptomatic and unobserved; however, these asymptomatic carriers can still facilitate stealth transmission through environmental contamination or direct contact. Indeed, modeling has indicated that failure to account for asymptomatic colonized patients can result in long-term control issues[35]. While AR pathogen colonization in clinical settings has been linked to a number of risk factors[62-64,74-81], such as recent

hospitalization and antibiotic use, asymptomatic carriers without these risk factors may still be under-detected and shed AR pathogens[75]. Without better identification methods, asymptomatic carriage of AR pathogens remains a major barrier to effective containment/elimination of HAIs[35].

The movement and assortment of patients within and among healthcare facilities also strongly affect HAI pathogen transmission dynamics and complicate HAI intervention and control. Contacts between patients, patients and health care workers (HCWs) and among HCWs have been tracked using wearable proximity sensors[82,83] and reconstructed from electronic medical records (EMR)[44,84]. The time-varying person-to-person contact networks exhibit large heterogeneity in the number and duration of contacts[44,82-84]. In particular, a small number of individuals account for a disproportionate fraction of recorded contacts and have the potential to act as super-spreaders[85-89]. At the facility scale, patient transfer and referral patterns between hospitals have been analyzed in the Netherlands[90], France[91] and California[92,93]. Modeling studies indicate that such inter-hospital connections may facilitate the regional spread of AR pathogens. A coordinated approach implemented in multiple facilities thus may be necessary to interrupt HAI spread[41,45].

In addition, the evolution of existing AR pathogens and the potential emergence of novel AR organisms pose a pressing threat. For example, community-associated MRSA (CA-MRSA) USA300 genotype has emerged as a major cause of healthcare-associated blood stream infections in the United States[94,95]. While most other AR pathogens remain restricted to healthcare settings, this risk of emergence as a community infection cannot be ignored. Further, antimicrobial resistance in one strain may contribute to the spread to other currently antibiotic-sensitive pathogens. An alarming example is the acquisition of carbapenem resistance in CRE; *Klebsiella pneumonia* carbapenemase (KPC), the most common carbapenemase in the United States, was transmitted through mobile gene vehicles such as plasmids during horizontal gene transfer (HGT)[96-98]. Although CRE spreads mostly as stable clones, the coexistence of both resistant and sensitive pathogens within one host could facilitate the spread of resistance[99-101], and, of course, the selective pressure for resistance exerted by antibiotic use may expedite the emergence of new AR agents[102-105].

In hospital settings, multiple intervention options exist to combat HAIs, each of which targets a specific route of transmission. 1) **Hand hygiene and barrier precautions** reduce the probability of transmission during healthcare contact. Indeed, available evidence indicates that improved hand hygiene and barrier precautions lead to a reduction in HAIs[106,107]. Contact precautions such as the use of gloves and gowns have also been found important for reducing endemic MRSA and VRE transmission[108]. 2) **Isolation of infections** reduces the contact frequency of infected individuals with other patients and HCWs. Nevertheless, a modeling study on MRSA interventions indicates that isolation of MRSA carriers identified by clinical cultures alone is insufficient to control its spread[40]. Moreover, reduced HCW-patient contact may lower the quality of patient care[109]. 3) **HCW staffing practices** can reduce the number of contacts between patients and healthcare providers. Associations between high nurse-to-patient ratios and reduced HAIs have been identified in several observational studies[110-112]. 4) **Environmental cleaning** reduces the risk of transmission from contaminated surfaces and devices. Disinfection of high-touch surfaces and equipment has been demonstrated effective in reducing infections with *C. diff*, VRE and MRSA[113-116]. 5) **Active screening and contact tracing** can locate otherwise unobserved carriers of AR pathogens[74,117], which helps to disrupt subsequent transmission from these patients. This proactive measure, though rarely implemented, could be of particular importance when large numbers of carriers are imported from the community[44]. 6) **Use of rapid tests** to identify pathogens and their antimicrobial susceptibility can reduce the period of unnoticed shedding of AR pathogens[40]. 7) **Antimicrobial stewardship (AS)** programs can slow the evolution of resistance and the appearance of MDR pathogens. A recent modeling analysis indicates that the use of an AS program in conjunction with multifaceted infection control could avert up to 600,000 *C. diff* infections over five years in the US[118,119].

In practice, a single intervention strategy is unlikely to achieve optimal HAI control. As a result, the effectiveness of combinations of interventions, i.e., bundles, should be evaluated and compared so that the optimal bundle approach can be identified.

Here we will use a hierarchy of models to simulate HAI outbreaks in healthcare systems. Specifically, we will leverage two model forms: 1) a metapopulation model that represents connectedness between multiple healthcare facilities at a specified resolution (e.g. clinics or hospitals); and 2) an agent-based model, described using a modified Overview, Design concepts, and Details (ODD) framework[120,121], that depicts individual-level patient movement among multiple hospital locations. Both models will simulate person-to-person transmission, environmental contamination and community importation. The model structures will be designed to represent the real-world hospitalization records of four major hospitals in New York City over a period of 11 years. Both

models will be validated using diagnostic records for **six** AR pathogens available in this dataset. Key epidemiological parameters, including transmission rates, environmental contamination rates and importation rates, will be inferred using these hospitalization and diagnostic records.

Following model validation, we will perform a comprehensive evaluation of the effects of various individual-level, facility-level and regional-level intervention strategies, both in isolation and in combination. The availability of data for multiple AR pathogens within the same healthcare system during the same period will enable simultaneous evaluation of the impact of interventions on a range of organisms. To improve the utility of model findings, we will perform uncertainty analyses and generate distributions of possible outcomes (e.g., the number of averted cases/colonizations) and thus account for imperfect identification within each model. Note, both models will be run **stochastically**. Sensitivity analyses will also be performed to identify the parameters and model structures that have the greatest impact on system outcomes.

To improve the ability to detect AR pathogen carriers, we will develop an efficient individual-level inference framework for use with large-scale agent-based models and sparse observations. The colonization probability for all patients through time will be inferred using sparse, **individual-level diagnostic records**; this inference will enable identification of high-risk and unobserved potential AR spreaders. Currently, carrier detection in healthcare systems mainly relies on contact tracing and clinical risk factors. Contact tracing attempts to identify colonized individuals who previously had contact with observed infections; however, this tracing approach may miss considerable numbers of contagious asymptomatic carriers, as infections are only observed sporadically. Risk factor-based approaches may also overlook carriers without identified risk factors. The mathematical modeling we will pursue here offers an alternative approach for locating individuals with a high probability of colonization, which then can be used to guide targeted deployment of laboratory testing.

We will also leverage the metapopulation model to forecast the progression of HAI spread across different wards, clinics or hospitals. Accurate prediction of the spread of AR pathogens within healthcare systems would improve preparedness for HAI outbreaks and even enable preemptive interventions. Healthcare facilities forecast to experience a higher risk of colonization/infection could apply additional control, and patients in these wards could be screened and treated in order to disrupt probable transmission pathways. For this effort, we will design and validate a forecasting system using both simulated HAI outbreaks and historical diagnostic records of multiple AR pathogens.

Using both the improved identification of AR carriers and accurate projection of transmission risk, we will develop a targeted screening strategy and evaluate its performance in HAI control. More specifically, we will design a model-inference framework for optimally identifying intervention bundles that achieve maximum HAI reductions under real-world cost and logistical constraints. In particular, we will identify intervention strategies that could be applied and tested in actual hospital settings.

As intervention findings depend on the assumptions and parameters used within a given model, it is important to compare results from different models[122]. Here, if feasible, we will collaborate with the CDC and other groups in the MInD Healthcare Network to compare and communicate competing model findings. The comparison across these diverse models will provide a distribution of possible outcomes that better inform operational intervention efforts.

B. Innovation

The proposed work is innovative in at least six ways. **1)** We will use comprehensive electronic health records and laboratory test results for six AR pathogens from four major hospitals in New York City spanning an 11-year period, as well as detailed genomic data for all CRE isolates collected over a 7-year period within those same hospitals. The patient records also include information on admissions and discharge, room and bed number, antibiotic use, medical procedures, etc., which will facilitate in-depth study of HAI transmission dynamics. **2)** We will use a hierarchy of models at both facility and individual levels to evaluate an array of intervention strategies. In particular, we will evaluate the effectiveness of a common set of interventions against multiple AR pathogens. **3)** We will develop a novel Bayesian inference method that can more precisely infer the colonization probability for all patients using individual-level diagnostic information. This method will improve identification of asymptomatic carriers and inform an optimal screening strategy for selecting, testing and treating patients at high risk of colonization. **4)** We will develop a system that can predict the spread of AR pathogens within and among healthcare facilities in real time. Accurate forecasts of projected transmission risk will inform improved preemptive interventions. **5)** We will optimize intervention bundles – combinations of interventions – in order to achieve maximal reductions of multiple AR pathogens. This optimization will be

subject to real-world economic constraints with additional considerations of applicability to actual hospital settings. **6)** We will compare projections of intervention outcomes across different modeling approaches. The competing outcomes will inform more robust public health decision-making in response to HAI outbreaks.

Together, these efforts will provide a framework for quantitatively evaluating the potential benefits of differing AR pathogen prevention and control strategies and for supporting the development of optimal interventions in response to HAI outbreaks.

C. Work Plan

C.1. Project Data

In this study, we will use electronic hospitalization records (EHRs), personal medical information and cross-linked laboratory test results for six organisms from four hospitals in New York City (NYC). The dataset contains admission and discharge records for 640,703 distinct patients spanning 11 years from Jan 1, 2006 to Dec 31, 2016. In total, 1,051,098 admission events are recorded, providing information on the clinics, rooms and beds where patients stayed. The hospitalization records cover patients in 424 clinics, 2,750 rooms and 5,636 beds. Patient room and bed assignments for each day of hospitalization are derived from the EHR admission–discharge–transfer system. Personal medical information includes medical conditions (diabetes mellitus, wounds, trauma, etc.), procedures (catheterization, ventilation, intubation, operation, etc.), medication (antibiotic use) and other potential risk factors (history of transplant, substance abuse, admission to the ICU, etc.). Cases include all patients who developed a hospital acquired bloodstream infection, urinary tract infection, surgical site infection, or pneumonia with 1 of the following organisms: oxacillin-sensitive *Staphylococcus aureus*, oxacillin-resistant *S. aureus*, ampicillin-sulbactam-sensitive *Acinetobacter baumannii*, ampicillin-sulbactam-resistant *A. baumannii*, penicillin-sensitive *Streptococcus pneumonia*, penicillin-resistant *S. pneumonia*, levofloxacin-sensitive *Pseudomonas aeruginosa*, levofloxacin-resistant *P. aeruginosa*, imipenem-sensitive *Klebsiella pneumoniae*, imipenem-resistant *Klebsiella pneumoniae*, vancomycin-sensitive *Enterococcus faecalis* and *E. faecium*, and vancomycin-resistant *E. faecalis* and *E. faecium*. For each bed occupied during the 3–5-day period prior to infection, microbiology results for assigned roommates and the patient who occupied the bed immediately prior to the case were collected. Culture results and antibiogram data including date and site of culture collection were compiled from clinical microbiology records. These organisms were chosen to represent a broad range of pathogens commonly seen in inpatient settings with various modes of transmission, preferential body sites of infection, and differing viability on healthcare surfaces. Total numbers of resistant cases, sensitive cases and laboratory tests are summarized in **Table 1**. The dataset also contains 10,449 positive cases of *C. diff* sampled from 8,663 unique patients. Use of this dataset has been approved by Columbia University Irving Medical Center under IRB-AAAQ9409 (Nursing Intensity of Patient Care Needs and Rates of Healthcare-associated Infections).

Table 1. Numbers of resistant cases (the first number), sensitive cases (the second number) and total laboratory tests (the third number) listed by organism.

AR organisms	Resistance	Blood stream	Urinary tract	Surgical site	Pneumonia
Acinetobacter baumannii	ampicillin-sulbactam	114 300 10903	165 302 43048	28 45 3717	441 696 15131
Enterococcus faecalis & faecium	vancomycin	1365 1810 10903	3490 6608 43048	346 878 3717	295 230 15131
Klebsiella pneumoniae	imipenem	229 1872 10903	1199 8593 43048	70 256 3717	557 2397 15131
Pseudomonas aeruginosa	levofloxacin	143 508 10903	1133 2774 43048	114 360 3717	1872 3125 15131
Staphylococcus aureus	oxacillin	1502 2681 10903	732 963 43048	729 1106 3717	2894 3605 15131
Streptococcus pneumoniae	penicillin	139 468 10903	1 8 43048	0 5 3717	227 494 15131

Nested within this dataset we have detailed genomic data on all CRE isolates collected over a 7-year period[123-125]. These were carried out as part of ongoing MDR surveillance studies. All isolates underwent comparative Illumina sequencing. Sequences were matched to reference genomes from the same multi-locus sequence type (MLST) and single nucleotide polymorphisms (SNPs) were extracted. For isolates without a close publicly available reference single contig reference genomes were generated using Nanopore long-read sequencing. Multiple isolates per patient were examined to define the individual "cloud of diversity" to define cut-offs for isolate relatedness and transmission. Likewise, all carbapenem resistant *A. baumanii* and *P. aeruginosa* cultured during this time period are available (in the lab of co-I Uhlemann) and could be further sequenced during the project period (using funds from other projects). Such detailed genomic data can support molecular tracing of the emergence, diversification, and transmission of these AR orgamisms, which is the gold standard for defining AR transmission[126].

In addition to this NYC hospital dataset, we also have access to hospitalization records, including MRSA diagnostic records, from multiple Swedish hospitals. This dataset contains admission and discharge records for 743,599 distinct patients from 66 hospitals (271 clinics, 1041 wards) in Stockholm County, Sweden, spanning more than 3500 continuous days during the 2000s. The exact dates and ward types are confidential for the protection of patient privacy. In total, 2,041,531 admission records are available. Patients in the hospitalization dataset constitute over one third of the total 2.2 million population of Stockholm County. In addition, the dataset also contains individual diagnostic records of MRSA, which provide the relative date of diagnosis and strain of MRSA. A total of 991 positive cases from 172 different strains were confirmed; the most prevalent strain was UK EMRSA-15 (289 cases). This dataset was approved for use by the Regional Ethical Review Board in Stockholm (Record Number 2004/5:8). Although not collected within the United States, the dataset can be used as a supplement to the NYC hospital data for purposes of construction and testing patient transfer networks among multiple facilities in developed countries.

C.2. Overarching Aims and Responsiveness to RFA

For this project our overarching aims are to:

- Aim 1: Model the spread of AR pathogens in healthcare systems at individual and facility levels. This aim addresses the thematic areas of ***Antimicrobial Resistance*** and ***Connectedness of Patients Within and/or Among Healthcare Facilities***.
- Aim 2: Infer key epidemiological parameters and asymptomatic carriage using sparse observations. This aim addresses the thematic areas of ***Surveillance*** and ***Simulations of Epidemiologic Studies***.
- Aim 3: Predict outbreaks of HAI pathogens in healthcare facilities. This aim addresses ***Connectedness of Patients Within and/or Among Healthcare Facilities*** and ***Simulations of Epidemiologic Studies***.
- Aim 4: Evaluate and optimize intervention combinations given cost and logistical constraints. This aim addresses ***Simulations of Epidemiologic Studies*** and ***Economic Modeling***.
- Aim 5: Work with the CDC and the network to test interventions in different modeling approaches. This aim addresses ***Simulations of Epidemiologic Studies*** and ***Outbreak Response***. This aim also provides plans for ***collaboration/partnerships*** with other MInD-Healthcare grantees.

C.3. Aim 1: Model the spread of AR pathogens in healthcare systems at individual and facility levels

Two model structures will be utilized for this project: 1) a metapopulation model capable of simulating HAI transmission dynamics across multiple healthcare facilities; and 2) an agent-based model representing individual-level patient movements within multiple hospitals. Both model structures will be designed to simulate person-to-person contact, environmental contamination and importation.

C.3.1 Metapopulation model

The metapopulation model describes the HAI transmission dynamics among different subpopulations within a healthcare system or systems. Each subpopulation represents patients in a specific clinic or hospital, for which a well-mixed population is assumed. The evolution of colonized and susceptible populations is described by a set of ordinary differential equations:

$$\frac{dC_i}{dt} = \frac{\beta S_i C_i}{n_i} + \varepsilon_i S_i - \alpha C_i - \delta_i C_i + \sum_j M_{ij}\frac{C_j}{n_j} + \gamma a_i, \qquad (1)$$

$$\frac{dS_i}{dt} = -\frac{\beta S_i C_i}{n_i} - \varepsilon_i S_i + \alpha C_i - \delta_i S_i + \sum_j M_{ij}\frac{S_j}{n_j} + (1-\gamma)a_i, \qquad (2)$$

$$\frac{d\varepsilon_i}{dt} = -\lambda\varepsilon_i + \frac{\theta}{n_i}C_i. \qquad (3)$$

Here C_i, S_i and n_i are the numbers of colonized, susceptible and total patients in subpopulation i; β is the transmission rate; ε_i is the environmental force of infection in facility i; α is the decolonization rate of carriers; δ_i is the discharge rate for patients in facility i; M_{ij} is the number of patients transferred from facility j to facility i; γ is the importation rate of colonization from the community; a_i is the number of new admissions from the community in facility i; λ is the environmental decolonization rate; and θ is the environmental contamination rate. Note that, on the right-hand-side of equation (1), the six terms represent contact transmission, environmental contamination, decolonization of carriers, patient discharge, patient transfer from other facilities, and importation from the community. Facility and patient movement information, such as the number of patients in each facility n_i, discharge rate δ_i, patient transfer matrix $\{M_{ij}\}$ and number of new admissions from the community a_i, is available from the hospitalization dataset. Equation (3) describes the evolution of the environmental force of infection.

Similar model constructs have been used to simulate the spatial transmission of other infectious diseases[127-136]. In our own work, we have used a metapopulation framework to develop a forecasting system capable of generating accurate real-time predictions of influenza outbreak onset up to six weeks in advance of these events[135]. This same metapopulation model has also been used to support the development of a novel method for optimizing surveillance networks that improves forecast accuracy while minimizing surveillance costs[136].

C.3.2 Agent-based model

We will also use an agent-based model informed by real-world patient movements within multiple healthcare facilities to simulate HAI outbreaks. We have used a similar model construct to infer the transmission dynamics of MRSA in 66 Swedish hospitals[44]. In this project, we will extend this model to additionally include environmental contamination. The ODD description of the agent-based model is listed as follows[120,121].

Purpose: The purpose of the model is to simulate the spread of AR pathogens in healthcare systems.

State variables and scales: The model spans four hierarchical levels in hospital: individual, room, clinic, and total population. Model dynamics are defined at the individual level, but outcomes can be aggregated to any of the four scales. Individuals are characterized by state variables: colonization status (susceptible or colonized), hospitalization time, and the location (clinic and room) where the patient resides. Hospitalization time and location data provide information on patient transfer within healthcare systems. For each room, an additional state variable is defined: the force of infection attributed to environmental contamination ε. Six parameters are introduced: 1) the transmission probability upon contact, β; 2) the importation probability of colonization, γ; 3) the environmental contamination rate, θ; 4) the environmental decolonization rate, λ; 5) the patient decolonization probability, α; and 6) the probability of laboratory testing, ρ.

Process overview and scheduling: The model proceeds in daily time steps. Within each time step, five modules are processed: transmission between patients staying in the same room, environmental contamination within each room, transmission from the environment to patients, importation of colonized patients from the community, and laboratory testing as surveillance. Transmission in the community is not explicitly simulated but is reflected by the importation rate of colonization.

Design concepts: *Emergence*: HAI outbreaks emerge from the movement of individuals within healthcare systems. *Interaction*: HCW-mediated contacts between patients facilitate transmission of AR pathogens. We model HCW-mediated transmission indirectly by assuming that AR pathogens can spread between all pairs of patients staying in the same room at the same time. Colonization in patients may spillover to contaminate the environment resulting in indirect transmission to patients admitted to the same room at a later time. Nosocomial transmission interacts with the community through the admission and discharge of colonized patients. *Stochasticity*: Transmission, importation, decolonization and laboratory testing are all run stochastically according to predefined probabilities. Distributions of outcomes can be generated through

repeated model simulations. *Observation*: A randomly selected proportion of patients undergo laboratory testing for AR carriage. The number/fraction of carriers in each clinic/total population indicates the prevalence of AR pathogens.

Initialization: A certain percentage (e.g., 5%) of patients within hospitals at the beginning of a simulation are randomly assigned as carriers. The environmental force of infection in each room is set to zero initially.

Input: We will use the 11-year daily admission-discharge-transfer record from the four NYC hospitals to inform patient movement within the model.

Submodels: The contact network within a collection of hospitals is represented by a time-varying graph *G* constructed using the actual hospitalization records (see an example in **Fig. 1a**). In this contact network, nodes represent uniquely labeled patients, connected by undirected links among individuals sharing a room at a given time. Individuals are classified into two categories: Susceptible (*S*) and Colonized (*C*). Within hospital, transitions between these states are governed by model transmission dynamics. ***Contact transmission*****:** A susceptible individual *i* can be colonized, with probability $\beta/(n_{r_i}-1)$, upon contact with a colonized person *j* who is directly linked to *i* in the contact network *G*. Here n_{r_i} is the capacity of the room in which patient i resides. We use a frequency dependent transmission model here as the chance of person-to-person contact decreases in larger rooms[137]. ***Environmental contamination*****:** Each colonized patient in a given room contributes a daily $\theta/(n_{r_i}-1)$ increment to the environmental force of infection ε_{r_i}. Meanwhile, ε_{r_i} decays to a fraction λ of its prior value per day. A susceptible individual in room r_i becomes colonized with probability ε_{r_i} due to environmental contamination. ***Importation*****:** For new admissions, patients are colonized with a probability γ. ***Surveillance*****:** Each patient in hospital or certain clinics is selected with probability ρ for laboratory testing to confirm whether he/she is colonized with AR pathogens.

A schematic illustration of these HAI transmission dynamics is shown in **Fig. 1b**. The network framework can be used to assess the evolution of an outbreak and the effects of interventions. More complex model constructs with additional features, such as personal medical information, can be developed from this basic framework.

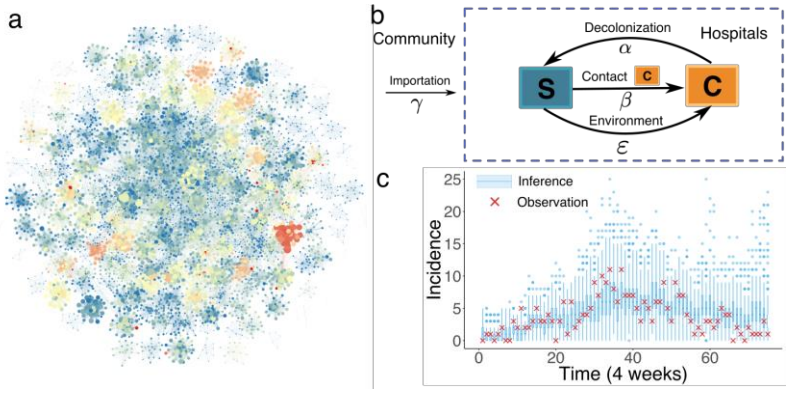


Fig.1 (a) A snapshot of the contact network for Stockholm County, Sweden hospitals in a one-week time window. Color indicates the colonization risk obtained from a large number of free simulations. (b) The transmission dynamics of HAI in hospital and introduction from the community. (c) Free simulations using inferred parameters reproduce observed incidence over a 6-year outbreak period. Red crosses are observed infections summed every 4 weeks and blue boxes are the distributions of infections obtained from simulations.

C.3.3 Model validation

Model validation is central for establishing that a given model can reliably represent previously observed HAI transmission dynamics and be used to investigate intervention effects. For this project, both model forms will be extensively validated for the multiple AR pathogens observed in the NYC hospitals. Model validation will utilize both free simulation and data assimilation approaches vetted against historical records of HAI incidence at a variety of scales (e.g. room, clinic, hospital and total population). Specifically, key epidemiological parameters including transmission rate β, importation rate γ and environmental contamination rate θ will be inferred using techniques developed in Aim 2. With these inferred parameters, we will run stochastic free simulations of AR pathogen transmission and compare these outcomes with observed incidence. **Figure 1c** shows an example of model validation for an agent-based model simulating a MRSA outbreak across 66 Stockholm County hospitals. The model-generated outbreaks agree well with observed MRSA incidence.

C.3.4 Potential Pitfalls and Alternative Approaches

While it is highly likely that the proposed Aim 1 work will proceed smoothly, some difficulties could arise. First, if we find that certain AR pathogens concentrate in a subset of locations, we will restrict modeling to these relevant facilities. Second, should the use of constant parameters not generate sufficiently large uncertainty,

we will use disease progression parameters (e.g., the patient decolonization probability, α) randomly drawn from broad distributions to account for the heterogeneity among different individuals. Different sets of parameters could also be estimated for different ward types (e.g., ICUs, EDs, etc.).

C.4. Aim 2: Infer key epidemiological parameters and asymptomatic carriage using sparse observations

To validate each model structure, we will fit the models to observations of AR pathogens in NYC hospitals. Specifically, we will infer key epidemiological parameters including transmission rate, importation rate and environmental contamination rate using a model-data assimilation (M/D/A) framework, a technique we have applied successfully to a range of infectious diseases[138-150]. This estimation of parameters will help understanding of the burden of disease in healthcare systems. In addition, we propose to develop an individual-level inference method that can more accurately identify asymptomatic carriers using sparse observations. Identification of asymptomatic colonization is critical for HAI control; however, this individual-level inference problem remains challenging for a number of reasons: the high dimensionality of agent-based models, the computational cost of running inference, and the sparseness of observations. We will solve these difficulties by developing and implementing a novel and efficient algorithm.

C.4.1. Model-data assimilation (M/D/A) framework

An M/D/A system has three basic components: 1) a mathematical model describing the propagation of an infectious agent through a given population; 2) observations of disease incidence; and 3) a data assimilation (DA) methodology. In practice, the observations and data assimilation are used to iteratively optimize an ensemble of model simulations.

Data assimilation, or sequential filtering, can be viewed as the problem of estimating the probability of a system state (including state variables and parameters) at a given time (Z_t) conditional on observations (y_t) measured up until and including time t. Bayes' rule provides a target for the update of the system state given an observation: $P(Z_t|y_t, y_{t-1}, \dots) \propto P(y_t|Z_t)P(Z_t|y_{t-1}, \dots)$. Here, the first term on the right-hand side is the likelihood of observing the data given the state, and the second term is the prior distribution of the system state. The updated distribution, $P(Z_t|y_t, y_{t-1}, \dots)$, is called the posterior, and in practice it is calculated by halting the integration of an ensemble of model simulations at each new observation and using that observation to calculate a new posterior for Z_t. The model is then integrated forward to the next observation and the process is repeated. Through this iterative updating of state variables and parameters, the model is optimized. A variety of DA algorithms exist for calculating the posterior[151-156], and the exact method of transforming the ensemble members of the prior into ensemble members of the posterior is what distinguishes these algorithms. Commonly used DA algorithms include ensemble Kalman filters[151,152], particle filters[153], and iterated filtering[154,155].

For the metapopulation model, we will use an efficient DA algorithm – the Ensemble Adjustment Kalman Filter (EAKF)[151]. Unlike particle filters[153], which require a larger number of particles and are more suitable for low-dimension systems[156], the EAKF uses a limited number of ensemble members and is thus conducive for use with high-dimension systems. In a recent study, we showed that the EAKF can accurately infer key epidemiological parameters in a metapopulation model of influenza transmission with over 3,000 subpopulations (see **Fig. 2a**)[135]. In this project, we will use the EAKF to infer parameters in the metapopulation model of HAI spread.

For agent-based models, several studies have generated parameter estimates using observations[57-59]; however, most existing methods rely on the Markov Chain Monte Carlo (MCMC) algorithm, which requires a large number of model simulations. These approaches are feasible for small or medium-scale models representing a single or a few facilities, but less viable for the large-scale models we encounter here. Recently, leveraging an equation-free approach[158], we developed a statistical filtering technique capable of inferring epidemiological parameters within large-scale agent-based models using aggregated incidence numbers[44]. The algorithm applies the EAKF repeatedly within an iterated filtering (IF) framework[154,155], such that the system parameters are gradually adjusted toward their true values. In a recent study, we showed that this inference algorithm can estimate parameters for an agent-based model simulating over 750,000 patients (see **Fig. 2b**)[44]. In this project, we will use this IF-EAKF framework to infer parameters in the agent-based model.

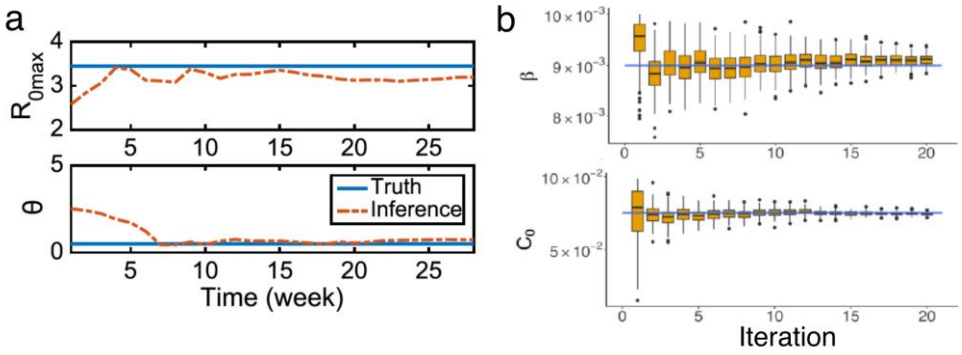


Fig.2 (a) Inference of parameters (the maximum reproductive number, R_{0max}, and random movement rate, θ) using the EAKF and a metapopulation model depicting influenza transmission. The parameter estimates (red dotted lines) agree well with the true parameters (blue lines) used in generating synthetic outbreaks. (b) Inference of parameters (the transmission rate, β, and importation rate, C_0) in an agent-based model for MRSA transmission in Stockholm County hospitals. The distributions of inferred parameters (yellow boxes) gradually converge to the true values (blue lines) after a few iterations.

C.4.2. Inference of asymptomatic colonization using sparse observations

Identification of asymptomatic carriers of AR pathogens is important for HAI containment. Using agent-based models and parameters estimated from incidence numbers aggregated across all patients (i.e., the macro scale), we can reproduce observed population-level epidemic curves in free simulations (**Fig. 1c**). However, to obtain a more precise colonization probability for each patient, an inference system that utilizes **individual-level** diagnostic information is needed. This task is particularly challenging, as observations of colonization are highly sparse relative to the total number of patients in hospital.

In infectious disease modeling, several techniques have been developed to infer transmission risk factors and reconstruct transmission trees using incomplete individual-level infection data for diseases such as H1N1[159] and MERS-CoV[160]. However, these approaches are not readily applicable to HAI transmission dynamics, partly due to the co-occurring dynamics of transmission, importation and environmental contamination. In network science, a Bayesian inference framework for network epidemics has been proposed using the belief propagation (BP) technique[161]. BP is a sophisticated graphic-based algorithm developed based on message passing theories in statistical physics[162]. In the context of infectious disease spread, BP has been used to identify the origin of epidemic outbreaks with high accuracy[163]. Despite its success in these applications, the BP algorithm only works for irreversible transmission dynamics such as a susceptible-infected-recovered model; unfortunately, these dynamics do not hold for HAIs. Moreover, with sparse observations, the algorithm typically encounters convergence problems during model iteration. No alternate methods presently exist that are capable of solving the individual-level inference problem; we therefore here propose to develop a new algorithm that can accurately identify individual patient colonization risk in hospital.

To begin developing this system, we propose the following structure. The evolution of the probability of individual *i* currently being in each state in the network model can be described by the following equations:

$$S_i^{t+1} = S_i^t + \alpha C_i^t - \frac{\beta}{n_{r_i} - 1} S_i^t \sum_{j \in \partial i} C_j^t - \varepsilon_{r_i}^t S_i^t, \qquad (4)$$

$$C_i^{t+1} = C_i^t + \frac{\beta}{n_{r_i} - 1} S_i^t \sum_{j \in \partial i} C_j^t + \varepsilon_{r_i}^t S_i^t - \alpha C_i^t, \qquad (5)$$

$$\varepsilon_{r_i}^{t+1} = \varepsilon_{r_i}^t - \lambda \varepsilon_{r_i}^t + \frac{\theta}{n_{r_i} - 1} \sum_{j \text{ in } r_i} C_j^t. \qquad (6)$$

Here S_i^t and C_i^t denote the **probability** of individual i belonging to the susceptible and colonized populations at time t; all other parameters and state variables are as defined in **C.3.** For each individual i, our objective is to estimate the probability C_i^t.

Due to the large number of patients in the agent-based model, this inference problem is high dimensional, yet observations of colonized patients, including infected patients, are quite sparse. To supplement these data and thus better constrain the model, we will use a backward inference process to augment the limited observed

information by inferring the colonization probability of infected individuals prior to their diagnoses. This additional information allows an updating of model states prior to the observation of an infection. Using this augmented information, we propose to solve the high-dimensional inference problem by developing a Sequential Individual-Level Inference (SILI) algorithm. We begin at the start of the observed record and sequentially update the model states using the augmented data. During this process, information from future observations will propagate in the contact network and reach individuals with no observations. In this way, information is expanded and used to constrain the states (S_i^t and C_i^t) of unobserved individuals. **Figure 3** shows an illustration of the proposed inference framework.

Specifically, an ensemble of system states, which represent the distribution of probabilities S_i^t and C_i^t for all patients, will be iteratively adjusted using individual-level diagnostic information so that the model can better fit the observations. This adjustment, developed based on the Bayes' rule, is applied weekly starting from the model initialization. At each week t, three procedures will be used to constrain the model state. 1) *Backward inference*: for each patient i tested for AR carriage at some later week, we will infer his/her colonization probability at current week t. Using model simulations, we can calculate the likelihood of observing the reported cases at some later week given that patient i is colonized at week t. Together with the prior probability for patient i being colonized at week t, the posterior colonization probability for patient i at week t can be computed. 2) *Covariability adjustment*: using the inferred colonization probability, we will adjust the states of observed carriers' neighbors based on the covariability between their colonization probabilities, which arises from their dynamical coupling before week t. 3) *Model integration*: we integrate the model to week $t+1$ using the updated states. This sequence of procedures uses inference and dynamic simulation to augment the sparse observations back in time and propagate information to individuals without observations, thus covering the entire population after a sufficient number of updates.

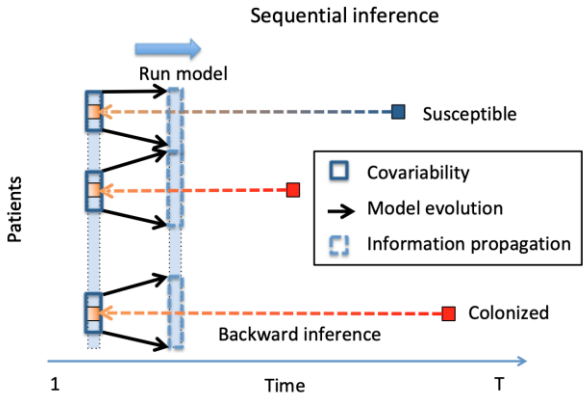


Fig.3 An illustration of the sequential inference framework. Red and blue squares are laboratory-tested patients; horizontal dotted lines indicate backward inference of colonization probability prior to observation; vertical blue bands represent the state of patients. When augmented data are available at a given time (orange squares intersect with blue bands); covariability adjustment is performed to update the states of neighbors of these patients (solid blue boxes). As the model runs forward in time, information from augmented (i.e., later observed) patients propagates to other individuals with no observations (black arrows). The blue dotted boxes highlight individuals whose states are affected by information from observations made in the future. After a sufficient number of updates, information from sparse observations can propagate to the entire population.

C.4.3. System verification

We will validate the inference system using model-generated outbreaks. Specifically, we will generate a synthetic outbreak using the agent-based model and record patient states (i.e. colonized or susceptible), as well as testing and observed infection cases, at all times. We will then run the inference algorithm *using only the observed infection cases* and determine whether the inferred colonization probabilities identified by the algorithm match the synthetic record of colonized patients. Recovery of the synthetic colonization record will indicate that the inference algorithm works well and can be used with real records. A receiver operating characteristic (ROC) curve will be used to quantify and evaluate the performance of the inference algorithm.

C.4.4. Preliminary results

We performed a comparison between a preliminary version of the SILI algorithm and four alternative strategies (i.e., free simulation, length of stay, number of contacts and contact tracing) in a synthetic outbreak. We generated an outbreak in a time-varying contact network over 24 weeks, which simulated 29 'observed' positive patients. We next applied the inference algorithm and observations to infer patient colonization probabilities. We then ranked patients by their mean colonization probability from a large number of free simulations, length of stay in hospital, total number of contacts with other patients, contact with observed carriers, and inferred colonization probability. In **Fig. 4a**, we display the ROC curves of the five methods in locating colonized patients. The inference algorithm outperforms the other four heuristic methods with a higher AUC (Area Under Curve) value. Note contact tracing only reaches about 200 patients, as other patients have no direct contact with observed carriers. The number of carriers identified by screening top-ranked patients

according to different methods is shown in **Fig. 4b**. The inference algorithm again outperforms competing methods. These preliminary results indicate that the SILI algorithm is a promising approach. Additional experiments in more realistic scenarios and comparisons with other strategies still need to be performed.

C.4.5. Potential Pitfalls and Alternative Approaches

The model structures and inference algorithms outlined above have been tested and validated in previous studies, so we do not anticipate serious pitfalls. However, the metapopulation model and agent-based model may be mis-specified for real-world transmission. For instance, heterogeneity of transmission rates among different types of wards is simplified in the model due to a lack of relevant information. If certain wards have significantly larger observed outbreaks, a separate set of parameters can be defined and estimated for these wards. Another possible pitfall is that one round of data assimilation at each time step may not sufficiently constrain the model. Should this occur, we will perform multiple updating rounds using an iterated filtering framework. Another possible approach would be to use Approximate Bayesian Computation (ABC)[164] to estimate model parameters, though this method is more time-consuming than sequential inference approaches.

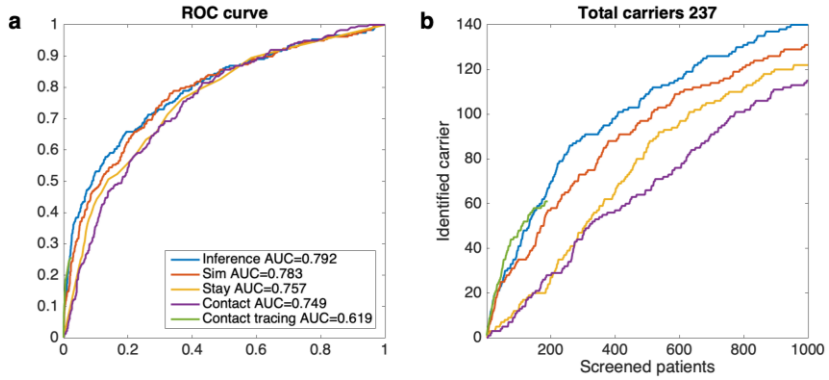


Fig.4 We ran the SILI algorithm for a synthetic outbreak generated in a time-varying network over 24 weeks (transmission rate $\beta = 2\%$; importation rate $\gamma = 1\%$; environmental contamination rate $\theta = 0.02\%$). This figure shows the inference results for patients in hospital at week 24. (a) The ROC curves for identifying asymptomatic colonization for five different methods (the SILI inference algorithm, free simulation, length of stay, number of contact and contact tracing) in identifying asymptomatic colonization. AUC values are reported in the legend. (b) The number of positive patients identified by screening top-ranked patients by methods. The total number of carriers in hospital is 237.

C.5. Aim 3: Predict outbreaks of HAI pathogens in healthcare facilities

Projection of outbreak risk in healthcare facilities is of paramount importance for effecting HAI control. Prediction of the spread of emerging AR pathogens within healthcare systems and their invasion into the community can inform preemptive interventions that impede the establishment and further transmission of these organisms. For this aim, we will develop a forecasting system using the metapopulation model that is able to predict the emergence and progression of local outbreaks of AR agents on a facility-by-facility basis.

C.5.1. Ensemble prediction of HAI transmission in healthcare systems

In recent years, a number of different infectious disease forecasting methods have been developed[135,138,165-172], and applied to forecast diseases such as influenza[173-175], dengue[176] and Ebola[177,178] at population scales. However, the localized spatiotemporal transmission dynamics of HAIs are more difficult to predict. First, the patient transfer network is time-varying, exhibiting high levels of complexity in both space and time. It is therefore challenging to make predictions for an ever-changing system. Second, the system is poorly observed. Only a small fraction of patients are typically tested for AR pathogen carriage, with strong bias to patients with clinical manifestations, while the states of other patients remain unknown. Third, the transmission dynamics are highly stochastic; an observed outbreak is just one realization of these stochastic dynamics. Thus, the future course of an outbreak has substantial uncertainty.

To overcome these difficulties, we propose the following approaches. Despite the time-varying nature of the patient transfer network, we have found that patient movement within healthcare systems is fairly regular with a roughly weekly periodicity[44]. Leveraging this regularity, we will develop a stochastic patient flow model to simulate future patient movement across facilities. To supplement limited observations of AR pathogen carriage, we will use an observation model that rescales observed cases to the total number of carriers in each facility. For instance, an observational study found 3 colonized patients for each infection of CRE in liver patients[126]. The observation model will consider the heterogeneity of laboratory testing coverage in different types of wards and map the total number of colonized patients to the number of observed positive patients. Finally, key model components (e.g., model dynamics, patient transfer and the observation process) will be run stochastically, using an ensemble of possible model states, to warrant well-calibrated uncertainty of forecast outcomes.

We will generate probabilistic predictions for a number of targets including onset week (i.e., the week that local observed incidence increases above a baseline threshold), near-term observed incidence and the estimated number of asymptomatic carriers, using only information that would be available in real time. Forecasts will be aggregated to the clinic, hospital and population levels. These forecasts will be validated using both model-generated outbreaks and the historical records of multiple AR pathogens in NYC hospitals. Standard metrics for quantifying forecast accuracy such as mean absolute error (MAE) for point predictions and log-score[173-175] (the logarithmic value of the weight assigned to the observed target) for probabilistic predictions will be used to evaluate forecast performance.

C.5.2. Potential Pitfalls and Alternative Approaches

We have extensive experience working with M/D/A frameworks to develop predictions of infectious disease spread[135,136,138-150]. As a result, we do not expect serious problems pursuing this aim. One possible shortcoming of the proposed approach is the homogenization of the heterogeneity among different types of wards (i.e., ICUs versus neurology). Should this simplification affect forecast accuracy, we can either distinguish ward types by defining distinct sets of parameters for each ward (e.g., different β) or limit forecasting to certain subsets of clinics.

C.6. Aim 4: Evaluate and optimize intervention combinations given cost and logistical constraints

Using the models described, developed and tested in Aim 1, we will implement interventions *in silica* both singly and in combination. Further, supported by the improved carrier identification method designed and validated in Aim 2, we will develop a targeted screening strategy to proactively search for asymptomatic colonized patients. Additionally, we will use optimization approaches to find best intervention solutions given a particular HAI outbreak scenario in the context of intervention costs and availability constraints.

C.6.1. Simulation and evaluation of intervention bundles against AR pathogens

A suite of intervention experiments will be undertaken in coordination with the CDC. Specifically, we will test six interventions that can be properly represented in both models: 1) hand hygiene and barrier precautions; 2) isolation of infections; 3) environmental cleaning; 4) active patient screening within hospital; 5) contact tracing; and 6) screening at admission. Each intervention is designed to slow the progression of HAIs through modulation of different components of the transmission cycle (**Table 2**). Other interventions such as HCW staffing and use of rapid testing could also be tested. Patient transfer restriction is typically not realistic in hospital settings so we will not test it here. The timing, compliance and efficacy of these interventions will be varied and multiple interventions will be used in combination. The impact of these interventions will be assessed by quantifying changes in the timing, magnitude and duration of incidence relative to simulations without interventions.

In practice, multiple, layered interventions can be bundled and implemented simultaneously. As the number of combinations of interventions is essentially infinite—given that the timing of implementation, uptake, compliance and effectiveness of each can be varied—we will work with the CDC and clinicians to identify a limited suite of intervention combinations to be assessed. A robust approach will likely use a combination of mitigation methods. These effects will be explored for different outbreak scenarios. Similar approaches have been used to evaluate control measures for pandemic influenza[179-183].

With the availability of surveillance data for six AR pathogens, our project is unique in its capacity to test the effectiveness of a common intervention strategy on multiple organisms. This advantage allows the identification of intervention bundles that work for both existing and unforeseen emerging AR agents in healthcare systems.

C.6.2. Design targeted screening strategy to better control HAIs

The individual-level inference in Aim 2 could inform active screening strategies by targeting the testing of more likely asymptomatic AR carriers. Specifically, patients currently staying in hospital with estimated higher probabilities of colonization could be screened with priority. Targeted intervention would screen in-hospital patients whose colonization probabilities exceeded a certain threshold or who rank in a top percentage tier. Patients who are confirmed colonized would then be treated to prevent further shedding of AR pathogens in hospital. Within our agent-based model, we will implement and test a variety of screening strategies and optimize the screening criterion (i.e., the threshold value) in terms of cost and benefit.

Table 2. Possible interventions, their effects, and their representation in project model structures

Intervention	Effects	Model Representation
Hand hygiene and barrier precautions	Reduces contact transmission rate	Lower β
Isolation of infections	Reduces contact between observed cases and other patients	Lower contacts between susceptible and observed colonized patients
Environmental cleaning	Reduces environmental contamination	Lower θ
Active patient screening within hospital	Identifies asymptomatic carriers by selectively screening a fraction of patients according to a certain criterion (e.g., the SILI inference, length of stay, etc.)	Identification and treatment of colonized patients – i.e., change of the model state variables for discovered positive patients from colonized to susceptible
Contact tracing	Identifies asymptomatic carriers who have had contact with observed cases	Identification and treatment of colonized patients – i.e., change of the model state variables for discovered positive patients from colonized to susceptible
Screen at admission	Reduces importation from the community	Lower γ

The screening method will be evaluated using synthetic outbreaks and historical records. We will compare the M/D/A inferred targeted screening strategy with traditional approaches based on length of stay, number of contacts and contact tracing. Given the same number of screened patients, the impact of isolating AR carriers identified by these approaches will be compared. If the number of reduced infections achieved by the inference-based screening is significantly higher than those achieved by alternative methods, this finding would support the use of the model-inference framework in practice.

C.6.3. Optimization of intervention combinations under constraints

As discussed in **C.6.1**, interventions will be evaluated by quantifying changes in the timing, duration and magnitude of incidence relative to reference simulations without intervention. Cost effectiveness will also be assessed by constraining the implementation of interventions to reflect resource restrictions. In addition, we will work with the CDC to identify objective functions based on intervention cost and availability. Once specified, we will use optimization approaches (e.g. MCMC methods[184] or simulated annealing[185]) to find an optimal intervention solution given a particular objective function.

For example, patient isolation is limited by the availability of rooms; laboratory tests are limited by medical cost; barrier precautions and environmental cleaning are limited by the costs of gloves, gowns and labor. An optimization framework can be formulated for each of these constraints (i.e., objective functions). As interventions can be targeted to specific locations and time windows, this optimization takes the form of a continuous optimal control problem[186-188]. Given the complexity of the transmission model construct, analytic solutions are not possible; consequently, we will instead use iterative optimization algorithms to find the optimal intervention bundle. Previous research has shown that optimal control is possible for networked systems[189-192]. Thus, a suite of strategies and corresponding best outcomes can be generated to support decision-making by hospital and public health officials. These findings will enable estimation of a best intervention approach (e.g. most cases averted, or greatest delay in outbreak onset) given cost and logistical constraints.

C.6.4. Potential Pitfalls and Alternative Approaches

We have used the proposed model structures and optimization techniques in previous studies. As a result, their generalization to HAI outbreaks should be relatively straightforward. However, for the targeted screening, heterogeneity in colonization risk among individuals due to differing host characteristics is not considered. If the targeted screening does not yield significant improvements, we will incorporate personal medical information (procedures, antibiotic use, etc.) to inform individual prior probability of colonization. Also, if the iterative optimization is computationally infeasible for the agent-based model, we will run a grid search using a finite number of plausible intervention combinations in order to identify an optimal intervention strategy.

C.7. Aim 5: Work with the CDC and the network to test interventions in different modeling approaches

For this aim, we will work in conjunction with the CDC and other network centers willing to participate in a multi-model comparison effort. We propose to design both a suite of reference free simulation scenarios, representing outbreaks of an emerging AR pathogen in healthcare systems, and a set of interventions to be implemented and evaluated within each reference scenario.

C.7.1. Defining reference scenarios and interventions

Here, the collective goal is to define a set of standardized scenarios in a synthetic healthcare system, representative of a real-world facility, with designated epidemiological characteristics for simulation by all models engaged in this collaborative effort. As models will likely differ radically across the network of participating groups (e.g. compartmental versus agent-based), standardization will define target free simulation behaviors but not fixed epidemiological characteristics (e.g. a precise β will not be defined). While such delineations are not as prescriptive, this approach will enable participation in each scenario by more model forms and a more robust and meaningful comparison of simulation and intervention outcomes across the network. This standardization also mirrors other inter-comparison efforts, such as the comparison of interventions against influenza[122], which similarly only defines general model scenarios. The reference scenarios, which will be determined in consultation with the CDC and other participant groups, will likely represent a range of transmission dynamics and outbreak outcomes.

We will use these standards to generate the reference simulations with our two model forms. The effectiveness of particular interventions will be evaluated relative to these reference simulations. In particular, our focus will be assessment of changes in the timing and magnitude of HAI incidence, both overall and within specific facilities, and the sensitivity of these changes to the timing, uptake, and efficacy of a particular intervention.

C.7.2. Comparison of intervention outcomes in different models

A central goal of Aim 5 is to enable a comparison of findings across all models. This aim motivates the standardization of both reference scenarios and interventions. By employing a common work plan across the participating groups, model outputs can be formally compared. *Most of the interventions that will be simulated for this project will not have been implemented in practice; as a consequence, there are no data available to validate model behavior.* There is thus a need to develop methods, collaboratively across the network, that objectively combine the projections of different models and provide a probabilistic, distributional estimate of the effectiveness of a particular intervention. We intend to work alongside the CDC and other participating centers of the network to develop harmonized model output forms that specify intervention effects as probabilistic distributions (not point estimates) and quantify the sensitivity of each effect as a function of the initiation timing, duration, uptake and effectiveness of each intervention. With this information, it should be possible to reconcile and combine the findings of all models used across the network.

C.8. Project Timeline, SMART Objectives, Rigor and Reproducibility

Our project team has the relevant disciplinary expertise in infectious disease epidemiology, mathematics, dynamic modeling, algorithmic statistics, observational analysis, model design and implementation, Bayesian methods, surveillance and clinical treatment of HAIs, and genomic analysis of AR pathogens needed to carry out the proposed work. All model systems will be tested and validated using DA methods, free simulation and comparison to historical records. We have used these approaches extensively[44,135,136,138-150], which has insured the rigor and reproducibility of the model frameworks we have designed. Validated methodological advances and findings will be disseminated through presentations at scientific meetings and conferences, publications in peer-reviewed journals, and the archiving and sharing of data and software.

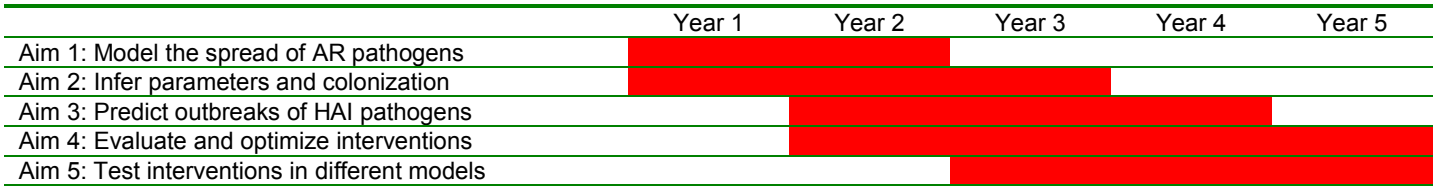
Five years of funding are requested for this project in order to address the specific aims thoroughly and to tackle any unforeseen issues. The project goals are accomplishable in the proposed time frame, as a large proportion of the proposed work constitutes an expansion of modeling efforts and inference methodologies we have pursued and published in peer-reviewed journals. Specific, measurable, achievable, realistic and time-bound (SMART) objectives for each of the aims will be utilized to evaluate project success and ensure that findings are rigorous and reproducible. These **SMART** objectives are:

- **Aim 1:** We will develop the metapopulation and agent-based models during Years 1 and 2. By the end of Year 2 all model forms will be tested and validated. Forms found to be capable of validated, accurate simulation of historical outbreaks will be used.
- **Aim 2:** During Years 1 and 2, we will infer key epidemiological parameters within both models. By the end of Year 2, we will be able to generate free simulations using these inferred parameters and reproduce key features of historical outbreaks. During Years 1-3, the individual-level inference framework will be developed and validated. By Year 3, this inference algorithm will be ready for use designing targeted screening strategies.

- **Aim 3:** By Year 3, we will finish the development of the forecasting system for HAI outbreaks. During Years 3 and 4, we will generate predictions for both synthetic and historical outbreaks. By the end of Year 4, the forecasting system will be validated and ready for use.
- **Aim 4:** During Years 2-5, we will generate and test proposed intervention combinations. By the end of Year 4, the targeted screening strategy will be developed and validated. During Years 4 and 5, we will develop and implement an optimization framework to identify maximally effective interventions given cost and logistical constraints, which will be identified in conjunction with the CDC. By the end of Year 5, we will have finalized and implemented this framework.
- **Aim 5:** During Years 3 and 4, in collaboration with the CDC and other participating groups, we will identify the reference scenarios, intervention scenarios, and cost and logistical constraints. Cross-site comparisons will be developed and delivered during Year 5.

A timeline of work for all specific aims is summarized below.

Table 3. Project Timeline



C.9. Team and Accomplishments

The investigators have the combined expertise in infectious disease epidemiology, mathematics, dynamic modeling, algorithmic statistics, observational analysis, model design and implementation, Bayesian methods, and surveillance and clinical treatment of HAIs needed to carry out the proposed work. The MPI, Dr. Jeffrey Shaman, has designed, coded and experimented with a spectrum of disease, climate, hydrology, and ecology models. These models range from simple oscillators[193,194] to models of intermediate complexity[195-201], up to large-scale global climate models[202-204]. Dr. Shaman has also developed and worked with the modeling and inference methods proposed for use in this project, as well as a number of network disease modeling approaches[205-210]. He has led NIH-, NSF-, DoD- and CDC-sponsored projects and is experienced mentoring both doctoral students and post-doctoral scientists.

The MPI, Dr. Sen Pei, is an applied mathematician with interdisciplinary experience in infectious disease modeling, statistical inference, nonlinear dynamical system analysis, and network science. He has developed a model-inference system informed by human movement and capable of simulating and forecasting the spatial spread of influenza at different scales[135], a principled mathematical framework for optimizing surveillance networks in order to save resources and maximize forecast utility[136], an improved influenza forecasting method that corrects nonlinear error growth in dynamic transmission models[145], and a novel Bayesian inference framework for modeling the transmission of antibiotic-resistant pathogens in large-scale contact networks[44]. Dr. Pei also has research experience in information diffusion modeling[210,211], big data analysis[86,212-214] and optimal control of networked biological systems[215-219]. Dr. Pei and Dr. Shaman have closely collaborated during the last 4 years.

Co-I, Dr. Anne-Catrin Uhlemann, is a clinician and Associate Professor of Medicine in the Division of Infectious Diseases in the College of Physicians and Surgeons at Columbia University. Her laboratory investigates the evolution and transmission of MDR bacterial pathogens such as *S. aureus* and CRE in community and hospital settings. Her studies have provided important frameworks on how to leverage genomic data to define the transmission of MDR pathogens. Dr. Uhlemann has published extensively on bacterial evolution and MDR infections[95,123-125,220-228] and has led many NIH-sponsored research projects.

Our team has a track record of working effectively with local and federal organizations to address public health problems. Over the last 7 years, we have provided weekly real-time forecasts of ILI and hospitalizations to the CDC *FluSight* effort[173-175] and have contributed component models to the *FluSightNetwork* collaborative ensemble[172], which is now being used operationally by the CDC. Additionally, under a one-year CDC contract (BAA 75D301-19-R-67835), we currently serve as the coordinating center for a modeling network consisting of five institutions that is systematically testing pandemic influenza response and intervention practices. We have also participated in multi-institution modeling efforts for Dengue[176] and *Aedes* mosquitoes.

Reference

1. Magill, S.S. *et al.* Multistate point-prevalence survey of health care–associated infections. *N. Engl. J. Med*. **370,** 1198–1208 (2014).
2. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2019*. (2019). https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf
3. Weiner, L. M. *et al.* Infection Control & Hospital Epidemiology Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011–2014. *Infect. Control Hosp. Epidemiol. / FirstView Artic.* **37,** 1–14 (2016).
4. Weiner, L. M. *et al.* Vital Signs: Preventing Antibiotic-Resistant Infections in Hospitals - United States, 2014. *MMWR. Morb. Mortal. Wkly. Rep.* **65,** 235–241 (2016).
5. Higgins, P.G., Dammhayn, C., Hackel, M. & Seifert, H. Global spread of carbapenem-resistant Acinetobacter baumannii. *J. Antimicrob. Chemother.* **65,** 233–238 (2010).
6. Perez, F., Hujer, A.M., Hujer, K.M., Decker, B.K., Rather, P.N. & Bonomo, R.A. Global challenge of multidrug-resistant Acinetobacter baumannii. *Antimicrob. Agents Chemother.* **51,** 3471–3484 (2007).
7. Dijkshoorn, L., Nemec, A. & Seifert, H. An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii. *Nat. Rev. Microbiol.* **5,** 939 (2007).
8. Abbo, A., Navon-Venezia, S., Hammer-Muntz, O., Krichali, T., Siegman-Igra, Y. & Carmeli, Y. Multidrug-resistant Acinetobacter baumannii. *Emerg. Infect. Dis.* **11,** 22 (2005).
9. Centers for Disease Control and Prevention. Drug-Resistant *Candida auris*, Fact sheet. https://www.cdc.gov/drugresistance/pdf/threats-report/candida-auris-508.pdf
10. Chowdhary, A., Sharma, C. & Meis, J.F. Candida auris: a rapidly emerging cause of hospital-acquired multidrug-resistant fungal infections globally. *PLoS Pathog.* **13,** e1006290 (2017).
11. Guh, A.Y. & Kutty, P.K. Clostridioides difficile infection. *Ann. Intern. Med.* **169,** ITC49-ITC64 (2018).
12. Schwaber, M. J. & Carmeli, Y. Carbapenem-Resistant Enterobacteriaceae. *JAMA* **300,** 2911 (2008).
13. Gupta, N., Limbago, B. M., Patel, J. B. & Kallen, A. J. Carbapenem-resistant enterobacteriaceae: Epidemiology and prevention. *Clin. Infect. Dis.* **53,** 60–67 (2011).
14. Guh, A. Y., Limbago, B. M. & Kallen, A. J. Epidemiology and prevention of carbapenem-resistant Enterobacteriaceae in the United States. *Expert Rev. Anti. Infect. Ther.* **12,** 565–80 (2014).
15. Tapsall, J.W., Ndowa, F., Lewis, D.A. & Unemo, M. Meeting the public health challenge of multidrug- and extensively drug-resistant Neisseria gonorrhoeae. *Expert Rev. Anti-infect. Ther.* **7,** 821–834 (2009).
16. Unemo, M. & Nicholas, R.A. Emergence of multidrug-resistant, extensively drug-resistant and untreatable gonorrhea. *Future Microbiol.* **7,** 1401–1422 (2012).
17. Palavutitotai, N., Jitmuang, A., Tongsai, S., Kiratisin, P. & Angkasekwinai, N. Epidemiology and risk factors of extensively drug-resistant Pseudomonas aeruginosa infections. *PLoS ONE* **13,** e0193431 (2018).
18. Obritsch, M.D., Fish, D.N., MacLaren, R. & Jung, R. Nosocomial infections due to multidrug- resistant Pseudomonas aeruginosa: epidemiology and treatment options. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **25,** 1353–1364 (2005).
19. Driscoll, J.A., Brody, S.L. & Kollef, M.H. The epidemiology, pathogenesis and treatment of Pseudomonas aeruginosa infections. *Drugs* **67,** 351–368 (2007).
20. Murray, B.E. Vancomycin-resistant enterococcal infections. *N. Engl. J. Med*. **342,** 710–721 (2000).
21. Cetinkaya, Y., Falk, P. & Mayhall, C.G. Vancomycin-resistant enterococci. *Clin. Microbiol. Rev*. **13,** 686–707 (2000).
22. Reyes, K., Bardossy, A.C. & Zervos, M. Vancomycin-resistant enterococci: epidemiology, infection prevention, and control. *Infect. Dis. Clin*. **30,** 953–965 (2016).
23. Rupp, M.E. & Fey, P.D. Extended spectrum β-lactamase (ESBL)-producing Enterobacteriaceae. *Drugs* **63,** 353–365 (2003).
24. Logan, L.K., Braykov, N.P., Weinstein, R.A. & Laxminarayan, R. Extended-spectrum β-lactamase–producing and third-generation cephalosporin-resistant Enterobacteriaceae in children: trends in the United States, 1999–2011. *Journal of the Pediatric Infectious Diseases Society* **3,** 320–328 (2014).
25. Lowy, F. D. Staphylococcus aureus infections. *N. Engl. J. Med*. **339**, 520–532 (1998).
26. Dantes, R. National Burden of Invasive Methicillin-Resistant *Staphylococcus aureus* Infections, United States, 2011. *JAMA Intern. Med.* **30333,** 1970–1978 (2013).
27. Klein, E. Y., Sun, L., Smith, D. L. & Laxminarayan, R. The changing epidemiology of methicillin-

resistant Staphylococcus aureus in the United States: A national observational study. *Am. J. Epidemiol.* **177,** 666–674 (2013).
28. Breiman, R.F., Butler, J.C., Tenover, F.C., Elliott, J.A. & Facklam, R.R. Emergence of drug-resistant pneumococcal infections in the United States. *JAMA* **271,** 1831–1835 (1994).
29. Campbell Jr, G.D. & Silberman, R. Drug-resistant Streptococcus pneumoniae. *Rev. Infect. Dis.* **26,** 1188–1195 (1998).
30. Grundmann, H. & Hellriegel, B. Mathematical modelling: a tool for hospital infection control. *Lancet Infect. Dis.* **6,** 39–45 (2006).
31. Weinstein, R.A., Bonten, M.J., Austin, D.J. & Lipsitch, M. Understanding the spread of antibiotic resistant pathogens in hospitals: mathematical models as tools for control. *Clin. Infect. Dis*. **33,** 1739–1746 (2001).
32. van Kleef, E., Robotham, J.V., Jit, M., Deeny, S.R. & Edmunds, W.J. Modelling the transmission of healthcare associated infections: a systematic review. *BMC Inf. Dis.* **13,** 294 (2013).
33. Doan, T.N., Kong, D.C., Kirkpatrick, C.M. & McBryde, E.S. Optimizing hospital infection control: the role of mathematical modeling. *Infect. Control Hosp. Epidemiol*. **35,** 1521–1530 (2014).
34. Spicknall, I.H., Foxman, B., Marrs, C.F. & Eisenberg, J.N. A modeling framework for the evolution and spread of antibiotic resistance: literature review and model categorization. *Am. J. Epidemiol*. **178,** 508–520 (2013).
35. Cooper, B. S. *et al.* Methicillin-resistant Staphylococcus aureus in hospitals and the community: stealth dynamics and control catastrophes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10223–10228 (2004).
36. Kajita, E., Okano, J. T., Bodine, E. N., Layne, S. P. & Blower, S. Modelling an outbreak of an emerging pathogen. *Nat. Rev. Microbiol.* **5,** 700–709 (2007).
37. Wang, X., Panchanathan, S. & Chowell, G. A Data-Driven Mathematical Model of CA-MRSA Transmission among Age Groups: Evaluating the Effect of Control Interventions. *PLoS Comput. Biol.* **9,** (2013).
38. Chamchod, F. & Ruan, S. Modeling the spread of methicillin-resistant Staphylococcus aureus in nursing homes for elderly. *PLoS ONE* **7,** e29757 (2012).
39. Toth, D.J., Khader, K., Beams, A. & Samore, M.H. Model-based Assessment of the Effect of Contact Precautions Applied to Surveillance-detected Carriers of Carbapenemase-producing Enterobacteriaceae in Long-term Acute Care Hospitals. *Clin. Infect. Dis*. **69,** S206–S213 (2019).
40. Bootsma, M. C. J., Diekmann, O. & Bonten, M. J. M. Controlling methicillin-resistant Staphylococcus aureus: Quantifying the effects of interventions and rapid diagnostic testing. *Proc. Natl. Acad. Sci. U. S. A.* **103,** (2006).
41. Smith, D.L., Levin, S.A. and Laxminarayan, R. Strategic interactions in multi-institutional epidemics of antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3153–3158 (2005).
42. Smith, D.L., Dushoff, J., Perencevich, E.N., Harris, A.D. & Levin, S.A. Persistent colonization and the spread of antibiotic resistance in nosocomial pathogens: resistance is a regional problem. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3709-3714 (2004).
43. Paul, P., Slayton, R.B., Kallen, A.J., Walters, M.S. & Jernigan, J.A. Modeling regional transmission and containment of a healthcare-associated multidrug-resistant organism. *Clin. Infect. Dis*. ciz248 (2019).
44. Pei, S., Morone, F., Liljeros, F., Makse, H. & Shaman, J. L. Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. *eLife* **7**, e40977 (2018).
45. Slayton, R.B. et al. Vital signs: estimated effects of a coordinated approach for action to reduce antibiotic-resistant infections in health care facilities—United States. *MMWR. Morb. Mortal. Wkly. Rep.* **64,** 826 (2015).
46. Lee, B. Y. *et al.* Modeling the Spread of Methicillin-Resistant Staphylococcus aureus (MRSA) Outbreaks throughout the Hospitals in Orange County, California. *Infect. Control Hosp. Epidemiol.* **32,** 562–572 (2011).
47. Lee, B.Y., *et al.* How Introducing a Registry With Automated Alerts for Carbapenem-resistant Enterobacteriaceae (CRE) May Help Control CRE Spread in a Region. *Clin. Infect. Dis*. ciz300 (2019).
48. Harris, A.D., Morgan, D.J., Pineles, L., Perencevich, E.N. & Barnes, S.L., Deconstructing the relative benefits of a universal glove and gown intervention on MRSA acquisition. *J. Hosp. Infect*. **96,** 49-53 (2017).
49. Toth, D.J. *et al.* The potential for interventions in a long-term acute care hospital to reduce transmission of carbapenem-resistant Enterobacteriaceae in affiliated healthcare facilities. *Clin. Infect. Dis*. **65,** 581–587 (2017).

50. Khader, K. *et al.* Variation and Trends in Transmission Dynamics of Methicillin-Resistant Staphylococcus Aureus in Veterans Affairs Hospitals and Nursing Homes. *Epidemics* **28,** 100347 (2019).
51. Boyce, J.M. Environmental contamination makes an important contribution to hospital infection. *J. Hosp. Infect*. **65,** 50–54 (2007).
52. Kumar, J. *et al.* Environmental Contamination with Candida Species in Multiple Hospitals Including a Tertiary Care Hospital with a Candida auris Outbreak. *Pathogens and Immunity* **4,** 260 (2019).
53. Skov, R. L. & Jensen, K. S. Community-associated meticillin-resistant Staphylococcus aureus as a cause of hospital-acquired infections. *J. Hosp. Infect*. **73,** 364–370 (2009).
54. D'Agata, E. M. C. *et al.* Modeling the Invasion of Community Acquired Methicillin Resistant Staphylococcus aureus into Hospitals. *Clin. Infect. Dis.* **48,** 274–284 (2009).
55. Stenhem, M. *et al.* Epidemiology of methicillin-resistant Staphylococcus aureus (MRSA) in Sweden 2000–2003, increasing incidence and regional differences. *BMC Infect. Dis.* **6,** 30 (2006).
56. Larsson, A.K., Gustafsson, E., Johansson, P.J.H., Odenholt, I., Petersson, A.C. & Melander, E. Epidemiology of MRSA in southern Sweden: strong relation to foreign country of origin, health care abroad and foreign travel. *Eur. J. Clin. Microbiol. Infect. Dis.* **33,** 61–68 (2014).
57. Cooper, B.S., Medley, G.F., Bradley, S.J. & Scott, G.M. An augmented data method for the analysis of nosocomial infection data. *Am. J. Epidemiol*. **168,** 548–557 (2008).
58. Thomas, A., Redd, A., Khader, K., Leecaster, M., Greene, T. & Samore, M. Efficient parameter estimation for models of healthcare-associated pathogen transmission in discrete and continuous time. *Math. Med. Biol*. **32,** 81–100 (2013).
59. Thomas, A. *et al.* Extended models for nosocomial infection: parameter estimation and model selection. *Math. Med. Biol*. **35,** i29–i49 (2017).
60. Scanvic, A., Denic, L., Gaillon, S., Giry, P., Andremont, A. & Lucet, J.C. Duration of colonization by methicillin-resistant Staphylococcus aureus after hospital discharge and risk factors for prolonged carriage. *Clin. Infect. Dis.* **32,** 1393–1398 (2001).
61. MacKinnon, M.M. & Allen, K.D. Long-term MRSA carriage in hospital patients. *J. Hosp. Infect*. **46,** 216–221 (2000).
62. Schechner, V. *et al.* Asymptomatic rectal carriage of blaKPC producing carbapenem-resistant Enterobacteriaceae: who is prone to become clinically infected? *Clin. Microbiol. Infect.* **19,** 451–456 (2013).
63. Wiener-Well, Y. *et al.* Carriage rate of carbapenem-resistant Klebsiella pneumoniae in hospitalised patients during a national outbreak. *J. Hosp. Infect*. **74,** 344–349 (2010).
64. Zimmerman, F.S., Assous, M.V., Bdolah-Abram, T., Lachish, T., Yinnon, A.M. & Wiener-Well, Y. Duration of carriage of carbapenem-resistant Enterobacteriaceae following hospital discharge. *Am. J. Infect. Control* **41,** 190–194 (2013).
65. Eyre, D.W. *et al.* Asymptomatic Clostridium difficile colonisation and onward transmission. *PLoS ONE* **8,** e78445 (2013).
66. Furuya-Kanamori, L. *et al.* Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. *BMC Infect. Dis.* **15,** 516 (2015).
67. Ozaki, E. *et al.* Clostridium difficile colonization in healthy adults: transient colonization and correlation with enterococcal colonization. *J. Med. Microbiol.* **53,** 167–172 (2004).
68. Rea, M.C. *et al.* Clostridium difficile carriage in elderly subjects and associated changes in the intestinal microbiota. *J. Clin. Microbiol.* **50,** 867–875 (2012).
69. Donskey, C.J., Kundrapu, S. & Deshpande, A. Colonization versus carriage of Clostridium difficile. *Infect. Dis. Clin.* **29,** 13–28 (2015).
70. Galdys, A.L. *et al.* Prevalence and duration of asymptomatic Clostridium difficile carriage among healthy subjects in Pittsburgh, Pennsylvania. *J. Clin. Microbiol.* **52,** 2406–2409 (2014).
71. Agodi, A., Barchitta, M., Cipresso, R., Giaquinta, L., Romeo, M.A. & Denaro, C. Pseudomonas aeruginosa carriage, colonization, and infection in ICU patients. *Intensive Care. Med*. **33,** 1155–1161 (2007).
72. Harris, A.D. *et al.* Pseudomonas aeruginosa colonization in the intensive care unit: prevalence, risk factors, and clinical outcomes. *Infect. Control Hosp. Epidemiol*. **37,** 544–548 (2016).
73. Centers for Disease Control and Prevention. *Methicillin-resistant Staphylococcus aureus (MRSA).* https://www.cdc.gov/mrsa/community/index.html Accessed in December 2019.
74. Girou, E., Pujade, G., Legrand, P., Cizeau, F. & Brun-Buisson, C. Selective screening of carriers for

control of methicillin-resistant Staphylococcus aureus (MRSA) in high-risk hospital areas with a high level of endemic MRSA. *Clin. Infect. Dis.* **27,** 543–550 (1998).
75. Harbarth, S., Sax, H., Fankhauser-Rodriguez, C., Schrenzel, J., Agostinho, A. & Pittet, D. Evaluating the probability of previously unknown carriage of MRSA at hospital admission. *Am. J. Med.* **119,** 275 (2006).
76. Hidron, Alicia I., et al. Risk factors for colonization with methicillin-resistant Staphylococcus aureus (MRSA) in patients admitted to an urban hospital: emergence of community-associated MRSA nasal carriage. *Clin. Infect. Dis.* **41,** 159–166 (2005).
77. Grundmann, H., Hori, S., Winter, B., Tami, A. & Austin, D. J. Risk factors for the transmission of methicillin-resistant Staphylococcus aureus in an adult intensive care unit: fitting a model to the data. *J. Infect. Dis.* **185,** 481–488 (2002).
78. Alasmari, F., Seiler, S.M., Hink, T., Burnham, C.A.D. & Dubberke, E.R. Prevalence and risk factors for asymptomatic Clostridium difficile carriage. *Clin. Infect. Dis.* **59,** 216–222 (2014).
79. McFarland, L.V., Surawicz, C.M. & Stamm, W.E. Risk factors for Clostridium difficile carriage and C. difficile-associated diarrhea in a cohort of hospitalized patients. *J. Infect. Dis.* **162,** 678–684 (1990).
80. Torres-Gonzalez, P. *et al.* Factors associated to prevalence and incidence of carbapenem-resistant Enterobacteriaceae fecal carriage: a cohort study in a Mexican tertiary care hospital. *PLoS ONE* **10,** e0139883 (2015).
81. Defez, C. *et al.* Risk factors for multidrug-resistant Pseudomonas aeruginosa nosocomial infection. *J. Hosp. Infect*. **57,** 209–216 (2004).
82. Vanhems, P. *et al.* Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8,** e73970 (2013).
83. Hornbeck, T., Naylor, D., Segre, A.M., Thomas, G., Herman, T. & Polgreen, P.M. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *J. Infect. Dis.* **206,** 1549–1557 (2012).
84. Curtis, D.E., Hlady, C.S., Kanade, G., Pemmaraju, S.V., Polgreen, P.M. & Segre, A.M. Healthcare worker contact networks and the prevention of hospital-acquired infections. *PLoS ONE* **8,** e79906 (2013).
85. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6,** 888 (2010).
86. Pei, S., Muchnik, L., Andrade Jr, J.S., Zheng, Z. & Makse, H.A. Searching for superspreaders of information in real-world social media. *Sci. Rep*. **4,** 5547 (2014).
87. Pei, S. & Makse, H.A. Spreading dynamics in complex networks. *J. Stat. Mech.* **2013,** P12002 (2013).
88. Pei, S., Morone, F. & Makse, H.A. Theories for influencer identification in complex networks. In *Complex Spreading Phenomena in Social Systems* (pp. 125-148). Springer, Cham (2018).
89. Pei, S., Wang, J., Morone, F. & Makse, H.A. Influencer identification in dynamical complex systems. *J. Complex Netw.* In press (2019).
90. Donker, T., Wallinga, J. & Grundmann, H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLoS Comput. Biol.* **6,** e1000715 (2010).
91. Nekkab, N., Astagneau, P., Temime, L. & Crepey, P. Spread of hospital-acquired infections: A comparison of healthcare networks. *PLoS Comput. Biol.* **13,** e1005666 (2017).
92. Lee, B.Y. *et al.* Social network analysis of patient sharing among hospitals in Orange County, California. *Am. J. Public Health* **101,** 707–713 (2011).
93. Simmering, J.E., Polgreen, L.A., Campbell, D.R., Cavanaugh, J.E. & Polgreen, P.M. Hospital transfer network structure as a risk factor for Clostridium difficile infection. *Infect. Control Hosp. Epidemiol.* **36,** 1031–1037 (2015).
94. Seybold, U. *et al.* Emergence of community-associated methicillin-resistant Staphylococcus aureus USA300 genotype as a major cause of health care–associated blood stream infections. *Clin. Infect. Dis.* **42,** 647–656 (2006).
95. Uhlemann, A. C. et al. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. *Proc Natl Acad Sci U S A* **111**, 6738-6743 (2014).
96. Carattoli, A. Resistance plasmid families in Enterobacteriaceae. *Antimicrob. Agents Chemother.* **53,** 2227–38 (2009).
97. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3,** 711–21 (2005).
98. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer in situ: a critical review. *Nat. Rev. Microbiol.* **3,** 700–10 (2005).

99. Berg, O. G. & Kurland, C. G. Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.* **19,** 2265–76 (2002).
100. Stecher, B. *et al.* Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 1269–74 (2012).
101. Blake, D. P., Hillman, K., Fenlon, D. R. & Low, J. C. Transfer of antibiotic resistance between commensal and pathogenic members of the Enterobacteriaceae under ileal conditions. *J. Appl. Microbiol.* **95,** 428–436 (2003).
102. Gustafsson, I. *et al.* Bacteria with increased mutation frequency and antibiotic resistance are enriched in the commensal flora of patients with high antibiotic usage. *J. Antimicrob. Chemother*. **52,** 645–650 (2003).
103. Lindgren, M., Löfmark, S., Edlund, C., Huovinen, P. & Jalava, J. Prolonged impact of a one-week course of clindamycin on Enterococcus spp. in human normal microbiota. *Scand. J. Infect. Dis.* **41,** 15–219 (2009).
104. Nyberg, S.D. *et al.* Long-term antimicrobial resistance in Escherichia coli from human intestinal microbiota after administration of clindamycin. *Scand. J. Infect. Dis.* **39,** 514–520 (2007).
105. Tedijanto, C., Olesen, S.W., Grad, Y.H. & Lipsitch, M. Estimating the proportion of bystander selection for antibiotic resistance among potentially pathogenic bacterial flora. *Proc. Natl. Acad. Sci. U. S. A.* **115,** E11988–E11995 (2018).
106. Allegranzi, B. & Pittet, D. Role of hand hygiene in healthcare-associated infection prevention. *J. Hosp. Infect.* **73,** 305–315 (2009).
107. Aboelela, S.W., Saiman, L., Stone, P., Lowy, F.D., Quiros, D. & Larson, E. Effectiveness of barrier precautions and surveillance cultures to control transmission of multidrug-resistant organisms: a systematic review of the literature. *Am. J. Infect. Control* **34,** 484–494 (2006).
108. Rubin, M.A., Samore, M.H. & Harris, A.D. The importance of contact precautions for endemic methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci. *JAMA* **319,** 863–864 (2018).
109. Assadian, O., Toma, C.D. & Rowley, S.D. Implications of staffing ratios and workload limitations on healthcare-associated infections and the quality of patient care. *Crit. Care Med.* **35,** 296–298 (2007).
110. Weinstein, R.A., Stone, P.W., Pogorzelska, M., Kunches, L. & Hirschhorn, L.R. Hospital staffing and health care–associated infections: a systematic review of the literature. *Clin. Infect. Dis.* **47,** 937–944 (2008).
111. Cimiotti, J.P., Aiken, L.H., Sloane, D.M. & Wu, E.S. Nurse staffing, burnout, and health care–associated infection. *Am. J. Infect. Control* **40,** 486–490 (2012).
112. Shang, J., Stone, P. & Larson, E. Studies on nurse staffing and health care–associated infection: Methodologic challenges and potential solutions. *Am. J. Infect. Control* **43,** 581–588 (2015).
113. Donskey, C.J. Does improving surface cleaning and disinfection reduce health care-associated infections? *Am. J. Infect. Control* **41,** S12–S19 (2013).
114. Alhmidi, H. *et al.* Shedding of methicillin-resistant Staphylococcus aureus by colonized patients during procedures and patient care activities. *Infect. Control Hosp. Epidemiol.* **40,** 328–332 (2019).
115. Dancer, S.J. The role of environmental cleaning in the control of hospital-acquired infection. *J. Hosp. Infect.* **73,** 378–385 (2009).
116. Weber, D.J., Anderson, D. & Rutala, W.A. The role of the surface environment in healthcare-associated infections. *Curr. Opin. Infect. Dis*. **26,** 338–344 (2013).
117. Ciccolini, M., Donker, T., Grundmann, H., Bonten, M. J. M. & Woolhouse, M. E. J. Efficient surveillance for healthcare-associated infections spreading between hospitals. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 2271–2276 (2014).
118. Slayton, R.B. *et al.* The cost–benefit of federal investment in preventing Clostridium difficile infections through the use of a multifaceted infection control and antimicrobial stewardship program. *Infect. Control Hosp. Epidemiol.* **36,** 681–687 (2015).
119. Scott, R.D. *et al.* Assessing the social cost and benefits of a national requirement establishing antibiotic stewardship programs to prevent Clostridioides difficile infection in US hospitals. *Antimicrob. Resist. Infect. Control* **8,** 17 (2019).
120. Grimm, V. *et al.* A standard protocol for describing individual-based and agent-based models. *Ecol. Model*. **198,** 115–126 (2006).
121. Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J. & Railsback, S.F. The ODD protocol: a review and first update. *Ecol. Model*. **221,** 2760–2768 (2010).

122. Halloran, M.E. *et al.* Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4639–4644 (2008).
123. Gomez-Simmonds. *et al.* Genomic and geographic context for the evolution of high-risk carbapenem-resistant Enterobacter cloacae complex clones ST171 and ST78. *MBio* **9,** e00542–18 (2018).
124. Macesic, N. *et al.* Emergence of polymyxin resistance in clinical Klebsiella pneumoniae through diverse genetic adaptations: A genomic, retrospective cohort study. *Clin. Infect. Dis*. doi: 10.1093/cid/ciz623 (2019).
125. Macesic, N. et al. Genomic surveillance reveals diversity of multidrug-resistant organism colonization and infection: a prospective cohort study in liver transplant recipients. *Clin. Infect. Dis*. **67**, 905–912 (2018).
126. Croucher, N.J. & Didelot, X. The application of genomics to tracing bacterial pathogen transmission. *Curr. Opin. Microbiol*. **23,** 62–67 (2015).
127. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316,** 1298–1301 (2007).
128. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J. & Vespignani, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21484–21489 (2009).
129. Belik, V., Geisel, T. & Brockmann, D. Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1,** 011001 (2011).
130. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 2015–2020 (2006).
131. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342,**1337–1342 (2013).
132. Wang, L. & Wu, J.T. Characterizing the dynamics underlying global spread of epidemics. *Nat. Commun.* **9,** 218 (2018).
133. Wesolowski, A. et al. Quantifying the impact of human mobility on malaria. *Science* **338,** 267–270 (2012).
134. Wesolowski, A. et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 11887–11892 (2015).
135. Pei, S., Kandula, S., Yang, W. & Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 2752–2757 (2018).
136. Pei, S., Teng, X., Lewis, P. & Shaman, J. Optimizing respiratory virus surveillance networks using uncertainty propagation. *Under Review* (2019).
137. Begon, M., Bennett, M., Bowers, R.G., French, N.P., Hazel, S.M. & Turner, J. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology & Infection* **129,** 147–153 (2002).
138. Shaman, J. & Karspeck, A. Forecasting seasonal outbreaks of influenza. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 20425–30 (2012).
139. Shaman, J. *et al.* Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* **4,** 1057–1062 (2013).
140. DeFelice, N. B., Little, E., Campbell, S. R. & Shaman, J. Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nat. Commun*. **8,** 14592 (2017).
141. DeFelice, N.B., *et al.* Use of temperature to improve West Nile virus forecasts. *PLoS Comput. Biol.* **14,** e1006047 (2018).
142. Shaman, J., Yang, W. & Kandula, S. Inference and Forecast of the Current West African Ebola Outbreak in Guinea, Sierra Leone and Liberia. *PLoS Curr.* **6,** ecurrents.outbreaks.3408774290b1a0f2dd7cae877c8b8f (2014).
143. Yamana, T. K., Kandula, S. & Shaman, J. Superensemble forecasts of dengue outbreaks. *J. Royal Soc. Interface* **13,** 20160410 (2016).
144. Reis, J. & Shaman, J. Retrospective parameter estimation and forecast of respiratory syncytial virus in the United States. *PLoS Comput. Biol.* **12,** e1005133 (2016).
145. Pei, S. & Shaman, J. Counteracting structural errors in ensemble forecast of influenza outbreaks. *Nat. Commun.* **8,** 925 (2017).
146. Pei, S., Cane, M. A. & Shaman, J. Predictability in process-based ensemble forecast of influenza. *PLoS*

Comput. Biol. **15,** e1006783 (2019).
147. Pei, S. & Shaman, J. Aggregating forecasts of multiple respiratory pathogens supports more accurate forecasting of influenza-like-illness. *Under Review* (2019).
148. Kandula, S., Yamana, T., Pei, S., Yang, W., Morita, H. & Shaman, J. Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *J. Roy. Soc. Interface* **15,** 20180174 (2018).
149. Kandula, S., Pei, S. & Shaman, J. Improved forecasts of influenza-associated hospitalization rates with Google Search Trends. *J. Roy. Soc. Interface* **16,** 20190080 (2019).
150. Heaney, A.K., Alexander, K.A. and Shaman, J. Ensemble forecast and parameter inference of childhood diarrhea in Chobe District, Botswana. *Epidemics* In Press (2019).
151. Anderson, J. L. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Weather Rev.* **129,** 2884–2903 (2001).
152. Evensen, G. *Data Assimilation: The ensemble Kalman filter* (Springer, 2008).
153. Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal. Process* **50,** 174–188 (2002).
154. Ionides, E. L., Bhadra, A., Atchadé, Y. & King, A. Iterated filtering. *Ann. Stat.* **39,** 1776–1802 (2011).
155. Ionides, E. L., Breto, C. & King, A. A. Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 18438–18443 (2006).
156. Andrieu, C., Doucet, A. & Holenstein, R. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. B* **72,** 269–342 (2010).
157. Snyder, C., Bengtsson, T., Bickel, P. & Anderson, J. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136,** 4629–4640 (2008).
158. Kevrekidis, I.G., Gear, C.W., Hyman, J.M., Kevrekidid, P.G., Runborg, O. & Theodoropoulos, C. Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. *Commun. Math. Sci*. **1,** 715–762 (2003).
159. Cauchemez, S. *et al.* Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2825–2830 (2011).
160. Cauchemez, S. *et al.* Unraveling the drivers of MERS-CoV transmission. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 9081–9086 (2016).
161. Altarelli, F., Braunstein, A., Dall'Asta, L., Lage-Castellanos, A. & Zecchina, R. Bayesian inference of epidemics on networks via belief propagation. *Phys. Rev. Lett.* **112,** 118701 (2014).
162. Mezard, M. & Montanari, A. *Information, physics, and computation*. (Oxford University Press, 2009).
163. Altarelli, F., Braunstein, A., Dall'Asta, L., Ingrosso, A. & Zecchina, R. The patient-zero problem with noisy observations. *J. Stat. Mech. Theory Exp.* **2014,** P10016 (2014).
164. Beaumont, M.A., Zhang, W. & Balding, D.J. Approximate Bayesian computation in population genetics. *Genetics* **162,** 2025–2035 (2002).
165. Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med.* **10,** 165 (2012).
166. Axelsen, J. B., Yaari, R., Grenfell, B. T. & Stone, L. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9538–9542 (2014).
167. Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J. & Rosenfeld, R. Flexible modeling of epidemics with an empirical Bayes framework. *PLoS Comput. Biol.* **11,** e1004382 (2015).
168. Ben-Nun, M., Riley, P., Turtle, J., Bacon, D. P. & Riley, S. Forecasting national and regional influenza-like illness for the USA. *PLoS Comput. Biol.* **15,** e1007013 (2019).
169. Du, X., King, A. A., Woods, R. J. & Pascual, M. Evolution-informed forecasting of seasonal influenza A (H3N2). *Sci. Transl. Med.* **9,** eaan5325 (2017).
170. Osthus, D., Gattiker, J., Priedhorsky, R. & Del Valle, S. Y. Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy. *Bayesian Anal.* 10.1214/18-BA1117 (2018).
171. Ray, E. L. & Reich, N. G. Prediction of infectious disease epidemics via weighted density ensembles. *PLoS Comput. Biol.* **14,** e1005910 (2018).
172. Reich, N. G. *et al.* A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 3146–3154 (2019).
173. Biggerstaff, M. *et al.* Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge. *BMC Infect. Dis.* **16**, 357 (2016).
174. Biggerstaff, M. *et al.* Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics* **24,** 26–33 (2018).

175. McGowan, C. J. *et al.* Collaborative efforts to forecast seasonal influenza in the United States, 2015-2016. *Sci. Rep.* **9,** 683 (2019).
176. Johansson, M.A. *et al.* An open challenge to advance probabilistic forecasting for dengue epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 24268–24274 (2019).
177. Viboud, C. *et al.* The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics* **22,** 13–21 (2018).
178. Ajelli, M. *et al.* The RAPIDD Ebola forecasting challenge: Model description and synthetic data generation. *Epidemics* **22,** 3–12 (2018).
179. Longini, I.M. *et al.* Containing pandemic influenza at the source. *Science* **309,** 1083–1087 (2005).
180. Germann, T.C., Kadau, K., Longini, I.M. & Macken, C.A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5935–5940 (2006).
181. Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429,** 180 (2004).
182. Ferguson, N.M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437,** 209 (2005).
183. Ferguson, N.M., Cummings, D.A., Fraser, C., Cajka, J.C., Cooley, P.C. & Burke, D.S. Strategies for mitigating an influenza pandemic. *Nature* **442,** 448 (2006).
184. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. *Bayesian Data Analysis* (Chapman and Hall/CRC, 2013).
185. Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. Optimization by simulated annealing. *Science* **220,** 671–680 (1983).
186. Fleming, W.H. & Rishel, R.W. *Deterministic and Stochastic Optimal Control* (Springer Science & Business Media, New York, 2012).
187. Pontryagin, L.S. *Mathematical Theory of Optimal Processes* (CRC Press, New York, 1987).
188. Sharomi, O. & Malik, T. Optimal control in epidemiology. *Ann. Oper. Res.* **251,** 55–71 (2017).
189. Cornelius, S.P., Kath, W.L. & Motter, A.E. Realistic control of network dynamics. *Nat. Commun.* **4,** 1942 (2013).
190. Gao, J., Liu, Y.Y., D'souza, R.M. & Barabási, A.L. Target control of complex networks. *Nat. Commun*. **5,** 5415 (2014).
191. Wang, L.Z. *et al.* A geometrical approach to control and controllability of nonlinear dynamical networks. *Nat. Commun.* **7,** 11323 (2016).
192. Tang, E. & Bassett, D.S., Colloquium: Control of dynamics in brain networks. *Rev. Mod. Phys.* **90,** 031003 (2018).
193. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3243–3248 (2009).
194. Shaman, J. & Day, J. F. Reproductive phase locking of mosquito populations in response to rainfall frequency. *PLoS One* **2,** e331 (2007).
195. Shaman, J., Stieglitz, M., Engel, V., Koster, R. & Stark, C. Representation of subsurface storm flow and a more responsive water table in a TOPMODEL-based hydrology model. *Water Resour. Res.* **38,** 31-1-31–16 (2002).
196. Shaman, J., Spiegelman, M., Cane, M. & Stieglitz, M. A hydrologically driven model of swamp water mosquito population dynamics. *Ecol. Modell.* **194,** 395–404 (2006).
197. Shaman, J., Esbensen, S. K. & Maloney, E. D. The dynamics of the ENSO-Atlantic hurricane teleconnection: ENSO-related changes to the North African-Asian jet affect basin tropical cyclogenesis. *J. Clim.* **22,** 2458–2482 (2009).
198. Shaman, J., Samelson, R. M. & Tziperman, E. Complex Wavenumber Rossby Wave Ray Tracing. *J. Atmos. Sci.* **69,** 2112–2133 (2012).
199. Shaman, J. Amplification due to spatial clustering in an individual-based model of mosquito-avian arbovirus transmission. *Trans. R. Soc. Trop. Med. Hyg.* **101,** 469–483 (2007).
200. Shaman, J. Strategies for controlling the epizootic amplification of arboviruses. *J Med Entomol* **48,** 1189–1196 (2011).
201. Shaman, J. & Tziperman, E. Summertime ENSO-North African-Asian jet teleconnection and implications for the Indian monsoons. *Geophys. Res. Lett.* **34,** L11702 (2007).
202. Rydbeck, A. V., Maloney, E. D., Xie, S. P., Hafner, J. & Shaman, J. Remote forcing versus local feedback of east pacific intraseasonal variability during boreal summer. *J. Clim.* **26,** 3575–3596 (2013).
203. Shaman, J. The seasonal effects of ENSO on atmospheric conditions associated with European

precipitation: Model simulations of seasonal teleconnections. *J. Clim.* **27,** 1010–1028 (2014).
204. Shaman, J. & Tziperman, E. The superposition of eastward and westward Rossby waves in response to localized forcing. *J. Clim.* **29,** 7547–7557 (2016).
205. Yang, W., Lipsitch, M. & Shaman, J. Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl. Acad. Sci.* **112,** 201415012 (2015).
206. Yang, W., Cowling, B. J., Lau, E. H. Y. & Shaman, J. Forecasting Influenza Epidemics in Hong Kong. *PLoS Comput. Biol.* **11,** e1004383 (2015).
207. Shaman, J. & Kandula, S. Improved discrimination of influenza forecast accuracy using consecutive predictions. *PLoS Curr.* **7,** (2015).
208. Yang, W. *et al.* Transmission network of the 2014 – 2015 Ebola epidemic in Sierra Leone. *J. R. Soc. Interface* **12,** 1–9 (2015).
209. Yang, W., Olson, D. R. & Shaman, J. Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Comput. Biol.* **12,** e1005201 (2016).
210. Pei, S., Teng, X., Shaman, J., Morone, F. & Makse, H. A. Efficient collective influence maximization in cascading processes with first-order transitions. *Sci. Rep.* **7,** 45240 (2017).
211. Pei, S., Tang, S. & Zheng, Z. Detecting the influence of spreading in social networks with excitable sensor networks. *PLoS ONE* **10,** e0124848 (2015).
212. Muchnik, L., Pei, S., Parra, L.C., Reis, S.D., Andrade Jr, J.S., Havlin, S. & Makse, H.A. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Sci. Rep.* **3,** 1783 (2013).
213. Pei, S., Muchnik, L., Tang, S., Zheng, Z. & Makse, H.A. Exploring the complex pattern of information spreading in online blog communities. *PLoS ONE* **10,** e0126894 (2015).
214. Teng, X., Pei, S., Morone, F. & Makse, H.A. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Sci. Rep.* **6,** 36043 (2016).
215. Pei, S., Tang, S., Yan, S., Jiang, S., Zhang, X. & Zheng, Z. How to enhance the dynamic range of excitatory-inhibitory excitable networks. *Phys. Rev. E* **86,** 021909 (2012).
216. Zhang, R. & Pei, S. Dynamic range maximization in excitable networks. *Chaos* **28,** 013103 (2018).
217. Wang, J., Pei, S., Wei, W., Feng, X. & Zheng, Z. Optimal stabilization of Boolean networks through collective influence. *Phys. Rev. E* **97,** 032305 (2018).
218. Wang, J., Zhang, R., Wei, W., Pei, S. & Zheng, Z. On the stability of multilayer Boolean networks under targeted immunization. *Chaos* **29,** 013133 (2019).
219. Zhang, R., Quan, G., Wang, J. & Pei, S. Backtracking activation impacts the criticality of excitable networks. *New J. Phys.* In press (2019).
220. Uhlemann, A. C. et al. The environment as an unrecognized reservoir for community-associated methicillin resistant Staphylococcus aureus USA300: a case-control study. *PLoS ONE* **6**, e22407 (2011).
221. Knox, J. et al. Environmental contamination as a risk factor for intra-household Staphylococcus aureus transmission. *PLoS ONE* **7**, e49900 (2012).
222. Rojas, R., Macesic, N., Tolari, G., Guzman, A. & Uhlemann, A.C. Multidrug-Resistant Klebsiella pneumoniae ST307 in Traveler Returning from Puerto Rico to Dominican Republic. *Emerg. Infect. Dis.* **25**, 1583 (2019).
223. Uhlemann, A.C., Annavajhala, M. & Gomez-Simmonds, A. Multidrug-resistant Enterobacter cloacae complex emerging as a global, diversifying threat. *Front. Microbiol.* **10**, 44 (2019).
224. McConville, T. H., Sullivan, S. B., Gomez-Simmonds, A., Whittier, S. & Uhlemann, A. C. Carbapenem-resistant Enterobacteriaceae colonization (CRE) and subsequent risk of infection and 90-day mortality in critically ill patients, an observational study. *PLoS ONE* **12**, e0186195 (2017).
225. Uhlemann, A. C. et al. Evolutionary dynamics of pandemic methicillin-sensitive Staphylococcus aureus ST398 and its international spread via routes of human migration. *MBio* **8**, e01375-16 (2017).
226. Gomez-Simmonds, A. & Uhlemann, A. C. Clinical implications of genomic adaptation and evolution of carbapenem-resistant Klebsiella pneumoniae. *J. Infect. Dis.* **215**(suppl_1), S18-S27 (2017).
227. Uhlemann, A. C., Otto, M., Lowy, F. D. & DeLeo, F. R. Evolution of community-and healthcare-associated methicillin-resistant Staphylococcus aureus. *Infect. Genet. Evol.* **21**, 563-574 (2014).
228. Uhlemann, A.C. et al. Identification of a highly transmissible animal-independent Staphylococcus aureus ST398 clone with distinct genomic and cell adhesion properties. *MBio* **3**, e00027-12 (2012).