

Utilizing big data to track outbreaks of antimicrobial resistant pathogens

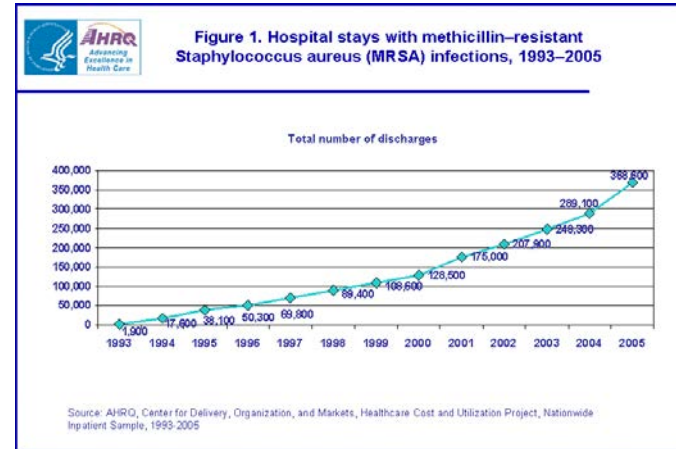
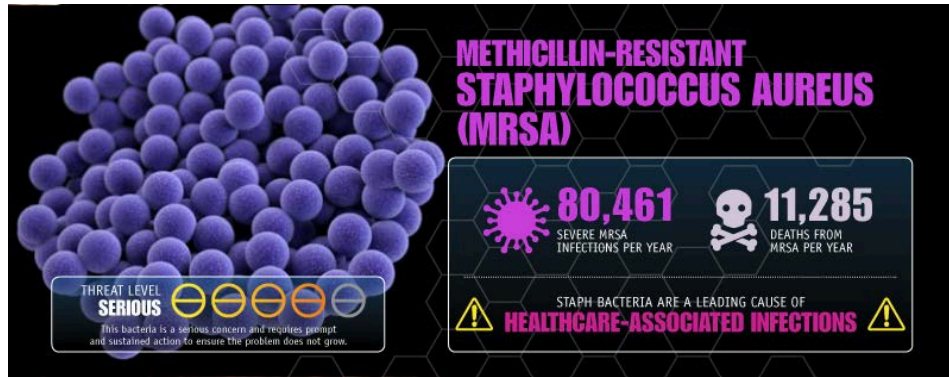
Jeffrey Shaman & Sen Pei

Mailman School of Public Health, Columbia University

July 10th 2019, McLean, VA

Methicillin-resistant Staphylococcus aureus (MRSA)

- ❖ “One of the most common causes of healthcare-associated infections” – US CDC
 - ❑ Antimicrobial resistance – lack of effective treatment, substantial morbidity and mortality
 - ❑ Skin and wound infections; pneumonia; bloodstream infections; sepsis; death



MRSA is transmitted through contact

- ❖ In hospitals, MRSA transmission is primarily mediated by healthcare workers
- ❖ “Approximately 5% of patients in US hospitals carry MRSA” – US CDC
 - ❑ Most of them have no symptoms – colonized
 - ❑ Colonized patients can still transmit MRSA
- ❖ Tracking MRSA spread is difficult
 - ❑ Invasion into the community
 - ❑ Asymptomatic colonization
 - ❑ Limited observation
- ❖ Stealth transmission dynamics



Figure shows schematic of MRSA transmission cycle

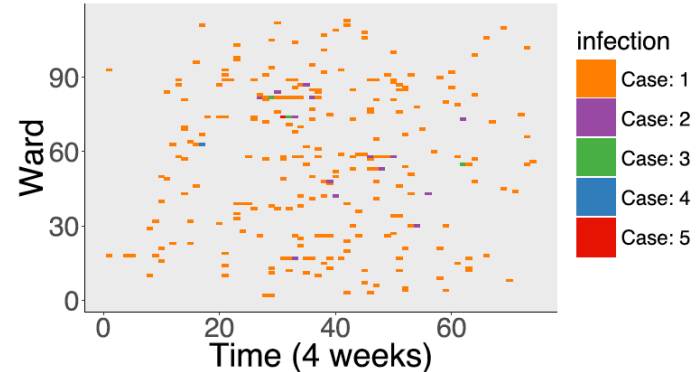
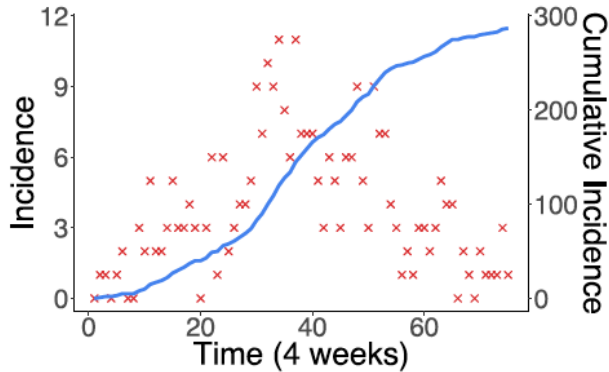
Big data in Swedish hospitals

❖ Hospitalization data

- ❑ Over 2M hospitalization records in 66 hospitals in Stockholm county
- ❑ Cover 744K distinct patients (1/3 of the total 2.2M population), spanning nearly 10 years

❖ Infection data

- ❑ 991 confirmed MRSA cases (172 strains)
- ❑ “UK E15” – 289 cases in 6 years



Left figure presents Time series of UK E15 MRSA incidence and cumulative incidence. Right figure presents time series of UK E15 MRSA cases by ward.

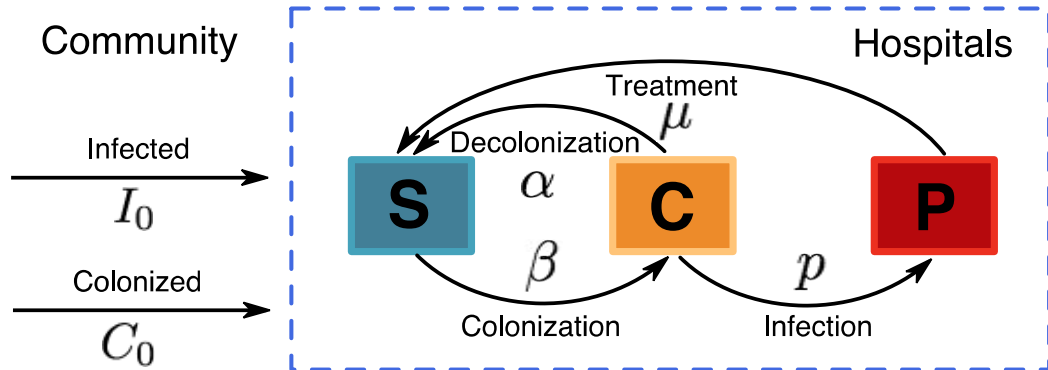
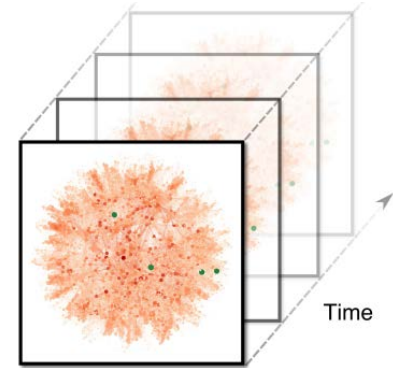
Contact network and transmission model

❖ Time-varying contact network

- ❑ Connect patients if they reside on the same ward at the same time
- ❑ Changing from day to day

❖ Agent-based MRSA transmission model

- ❑ Susceptible (S), colonized (C), tested positive (P)
- ❑ Interaction with community



Top right figure is illustration of the evolving contact network. Bottom figure is a schematic of the transmission model

Validation of Model-Inference Framework

Generate weekly incidence of infection, impose noise, and infer using the noisy observations.

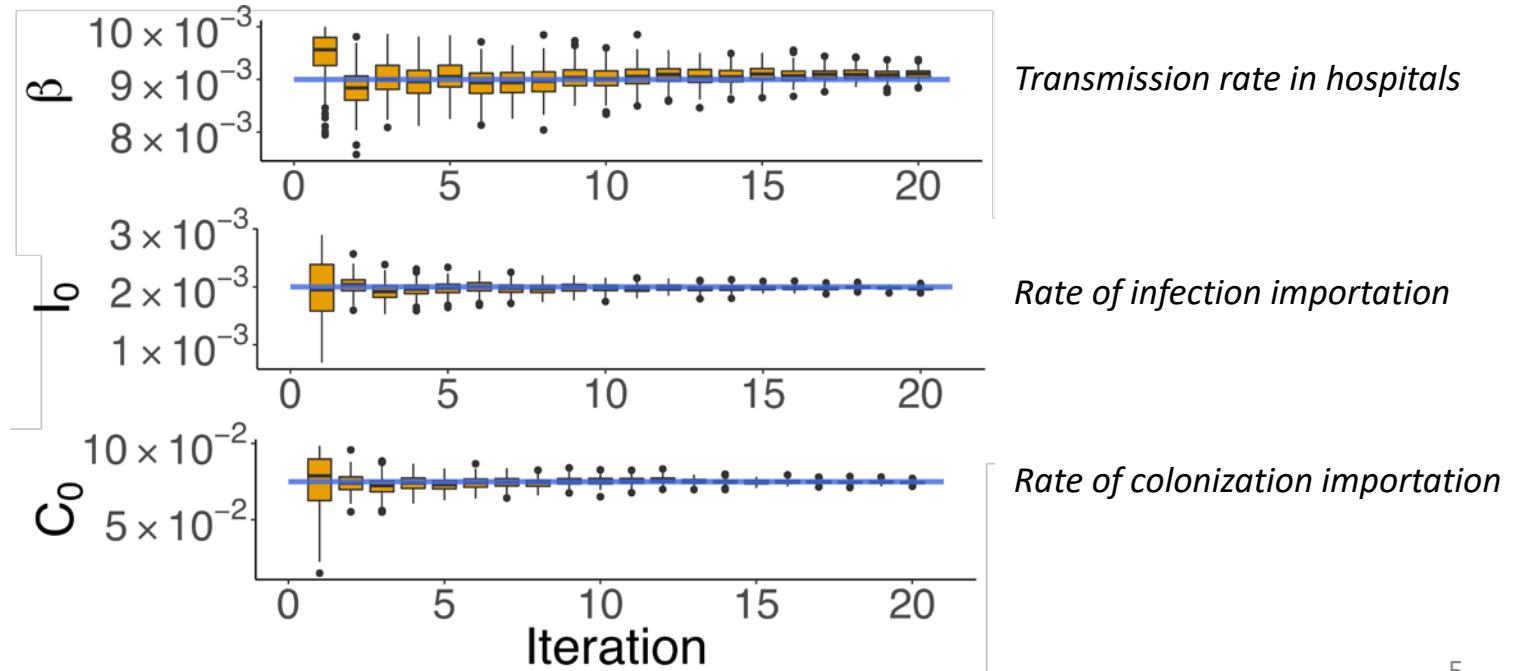


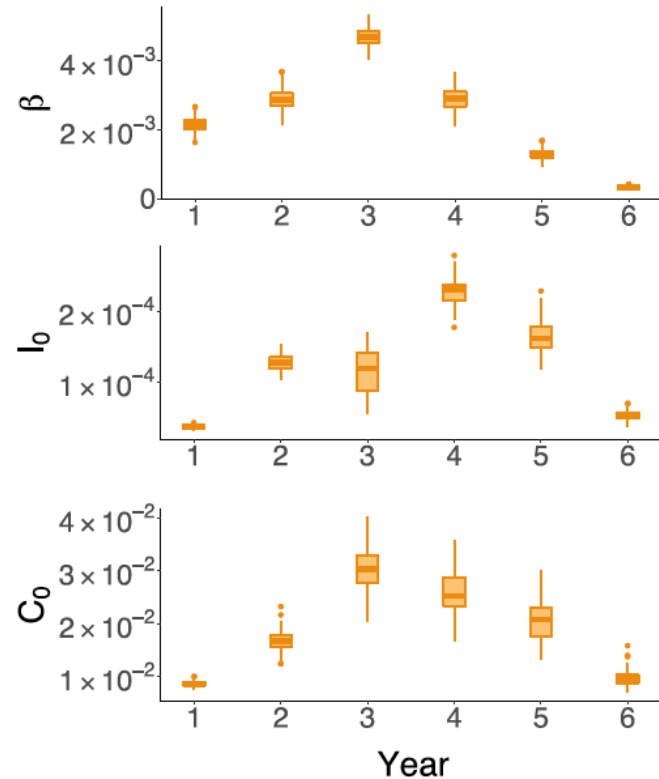
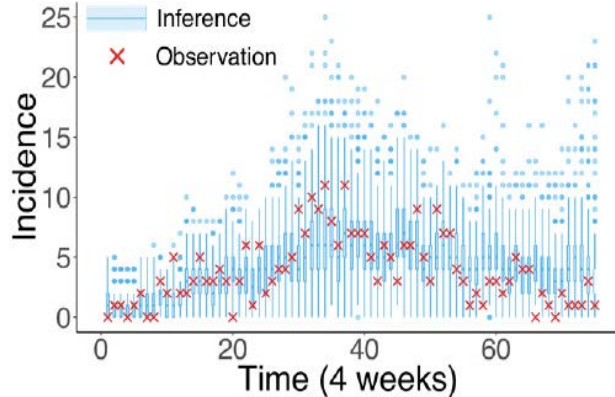
Figure shows synthetic validation of model-inference framework to known targets. The inference method iteratively loops through the time series of synthetic observations and with each iteration converges to the true values of unobserved parameters.

Calibrating model using actual infection data

❖ Infer unobserved information

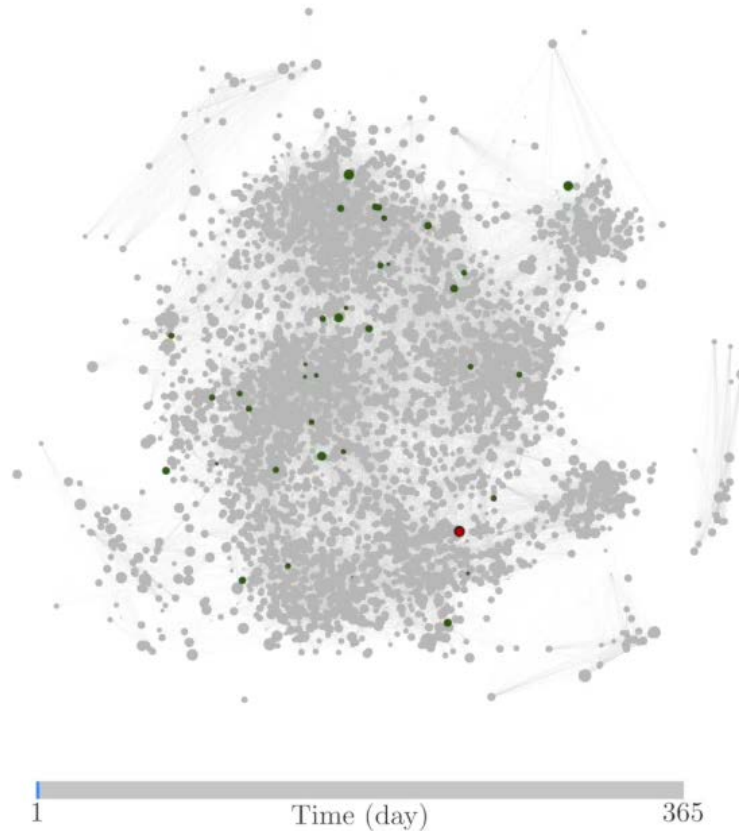
- Transmission rate in hospitals
- Rate of infection importation
- Rate of colonization importation

❖ Reproducing observed outbreaks



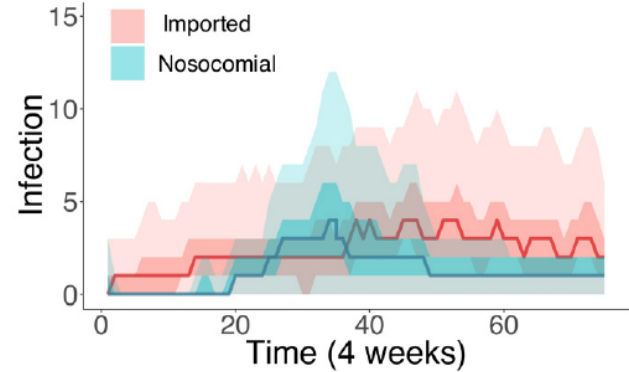
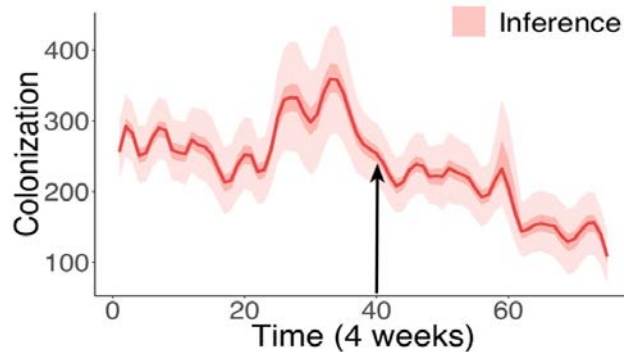
Left figure: time series of 1000 stochastic simulations of the fit model (shown as blue box and whiskers). Red 'x's are the observations of UK E15 MRSA. Right figure: Inferred distributions of key model parameters over 6 years, obtained from 100 independent realizations of the inference algorithm.

Agent-based model simulation



Importation and colonization

- ❖ Transmission versus Importation
 - ❑ Importation > Transmission
 - ❑ Invasion from community into hospitals
 - ❑ Increase screening!
- ❖ How many patients are colonized?



Who and where are they?

Are they subject to the same risk?

Left bottom figure: Inferred distributions of colonized patients through time. The dark areas mark the IQR; light areas show values within the range $[Q1 - 1.5 \text{ IQR}, Q3 + 1.5 \text{ IQR}]$.

Top right figure: Inferred distributions of infections transmitted in hospital (turquoise area) and imported from outside the study hospitals (pink area). The dark areas mark the IQR; light areas show values within the range $[Q1 - 1.5 \text{ IQR}, Q3 + 1.5 \text{ IQR}]$.

Heterogeneous colonization risk

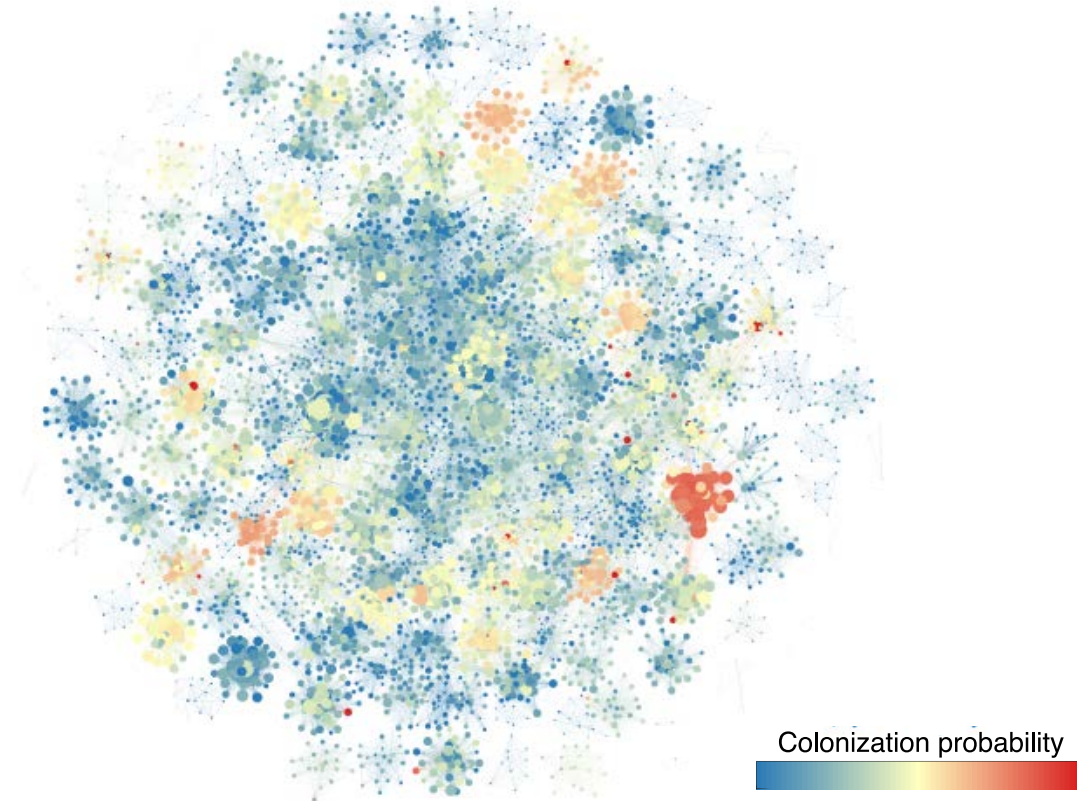
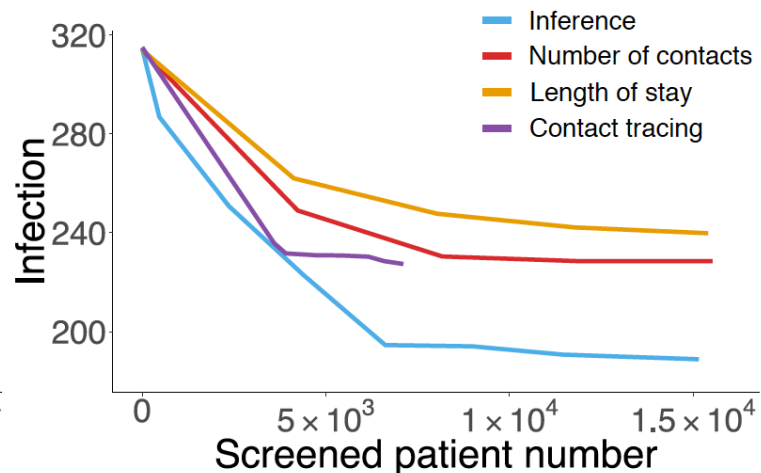
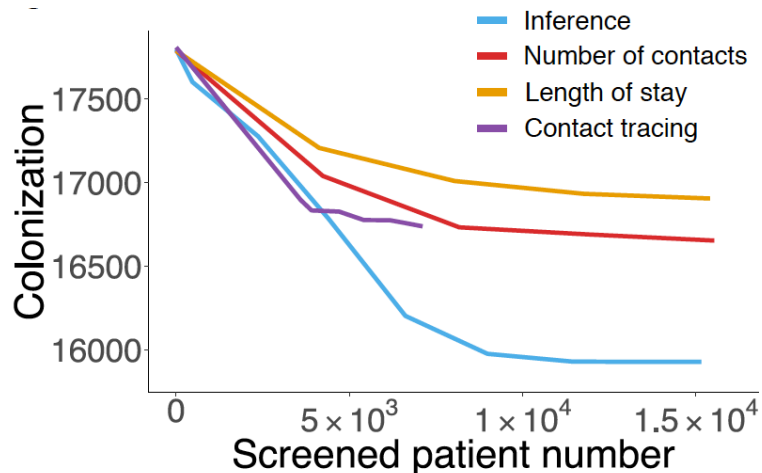


Figure: Visualization of individual- level colonization probability at $T = 40$. The probability is color-coded in a logarithmic scale. Node size reflects the number of connections.

Retrospective control experiment

❖ Targeted intervention on high-risk patients

- ❑ Screening 0.89% patients can avert up to 38% infection and 9% colonization
- ❑ Outperform heuristic strategies



Figures: Comparison of the inference-based intervention with heuristic control measures informed by number of contacts, length of stay and contact tracing. Curves are average cumulative cases obtained from 1000 experiments with a 100% decolonization success rate. Left: colonizations; Right: infections.

MRSA and *Candida auris* (*C. auris*)

- ❖ MRSA and *C. auris* share many transmission features

Similarity in transmission process	MRSA	<i>C. auris</i>
Spread in healthcare settings	✓	✓
Through contact with contaminated persons/objects	✓	✓
Most carriage has no symptom, hard to detect	✓	✓
Higher risk for patients with long-term stay, high antibiotic use and foreign objects in the body	✓	✓
Potential to enter the community	✓	✓

Adapting to track *C. auris* spread

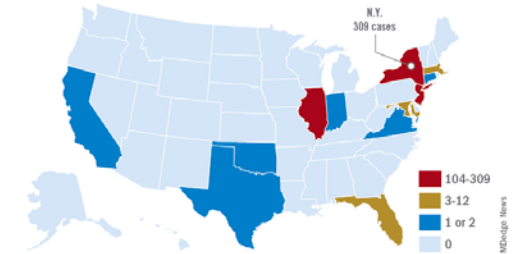
❖ Possible to extend the MRSA model to track *C. auris*

- ❑ Similar transmission process
- ❑ Framework can be adapted for other diseases
- ❑ Flexible to include personal characteristics

❖ Several challenges in tracking *C. auris*

- ❑ Limited data on infection diagnosis and contact information
- ❑ Personalized risk evaluation, e.g., medical procedures, medications
- ❑ Potential interaction with the community
- ❑ Spatial transmission over large geographical scale
- ❑ Validation in real-world settings

Confirmed clinical cases of *Candida auris* as of Feb. 28, 2019



Along with these 587 cases, *C. auris* colonization has been found in 1,056 other patients through targeted screening in seven states with clinical cases.

Notes: 2019 cases categorized by the patient's state of residence. Before 2019 they were categorized by the state where the specimen was collected.

Source: Centers for Disease Control and Prevention

Acknowledgements

❖ Funding support

- ❑ NIGMS Models of Infectious Disease Agent Study (MIDAS)
- ❑ Defense Threat Reduction Agency (DTRA)



❖ Contact

- ❑ Email: jls106@cumc.columbia.edu (JS), sp3449@cumc.columbia.edu (SP)
- ❑ Website: <https://cpid.iri.columbia.edu> - Real-time infectious disease forecast

A screenshot of the Columbia Mailman School of Public Health website. The top navigation bar is dark blue with the Columbia University logo and name on the left, and "MAILMAN SCHOOL OF PUBLIC HEALTH" on the right. Below this, a secondary navigation bar contains links for "Columbia Prediction of Infectious Diseases", "Publications", "Methods", and "Support". The main content area has a white background with the heading "Influenza Observations and Forecast" in blue. Below the heading is a dark blue bar with three buttons: "US Map", "Chart", and "World Map".

Thank you!