### Supplementary material

### Appendix: analytical details

The regional transmission model is a multifacility SIS (susceptible – infected/infectious – susceptible) model that assumes N constant occupancy facilities (hospitals, nursing homes, communities, etc.) linked through patient transfer. The dynamics of the prevalence at facility *a* are governed by

Here is the transmissibility at facility *a* (the number infected per unit time by an infectious person introduced into facility *a* when all others at the facility are susceptible), is the carriage clearance rate, is the average length of stay at facility *a*, and represents the fraction of admissions at facility *a* that are transfers from facility *b*. The first term on the right-hand side represents incident cases at the facility; the second, loss of cases through clearance and discharge; and the third, introduction of cases at admission through transfers from other facilities in the network.

A pair of approximate results that follow from (A1) are particularly useful in this analysis. They apply to the endemic or steady state prevalence at a facility with a fixed and known admission prevalence :

This approximates for small (at a short-stay facility) to

and for large (at a long-stay facility) to

### Patient flow network

The patient flow network is characterized by the facility-to-facility transfer tallies , the number of admissions (and discharges) aggregated over a given time interval, and the average lengths of stay . The principal source for these quantities are the Center for Medicare and Medicaid Services (CMS) patient-level fee-for-service claims data for CMS beneficiaries. The dataset allows every claim to be tracked by an anonymized but unique patient identifier, a facility identifier and admission and discharge dates, to weave together the patient flow network, either across the United States, or by state or group of states. In practice, the tallies were aggregated over a single calendar year. For transfers among facilities, tallies were aggregated by the date of the second admission and by the time from the first discharge to the second admission (a day or less for direct transfers, longer for indirect). Discharge and admission dates were used to estimate average lengths of stay, by facility.

Most patients admitted to healthcare facilities are discharged to the community and will not be readmitted to another healthcare facility within a year. Thus, the communities are important reservoirs in the network. In order to capture the clustering reflected in the communities, the Dartmouth Atlas of Health Care Hospital Referral Regions (HRRs) were used as the community components of the patient flow network: admission and discharge tallies at a healthcare facility unaccounted for by transfers from or to other healthcare facilities were ascribed to the HRR associated with the facility or its ZIP code.

The patient flow network is built self-consistently from CMS fee-for-service claims data that track CMS beneficiaries only, and a census of persons 65 years and older in the HRRs. The occupancy and transfer numbers used are expected to underestimate their true values. However, in our model (1), it is the relative values of these numbers, rather than their absolute values, that determine the dynamics.

### Disease characteristics

The characteristics of the infectious disease that are relevant to our simplified model are the clearance rate and the setting specific transmissibility . Since the model implicitly assumes a constant population where births and deaths balance, the death of an infectious person and clearance – both resulting in the replacement of an infectious person by a susceptible person – are quantitatively indistinguishable. Thus, captures both mortality and clearance rate. Among the diseases we have modeled, clearance rate is understood to dominate mortality for CRE, while mortality is suspected to dominate clearance rate for *Candida auris*. For CRE, the clearance rate is an input parameter in our model, estimated from a review of the literature.

The estimation of transmissibility, which is disease and setting specific and may change due to changes in infection control practices, etc., is particularly challenging. To estimate the transmissibility for CRE, we used data on positive laboratory tests for CRE reported to the National Healthcare Safety Network (NHSN) in 2015, tallied by the hospital where the specimen was collected. In addition, we made several assumptions: First, that the situation in 2015 was approximately a steady, or endemic, state for CRE. Second, that the proportion of laboratory positive cases among all infected persons is a constant. Third, that the admission prevalence at all hospitals within each HRR were identical. And, finally, that the transmissibility depends only on the type of hospital (short-stay versus long-stay). Denoting the number of positive laboratory tests for CRE from facility *a* by (and the number of admissions by ), and the proportion of laboratory positives among all infected as *p*, we may write

From the short-stay approximation (A3), we get

where denotes the average of *x*. This motivates the regression model

which yields, for , the transmissibility at short-stay hospitals,

where is the coefficient of in (A7). From the long-stay approximation (A4) we get

motivating the regression model

If and are the intercept and coefficient of , respectively, in the model (A10), then we obtain, for , the transmissibility at long-stay hospitals,

As a bonus, we obtain an estimate for *p*, the fraction of all infected cases that are laboratory positive for CRE: