Supporting Appendix: Efficient surveillance for healthcare-associated infections spreading between hospitals

M. Ciccolini*[1,2], T. Donker[2,3], H. Grundmann[2,3], M. J. Bonten[4], M. E. J. Woolhouse[1]

1. Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, United Kingdom
2. Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
3. Department of Medical Microbiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
4. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

* Corresponding author.
E-mail: m.ciccolini@umcg.nl

Supporting Figures

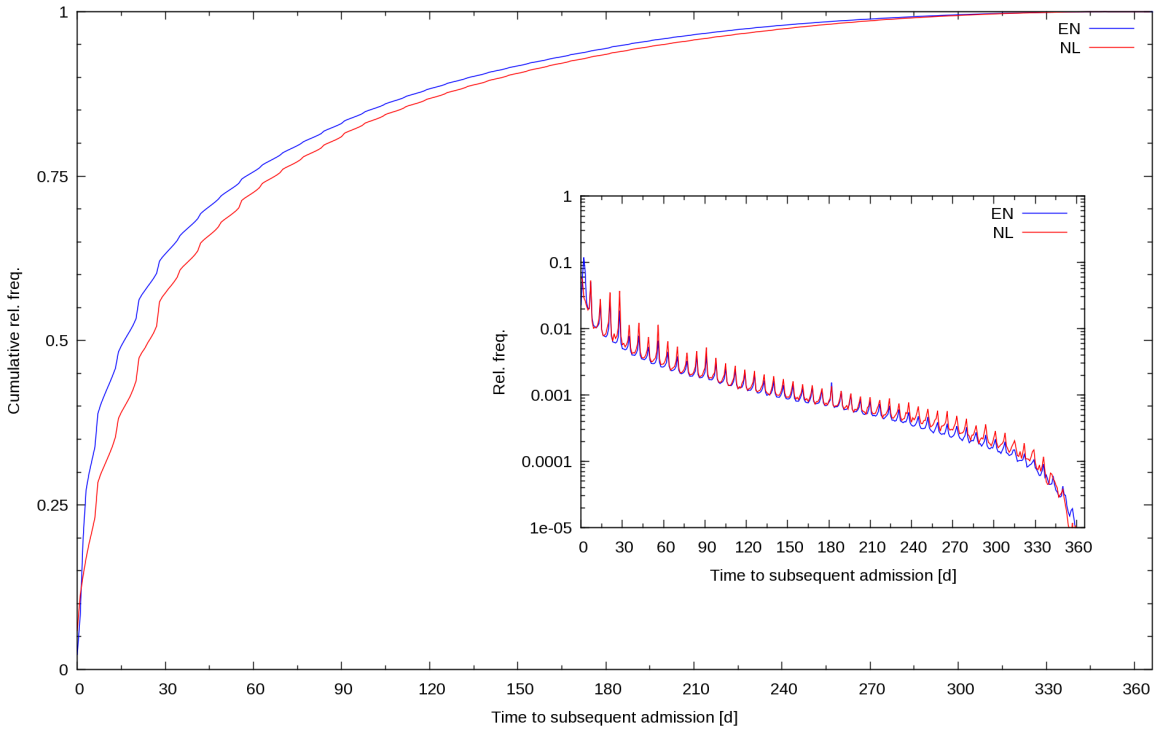

Fig. S1. Time to subsequent admission. Cumulative probability, estimated as cumulative relative frequency, of time to subsequent admission in the English(EN) and Dutch(NL) healthcare networks. The insert shows the corresponding probability distributions.

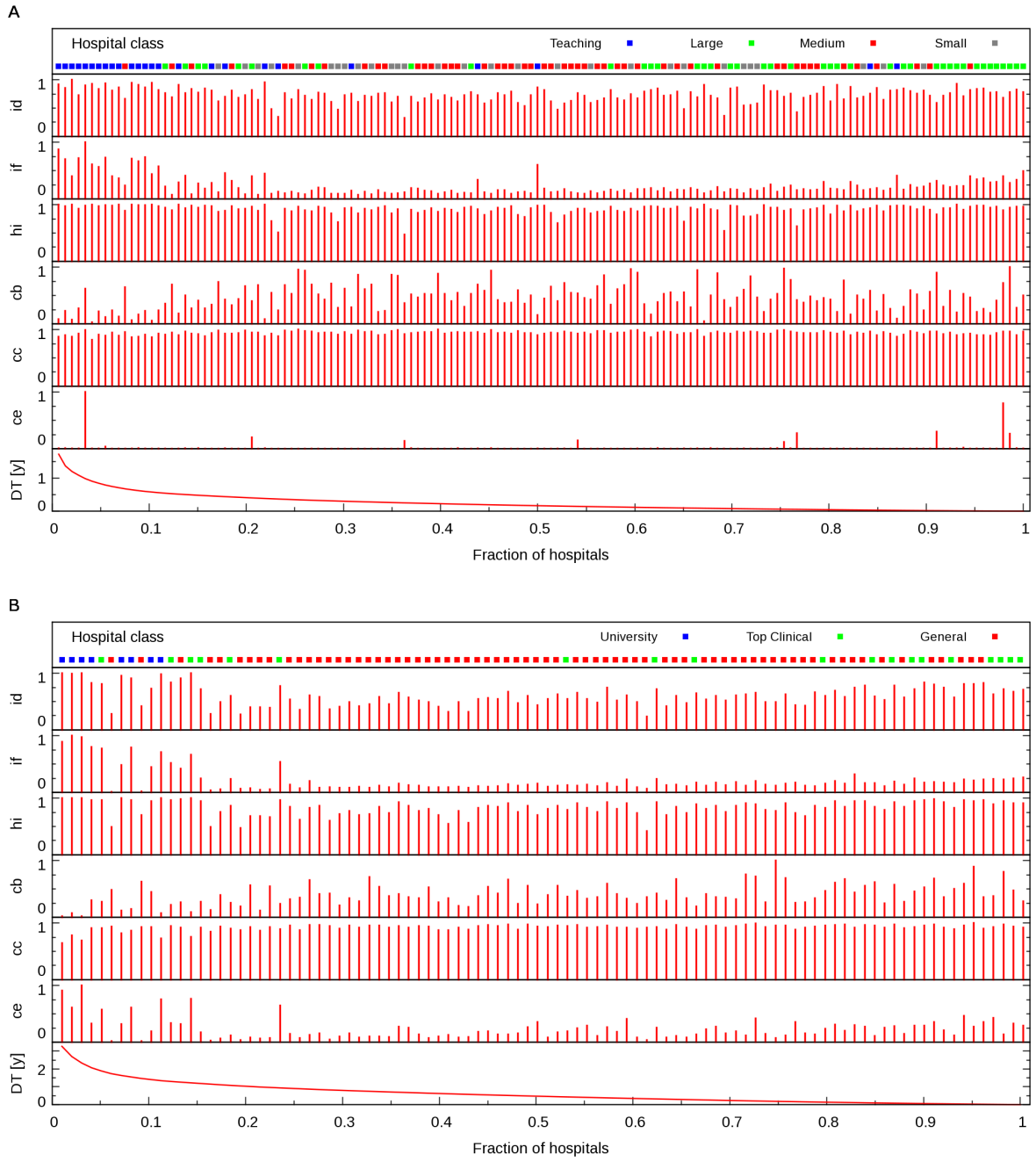

Fig. S2. Hospital category and network metrics in the optimal sentinel list. Hospital category and network metrics associated with each hospital in England (panel A) and The Netherlands (panel B). The i-th hospital in the priority list obtained with the greedy algorithm is displayed at position i/N along the x-axis, with N the total number of hospitals in the country. The values for each network metric have been normalized dividing by the maximum value among all hospitals (i.e. the range of the y-axis is 0–1 in all sub-panels). Each sub-panel corresponds to a different normalized network metric: in-degree (id), in-flux (if), h-index (hi), betweenness centrality (cb), closeness centrality (cc), and eigenvector centrality (ce). For reference, mean detection time in years (DT) obtained with the greedy algorithm is shown in the lowermost sub-panels (cf. Fig. 2A).

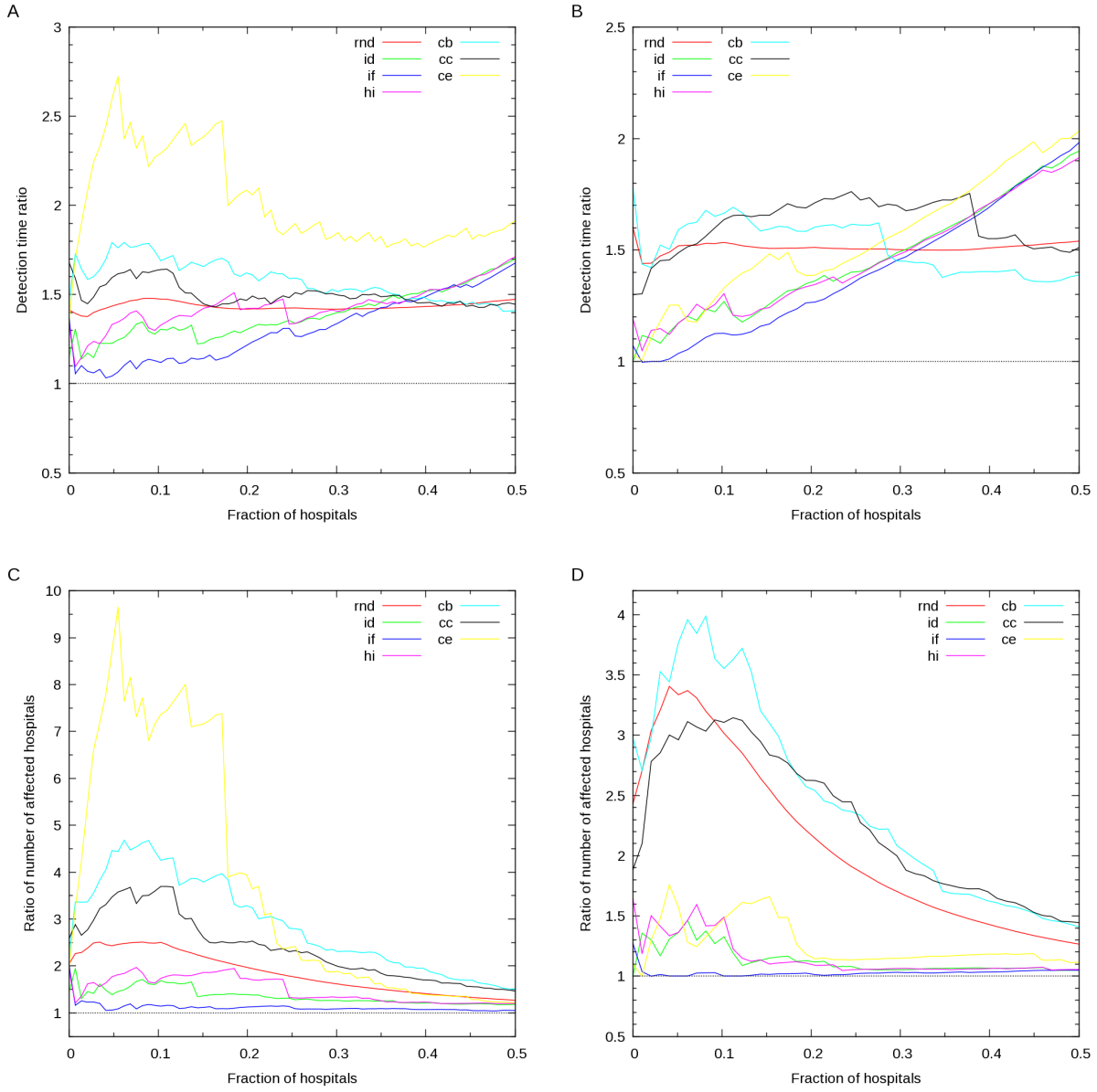

Fig. S3. Comparative performance of different selection schemes. Ratios of mean detection time obtained with different selection schemes to detection time obtained with the greedy algorithm in England (panel A) and The Netherlands (panel B). Ratios of mean number of affected hospitals at detection time obtained with different selection schemes to number of affected hospitals obtained with the greedy algorithm in England (panel C) and The Netherlands (panel D). Hospitals have been selected randomly (rnd), and according to in-degree (id), number of admitted patients previously discharged from a different hospital (if), h-index (hi), betweenness centrality (cb), closeness centrality (cc), and eigenvector centrality (ce). The curve corresponding to random selection was calculated as the mean of 1000 random re-orderings of the hospital priority list. All curves obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$). As an aid to comparison, the curve of equality with the gold standard (i.e. ratio $= 1.0$) is plotted with a dotted line.

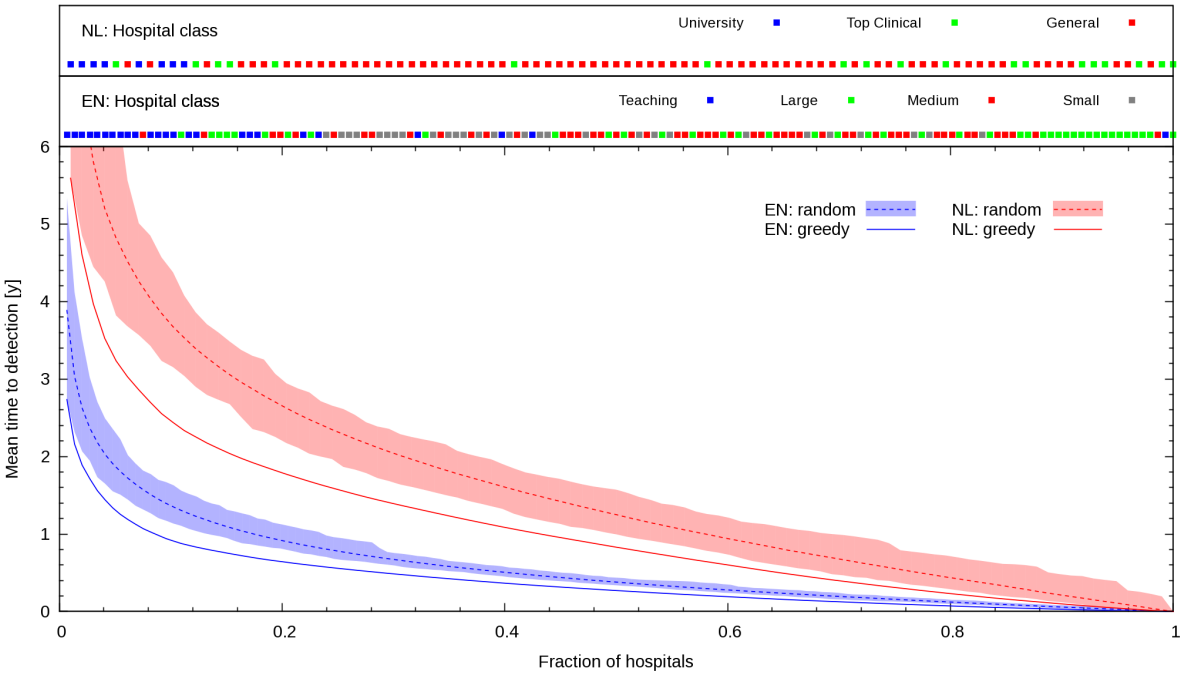

Fig. S4. Mean detection time in a sentinel surveillance programme with time to subsequent admission cut-off. Mean detection time of a novel nosocomial pathogen following emergence in a single, randomly selected hospital, versus fraction of hospitals participating in a sentinel surveillance programme. The continuous lines correspond to results obtained using the greedy algorithm with the English (EN: greedy) and Dutch (NL: greedy) data sets. The shaded region and the dash lines (EN: random, and NL: random) correspond to 1000 random selections of sentinel hospitals and their mean, respectively. The two upper panels show information on hospital category for England (EN) and The Netherlands (NL): the symbol corresponding to the i-th element in the priority list obtained with the greedy algorithm is displayed at position i/N along the x-axis, with N the total number of hospitals in the country. All curves obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$), introducing a cut-off of time to subsequent admission of $T_r = 30$ days.

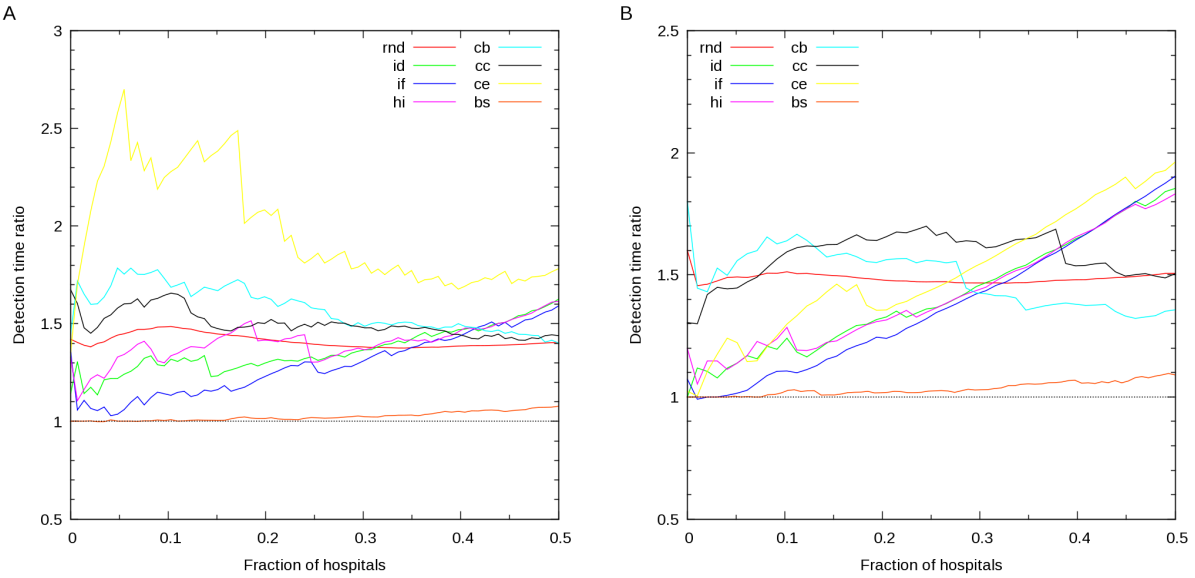

Fig. S5. Detection time ratios with time to subsequent admission cut-off. Ratios of mean detection time obtained with different selection schemes to detection time obtained with the greedy algorithm in England (panel A) and The Netherlands (panel B). Hospitals have been selected randomly (rnd), and according to in-degree (id), number of admitted patients previously discharged from a different hospital (if), h-index (hi), betweenness centrality (cb), closeness centrality (cc), and eigenvector centrality (ce). The (bs) curve shows the result of selecting the hospitals according to the greedy algorithm in the baseline scenario (i.e. without discharge-admission interval cut-off). The curve corresponding to random selection was calculated as the mean of 1000 random re-orderings of the hospital priority list. All curves obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$), introducing a cut-off of time to subsequent admission of $T_r = 30$ days. As an aid to comparison, the curve of equality with the gold standard (i.e. ratio = 1.0) is plotted with a dotted line.

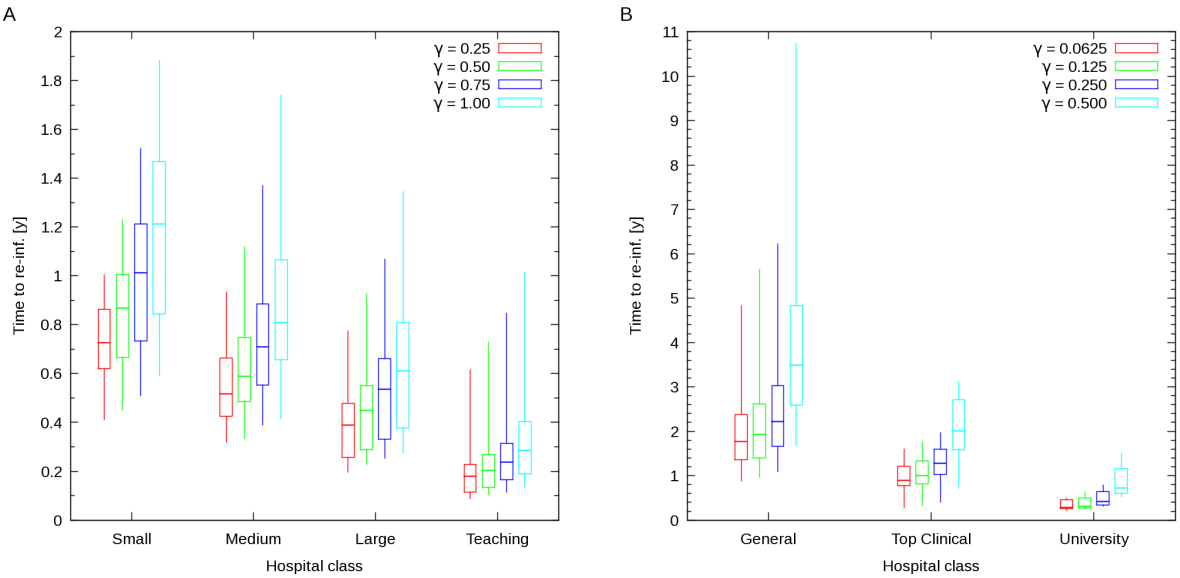

Fig. S6. Time to re-acquisition with time to subsequent admission cut-off. Box plots of median HCAI-free time in the endemic regime in England(panel A) and The Netherlands (panel B). Median times have been grouped according to type of hospital. Different colours correspond to different values of γ. Box bottom and top represent the lower and upper quartiles, respectively. Lower and upper line ends represent 5^{th} and 95^{th} percentiles, respectively. The horizontal bar corresponds to the median. Elimination rates (γ) are expressed in units of year^{-1}. Plot obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$), introducing a cut-off of time to subsequent admission of $T_r = 30$ days.

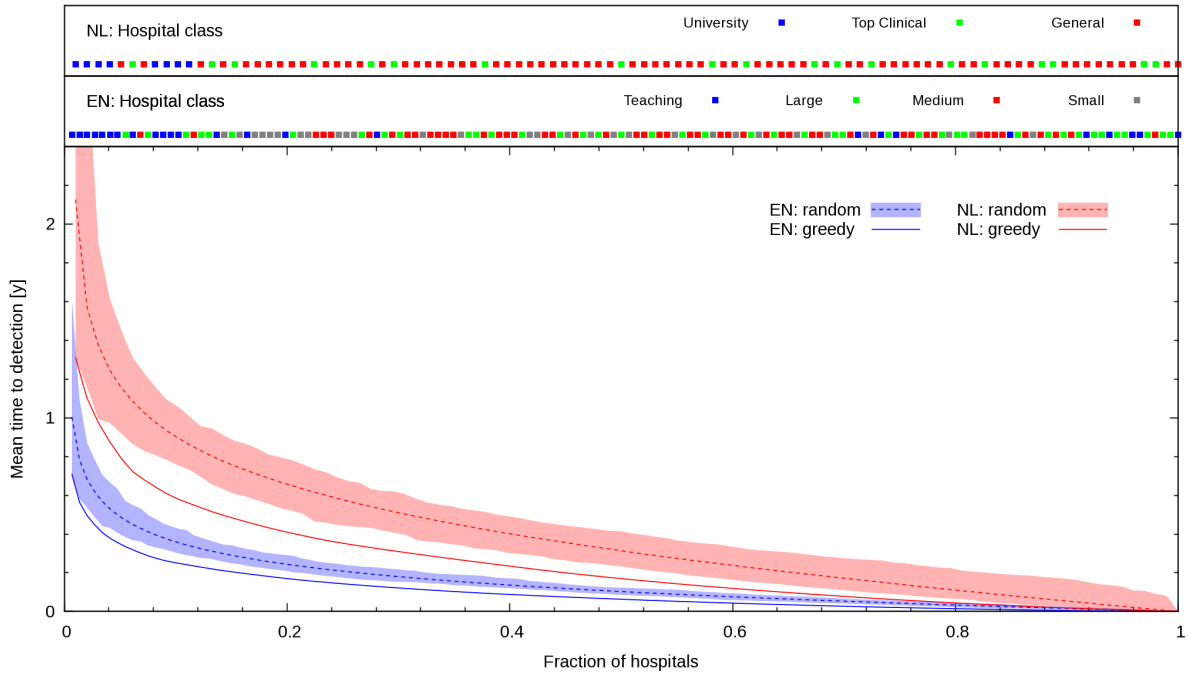

Fig. S7. Mean detection time in a sentinel surveillance programme with heterogeneous β. Mean detection time of a novel nosocomial pathogen following emergence in a single, randomly selected hospital, versus fraction of hospitals participating in a sentinel surveillance programme. The continuous lines correspond to results obtained using the greedy algorithm with the English (EN: greedy) and Dutch (NL: greedy) data sets. The shaded region and the dash lines (EN: random, and NL: random) correspond to 1000 random selections of sentinel hospitals and their mean, respectively. The two upper panels show information on hospital category for England (EN) and The Netherlands (NL): the symbol corresponding to the i-th element in the priority list obtained with the greedy algorithm is displayed at position i/N along the x-axis, with N the total number of hospitals in the country. The per-patient transmission probability associated with each pair of hospitals has been randomly sampled as described in the Supporting Methods.

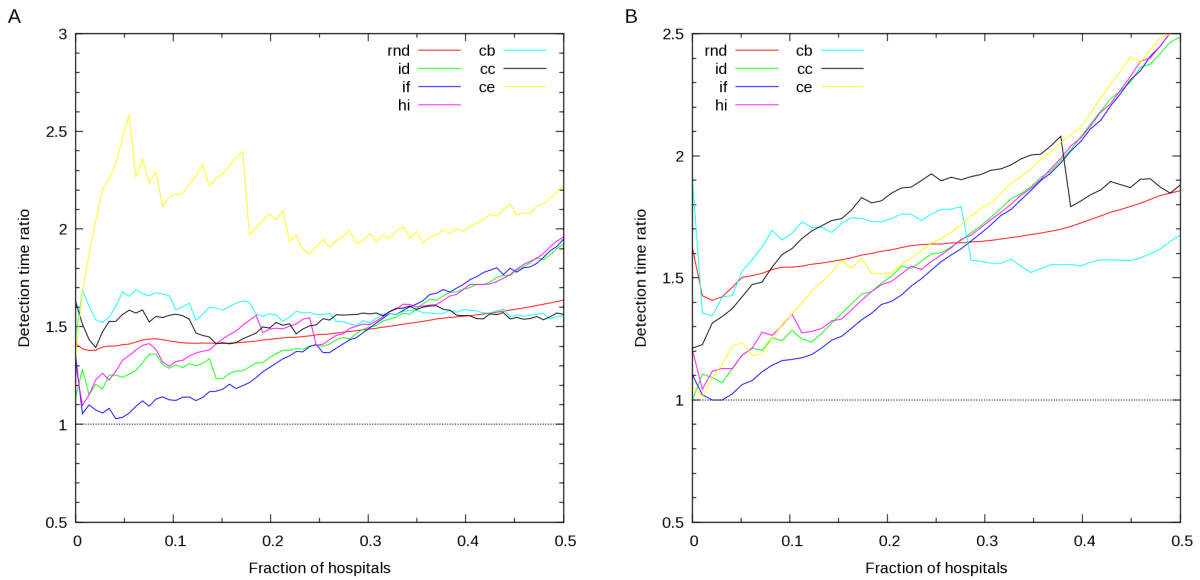

Fig. S8. Detection time ratios with heterogeneous β. Ratios of mean detection time obtained with different selection schemes to detection time obtained with the greedy algorithm in England (panel A) and The Netherlands (panel B). Hospitals have been selected randomly (rnd), and according to in-degree (id), number of admitted patients previously discharged from a different hospital (if), h-index (hi), betweenness centrality (cb), closeness centrality (cc), and eigenvector centrality (ce). The curve corresponding to random selection was calculated as the mean of 1000 random re-orderings of the hospital priority list. The per-patient transmission probability associated with each pair of hospitals has been randomly sampled as described in the Supporting Methods. As an aid to comparison, the curve of equality with the gold standard (i.e. ratio = 1.0) is plotted with a dotted line.

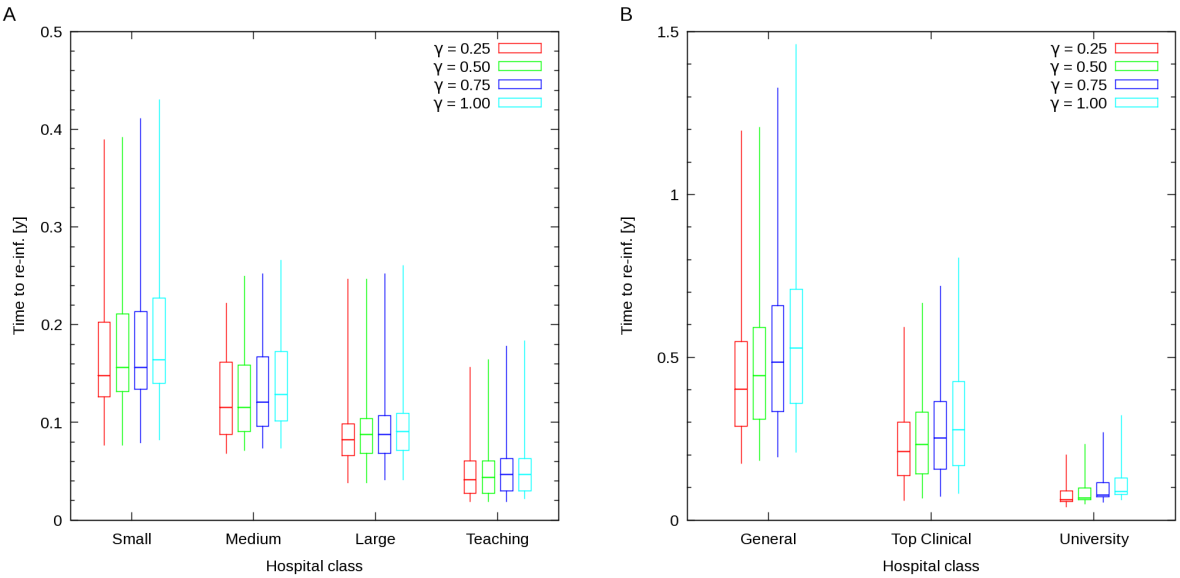

Fig. S9. Time to re-acquisition with heterogeneous β. Box plots of median HCAI-free time in the endemic regime in England(panel A) and The Netherlands (panel B). The per-patient transmission probability associated with each pair of hospitals has been randomly sampled as described in the Supporting Methods. Median times have been grouped according to type of hospital. Different colours correspond to different values of γ. Box bottom and top represent the lower and upper quartiles, respectively. Lower and upper line ends represent 5^{th} and 95^{th} percentiles, respectively. The horizontal bar corresponds to the median. Elimination rates (γ) are expressed in units of year^{-1}.

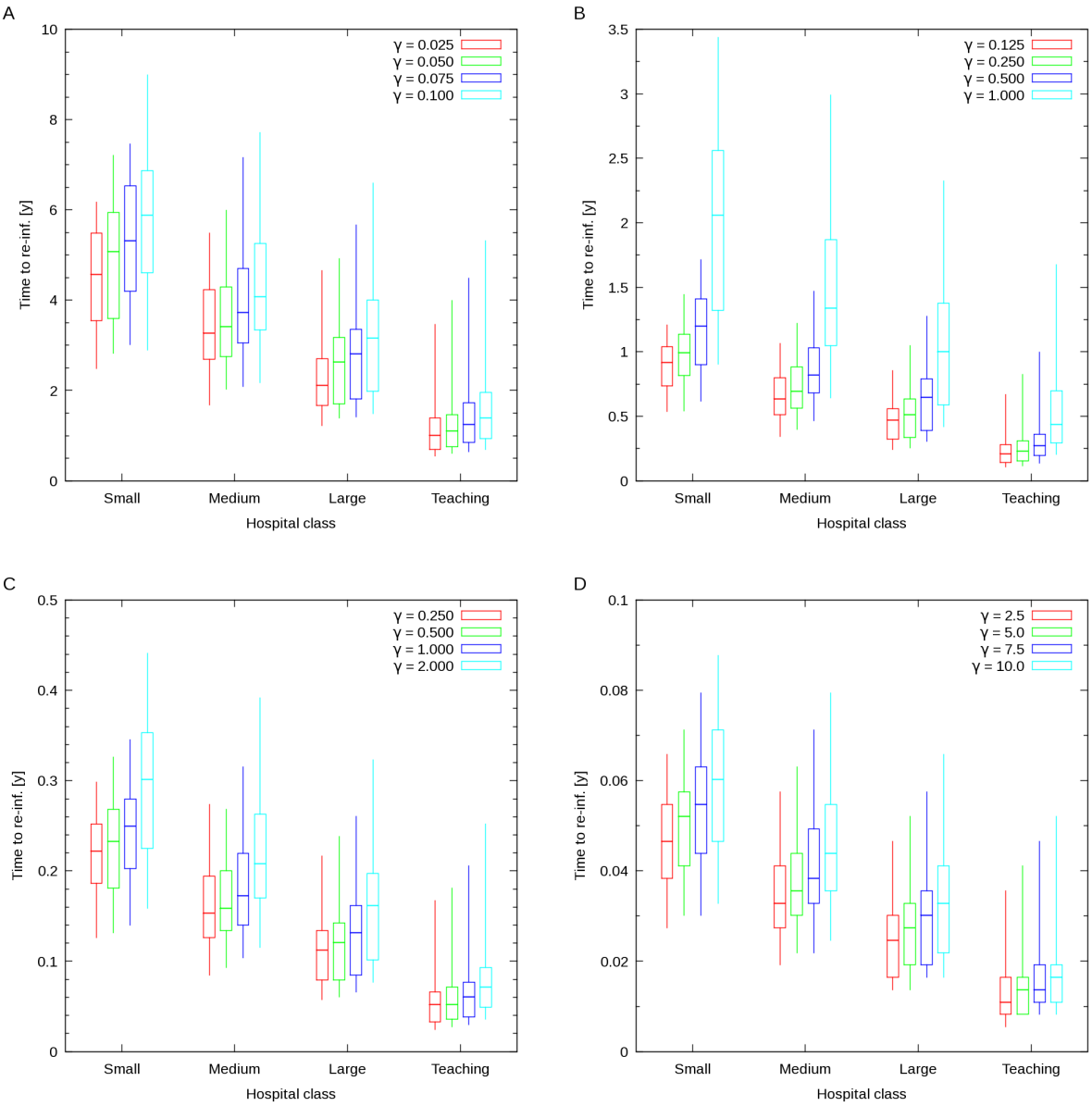

Fig. S10. Time to re-acquisition with rescaled β and γ. Box plots of median HCAI-free time in the endemic regime in England. The transmission probability β is 0.1 x (A), 0.5 x (B), 2.0 x(C) and 10.0 x (D) of the baseline value $\beta_0 = 0.001$. Median times have been grouped according to type of hospital. Different colours correspond to different values of γ. Elimination rates (γ) are expressed in units of year^{-1}.

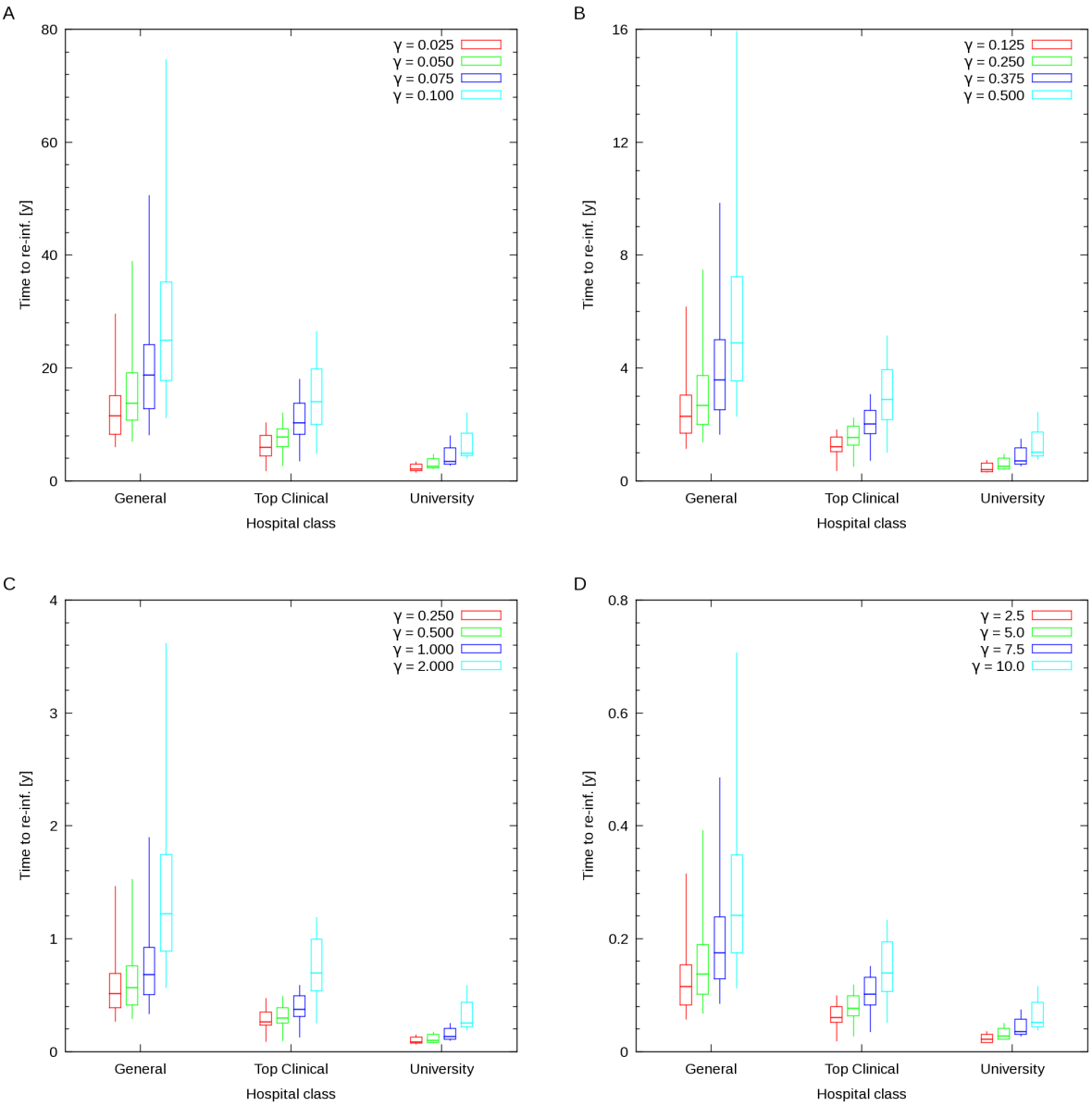

Fig. S11. Time to re-acquisition with rescaled β and γ. Box plots of median HCAI-free time in the endemic regime in The Netherlands. The transmission probability β is 0.1 x (A), 0.5 x (B), 2.0 x(C) and 10.0 x (D) of the baseline value $\beta_0 = 0.001$. Median times have been grouped according to type of hospital. Different colours correspond to different values of γ. Elimination rates (γ) are expressed in units of year^{-1}.

Supporting Tables

	England			The Netherlands*	
Selection scheme	Mean [years]	90^{th} percentile [years]	Fraction of beds	Mean [years]	90^{th} percentile[years]
Greedy	0.413	1.027	0.265	1.019	2.553
Random	0.587	1.386	0.199	1.533	3.369
In-degree	0.531	1.321	0.299	1.374	3.153
In-flux	0.494	1.252	0.310	1.286	2.970
h-index	0.585	1.458	0.294	1.360	3.121
Betweenness	0.662	1.468	0.161	1.613	3.490
Closeness	0.610	1.386	0.159	1.720	3.729
Eigenvector	0.853	1.899	0.280	1.411	3.219

*Hospital size data were not available in the case of The Netherlands.

Table S1. Mean detection time, and 90^{th} percentile, when 20% of all hospitals are recruited as surveillance sentinels. Results were obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$). Each row corresponds to a different sentinel selection scheme, namely greedy algorithm with detection time optimization, random selection, and prioritisation according to network connectivity metrics: in-degree, in-flux, h-index, betweenness centrality, closeness centrality, and eigenvector centrality. The values corresponding to random selection were calculated as the average of mean and 90^{th} percentile detection time over 1000 random re-orderings of the hospital priority list. For England, column 4 shows the fraction of available hospital beds (out of the national total) that belong to the sentinel hospitals recruited with each selection scheme.

	England			The Netherlands*	
Selecion scheme	Mean [n.a.h.]	90^{th} percentile [n.a.h.]	Fraction of beds	Mean [n.a.h.]	90^{th} percentile [n.a.h.]
Greedy	2.445	4.00	0.288	2.131	3.00
Random	4.837	10.36	0.199	4.703	10.06
In-degree	3.399	7.00	0.299	2.406	4.00
In-flux	2.729	5.00	0.310	2.185	3.00
h-index	4.246	10.00	0.294	2.363	4.00
Betweenness	7.949	19.00	0.161	5.487	11.00
Closeness	6.162	14.00	0.159	5.596	12.00
Eigenvector	9.745	24.00	0.280	2.520	4.00

*Hospital size data were not available in the case of The Netherlands.

Table S2. Mean number of affected hospitals (n.a.h), and 90^{th} percentile, when 20% of all hospitals are recruited as surveillance sentinels. Results were obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$). Each row corresponds to a different sentinel selection scheme, namely greedy algorithm with optimization of number of affected hospitals at detection time, random selection, and prioritisation according to network connectivity metrics: in-degree, in-flux, h-index, betweenness centrality, closeness centrality, and eigenvector centrality. The values corresponding to random selection were calculated as the average of mean and 90^{th} percentile detection time over 1000 random re-orderings of the hospital priority list. For England, column 4 shows the fraction of available hospital beds (out of the national total) that belong to the sentinel hospitals recruited with each selection scheme.

	England			The Netherlands*	
Selection scheme	Mean [years]	90^{th} percentile [years]	Fraction of beds	Mean [years]	90^{th} percentile [years]
Greedy	0.645	1.611	0.279	1.770	4.430
Random	0.922	2.177	0.199	2.622	5.853
In-degree	0.828	2.055	0.299	2.332	5.488
In-flux	0.783	1.973	0.310	2.204	5.192
h-index	0.911	2.266	0.294	2.317	5.471
Betweenness	1.050	2.351	0.161	2.746	6.063
Closeness	0.971	2.225	0.159	2.905	6.488
Eigenvector	1.336	2.964	0.280	2.399	5.595

*Hospital size data were not available in the case of The Netherlands.

Table S3. Mean detection time, and 90^{th} percentile, when 20% of all hospitals are recruited as surveillance sentinels. Results were obtained in the baseline scenario ($\beta = 0.001$ and $\gamma = 0$), introducing a cut-off of time to subsequent admission of $T_r = 30$ days. Each row corresponds to a different sentinel selection scheme, namely greedy algorithm with detection time optimization, random selection, and prioritisation according to network connectivity metrics: in-degree, in-flux, h-index, betweenness centrality, closeness centrality, and eigenvector centrality. The values corresponding to random selection were calculated as the average of mean and 90^{th} percentile detection time over 1000 random re-orderings of the hospital priority list. For England, column 4 shows the fraction of available hospital beds (out of the national total) that belong to the sentinel hospitals recruited with each selection scheme.

	England			The Netherlands*	
Selection scheme	Mean [years]	90^{th} percentile [years]	Fraction of beds	Mean [years]	90^{th} percentile [years]
Greedy	0.171	0.427	0.249	0.404	1.093
Random	0.243	0.580	0.199	0.649	1.402
In-degree	0.225	0.559	0.299	0.598	1.356
In-flux	0.217	0.553	0.310	0.562	1.288
h-index	0.251	0.636	0.294	0.595	1.351
Betweenness	0.265	0.595	0.161	0.699	1.488
Closeness	0.256	0.600	0.159	0.734	1.556
Eigenvector	0.349	0.786	0.280	0.613	1.384

*Hospital size data were not available in the case of The Netherlands.

Table S4. Mean detection time, and 90^{th} percentile, when 20% of all hospitals are recruited as surveillance sentinels. Results were obtained with unconstrained transmission ($\gamma = 0$), and per-patient transmission probability associated with each pair of hospitals randomly sampled as described in the Supporting Methods. Each row corresponds to a different sentinel selection scheme, namely greedy algorithm with detection time optimization, random selection, and prioritisation according to network connectivity metrics: in-degree, in-flux, h-index, betweenness centrality, closeness centrality, and eigenvector centrality. The values corresponding to random selection were calculated as the average of mean and 90^{th} percentile detection time over 1000 random re-orderings of the hospital priority list. For England, column 4 shows the fraction of available hospital beds (out of the national total) that belong to the sentinel hospitals recruited with each selection scheme.

Supporting Equations

In the emergent pathogen model, the probability of a susceptible hospital i becoming affected by the pathogen during one time step is

$$\mathcal{P}_i^{S\to I}(t) = 1 - \prod_{j:\ H_j=I} (1-\beta)^{w_{ij}\,\delta t}\,. \tag{S1}$$

The expression $(j:\ H_j = I)$ under the product symbol indicates that the product should be calculated over all hospitals j in which the pathogen is present at time t. w_{ij} is the patient movement rate from hospital j to hospital i, as given by the movement matrices. β is the probability of a moved patient being infected or colonised when discharged from hospital j, remaining infectious until the next admission, and successfully introducing the pathogen in hospital i. δt is the time-step size (1 day).

In the SIS model of endemically established pathogens, the probability of an affected hospital recovering a susceptible status during one time step, following the successful implementation of infection control measures, is

$$\mathcal{P}_i^{I\to S}(t) = 1 - \exp\left(-\gamma\delta t\right)\,. \tag{S2}$$

γ is the elimination rate, and $1/\gamma$ is the average time during which a hospital remains affected by the pathogen.

Note that whereas the probability of a hospital recovering a susceptible status is constant, the probability of a hospital becoming affected by the nosocomial micro-organism is time dependent, and is unique to each individual hospital.

Supporting Methods

In this section we assess the suitability of our sentinel selection scheme for the surveillance of different nosocomial pathogens by studying the impact of variations in model assumptions and parameters. First, we explore the effect of a reduction in the per-patient transmission probability associated with patients that remain in the community for extended periods of time between subsequent hospital admissions. Next, we evaluate the impact of heterogeneous variations in transmission probabilities, including the case of β positively correlated with hopital category, relaxing the assumption of homogeneous β. Lastly, we assess the effect of homogeneous variations in both the transmission probability and hospital elimination rate.

In the baseline model it is assumed that the per-patient probability of successfully introducing the pathogen during a subsequent hospital admission (β) does not depend on the discharge-admission interval. However, some pathogens may only transiently colonise patients reducing the period of infectiousness. To simulate this effect, we have introduced a cut-off readmission time T_r, such that β becomes 0 after a discharge-admission interval of length T_r. For $T_r > 180$ days, almost no impact is observed on all model results. This can be easily explained by noting that 93-94% of patients spent less than half a year in the community before being readmitted to hospital (cf. Fig. S1). Further decreasing the cut-off T_r slows the dissemination rate, increasing overall detection times. However, our main results remain unaffected. This can be seen in further detail in Fig. S4, S5, S6, and Table S3, which were obtained in the baseline configuration ($\beta = 0.001$, $\gamma = 0$) together with a cut-off $T_r = 30$ days.

To study how the introduction of non-homogeneous transmission probabilities affects our results, we employ a scenario in which we associate a different transmission probability to each pair of connected hospitals in the healthcare network. These transmission probabilities are randomly sampled, with uniform probability distribution, from the range $0.2\,\beta_0$ – $5.0\,\beta_0$. We sample these values only once, using the same set of probabilities in all simulation replicates.

Results of repeating the baseline analyses when employing heterogeneous transmission probabilities are shown in Fig. S7, S8, S9, and Table S4. Detection times are increased, but the ratio of different selection schemes to our gold standard do not show much variation. As in the baseline scenario, this is the best performing method, and selecting sentinels according to number of patients admitted after a previous discharge from other hospital is the next best choice.

To evaluate the impact of a positive correlation between probability of transmission and hospital category, we repeat our analysis associating higher β with patient movements from/to tertiary hospitals.

In particular, starting from the baseline scenario we run our model in two additional configurations, scaling up transmission probability associated with these referrals as $\beta = 2\,\beta_0$ and $\beta = 5\,\beta_0$. We then compute the difference in detection times obtained with the greedy algorithm (detection time optimization) and random selection when 20% of hospitals are recruited as sentinels. This time difference is a measure of the efficiency gain of our approach. We observe that in the English hospital network the introduced correlations increase the difference in detection time by 6–16%. No significant change in efficiency gain was observed in the Dutch healthcare network.

To study the effect of simple homogeneous variations in the model parameters we scaled the probability of transmission by a constant factor, repeating our analyses with β equal to 0.1 x, 0.5 x, 2.0 x, and 10 x the baseline value β_0.

In the emergent pathogen model, scaling the transmission probability by a constant factor yields scaled detection times, with only a small impact on detection times ratios. For example, increasing β by a factor of 2.0 we observe that mean detection times decrease by 50%. Conversely, decreasing β by a factor of 2.0 doubles mean detection times. The gold-standard method remains the best performing selection scheme, and ratios of detection times are unaffected.

Results of the endemic pathogen model with re-scaled β and γ are shown in Fig. S10 and S11. Re-infection times associated with tertiary hospitals are, also in this case, the lowest among the different hospital types.

In conclusion, the above analyses show that our results are robust under changes in model parameters and assumptions, despite the expected impact on rates of pathogen spread and absolute detection times. In the surveillance of nosocomial pathogens that predominantly spread through inter-hospital patient transfers, our gold-standard computational approach to sentinel selection is the best performing method. Moreover, detection times improve rapidly as up to 20% hospitals are included in the surveillance system, but much less rapidly thereafter. A near optimal solution is to prioritise hospitals simply on the basis of the number of admitted patients that have previously been discharged from a different hospital, which corresponds well to prioritising tertiary hospitals.