A passive monitoring tool using hospital administrative data enables earlier specific detection of healthcare-acquired infections

Jeffrey Rewley, PhD[a,b,c], Laura Koehly, PhD[a], Christopher Steven Marcum, PhD[a], Felix Reed-Tsochas, PhD[b]

[a] *National Human Genome Research Institute, National Institutes of Health*
[b] *CABDyN Complexity Centre, Saïd Business School, University of Oxford*
[c] *Corresponding author: jeffrey.rewley@pennmedicine.upenn.edu*

Abstract

Background: Healthcare-associated infections impose a significant burden on the health care system. Current methods for detecting these infections are constrained by combinations of high cost, long processing times, and imperfect accuracy, reducing their effectiveness.

Methods: We examine whether the quantity of time a patient spends in a ward with other patients clinically-suspected of infection, which we call co-presence, can be used as a tool to predict subsequent healthcare-associated infection. Compared to contact tracing, this leverages passively-collected electronic data rather than manually-collected data, allowing for improved monitoring. We abstracted all 133,304 inpatient records between 2011 and 2015 from a healthcare system in the UK. We calculate the AUROC for each of five pathogens based on co-presence time, the sensitivity and specificity for the test, and how much earlier co-presence would have predicted infection for the true positives.

Findings: Across the five pathogens, AUROC ranged from 0.92 to 0.99, and was 0.52 for the negative control. Optimal cut-points of co-presence ranged from 25 to 59 hours, and would have led to detection of true positives up to an average of one day earlier.

Interpretation: These findings show that co-presence time would help predict healthcare-acquired infection, and would do so earlier than the current standard of care. Using this measure prospectively in hospitals based on real-time data could limit the consequences of infection, both by being able to treat individual infected patients earlier, and by preventing potential secondary infections stemming from the original infected patient.

Funding: This research was funded via NHGRI, NIH (ZIA HG200335) and the Oxford Martin School, University of Oxford (LC1213-006).

Keywords:
Big data, proximity, contact tracing, co-presence, electronic medical records

---

Research in Context

Evidence Before this Study

In searching the literature via PubMed and Google scholar for studies including terms such as "contact tracing", "ward co-location", "predicting nosocomial infection", "administrative data", "electronic medical records", "outbreak", "healthcare-acquired infection", "healthcare-associated infection", and "nosocomial", we found that there was significant evidence for an increased risk of infection pursuant to spatial proximity with an infected individual. We also found that this knowledge has been utilized successfully to limit spread in the health care setting.

Added Value of this Study

This study probes deeper into the relationship between spatial proximity with persons suspected of infection and subsequent risk of infection. We find that there is nuance in the exact amount of time spent around those suspected of infection and infection risk, and that it differs based on the contagion in question. Importantly, we were able to do so primarily using hospital administrative data. Electronic medical record information commonly used for prediction models was not necessary for strong prediction here.

Implications of all the available evidence

Hospital administrative data should be enabled to passively monitor co-presence with suspected infected patients, and subsequently used to recommend increased screening of those crossing the threshold. Studies should be conducted to evaluate whether this does indeed leads to improved individual outcomes and reduces the size of health care outbreaks.

1. Introduction

Healthcare-associated infections (HCAI) are a burden on the health care system. Despite advancing medical technology and standards of care, the attributable costs of each case of HCAI across the globe continue to range from $2,992 to $29,000 (about £2,367 to £22,945) [1, 2, 3]. In the UK, each patient with an HCAI costs an additional £3,154, for an estimated total of £930.62 million per year (roughly $1.2B) [4]. As well, HCAIs adversely impact patient health outcomes such as length of hospital stay [5] and mortality [6] directly.

In this paper we focus on identifying infections earlier and more accurately. Doing so requires timely diagnostic tests, but current diagnostics (both those which measure biomarkers or the microbiological vector directly) are limited by some combination of long processing times, high costs, lack of specificity to pathogen, or imperfect accuracy [7, 8, 9, 10]. These limitations make it more difficult to control infectious outbreaks.

Non-biological methods exist for timely and accurate detection of secondary infections. A wide variety of factors predict subsequent infection likelihood [11]. Being

around an infected person increases one's risk of infection for many pathogens. In a hospital setting, Murray et al. [12] found that after an infected individual occupies a hospital ward, the risk of infection for subsequent occupants of the ward is significantly increased for the next 24 hours [12].

This knowledge partly informs contact tracing, which has been used in the past to find those most at risk of acquiring an infection. Contact tracing identifies those persons who have come in close physical proximity (co-present) with infected persons [13**?**]. Co-presence has been used in previous outbreaks in health care settings to trace a pathogen's source, and the subsequent path by which it spreads through the hospital population, and has been used in the recent COVID-19 outbreak [14, 15]. Although contact tracing functions on the knowledge that co-presence with an infected individual is a risk factor for infection, full information regarding the amount of co-presence and its association with infection risk has not been thoroughly examined. However, contact tracing is typically binary in nature, reflecting only whether or not a threshold of co-presence has been crossed. This threshold is often based either on whether *any co-presence* has occurred vs. none, or, if there is a non-zero threshold, it is based on simulation rather than empirical data [16]. As well, contact tracing relies on positively identifying a case and then retroactively identifying individuals who had crossed the threshold with the infected case, but this makes it of limited use in identifying the earliest-infected patients [17]. Recent work has shown that traditional contact tracing may be inadequate to fully contain outbreaks, instead requiring digital contact tracing and the use of mobile phone proximity [18]. Finally, the *patterns* of co-presence rather than the simple threshold may be more strongly

related to infection risk. The increasing ubiquity of hospital administrative data (HAD) allows co-presence to be passively monitored and quantified. When retrospectively asking infected persons to recollect their interpersonal interactions, many biases present when recalling quantitative amounts, rather than binary interactions [19]. Passively-collected HAD are not impacted by these same biases. Co-presence has been used in other hospital settings to predict mortality and readmission risk [20, 21]. In addition to formally testing how strongly co-presence predicts subsequent HCAI, HAD has the advantages of providing near-real-time information that can be used to identify those at risk of infection, ameliorating many of the limitations of standard contact tracing. Once the infrastructure is in place, the costs are minimal, both in terms of time and financial resources.

In this paper, we therefore show that co-presence measured by HAD functions as an effective monitoring tool for HCAI risk. Specifically, we use the amount of time a patient is in the same bay of a hospital ward as patients who are clinically-suspected of an infection, defined as either a diagnosis or ordering of a microbiological test (irrespective of whether the eventual result is positive or negative), whichever comes first. Using clinically-suspected infections, rather than confirmed infections, means this tool can potentially detect pathogens earlier than methods which require confirmed infections, such as contact tracing. Moreover, clinically-suspected infections reflect the information that would prospectively be available in real-time in a clinical setting, increasing the potential benefits of this tool. We will refer to this co-presence time as the "index test" [22]. Co-presence with individuals with clinically-suspected infections would ideally be used for surveillance. As many hospitals already have

administrative data and electronic medical records (EMR) that could be monitored for patient-patient co-presence, implementing this tool would likely be inexpensive and efficient. The clinical role of the index test would be for screening; a result indicating high risk of infection would lead to additional tests or increased monitoring of those patients during the incubation period of the vector.

2. Methods

2.1. Study design and population

The study population comprised all 133,304 patients with NHS hospital stays of at least 48 hours in a single county in the UK from 1 January, 2011 to 1 January, 2015. This NHS trust comprises multiple research hospitals and nine smaller community hospitals, covering a catchment area with approximately 688,000 inhabitants. The research hospitals are largely complementary rather than duplicative with respect to the specialties present.

We subset to 48 hour stays because we define a HCAI as one occurring more than 48 hours after a patient enters the hospital [23]. These patients comprise a consecutive series, where the index test was assessed retrospectively. We assess the index test on the following pathogens: MRSA, *Escheria coli*, *Pseudomonas aeruginosa*, *Clostridium difficile*, and norovirus. It is important to note that the HCAI itself is a syndrome - a cluster of symptoms stemming from a high level of a pathogen in a patient - rather than merely the presence of the pathogen. Here, due to the available data, we generally treat the presence of the pathogen and syndrome as equivalent [24]. The main exception is that we were able to exclude ICD-10 codes indicative of

infections not transmitted interpersonally (see below).

This study follows STARD guidelines for the reporting of a new diagnostic [22]. Ethics committee approval was gained from the University of Oxford IRB, and the database has been approved by the South Central Research Ethics Committee (19/SC/0403) and the Confidentiality Advisory Group of the Health Research Authority (19CAG0144).

2.2. Reference and index test methods

The reference test was either a diagnosis of the infection in question based on the EMR, or a positive microbiological test based on the test used by the NHS at the time for the disease in question, which was pre-specified. Diagnoses were abstracted using ICD-10 codes. Codes indicative of spread not due to other patients were excluded (e.g. "A04.4" - "Other intestinal E. coli infection") as were codes indicative of sepsis to avoid conflation in a subsequent sensitivity analysis (e.g. "A45.51" - "Sepsis due to E. coli"). The hospital administrative data was not configured such that physicians or lab technicians could examine co-presence, and as such the index test results were not available to the performers of the reference test. There were no missing or indeterminate reference test results, and the data do not contain any reference to adverse events due to the reference test.

The index test was the number of hours a patient spent co-present with patients having a clinically-suspected infection, defined as either a diagnosis or ordering of a microbiological test, whichever comes first. Importantly, this includes ordered tests whether they eventually returned positive or negative as both constituted the *suspicion* of infection. Co-presence time was defined as the length of time both patients

were in the same hospital ward bay. The clinical information and microbiological test results were available to the researchers evaluating co-presence time, as it was all contained in the electronic medical record. There were no missing or indeterminate co-presence time information due to the administrative nature of the data; every patient's co-presence was precisely quantifiable. As co-presence time was calculated entirely *in silico*, there were no adverse events due to conducting the index test.

The index test of time co-present was based on tested *or* diagnosed patients rather than patients with confirmed infection to more accurately reproduce the knowledge that would be available were this a prospective study. Because co-presence time was assessed retrospectively, the data would allow us to perfectly calculate the hours of co-presence at the time of observation, but this scenario would not occur in practice. Tests for the presence of pathogens take time to return results, particularly microbiological cultures, where previous studies have shown the time for optimal results is five days [25]. Therefore, we only used the subset of real-time information that would be available in practice.

The total co-presence time was tallied for all individuals at a single point in their hospital stay. The time at which the quantity of co-presence was measured was determined in the following manner: for patients with a confirmed infection, co-presence time was quantified up to the moment when the microbial test was collected or the diagnosis recorded in the EMR, whichever came first. For patients who were neither tested nor diagnosed, no corresponding time existed. For these patients, their time of assessment was chosen randomly from their hospital stay such that the distribution of times matched the distribution of times for those who had

a confirmed infection. This ensured that no systematic differences existed between those with positive or negative reference tests.

Finally, the exact infectious period of infected patients was unknown, but the potential to transmit an infection from patient-to-patient only exists during the infectious period of a focal infected person. Therefore, we applied deterministic infection periods to model the infectious period [26]. We assumed that the time at which the microbial test was collected or the diagnosis recorded in the EMR (i.e. when a focal patient's co-presence was measured) was the midpoint of their infectious period, with the length of their infectious period equal to literature values for the mode of the infectious period length (Table 1) [27, 28, 29, 30, 31]. Importantly, these models are based on interpersonal infection modes, and do not allow for interpersonal infection (e.g. carriage of a pathogen progressing to an infection). This aligns with our removal of ICD-10 codes likely not pertaining to HCAI.

2.3. Analysis

To compare the measures of diagnostic accuracy, we use the Receiver-Operator Characteristic (ROC) curve and the area under the receiver-operator curve (AUROC). We calculated the 95% confidence intervals of the AUROC using bootstrapping. To determine the optimal cut-point for these curves, we assume that the clinical costs of false positives, false negatives, true positives and true negatives are all equal. Following this, the optimal cut-point is the point closest to a perfect test (100% sensitivity and specificity) in Euclidian space. We assess the sensitivity and specificity at optimal cut-points.

For this tool to benefit clinical practice, it would need to both predict infections

as well as or better than other methods, and also do so *earlier*. To assess this we calculated the number of hours between the time a true positive patient's microbiological test was administered, and when they first crossed the cut-point of co-presence time during their hospital stay. This represents how many hours earlier the pathogen could be detected if co-presence time were used as a screening test, relative to the current standard operating procedure.

Finally, we conduct a number of sensitivity analyses, which are outlined here. For more information, see the Supplemental Information (SI). The results may be due to class imbalance between infected and uninfected individuals. We remove patients in two ways to balance the classes and rerun the analyses. The strong predictive power of the tool may be due to latent characteristics of patients in the same hospital ward; we conduct the analysis using sepsis from *S. aureus* as a negative control because it is relatively unlikely to be spread horizontally from patient-to-patient [32]. To determine if the results generalize to hospitals with only ward-level co-presence data (rather than bay), we rerun the analysis at the ward level.

2.4. Patient and public involvement

Because this was a retrospective study analyzing EMR, we did not involve our study population in the study.

3. Results

The patients included in the study, their reasons for exclusion, and their reference tests can be seen in Figure 1. We observe that most patients were excluded due to

being outpatients or being tested for microbiological vectors within the first 48 hours of entering the hospital.

This left the study population of patients who could potentially contract an HCAI. The demographics of these patients are shown in Table 2. Patients were on average 56 years old, and 45% were male. On average, these patients spent 13 days in the hospital, and 5.40% of them died while in the hospital. In total, 4326 (3.24%) patients were infected with one of the five pathogens studied.

After applying the index test to this set of patients, we observed distinct distributions of co-presence time with patients clinically-suspected of infection for those whose reference test was negative compared to those whose reference test was positive (Figure 2). Irrespective of a specific cut-point, the distributions of co-presence times were strongly differentiated based on whether or not a patient had a positive reference test (positive test or diagnosis). For all five pathogens, patients with a negative reference test had a distribution of co-presence times (index test results) much lower than for patients with a positive reference test.

To quantify the performance of the index test, we created ROC curves based on multiple cutoffs of hours co-presence with infected individuals. The ROC curves have AUROCs ranging from 0.92 to 0.99 (Table 3). The optimal cut-point ranges from 29 to 59 hours depending on the pathogen in question. This means patients must spend over 24 hours co-present with infected patients before the number of false negatives is minimized. Although this dichotomizes co-presence, similarly to contact tracing, the optimal cutoff is not none vs. any. Instead it takes at least 29 hours before co-presence time is maximally discriminatory. Of note is that exposure

to multiple infected persons can increase co-presence time by more than one hour per hour of elapsed time (i.e. patients may have reached this point in less than 29 hours if around multiple infected patients).

For patients with a positive reference test and a positive index test (true positives), we examined how many hours earlier they would have been tested if the reference test was administered immediately upon crossing the threshold of co-presence (Table 3). We observe that on average, the amount of time saved ranges from 6 hours for *C. difficile* to 22 hours for *P. aeruginosa*. This means that if a microbiological test for *P. aeruginosa* were administered as soon as a patient crossed this threshold, that patient's infection may be detected as much as 22 hours earlier than it had been.

As previously mentioned, the infected and uninfected populations are highly unbalanced, which may have artificially increased the performance of co-presence time as a diagnostic. We used matching and removal of patients not at risk of infection to examine this possibility. Both analyses showed minimal effects of unbalanced samples (Figures S1 and S2). Additionally, co-presence time may have high predictive power based on latent characteristics between patients independently increasing both patients' risk. Our use of a negative control (Figure S3) shows that this is unlikely, as the tool had low predictive power in the case of sepsis, a non-communicable HCAI. Finally, an analysis restricted one ward at a time showed that co-presence time remained strong when looking at a single ward at a time (median AUC: 0.86; Table S2). However, this analysis showed that there was additional power in incorporating information about patient ward-to-ward movement, as the full analysis had an AUC of 0.96, a 10% difference.

4. Discussion

4.1. Key findings

In this paper, we have shown that the number of hours of co-presence with patients clinically-suspected of being infected is a measure that serves as a screening test of infection. Additionally, we show that if co-presence time is used as a screening test, patients' syndromes which stemmed from interpersonal transmission in the health care setting may be detected as many as 22 hours earlier, on average.

4.2. Relationship with previous studies

In this study, we used Hospital Administrative Data (HAD) to measure co-presence between patients. Previous studies have used similar data to implicitly construct the path of infection from patient-to-patient [14]. Similarly, contact tracing scrutinizes any person with *any* co-presence with a patient having a confirmed infection to effectively contain an infection [17]. Rather than dichotomize co-presence as none vs. any or at fifteen minutes as has been suggested for transmission of COVID-19 [16], we evaluate all potential thresholds with empirical data, selecting the threshold that best balances false-positives and false-negatives.

Our analyses indicate that the optimal cut-point of co-presence for all five pathogens tested is greater than 24 hours. This is important, as previous methods, such as contact tracing, typically use any contact vs none (or no co-presence vs. any) as a cutoff for potential infection [13, 17]. Our findings show that any vs. none may unnecessarily increase the number of false positives, when a more stringent criteria will limit false positives while not increasing the number of false negatives.

Murray et al. [12] used data of a similar nature to assess the increased risk of infection following occupation by an infected patient for 24 hours. Our work, however, shows that the maximal risk of interpersonal transmission is not only dependent on the amount of time that has passed, but also on the amount of *co-presence time* which has occurred; co-presence time can vary from patient to patient and from ward to ward.

4.3. Study implications

Patient-patient co-presence time will ideally be used as a screening tool; once a patient was co-present with a tested or diagnosed patient for at least the cutoff time of 29-59 hours, they would then be tested for the microbiological agent and subsequently monitored for signs of infection. We have shown that if this were done, infections could be identified earlier than current standard operating procedure. Earlier detection may lead to reduced infectious periods for patients, as treatment could be administered sooner. A recent study found that the most cost-effective method for diagnosis of *C. difficile* cost $54,500 per quality-adjusted life year (QALY) [33]. The low costs of using co-presence as a diagnostic tool would likely lead its cost per QALY to be even lower.

Further, if there were downstream effects, the values we estimate in Table 3 reflect the *minimum* amount of time that would be saved. Earlier detection of a patient's infection via co-presence time combined with subsequent tests could result in earlier quarantine for that patient, thereby preventing other patients from being infected through contact with the original patient. These downstream effects are not captured in the calculations of the person-hours of infection potentially saved, and

therefore what we show is a lower bound on the potential benefits of a tool based on co-presence time.

4.4. Strengths

As previously stated, using co-presence time as a screening tool has many advantages. First, by using HAD, we were able to study a population of more than 100,000 patients from a catchment area including almost 700,000 patients. Many hospitals already use some form of electronic medical record, so adapting these records to monitor co-presence with infected and tested individuals should carry minimal effort, allowing this method to be easily-integrated into many health systems. HAD can also be passively monitored for relevant amounts of co-presence, meaning the cost to calculate co-presence time is inexpensive once the passive monitoring is enabled. Determining co-presence time is also fast; a result is returned immediately when a patient crosses the threshold of co-presence.

These results also dovetail well with recent advances in microbiological detection; a test for MRSA was recently approved which can provide results within a few hours [34]. In tandem with the tool proposed here, a patient who crosses the threshold of co-presence can be immediately given a rapid diagnostic test, and decisions regarding their infection status can be made in rapid order. This increased pace of medical decision-making taking advantage of the large amounts of data available in the EMR and HAD answers a recent call for a "Deep Learning Healthcare" [35]. Because co-presence would potentially identify an individual at risk prior to their biomarkers and other physical symptoms of infection manifest, increased testing of these individuals may also help identify individuals who would have been asymptomatic and otherwise

gone unnoticed. This problem has been highlighted with the recent COVID-19 outbreak, where asymptomatic individuals missed by standard procedures have caused additional infections [36]. Monitoring co-presence and referring to increased surveillance would likely lead to catching a large proportion of these cases, reducing the scope of the outbreak of HCAIs.

It is worth noting that co-presence time is pathogen-specific: identified co-presence thresholds, co-presence accrued, and a positive result all reflect a single pathogen, exclusive of others. In other words, a positive result of co-presence with patients tested for *E. coli* only predicts an *E. coli infection*, but not other pathogens. This specificity is in contrast to some biomarker diagnostic tests which measure general indicators of infection, and must be used in concert with physician expertise to identify the specific pathogen [9]. All of these strengths indicate that co-presence time would supplement the microbial tests currently available.

4.5. Limitations

Because the index test is based on administrative data, there are inherent limitations that make this approach imperfect. First, this method cannot disentangle disease-specific modes of transmission, and does not perfectly capture all the methods by which pathogens can be transmitted. For example, some infections can be caused by patients' endogenous flora [37]. Although we excluded some ICD-10 codes which clearly indicated intra-personal aetiology, this was not possible in all cases. Conditional on person-to-person transmission, co-presence makes no assumption about how the pathogen got from one person to another, whether it be directly, carried by a third party, or deposited on an object and picked up by another person. This

precludes us from predicting HCAIs which do not stem from interpersonal transmission. These results therefore show the strength of association between co-presence with *other patients* and subsequent infection, and it is this association which can be leveraged for screening. Future work should investigate how modes of transmission might be incorporated into such a tool for increased disease-specific precision.

These results are based on the relatively crude measure of co-presence, and therefore likely represent the minimum strength of the association. This highlights how this tool would likely fare in hospitals with a relatively crude HAD system. Increased sophistication (e.g. bed-level information) would only increase the association between co-presence and infection, and therefore the predictive power of our approach.

Finally, this test was quantified using data from NHS hospitals in a single catchment area. Standard operating procedures for infection control exist that may make the results here non-generalizable. However, our series of robustness tests, such as removing a quarantine ward and altering the time at which a microbiological test is administered, showed that these differences would make a minimal impact of the effectiveness of this approach. Because the diagnosis of infection was based on electronic medical records rather than administrative data, generalizability of these results is likely increased (diagnostic criteria for infections are more conserved across health care systems than administrative coding)[38]. However, future work should explore whether similar results can be obtained in other settings.

5. Conclusions

In this paper, we have shown that using HAD-based co-presence time with patients clinically-suspected of infection is a strong candidate as an indicator of HCAI. Beyond the implications for spread within healthcare systems, these results suggest that co-presence in hospitals matters: patients are not truly isolated and independent from one another. Embeddedness of patients in hospital settings needs to be recognized and leveraged for better health care. On the balance of the strengths and potential caveats discussed herein, we have shown that co-presence is a powerful indicator of HCAI, which merits implementation in hospitals for the reduction of outbreaks in healthcare settings.

6. Author contributions

Study conception and design: JPL, CSM, LK

Acquisition of data: JPL, FRT

Analysis and interpretation of data: JPL, CSM, FRT, LK

Drafting of manuscript: JPL

Critical revision: JPL, CSM, FRT, LK

7. Conflicts of interest

The authors certify that they have no conflicts of interest.

8. Funding

The authors acknowledge funding from the National Human Genome Research Institute, National Institutes of Health (Grant number ZIA HG200335 to Laura Koehly) and the Oxford Martin School, University of Oxford (Grant number LC1213-006 to Felix Reed-Tsochas).

9. References

[1] L Gabriel and A Beriot-Mathiot. Hospitalization stay and costs attributable to clostridium difficile infection: a critical review. *Journal of Hospital Infection*, 88(1):12–21, 2014.

[2] William R Jarvis. Selected aspects of the socioeconomic impact of nosocomial infections: Morbidity , mortality , cost , and prevention. *Infection Control and Hospital Epidemiology*, 17(8):552–557, 1996.

[3] Kenton J Johnston, Kenneth E Thorpe, Jesse T Jacob, and David J Murphy. The incremental cost of infections associated with multidrug-resistant organisms in the inpatient hospital settinga national estimate. *Health Services Research*, 54(4):782–792, 2019.

[4] R Plowman, N Graves, MAS Griffin, JA Roberts, AV Swan, B Cookson, and L Taylor. The rate and cost of hospital-acquired infections occurring in patients admitted to selected specialties of a district general hospital in england and the national burden imposed. *Journal of Hospital Infection*, 47(3):198–209, 2001.

[5] Binila Chacko, Kurien Thomas, Thambu David, Hema Paul, Lakshmanan Jeyaseelan, and John Victor Peter. Attributable cost of a nosocomial infection in the intensive care unit: a prospective cohort study. *World Journal of Critical Care Medicine*, 6(1):79, 2017.

[6] Christophe Adrie, Maité Garrouste-Orgeas, Wafa Ibn Essaied, Carole Schwebel, Michael Darmon, Bruno Mourvillier, Stéphane Ruckly, Anne-Sylvie Dumenil, Hatem Kallel, Laurent Argaud, et al. Attributable mortality of icu-acquired bloodstream infections: impact of the source, causative micro-organism, resistance profile and antimicrobial therapy. *Journal of Infection*, 74(2):131–141, 2017.

[7] Gemma Johnson, Michael R Millar, Stuart Matthews, Margaret Skyrme, Peter Marsh, Emma Barringer, Stephen O'Hara, and Mark Wilks. Evaluation of BacLite rapid MRSA, a rapid culture based screening test for the detection of ciprofloxacin and methicillin resistant s. aureus (MRSA) from screening swabs. *BMC Microbiology*, 6(1):83, 2006.

[8] Longxiang Su, Bingchao Han, Changting Liu, Liling Liang, Zhaoxu Jiang, Jie Deng, Peng Yan, Yanhong Jia, Dan Feng, and Lixin Xie. Value of soluble trem-1, procalcitonin, and c-reactive protein serum levels as biomarkers for detecting bacteremia among sepsis patients with new fever in intensive care units: a prospective cohort study. *BMC Infectious Diseases*, 12(1):157, 2012.

[9] Liliana Simon, France Gauvin, Devendra K Amre, Patrick Saint-Louis, and Jacques Lacroix. Serum procalcitonin and c-reactive protein levels as markers

of bacterial infection: A systematic review and meta-analysis. *Clinical Infectious Diseases*, 39(2):206–217, 2004.

[10] Valeria Visconti, Grazia Brunetti, Alessandra Giordano, and Giammarco Raponi. Rt-pcr for the diagnosis of clostridium difficile infection: the final answer has yet to come. *Journal of Clinical Pathology*, pages jclinpath–2017, 2017.

[11] Michael R Britt, Charles J Schleupner, and Sego Matsumiya. Severity of underlying disease as a predictor of nosocomial infection: utility in the control of nosocomial infection. *Journal of the American Medical Association*, 239(11): 1047–1051, 1978.

[12] Sara G Murray, Joanne WL Yim, Rhiannon Croci, Alvin Rajkomar, Gabriela Schmajuk, Raman Khanna, and Russell J Cucina. Using spatial and temporal mapping to identify nosocomial disease transmission of clostridium difficile. *JAMA internal medicine*, 177(12):1863–1865, 2017.

[13] N Bagdasarian, HC Chan, S Ang, MS Isa, SM Chan, and DA Fisher. A stone in the pond approach to contact tracing: Responding to a large-scale, nosocomial tuberculosis exposure in a moderate tb-burden setting. *Infection Control & Hospital Epidemiology*, pages 1–3, 2017.

[14] Evan S Snitkin, Adrian M Zelazny, Pamela J Thomas, Frida Stock, David K Henderson, Tara N Palmore, and Julia a Segre. Tracking a hospital outbreak of carbapenem-resistant klebsiella pneumoniae with whole-genome sequencing. *Science Translational Medicine*, 4(148):148ra116, aug 2012.

[15] Matt J Keeling, T Deirdre Hollingsworth, and Jonathan M Read. The efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *medRxiv*, 2020.

[16] European Centre for Disease Prevention and Control. Contact tracing: Public health management of persons having had contact with cases of novel coronavirus in the european union. Stockholm: ECDC, 2020.

[17] Centers for Disease Control, Prevention, et al. CDC methods for implementing and managing contact tracing for ebola virus disease in less-affected countries. Available at `https://stacks.cdc.gov/view/cdc/26492`, 2017. Accessed: 2019-09-30.

[18] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491), 2020.

[19] Andrew Bell, Patrick Ward, Md Ehsanul Haque Tamal, and Mary Killilea. Assessing recall bias and measurement error in high-frequency social data collection for human-environment research. *Population and Environment*, 40(3):325–345, 2019.

[20] Jeffrey Lienert, Christopher Steven Marcum, John M Finney, Felix Reed-Tsochas, and Laura M Koehly. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. *Network Science*, 2017.

[21] Jeffrey Lienert, Felix Reed-Tsochas, Laura Koehly, and Christopher Steven Marcum. Using hospital administrative data to infer patient-patient contact via the consistent co-presence algorithm. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2756–2762. IEEE, 2019.

[22] Patrick M Bossuyt, Johannes B Reitsma, David E Bruns, Constantine A Gatsonis, Paul P Glasziou, Les Irwig, Jeroen G Lijmer, David Moher, Drummond Rennie, Henrica C W de Vet, Herbert Y Kressel, Nader Rifai, Robert M Golub, Douglas G Altman, Lotty Hooft, Daniël A Korevaar, and Jérémie F Cohen. Stard 2015: an updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal*, 351, 2015. doi: 10.1136/bmj.h5527. URL `https://www.bmj.com/content/351/bmj.h5527`.

[23] Teresa C Horan, Mary Andrus, and Margaret A Dudeck. CDC/NHSN surveillance definition of health care–associated infection and criteria for specific types of infections in the acute care setting. *American Journal of Infection Control*, 36(5):309–332, 2008.

[24] Hassan Ahmed Khan, Aftab Ahmad, and Riffat Mehboob. Nosocomial infections and their control strategies. *Asian Pacific Journal of Tropical Biomedicine*, 5 (7):509–514, 2015.

[25] Nita Pal, Rajni Sharma, Suman Rishi, and Leela Vyas. Optimum time to detection of bacteria and yeast species with BACTEC-9120 culture system from blood and sterile body fluids. *Journal of Laboratory Physicians*, 1(2):69–72, 2009.

[26] Roy M Anderson. *The population dynamics of infectious diseases: theory and applications*. Springer Science, 1982.

[27] Public Health Agency of Canada. Pseudomonas spp. pathogen safety data sheet. Available at `http://www.phac--aspc.gc.ca/lab--bio/res/psds--ftss/p`, 2012. Accessed: 2019-09-30.

[28] CB Dalton, ED Mintz, JG Wells, CA Bopp, and RV Tauxe. Outbreaks of enterotoxigenic escherichia coli infection in american adults : a clinical and epidemiologic profile. *Epidemiology and Infection*, 123(May):9–16, 1999.

[29] Public Health Agency of Canada. Staphylococcus aureus pathogen safety data sheet. Available at `http://www.phac--aspc.gc.ca/lab--bio/res/psds--ftss/s`, 2012. Accessed: 2019-09-30.

[30] Stuart H Cohen, Dale N Gerding, Stuart Johnson, Ciaran P Kelly, Vivian G Loo, L Clifford McDonald, Jacques Pepin, and Mark H Wilcox. Clinical practice guidelines for clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of america (shea) and the infectious diseases society of america (idsa). *Infection Control & Hospital Epidemiology*, 31(5):431–455, 2010.

[31] Rapid detection of a norovirus pseudo-outbreak by using real-time sequence based information. *Journal of Clinical Virology*, 58(1):245–8, sep 2013. URL `http://www.ncbi.nlm.nih.gov/pubmed/23880160`.

[32] Stéphane Hugonnet, Hugo Sax, Philippe Eggimann, Jean-Claude Chevrolet, and Didier Pittet. Nosocomial bloodstream infection and clinical sepsis. *Emerging Infectious Diseases*, 10(1):76, 2004.

[33] Si Xuan, Kenneth M Zangwill, Weiyi Ni, Junjie Ma, and Joel W Hay. Cost-effectiveness analysis of four common diagnostic methods for clostridioides difficile infection. *Journal of General Internal Medicine*, pages 1–9.

[34] Rebecca Voelker. New test detects mrsa in hours rather than days. *Journal of the American Medical Association*, 323(3):210–210, 2020.

[35] Beau Norgeot, Benjamin S Glicksberg, and Atul J Butte. A call for deep-learning healthcare. *Nature Medicine*, 25(1):14–15, 2019.

[36] Roy M Anderson, H Heesterbeek, D Klinkenberg, and T Dierdre Hollingsworth. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 2020.

[37] Dennis M Flynn, Robert A Weinstein, Catherine Nathan, Michael A Gaston, and Sherwin A Kabins. Patients' endogenous flora as the source of nosocomial enterobacter in cardiac surgery. *Journal of Infectious Diseases*, 156(2):363–368, 1987.

[38] Eamon P Raith, Andrew A Udy, Michael Bailey, Steven McGloughlin, Christopher MacIsaac, Rinaldo Bellomo, and David V Pilcher. Prognostic accuracy of the sofa score, sirs criteria, and qsofa score for in-hospital mortality among

adults with suspected infection admitted to the intensive care unit. *Journal of the American Medical Association*, 317(3):290–300, 2017.

10. Tables

Pathogen	Mode Infectious period (hours)	Infectious period (range)
MRSA [27]	72	48-96
E. coli [28]	120	80-160
P. aeruginosa [29]	48	12-72
C. difficile [30]	60	24-96
Norovirus [31]	44	12-72

Table 1: Values for the infectious period for each pathogen used. Microbiological tests were assumed to occur at the mode of the infectious period.

Variable **Mean (SD) or N (%)**	**Mean (SD) or N (%)**
Age (years)	56.4 (27.8)
Sex (male)	59,988 (44.80%)
Stay length (hours)	319.5 (568.8)
Died in hospital	7,180 (5.40%)
Infected with MRSA	474 (0.36%)
Infected with *E. coli**	2,594 (1.95%)
Infected with *P. aeruginosa*	1,109 (0.83%)
Infected with *C. difficle*	133 (0.10%)
Infected with norovirus	16 (0.01%)

Table 2: Baseline demographics and clinical characteristics of patients based on the set of patients who received both a reference and an index test for the pathogen in question.

Pathogen	AUC (95% CI)	Threshold (hours)	Sensitivity	Specificity	True positives	PPV	Average hours saved per patient (range)
MRSA	0.962 (0.96,0.964)	35.00	1.00	0.95	474	0.067	10.93 (7.05,16.18)
E. coli	0.966 (0.965,0.967)	59.00	0.95	0.90	2,472	0.159	8.35 (4.91,13.61)
P. aeruginosa	0.925 (0.923,0.927)	35.00	1.00	0.95	1,107	0.142	21.97 (13.98,32.96)
C. difficile	0.993 (0.992,0.994)	29.00	1.00	0.99	133	0.091	6.36 (4.08,10.14)
Norovirus	1 (1,1)	34.00	1.00	1.00	16	1.00	8.19 (5.12,10.31)

Table 3: Index test statistics for all five pathogens. The threshold, or optimal cut-point, for each test was the number of hours of co-presence that gave sensitivities and specificities which were closest in Euclidian space to the optimal test. Sensitivities, specificities, and the number of true positives were taken at these optimal cut-points. Finally, hours saved is the difference in time between when a patient first crosses the threshold of the index test and when they were actually tested for or diagnosed with the infection. This number then represents how much earlier a patient may be screened for infection when using the index test than when using the reference test. Ranges indicate the minimum and maximum numbers when infectious period lengths were stochastic rather than deterministic.

11. Figures

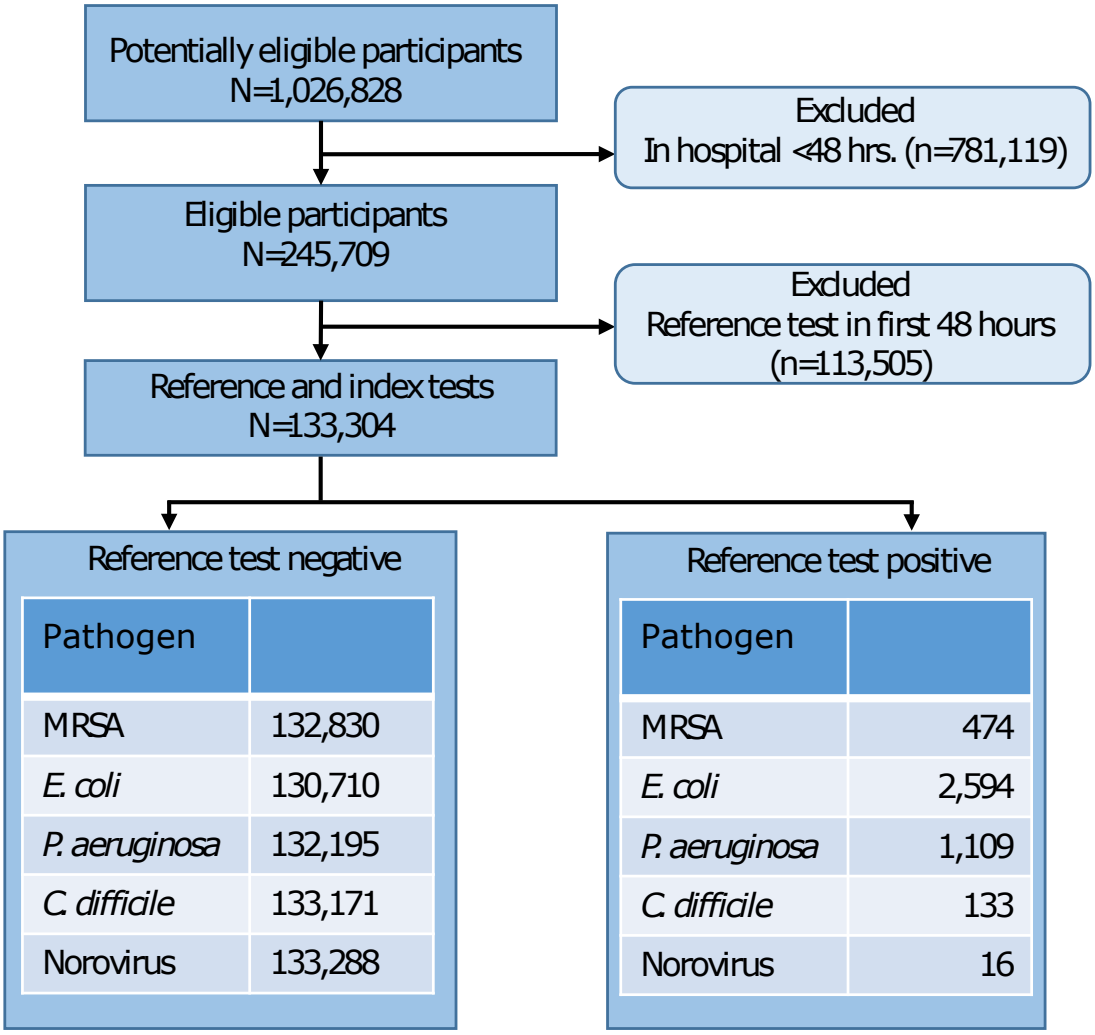

Figure 1: Patient flow diagram. Number of excluded eligible patients differs by pathogen because different numbers of patients had their reference test within the first 48 hours of their stay, and therefore were likely not healthcare-associated infections. The number of eligible participants excluded differs between pathogens because different sets of patients had their reference and index tests within the first 48 hours of their hospital stay. Importantly, the different populations for each pathogen are not exclusive; each patient is in all five populations, and only their results on the reference and index tests change.

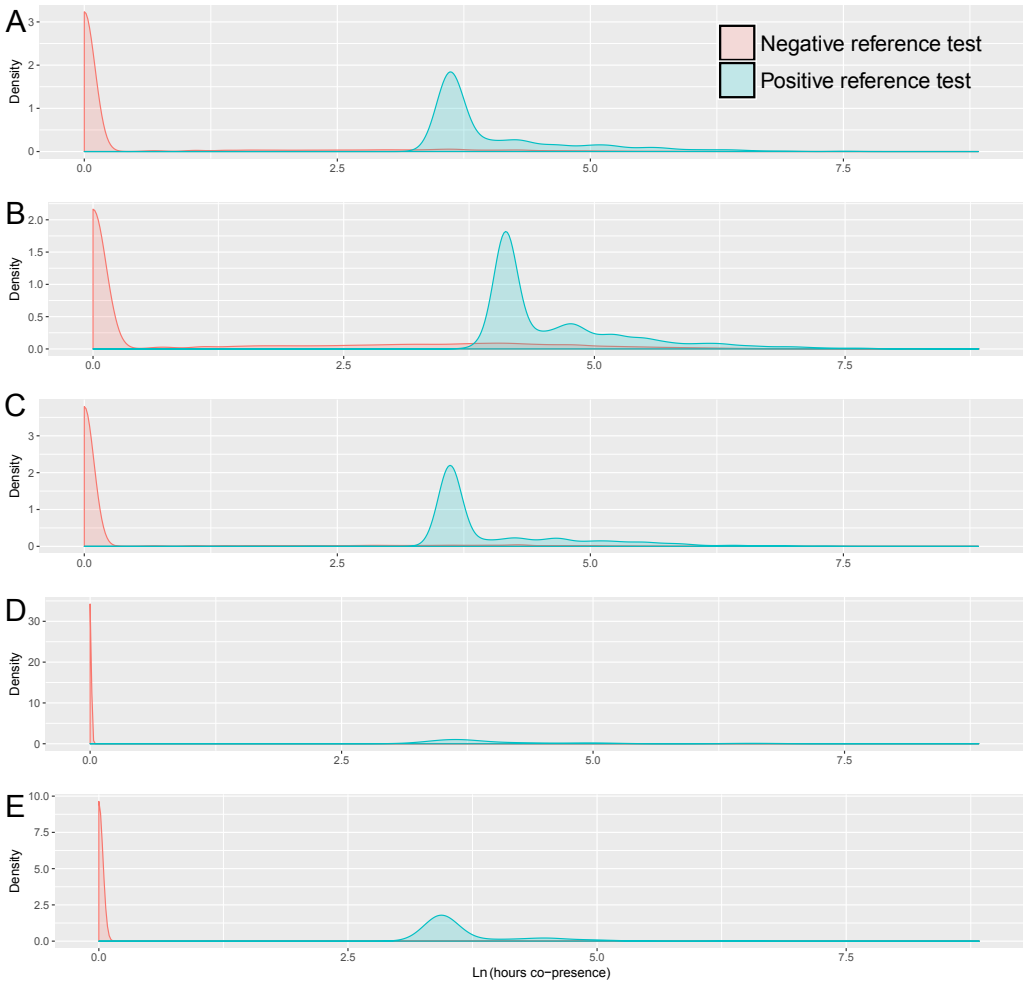

Figure 2: Empirical probability density functions of hours of co-presence with infected individuals (index test) stratified by the presence of a diagnosis or positive microbiological test (reference test). Each panel represents one of the pathogens tested: A) MRSA, B) *E. coli*, C) *P. aeruginosa* , D) *C. difficile* , and E) Norovirus.