Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Hospital Infection 106 (2020) 562–569

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevier.com/locate/jhin

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

J. Rewley [a,b,*], L. Koehly [a], C.S. Marcum [a], F. Reed-Tsochas [b]

[a] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
[b] CABDyN Complexity Centre, Saïd Business School, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 14 April 2020
Accepted 27 July 2020
Available online 1 August 2020

Keywords:
Big data
Proximity
Contact tracing
Co-presence
Electronic medical records

SUMMARY

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

Methods: This study examined whether the amount of time a patient spends on a ward with other patients clinically suspected of infection, termed 'co-presence', can be used as a tool to predict subsequent healthcare-associated infection. Compared with contact tracing, this leverages passively collected electronic data rather than manually collected data, allowing for improved monitoring. All 133,304 inpatient records between 2011 and 2015 were abstracted from a healthcare system in the UK. The area under the receiver-operator curve (AUROC) for each of five pathogens was calculated based on co-presence time, sensitivity and specificity of the test, and how much earlier co-presence would have predicted infection for the true-positive cases.

Findings: For 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 h, and would have led to detection of true-positive cases up to an average of 1 day earlier.

Interpretation: These findings show that co-presence time would help to 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.

Published by Elsevier Ltd on behalf of The Healthcare Infection Society.

Introduction

Healthcare-associated infections (HCAIs) are a burden on the healthcare system. Despite advancing medical technology and standards of care, the attributable costs of each case of HCAI across the globe continue to range from US$2992 to US$29,000 (approximately £2367 to £22,945) [1–3]. In the UK,

* Corresponding author. Address: National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.
E-mail address: jeffrey.rewley@pennmedicine.upenn.edu (J. Rewley).

https://doi.org/10.1016/j.jhin.2020.07.031
0195-6701/Published by Elsevier Ltd on behalf of The Healthcare Infection Society.

each patient with an HCAI costs an additional £3154, with an estimated total of £930.62 million per year (approximately US$1.2B) [4]. In addition, HCAIs have a direct adverse impact on patient health outcomes such as length of hospital stay [5] and mortality [6].

This study focused on identifying infections earlier and more accurately. This requires timely diagnostic tests, but current diagnostics (both those that measure biomarkers and those that measure the microbiological vector directly) are limited by various combinations of long processing times, high costs, lack of pathogen specificity or imperfect accuracy [7–10]. 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 the likelihood of subsequent infection [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 following occupancy of an infected individual on a ward, the risk of infection for subsequent occupants of the ward is increased significantly for the next 24 h [12].

This knowledge partly informs contact tracing, which has been used in the past to identify those at greatest risk of acquiring an infection. Contact tracing identifies those individuals who have come in close physical proximity (co-present) with infected persons [13]. Co-presence has been used in previous outbreaks in healthcare settings to trace the source of a pathogen, and the subsequent path by which it spreads through the hospital population, and has been used in the recent outbreak of coronavirus disease 2019 (COVID-19) [14,15]. Although contact tracing is based 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 the risk of infection has not been examined thoroughly. 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]. In addition, contact tracing relies on positive identification of a case followed by retroactive identification of individuals who had crossed the threshold with the infected case; as such, contact tracing is 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 related more strongly to the risk of infection. The increasing ubiquity of hospital administrative data (HAD) enables 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 the risk of mortality and re-admission [20,21]. In addition to formally testing how strongly co-presence predicts subsequent HCAI, HAD is able to provide 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 in terms of both time and financial resources.

This study shows that co-presence measured by HAD functions as an effective monitoring tool for the risk of HCAIs. The amount of time that a patient is in the same hospital bay as a patient who is clinically suspected of an infection, defined as either a diagnosis or ordering a microbiological test (irrespective of whether the eventual result is positive or negative), whichever comes first, was used in this study. The use of clinically suspected infection, rather than confirmed infection, means that this tool can potentially detect pathogens earlier than methods which require confirmation of infection, such as contact tracing. Moreover, clinically suspected infections reflect the information that would be available prospectively in real-time in a clinical setting, increasing the potential benefits of this tool. This co-presence time has been termed 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 (EMRs) that could be monitored for patient–patient co-presence, implementation of this tool would likely be inexpensive and efficient. The clinical role of the index test would be for screening, and 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.

Methods

Study design and population

The study population comprised all 133,304 patients with National Health Service (NHS) hospital stays of at least 48 h 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.

This study used hospital stays of at least 48 h because HCAI was defined as an infection occurring more than 48 h after a patient entered the hospital [23]. These patients comprise a consecutive series, where the index test was assessed retrospectively. The index test was assessed on the following pathogens: meticillin-resistant *Staphylococcus aureus* (MRSA), *Escherichia coli*, *Pseudomonas aeruginosa*, *Clostridioides 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. In this study, due to the available data, the presence of the pathogen and the syndrome were treated as equivalent [24]. The main exception is that the authors were able to exclude International Classification of Diseases-10 (ICD-10) codes indicative of infections not transmitted interpersonally (see below).

This study followed the Standards for Reporting Diagnostic Accuracy Studies guidelines for reporting a new diagnostic [22]. Ethics committee approval was gained from the University of Oxford Institutional Review Board, 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).

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 HAD was not configured such that physicians or laboratory technicians could examine co-presence; as such, the index test results were not available to those individuals performing the reference tests. 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 that a patient spent co-present with a patient with a clinically suspected infection (defined as either a diagnosis or ordering a microbiological test, whichever came first). Importantly, this included any tests ordered regardless of whether they eventually returned a positive or negative result, as both constituted the suspicion of infection. Co-presence time was defined as the length of time that both patients were in the same hospital bay. The clinical information and microbiological test results were available to the researchers evaluating co-presence time, as it was all contained in the EMR. There were no missing or indeterminate data for co-presence time 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 co-presence time was based on tested or diagnosed patients rather than patients with confirmed infection to more accurately reproduce the knowledge that would be available if this was a prospective study. As co-presence time was assessed retrospectively, the data would enable perfect calculation of 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 5 days [25]. Therefore, only the subset of real-time information that would be available in practice was used.

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 was 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 at random 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, deterministic infection periods were applied to model the infectious period [26]. It was assumed that the time at which the microbial test was collected or the diagnosis was 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 I) [27–31]. 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 the removal of ICD-10 codes likely not pertaining to HCAI.

Table I
Values for the infectious period for each pathogen used

Pathogen	Infectious period, mode (h)	Infectious period (range)
MRSA [27]	72	48–96
Escherichia coli [28]	120	80–160
Pseudomonas aeruginosa [29]	48	12–72
Clostridioides difficile [30]	60	24–96
Norovirus [31]	44	12–72

MRSA, meticillin-resistant *Staphylococcus aureus*.

Analysis

The receiver-operator characteristic (ROC) curve and the area under the receiver-operator curve (AUROC) were used to compare the measures of diagnostic accuracy. The 95% confidence intervals of AUROC were calculated using bootstrapping. To determine the optimal cut-point for these curves, it was assumed that the clinical costs of false-positive, false-negative, true-positive and true-negative cases are all equal. Following this, the optimal cut-point is the point closest to a perfect test (100% sensitivity and specificity) in Euclidian space. Sensitivity and specificity were assessed at optimal cut-points.

For this tool to benefit clinical practice, it would need to predict infections as well as or better than other methods, and also do so earlier. To assess this, 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 was calculated. This represents how many hours earlier the pathogen could be detected if co-presence time was used as a screening test, relative to the current standard operating procedure.

Finally, a number of sensitivity analyses were conducted, which are outlined here. For more information, see the online supplementary material. The results may be due to class imbalance between infected and uninfected individuals. Patients were removed 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; the analysis was conducted 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 ward-level co-presence data (rather than bay-level), the analysis was rerun at ward level.

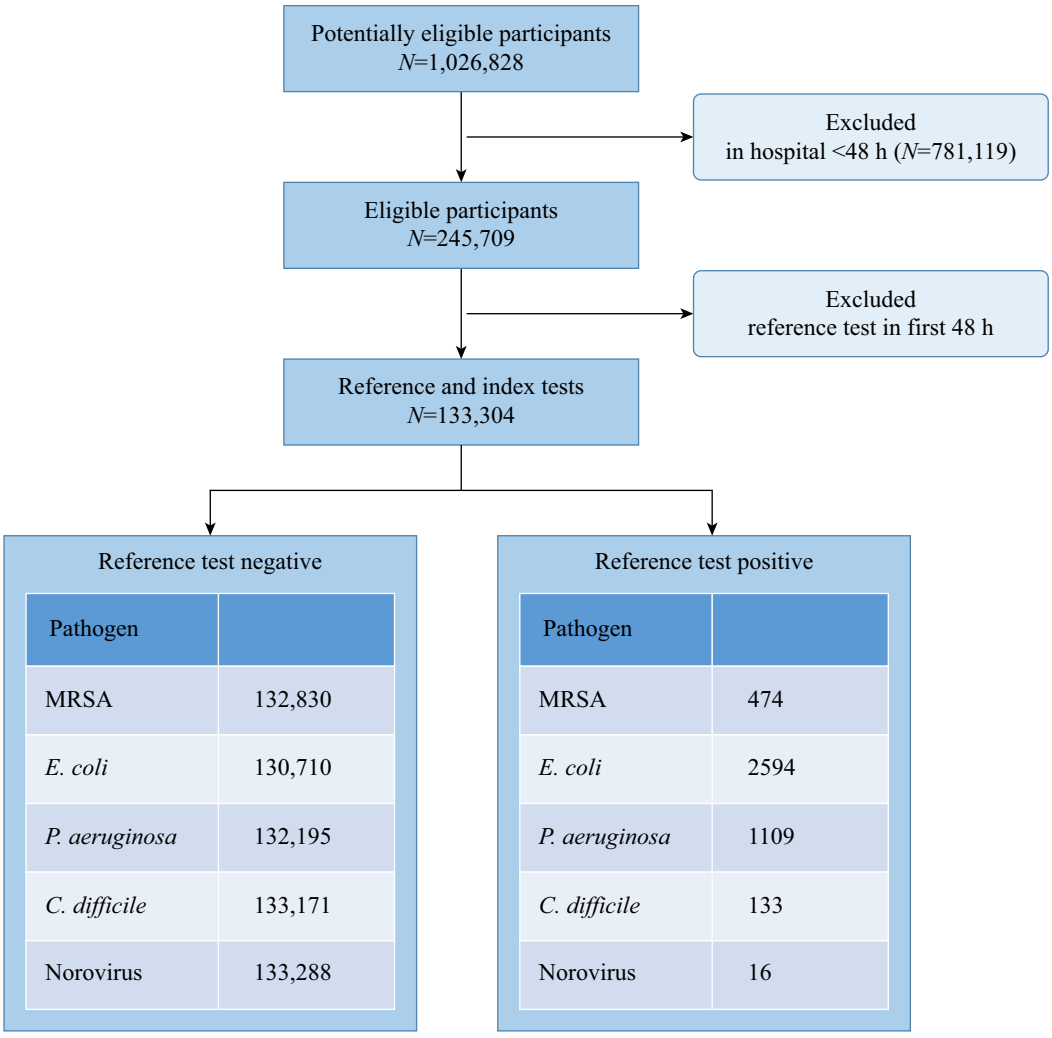

Figure 1. Patient flow diagram. Number of eligible patients excluded differs by pathogen because different numbers of patients had their reference test within the first 48 h of their hospital stay, and therefore were likely not healthcare-associated infections. Importantly, the different populations for each pathogen are not exclusive; each patient is in all five populations, and only their results on the reference and index tests change. MRSA, meticillin-resistant *Staphylococcus aureus*; *E. coli*, *Escherichia coli*; *P. aeruginosa*, *Pseudomonas aeruginosa*; *C. difficile*, *Clostridioides difficile*.

Patient and public involvement

As this was a retrospective study analysing EMRs, the study population was not involved in the study.

Results

The patients included in the study, their reasons for exclusion and their reference tests can be seen in Figure 1. Most patients were excluded due to being outpatients or being tested for microbiological vectors within the first 48 h of entering hospital.

This left the study population of patients who could potentially contract an HCAI. The demographics of these patients are shown in Table II. Patients were, on average, 56 years of age, and 45% were male. On average, these patients spent 13 days in hospital, and 5% of them died while in hospital. In total, 4326 (3%) patients were infected with one of the five pathogens studied.

Table II
Baseline demographics and clinical characteristics of eligible patients

Variable	Mean (SD) or *N* (%)
Age (years)	56.4 (27.8)
Sex (male)	59,988 (44.80%)
Length of stay (h)	319.5 (568.8)
Died in hospital	7180 (5.40%)
Infected with MRSA	474 (0.36%)
Infected with *Escherichia coli*	2594 (1.95%)
Infected with *Pseudomonas aeruginosa*	1109 (0.83%)
Infected with *Clostridioides difficile*	133 (0.10%)
Infected with norovirus	16 (0.01%)

MRSA, meticillin-resistant *Staphylococcus aureus*; SD, standard deviation.

After applying the index test to this set of patients, distinct distributions of co-presence time for patients with clinical suspicion of infection were observed for those whose reference test was negative compared with those whose reference test was positive (Figure 2). Irrespective of a specific cut-point, the distributions of co-presence time 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 time (index test results) much lower than that for patients with a positive reference test.

To quantify the performance of the index test, ROC curves were created based on multiple cut-offs of hours of co-presence with infected individuals. The ROC curves have AUROCs ranging from 0.92 to 0.99 (Table II). The optimal cut-point ranges from 29 to 59 h depending on the pathogen in question. This means that patients must spend over 24 h co-present with infected patients before the number of false-negative results is minimized. Although this dichotomizes co-presence, similarly to contact tracing, the optimal cut-off is not none vs any. Instead, it takes at least 29 h before co-presence time is maximally discriminatory. Of note is that exposure to multiple infected persons can increase co-presence time by more than 1 h per hour of elapsed time (i.e. patients may have reached this point in less than 29 h if they are in the same bay as multiple infected patients).

For patients with a positive reference test and a positive index test (true-positive case), the authors examined how

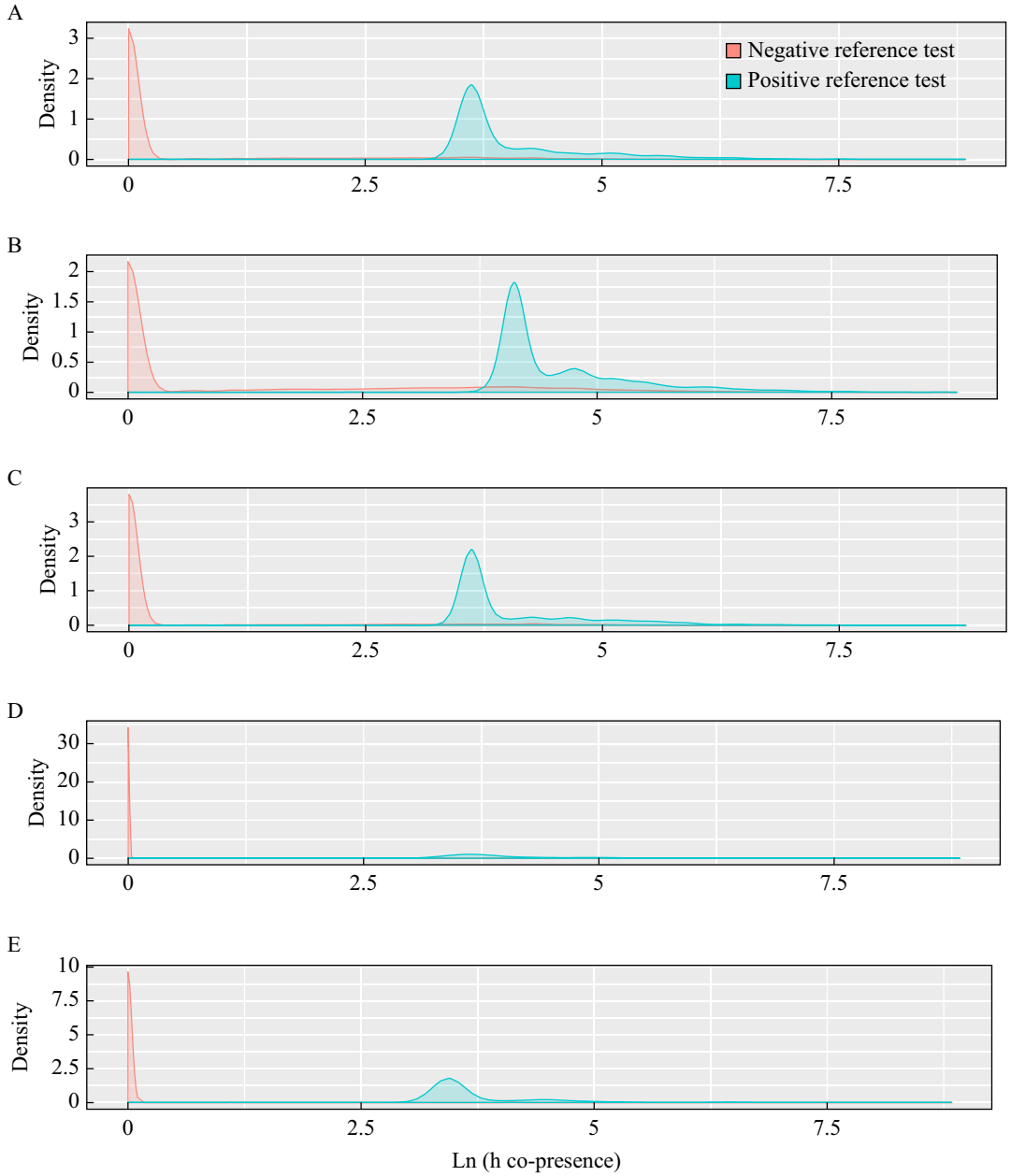

Figure 2. Empirical probability density functions of natural-logarithm of hours of co-presence with infected individuals (index test) stratified by the presence of a diagnosis or positive microbiological test (reference test). Each panel represents one of the pathogens tested: (A) meticillin-resistant *Staphylococcus aureus*; (B) *Escherichia coli*; (C) *Pseudomonas aeruginosa*; (D) *Clostridioides difficile*; and (E) norovirus.

many hours earlier they would have been tested if the reference test was administered immediately upon crossing the threshold of co-presence (Table III). On average, the amount of time saved was found to range from 6 h for *C. difficile* to 22 h for *P. aeruginosa*. This means that if a microbiological test for *P. aeruginosa* was administered as soon as a patient crossed this threshold, the patient's infection could be detected 22 h earlier. As mentioned previously, the infected and uninfected populations are highly unbalanced, which may have artificially increased the performance of co-presence time as a diagnostic. Matching and removal of patients not at risk of infection were used to examine this possibility. Both analyses showed minimal effects of unbalanced samples (Figures S1 and S2, see online supplementary material). Additionally, co-presence time may have high predictive power based on latent characteristics between patients independently increasing the risk to both patients. The use of a negative control (Figure S3, see online supplementary material) shows that this is unlikely, as the tool had low predictive power in the case of sepsis, a non-communicable HCAI. Finally, an analysis based on one ward at a time showed that co-presence time remained strong when looking at a single ward (median AUC 0.86; Table S2, see online supplementary material). 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.

Discussion

Key findings

This study showed 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, it showed that the use of co-presence time as a screening test can enable the detection of infection which stemmed from interpersonal transmission in the healthcare setting as much as 22 h earlier.

Relationship with previous studies

This study used 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 with a confirmed infection to contain an infection effectively [17]. Rather than dichotomize co-presence as none vs any, or at 15 min as has been suggested for transmission of COVID-19 [16], this study evaluated all potential thresholds with empirical data, selecting the threshold that best balanced false-positive and false-negative results.

These analyses indicate that the optimal cut-point of co-presence for all five pathogens tested is >24 h. This is important as previous methods, such as contact tracing, typically use any contact vs none (or no co-presence vs any) as a cut-off for potential infection [13,17]. The present findings show that any contact vs none may unnecessarily increase the number of false-positive results, when more stringent criteria will limit false-positive results while not increasing the number of false-negative results.

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 h. The present study, however, found 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.

Study implications

Patient–patient co-presence time will ideally be used as a screening tool; once a patient has been co-present with a tested or diagnosed patient for at least the cut-off time of 29–59 h, they would be tested for the microbiological agent and subsequently monitored for signs of infection. If this was done, infections could be identified earlier than when using the 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 estimated in Table III reflect the minimum amount of time that

Table III
Index test statistics for all five pathogens

Pathogen	AUC (95% CI)	Threshold (h)	Sensitivity	Specificity	True-positive results	PPV	Average hours saved per patient (range)
MRSA	0.962 (0.96–0.964)	35	1.00	0.95	474	0.067	10.93 (7.05–16.18)
Escherichia coli	0.966 (0.965–0.967)	59	0.95	0.90	2472	0.159	8.35 (4.91–13.61)
Pseudomonas aeruginosa	0.925 (0.923–0.927)	35	1.00	0.95	1107	0.142	21.97 (13.98–32.96)
Clostridioides difficile	0.993 (0.992–0.994)	29	1.00	0.99	133	0.091	6.36 (4.08–10.14)
Norovirus	1 (1–1)	34	1.00	1.00	16	1.00	8.19 (5.12–10.31)

MRSA, meticillin-resistant *Staphylococcus aureus*; AUC, area under curve; CI, confidence interval; PPV, positive predictive value. 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-positive results were taken at these optimal cut-points. Hours saved is the difference in time between when a patient first crossed the threshold of the index test and when they were actually tested for or diagnosed with the infection. This number 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 the lengths of infectious periods were stochastic rather than deterministic.

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 this study shows a lower bound on the potential benefits of a tool based on co-presence time.

Strengths

As stated previously, the use of co-presence time as a screening tool has many advantages. First, using HAD, the authors 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 EMR, so adapting these records to monitor co-presence with infected and tested individuals should involve minimal effort, allowing this method to be integrated easily into many health systems. HAD can also be monitored passively for relevant amounts of co-presence, meaning the cost to calculate co-presence time is inexpensive once passive monitoring is enabled. Determining co-presence time is also rapid; a result is returned immediately when a patient crosses the threshold of co-presence.

These results 'dovetail' well with recent advances in microbiological detection; a test for MRSA that can provide results within a few hours was approved recently [34]. In tandem with the tool proposed here, a patient who crosses the threshold of co-presence can be given a rapid diagnostic test immediately, and decisions regarding their infection status can be made rapidly. 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 health care' [35]. As co-presence would potentially identify an individual at risk prior to manifestation of biomarkers and other physical symptoms of infection, increased testing of these individuals may also help to 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 infection with *E. coli* — 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.

Limitations

As 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 some ICD-10 codes which clearly indicated intra-personal aetiology were excluded, 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 the prediction of 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 the 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, the 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 have a minimal impact on the effectiveness of this approach. As the diagnosis of infection was based on EMRs rather than administrative data, generalizability of these results is likely increased (diagnostic criteria for infections are more conserved across healthcare systems than administrative coding) [38]. However, future work should explore whether similar results can be obtained in other settings.

In conclusion, this study has shown that the use of HAD-based co-presence time with patients with clinical suspicion 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, this study has shown that co-presence is a powerful indicator of HCAI, which merits implementation in hospitals for the reduction of outbreaks in healthcare settings.

Conflict of interest statement
None declared.

Funding sources
The authors acknowledge funding from the National Human Genome Research Institute, National Institutes of Health (Grant No. ZIA HG200335 to Laura Koehly) and the Oxford Martin School, University of Oxford (Grant No. LC1213006 to Felix Reed-Tsochas).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhin.2020.07.031.

References

[1] Gabriel L, Beriot-Mathiot A. Hospitalization stay and costs attributable to *Clostridium difficile* infection: a critical review. J Hosp Infect 2014;88:12—21.

[2] Jarvis W. Selected aspects of the socioeconomic impact of nosocomial infections: morbidity, mortality, cost, and prevention. Infect Cont Hosp Epidemiol 1996;17:552—7.

[3] Johnston K, Thorpe K, Jacob J, Murphy D. The incremental cost of infections associated with multidrug-resistant organisms in the inpatient hospital setting: a national estimate. Health Serv Res 2019;54:782—92.

[4] Plowman R, Graves N, Griffin MAS, Roberts JA, Swan AV, Cookson B, et al. 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. J Hosp Infect 2001;47:198—209.

[5] Chacko B, Thomas K, David T, Paul T, Jeyaseelan L, Peter J. Attributable cost of a nosocomial infection in the intensive care unit: a prospective cohort study. World J Crit Care Med 2017;6:79.

[6] Adrie C, Garrouste-Orgeas M, Essaied W, Schwebel C, Darmon M, Mourvillier B, et al. Attributable mortality of ICU-acquired bloodstream infections: impact of the source, causative micro-organism, resistance profile and antimicrobial therapy. J Infect 2017;74:131—41.

[7] Johnson G, Millar M, Matthews S, Skyrme M, Marsh P, Barringer E, et al. 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 Microbiol 2006;6:83.

[8] Su L, Han B, Liu C, Liang L, Jiang Z, Deng J, et al. Value of soluble TREM-1, procalcitonin, and C-reactive protein serum levels as biomarkers for detecting bacteria among sepsis patients with new fever in intensive care units: a prospective cohort study. BMC Infect Dis 2012;12:157.

[9] Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clin Infect Dis 2004;39:206—17.

[10] Visconti V, Brunetti G, Giordano A, Raponi G. RT-PCR for the diagnosis of *Clostridium difficile* infection: the final answer has yet to come. J Clin Pathol 2017;70:1090—1.

[11] Britt M, Schleupner C, Matsumiya S. Severity of underlying disease as a predictor of nosocomial infection: utility in the control of nosocomial infection. JAMA 1978;239:1047—51.

[12] Murray S, Yim J, Croci R, Rajkomar A, Schmajuk G, Khanna R, et al. Using spatial and temporal mapping to identify nosocomial disease transmission of *Clostridium difficile*. JAMA Intern Med 2017;177:1863—5.

[13] Bagdasarian N, Chan HC, Ang S, Isa MS, Chan SM, Fisher DA. A stone in the pond approach to contact tracing: responding to a large-scale, nosocomial tuberculosis exposure in a moderate TB-burden setting. Infect Cont Hosp Epidemiol 2017;38:1—3.

[14] Snitkin E, Zelazny A, Thomas P, Stock F, Henderson D, Palmore T, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Translat Med 2012;4:148.

[15] Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health 2020. https://doi.org/10.1136/jech-2020-214051.

[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 and Prevention. CDC methods for implementing and managing contact tracing for Ebola virus disease in less-affected countries. Atlanta, GA: CDC; 2017. Available at: https://stacks.cdc.gov/view/cdc/26492 [last accessed September 2019].

[18] Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020;368:613—9.

[19] Bell A, Ward P, Tamal MD, Killilea M. Assessing recall bias and measurement error in high-frequency social data collection for human—environment research. Pop Environ 2019;40:325—45.

[20] Lienert J, Marcum C, Finney J, Tsochas T, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Net Sci 2017;5:308—27.

[21] Lienert J, Reed-Tsochas F, Koehly L, Marcum CS. 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). IEEE; 2019. p. 2756—62.

[22] Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Glasziou P, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:1446—52.

[23] Horan T, Andrus M, Dudeck M. CDC/NHSN surveillance definition of health care associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Cont 2008;36:309—32.

[24] Khan H, Ahmad A, Mehboob R. Nosocomial infections and their control strategies. Asian Pacif J Trop Biomed 2015;5:509—14.

[25] Pal N, Sharma R, Rishi S, Vyas L. Optimum time to detection of bacteria and yeast species with BACTEC-9120 culture system from blood and sterile body fluids. J Lab Phys 2009;1:69—72.

[26] Anderson MR. The population dynamics of infectious diseases: theory and applications. Springer; 2013.

[27] Public Health Agency of Canada. *Pseudomonas* spp. pathogen safety data sheet. 2012. Available at: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/pseudomonas.html [last accessed September 2019].

[28] Dalton CB, Mintz ED, Wells JG, Bopp CA, Tauxe RV. Outbreaks of enterotoxigenic *Escherichia coli* infection in American adults: a clinical and epidemiologic profile. Epidemiol Infect 1999;123:9—16.

[29] Public Health Agency of Canada. *Staphylococcus aureus* pathogen safety data sheet. 2012. Available at: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/staphylococcus-aureus.html [last accessed September 2019].

[30] Cohen S, Gerding D, Johnson S, Kelly C, Loo V, McDonald L, et al. 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). Infect Cont Hosp Epidemiol 2010;31:431—55.

[31] Rahamat-Langendoen JC, Lokate M, Schölvinck EH, Friedrich AW, Niesters HGM. Rapid detection of a norovirus pseudo-outbreak by using real-time sequence based information. J Clin Virol 2013;58:245—8.

[32] Hugonnet S, Sax H, Eggimann P, Chevrolet J, Pittet D. Nosocomial bloodstream infection and clinical sepsis. Emerg Infect Dis 2004;10:76.

[33] Xuan S, Zangwill K, Ni W, Ma J, Hay J. Cost-effectiveness analysis of four common diagnostic methods for *Clostridioides difficile* infection. J Gen Intern Med 2020;35:1—9.

[34] Voelker R. New test detects MRSA in hours rather than days. JAMA 2020;323:210.

[35] Norgeot B, Glicksberg B, Butte A. A call for deep-learning healthcare. Nat Med 2019;25:14—5.

[36] Anderson R, Heesterbeek H, Klinkenberg D, Hollingsworth T. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020;395:931—4.

[37] Flynn D, Weinstein R, Nathan C, Gaston M, Kabins S. Patients' endogenous flora as the source of nosocomial Enterobacteriaceae in cardiac surgery. J Infect Dis 1987;156:363—8.

[38] Raith E, Udy A, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. 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. JAMA 2017;317:290—300.