Inter-hospital transmission of hospital acquired infections: insights from a multi-level mathematical network model

Thesis

to obtain the academic degree of

Doctor rerum medicarum (Dr. rer. medic.)

in the field of Epidemiology

submitted to the Faculty of Medicine of

Martin Luther University Halle-Wittenberg

by Hanjue Xia

Supervisor: Prof. Dr. Rafael Mikolajczyk

Reviewers: Prof. Dr. Oliver Thews, Halle (Saale)

Prof. Dr. Vitaly Belik, Berlin

Date of the defense: 18.03.2024

Abstract

Hospital networks formed by patient movements between hospitals have a vital influence on transmission pathways of hospital-acquired infections (HAIs). More and more countries have studied HAI spread at the level of hospital networks. However, in Germany, this kind of study did not exist. Therefore, this thesis aims to model the spread of HAIs in inter-hospital networks in German federal states. Since patient movements inside a hospital also played an important role in HAI spread, another aim of the thesis is to analyze patient movement patterns between hospital departments.

To answer the thesis' aims, two studies were included in this cumulative doctoral project. Because it was very challenging to obtain data from all insurance companies in Germany and consequently any projects based on empirical data suffered from incomplete data access, the first study was to assess how different levels of data incompleteness affect estimates of network structures and modelled infection spread. Since the inter-hospital epidemic spread depends also strongly on the epidemic spread inside hospitals. The second study was to simulate epidemic spread inside hospitals and to assess how the movement patterns between hospital departments differ between patients at high or low risk of colonization.

In the first study, network measures derived from sampled, scaled-up and original AOK data were compared. It showed that the common network measures were affected by incompleteness, however, the simulated prevalence based on an individual based SIS (Susceptible-Infectious-Susceptible) model was robust over a large range of incompleteness proportions. Epidemics and the transmission of the infection diseases can be modelled on hospital data with a low population coverage (around 10%), whilst maintaining accuracy to within 10% of the true population prevalence. The obtained insurance data can be scaled up to access the full data in each federal state based on sex and age structure. Since the intra-hospital transfers played an important role in influencing prevalence at discharge, it is necessary to investigate the patient movement patterns between hospital departments, which can derive an underlying intra-hospital movement network. This work was done in the second study

The second study showed that the simulated department prevalence had a positive correlation to the network centrality measures. Risk stratification of patients according to their ICD-10 codes revealed that length of hospital stays, patient age, and average numbers of movements per admission were higher in the high-risk group. This study emphasized the importance of intra-hospital patient movements and their possible influence on pathogen spread. Implementing interventions in departments of higher (weighted) degree might slow down the MDR-E spread.

Hanjue, Xia: Inter-hospital transmission of hospital acquired infections: insights from a multi-level mathematical network model, Halle (Saale), Univ., Med. Fac., Diss., 73 Pages, 2023.

Referat

Die Patientenbewegung zwischen Krankenhäusern führt zu einer Vernetzung zwischen diesen und spielt damit eine wichtige Rolle bei der Übertragung von Krankenhausinfektionen (Hospital-acquired infections, kurz HAI). Während immer mehr Länder die Ausbreitung von HAIs aufgrund dieser Vernetzung untersuchen, fanden solche Studien in Deutschland bisher selten statt. Im Rahmen dieser Arbeit soll daher die Ausbreitung von HAIs in deutschen Krankenhausnetzwerken erforscht werden. Da Patientenbewegungen innerhalb eines Krankenhauses ebenfalls eine wichtige Rolle bei der Verbreitung spielen, ist ein weiteres Ziel, den Einfluss der Patientenbewegungen zwischen Krankenhausabteilungen zu analysieren. Diese Fragestellungen werden in dieser kumulativen Dissertation anhand von zwei Studien beantwortet.

Da es sehr schwierig ist, Daten von allen deutschen Versicherungen zu erhalten, sind Projekte, die auf empirischen Daten basieren, von einem unvollständigen Datensatz betroffen. In der ersten Studie wurde deshalb die Auswirkung von unterschiedlich vollständigen Datensätzen auf die Schätzung der Netzwerkstruktur und die Ausbreitung simulierter Infektionen betrachtet. Dabei wurden die Maßangaben der Netzwerke verglichen, welche sich aus gesampelten, erweiterten und rohen AOK-Daten ableiten lassen. Die Studie zeigte die Beeinträchtigung gemeinsamer Netzwerkmaße durch Unvollständigkeit während die simulierte Prävalenz des individualbasierten SIS-Modells jedoch auch für einen großen Anteil an Unvollständigkeit robust war. Übergangsmodelle für Epidemien und Infektionskrankheiten konnten auf Krankenhausdaten mit geringer Bevölkerungsabdeckung (ungefähr 10 %) aufgebaut werden, wobei die Genauigkeit innerhalb von 10 % der tatsächlichen Bevölkerungsprävalenz lag. Die Versicherungsdaten konnten basierend auf Geschlecht und Altersstruktur hochskaliert werden, um daraus Zugang auf die vollständigen Daten in jedem Bundesland zu erhalten.

Weil die Ausbreitung zwischen Krankenhäusern auch stark von der Ausbreitung der Epidemie innerhalb der Krankenhäuser abhängt, wurden in der zweiten Studie die Ausbreitung von Epidemien innerhalb von Krankenhäusern simuliert und untersucht, wie sich die Bewegungsmuster von Patienten mit hohem oder niedrigem Kolonisierungsrisiko unterscheiden. Dazu wurden die Bewegungsmuster der Patienten zwischen Krankenhausabteilungen betrachtet und analysiert, ob sich hieraus ein zugrundeliegendes Netzwerk der Bewegungen innerhalb eines Krankenhauses zeigen könnte. Dabei zeigte sich eine positive Korrelation zwischen der modellierten sektoralen Prävalenz und der Messung der Netzwerkzentralität. Die Risikostratifizierung der Patienten nach ihren ICD-10 Codes und die damit einhergehende Eingruppierung in eine Hochrisikogruppe ergaben, dass für diese die Dauer des Aufenthalts im Krankenhaus, das Alter der Patienten und die durchschnittliche Anzahl von Bewegungen pro Aufnahme höher waren. Die Studie unterstreicht die Bedeutung der Patientenmobilität in Krankenhäusern und ihre möglichen Auswirkungen auf die Übertragung von Krankheitserregern.

Hanjue, Xia: Inter-hospital transmission of hospital acquired infections: insights from a multi-level mathematical network model, Halle (Saale), Univ., Med. Fac., Diss., 73 Pages, 2023.

Table of Contents

1. Introduction and Objectives

1.1. Background

Healthcare-acquired infections (HAIs) threatened global patient health and healthcare systems. As an additional challenge, most of the HAIs were caused by drug-resistant pathogens. In Europe, 3.1-4.6 million people acquired a HAI each year in acute care hospitals in European Union (EU) countries, Iceland, Norway, and the United Kingdom (1). The hospital-related pathway was a key mechanism of HAI spread (2-8). Recently, inter-hospital patient movements between healthcare facilities have been considered as a vital route of transmitting pathogens between hospitals. Healthcare networks were constructed in previous work by using hospital discharge databases based on English, Dutch as well as French national medical registration datasets (2, 4, 7, 8). Based on those networks, various network measures and infection models were applied to contain the HAI spread.

Until year 2023, there were 96 statutory and several private health insurance companies in Germany with heterogeneous regional coverage (9). Over 90% of the population were insured by any one of the statutory health insurance companies (9) and everyone can freely choose any of these at any time. Obtaining data from all insurance companies was very challenging, such that any projects based on empirical data faced the issue of incomplete data. Nevertheless, the effects of incomplete patient transfers on network structures and modelling results have not been studied previously. Data from three German regional insurance companies covering four federal states: AOK (historically "general local health insurance company", but currently only the abbreviation is used) Lower Saxony (in Federal State of Lower Saxony), AOK Bavaria (in Bavaria), and AOK PLUS (in Saxony and Thuringia) with various local population coverage were used to assess how different levels of data incompleteness impact estimates of network structures and results of modelling infection spread. In addition, a method for constructing surrogate data which improves the accuracy of modelling results was proposed. The AOK data can be extrapolated to reach the corresponding federal state levels based on the scale-up methods which can deliver a better overview of patient transfer patterns, patient behavior and hospital networks in German federal states.

Besides impact of the inter-hospital transfers, patient movements between the departments of a single hospital (intra-hospital transfers) can also contribute to the epidemic spread. Many studies have reported the spread of multidrug-resistant Enterobacteriacea (MDR-E) pathogens inside one hospital (10-14). However, these studies were either limited to intensive care units (ICUs), or to some specific medical specialty departments, and intra-hospital transfers

were not assessed. Therefore, in this thesis, datasets from five hospitals in different countries were used to assess patient movement patterns inside hospitals. Based on that, the connections between intra-hospital patient transfers and epidemic spread were investigated.

1.2. German statutory health insurance

Statutory health insurance (GKV, abbreviation for "Gesetzliche Krankenversicherung") is the predominant (87.1 – 86.9%) form of health insurance in Germany (15). Together with pension, unemployment, accident and long-term care insurance, it forms the German social insurance system. There are more than 10% of Germans insured by private health insurance (PKV, abbreviation for "Private Krankenversicherung") (15). Unlike in statutory health insurance, the premium amount in private health insurance depends on the scope of the insured benefits and the individual insured risk. A change from GKV to PKV is only open to certain persons who are not subject to compulsory insurance in the GKV, such as self-employed persons, civil servants and employees with an income above the compulsory insurance limit. However, there are also narrow limits on switching from private health insurance to statutory health insurance. People in the same family without a job, for example spouses or children are automatically covered by the GKV in contrast to the PKV.

GKV was introduced in Germany by the Law on workers' health insurance as part of the German social insurance solidarity system since 15th June 1883 by Otto von Bismarck. It came into force after 1 December 1883. At that time, community nursing insurance already existed in Bavaria based on the law since 29 April 1869. These were the first laws ever proposed to regulate the social security for lower income people to handle their illness. The group of persons was limited to dependent employees with an annual income of no more than 2000 Reichsmark in the branches of mining, industry, railways, inland steam navigation, crafts and trades. In 1911, the law of insurance for normal employees was passed and the insurance memberships were expanded. Since 1996, the automatic assignment to a health insurance according to the types of employers and the professions has no longer existed in Germany, so that people could freely choose their insurance companies.

The main types of GKV in Germany are as follows (16):

- Ersatzkassen (EK) is developed from self-help associations and organized in the umbrella association (vdek, abbreviation for "Verband der Ersatzkassen").
- Allgemeine Ortskrankenkassen (AOK) exists for delimited regions and may be extended to different federal states.
- Betriebskrankenkassen (BKK) are founded by employers with at least 1000 insured persons. They can also be open to nonemployees.
- Innungskrankenkassen (IKK) are set up by craft guilds with at least 1000 insured persons.

- Landwirtschaftliche Krankenkasse (LKK) are for farmers and their families as well as those receiving a pension from the aging insurance for farmers.
- Knappschaft Bahn See (KBS) since 2008 was only for workers in the mining industry, originally.

In Table 1, more details about these GKV companies are showed.

Table 1. Statutory health insurance until 1st Jan 2019 (17).

Types	Number of insurance companies	Number of insures (including pensioners)
All statutory	109	72.84 Mill.
EK	6	27.97 Mill.
AOK	11	26.60 Mill.
BKK	84	10.88 Mill.
IKK	6	5.19 Mill.
LKK	1	0.62 Mill.
KBS	1	1.58 Mill.

1.3. AOK data description

According to the previous work, hospital network can be directly used as a basis for modelling of multidrug-resistant pathogen spread which was commonly used for studying the effectiveness of disease-control strategies. In order to derive a structure of healthcare network in German federal states, an analysis of anonymized data from patient hospitalization databases was presented, provided by three different German insurance companies, specific to each region: AOK (historically "general local health insurance company", but here only the abbreviation is used) Lower Saxony, AOK Bavaria, and AOK PLUS (merger of AOK Saxony and AOK Thuringia) for six years. In each dataset, following variables are available: anonymized patient ID, anonymized hospital facility ID, the federal state the hospital is located in, patient admission and discharge date, main diagnosis (ICD 10 GM code) as well as year of birth and sex of the patient.

The basic overall description of AOK data is presented in Table 2. The detailed distributions of population covered by AOKs according to the age and sex categories in given federal states in year 2015 are showed in Figure 1.

Table 2. Description of health insurance datasets.

Dataset	Time span	Population coverage*
AOK Bavaria	2010-2015	33.9%
AOK Lower Saxony	2010-2015	30.7%
AOK PLUS	2011-2016	44.0%

* Percentage of population in the federal state is based on the year 2015 (18).

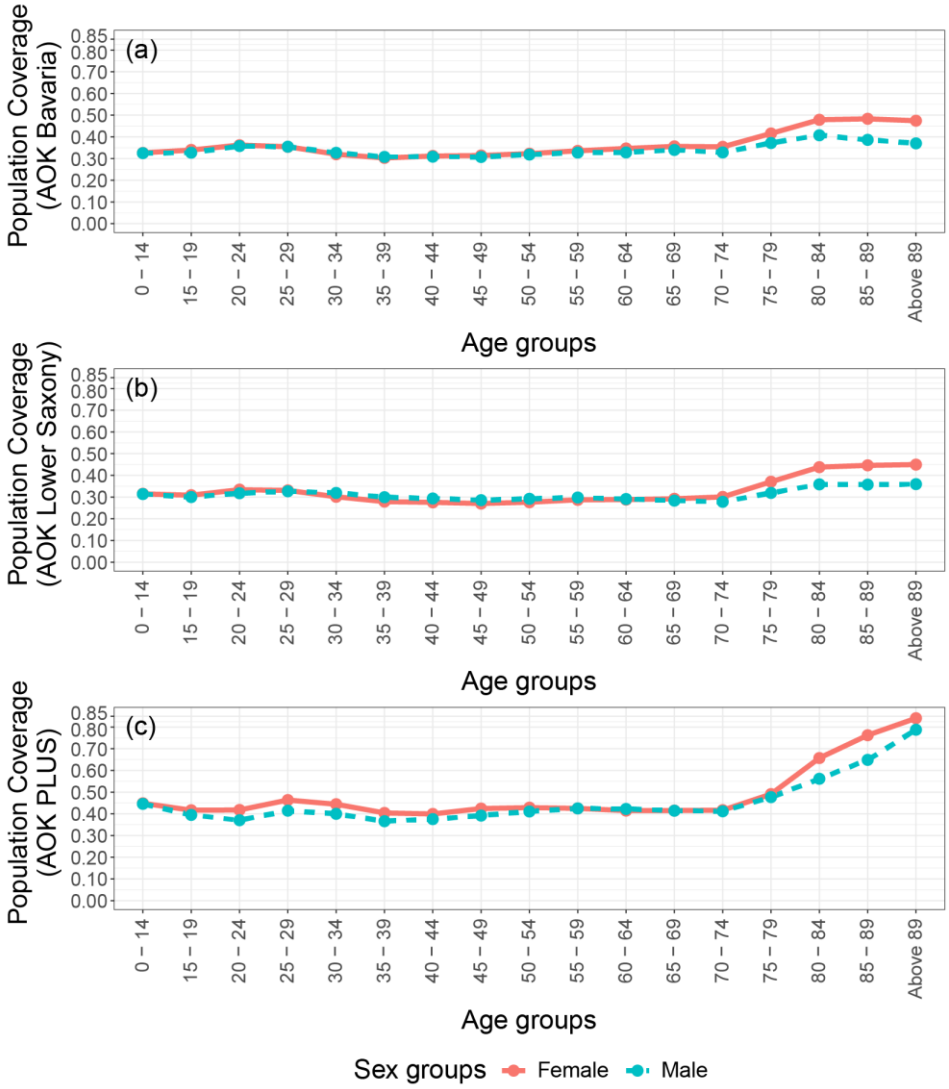

Figure 1. Proportions of population covered by AOK according to different sex and age groups in (a) Bavaria (b) Lower Saxony (c) Saxony and Thuringia in year 2015 (19).

1.4. Network measures

An adjacency matrix with entries which indicate the edge weights was always used to mathematically represent a weighted network: $A = (w_{ij})$ with w_{ij} indicating the edge weight from i to j. Here, a binary network (20, 21) with $w_{ij} = w_{ji}$ is used as an example and displayed in Figure 2 to explain the network measure.

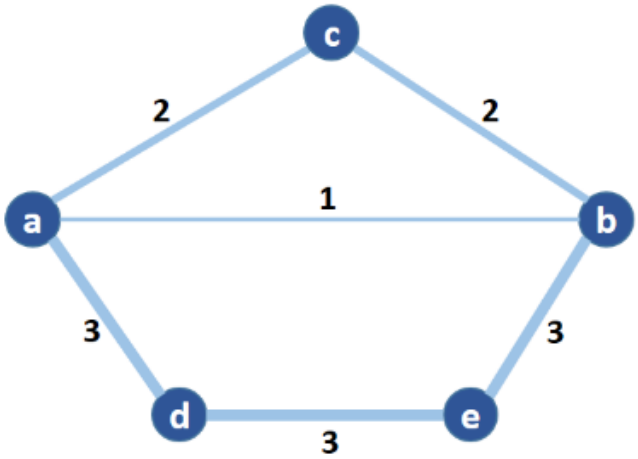

Figure 2. A binary network with five nodes and six edges. The numbers indicate the corresponding edge weights (own visualization)

In previous research, the network measure degree of a focal node was defined as the number of adjacencies in a network, i.e., the number of nodes that the focal node is connected to. To study a network, people often started with this network measure. Here, it can be formalized as $k_i = \sum_j^N sgn(w_{ij})$ where i is the focal node, j represents all other nodes, N is the total number of nodes, and sgn is the sign function with $sgn(x) = \begin{cases} -1, if\ x < 0; \\ 0\ , if\ x = 0; \\ 1\ , if\ x > 0. \end{cases}$ The degree can be extended to the measure strength, defined as sum of weights (20, 21) and formalized as $s_i = \sum_j^N w_{ij}$. However, these measures do not reflect the global structure of the network. For instance, a node might be connected to many others, but it might not be able to reach others quickly to assess resources, such as information or knowledge (22, 23). To capture this feature, network measure closeness was defined as the inverse sum of shortest distances to all other nodes from a focal node (21).

In a weighted network, edges with large weights might be considered to have greater impact than edges with small weights. For example. Infectious diseases are more likely to be transferred from one hospital to another if they have frequent patient transfers (2, 4, 7). The network in Figure 2 displays three different paths between nodes a and b, which are composed of different intermediate nodes and edges with heterogeneous weights. The path through the least intermediate nodes would be the direct connection $(\{a, b\})$, however, it might not be the quickest path for flow. The path $(\{a, d, e, b\})$ containing two intermediate nodes d and e could be quicker since it consists of stronger edges. For example, diseases might have higher probability of being

transmitted through a chain composed of more hospitals connected through more weighted edges than though a weak direct edge.

The methods for identifying shortest paths in weighted networks have been proposed in a massive amount of research (24, 25). Dijkstra (26) proposed an algorithm that detects the paths of least resistance and was defined for networks where the edges weights stood for costs of transmitting, for instance, distance in GPS devices or time route internet traffic. However, weights are operationalisations of edge strengths in most weighted networks and not the cost of them. Thus, the edges weights need to be reversed before directly utilizing Dijkstra's algorithm to detect the shortest paths in these networks. Newman (24) and Brandes (25) had independently proposed to invert the weights while extending network measures closeness and betweenness, respectively. Donker et al (7) had also applied this definition of shortest paths for analysing hospital networks. Following Opsahl et al and Donker et al, the length of the shortest path between two nodes i and j is defined as $d_{ij} = \min_h \left(\frac{1}{w_{ih}} + \cdots + \frac{1}{w_{hj}}\right)$. Table 3 shows the distance calculated by this algorithm for the three disparate paths in Figure 2. From this, it is easy to see that the distance between nodes a and b is not impacted by the number of nodes.

Table 3. Length of the paths in Figure 2.

Paths	Number of intermediate nodes	Distance
$(\{a, b\})$	0	1
$(\{a, c, b\})$	1	1
$(\{a, d, e, b\})$	2	1

The shortest path algorithm was used to define the network measure closeness with considering both the number of intermediary nodes and the edge weights: $C_i = \left[\sum_j^N d_{ij}\right]^{-1}$. Table 4 displays the network measures degree, strength, and closeness of each node. Some nodes have same values of degree or strengths, but their closeness is different. If the strength is only considered, the nodes a, b, d and e have the same connectivity, however, nodes d and e are more active than the others regarding higher closeness.

Table 4. Degree, strength, and closeness of network nodes in Figure 2.

Nodes	Degree	Strength	Closeness
a	3	6	0.40
b	3	6	0.40
c	2	4	0.38
d	2	6	0.46
e	2	6	0.46

1.5. Federal scaled-up datasets

To reduce the effect of distorted sex and age distribution (see Figure 1), AOK populations were extrapolated to fit to the sex and age distribution of the respective federal states by applying a sex and age stratified scale-up method. To compensate for patients which were not insured by AOK, for each sex and age group, patients were randomly resampled from the same sex and age group in the AOK datasets. For each resampling, once the patient was chosen, his/her information was duplicated and labelled as a new patient. The resampling procedure was repeated within each sex and age group until there was the same number of patients in the resampled dataset as there was in the corresponding total population of the federal state. These datasets were defined as federal scaled-up datasets. Here, 100 different federal scaled-up datasets were generated for each federal state. Table 5 shows some characteristics of hospitals and patient stays in three federal states for the scaled-up dataset.

Table 5. Basic characteristics of hospitals and patient stays in three federal states for the scaled-up datasets based on the AOK data (provided by German insurance companies: AOK Bavaria, AOK Lower Saxony and AOK PLUS).

Federal states	Bavaria	Lower Saxony	Saxony and Thuringia**
Time span (years)	2010-2015	2010-2015	2011-2016
Number of hospitals	357	211	126
Number of patients with at least one hospitalization during six years (CI*)	7 590 768 ([7 396 910,7 784 625])	4 287 973 ([4 204 486,4 371 461])	3 339 086 ([3 272 147,3 406 025])
Average number of occupied beds per hospital per day (CI*)	241.3 ([234.8,247.8])	223.2 ([218.7,227.7])	316.7 ([310.0,323.4])
Average length (days) of hospital stays (CI*)	9.1 ([9.1,9.1])	9.1 ([9.1,9.1])	9.7 ([9.7,9.7])
Average time (days) between two successive hospital stays (CI*)	255.6 ([255.4,255.8])	262.0 ([261.8,262.1])	242.9 ([242.8,243.0])
Average number of hospitalizations per patient during six years (CI*)	2.7 ([2.7,2.7])	2.7 ([2.6,2.7])	2.7 ([2.7,2.7])

* We generated 100 different federal scaled-up dataset samples and calculated the statistics above for each federal scaled-up dataset. Then we computed 95% confidence intervals (CIs) for each statistic.

** While we have data from two federal states provided by the AOK Plus, we cannot separate them in the dataset

The following histogram (Figure 3) shows how often patient were hospitalized within 6 years. It indicates that most patients (around 45%) were hospitalized only once.

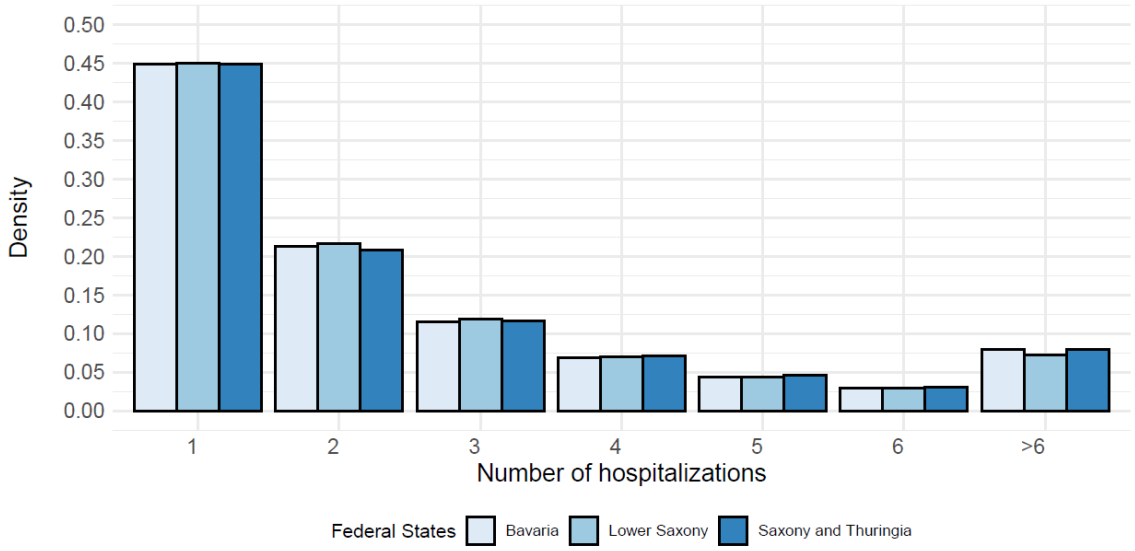

Figure 3. Distribution of the number of hospitalizations per person within six years by federal states.

1.6. Aims of the thesis

The aim of this work is to assess how different levels of data incompleteness affect the estimation of network measures and the spread of simulated infections (see publication 1). In order to better understand the local healthcare system and improve its structure to counteract the spread of epidemics. In the past, the presence of smaller subunits as hospital wards is overlooked in the transfer of patients between hospitals. However, it is possible for patients to transmit epidemics to patients in the same ward, rather than to all patients in the hospital. To fill this gap, the second aim of this study was to analyse patient flow patterns between hospital units to generate potential intra-hospital networks and to assess whether the flow patterns differed for patients at high or low risk (see publication 2).

2. Discussion

This doctoral thesis assesses the impact of data incompleteness and compares the network structures and epidemic spread in different German federal states. This dissertation is based on the publications that have addressed the research aims (Introduction 1.6).

2.1. Study main findings

The first publication demonstrates the studied network measures biased due to incomplete coverage of the population, however, the prevalence as a measure of epidemic spread is not much affected until the degree of incompleteness exceeds 90% (compared to the original dataset). Scale-up by "cloning" patients provides little improvement unless incompleteness is very high. Previous studies have shown that in the absence of a transmission model, the epidemiological risk of individual hospitals can be estimated by using their network measures (2, 4, 7, 27-29). However, in the first publication, it is proved that incomplete data can affect such estimates.

The degree and closeness show a strong deviation of around 25% at 50-60% incompleteness levels and close to 50% at 80% incompleteness levels. This effect is mainly due to the removal of weaker links, which are caused by missing patient records for transfers between specific hospitals. In terms of network measurement strength, this bias was reduced to 10% or less by using the scale-up method based on "cloned" patients. However, as the scale-up method cannot impute lost edges, it offers little benefit to other measurement methods.

Cosine similarity as a measure of affinity is relatively high even when incompleteness is high, possibly due to the fact that patient removal occurs randomly. The German health insurance system is fragmented across multiple companies, some of which are more similar to the others in terms of insurers, while some are characterised by more specific features. Nonetheless, there is a system of hospitals where patients are hospitalised. As a result, insurance data sets may differ in terms of average hospitalisation rates, but they will not produce different or differently weighted hospital networks.

Prevalence estimates as a measure of epidemic transmission in the network are robust until a high degree of incompleteness of over 90%. Considering the fact that the dataset itself is a subsample of the population of each state of the federation, a multiplication of 0.1 by 30-40% data incompleteness can be assumed, i.e., covering at least 3-4% of the local population can be used to represent the structure of the patient flow network in areas associated with pathogen transmission. In each area, there are several insurance companies which have a larger share of 5%, but more than 10% are usually only a few. In most federal states, AOKs insurance accounts for 30% or more of the local population, followed probably other more local insurers and the three largest national insurance companies. In this sense, each of the German federal states has 3-5 potential data with coverage requirements of more than 10%. They would provide the basis for a robust model of pathogen dissemination in regional hospital networks. In contrast, throughout Germany, there are 1-2 health insurance companies that meet this criterion in most federal states, excluding all local AOKs combined.

Since most patients were only hospitalized once and intra-hospital transfers played an important role in impacting prevalence at discharge, it is necessary to study the patient transfer patterns and epidemic spread inside hospitals.

The second publication described in detail the patterns of patient movement between hospital departments, based on data from several hospitals, and analyses how these movements affect the spread of bacterial pathogens within the hospital.

This study also included stratification by patient risk level and its impact on patient mobility and risk of infection. The risk of acquiring colonization may vary between patients and may be determined by several factors, such as functional status, immune response, chronic or severe disease. Taking these factors into account, patients were stratified into low- and high-risk groups based on certain ICD-10 codes. The results showed that high-risk patients were on average slightly older and had longer hospital stays compared to low-risk patients. High-risk patients also moved slightly more between wards per admission than low-risk patients. However, this higher number of moves may be due to their longer length of hospital stays. By visualizing the intra-hospital movement patterns of patients in participating hospitals as a network and using a discrete-event agent-based model, the impact of intra-hospital movements on pathogen transmission can be further assessed. The modelling results clearly showed that MDR-E can be easily spread out to all sections of one hospital from a highly centralized unit. The positive correlation between departmental centrality and mean departmental prevalence suggests that departments with high centrality end up with higher prevalence at higher transmission rates. Therefore, taking into account the central location of departments and the direction of patient flow in a given hospital can improve the efficiency of the intervention.

Patient flow within hospitals has been included in a small number of modelling studies for MRSA and MDR-E pathogens (27-29), but these studies have only considered a small number of special clinics and often ignored the structure of hospital departments (e.g., different departmental sizes). Rocha et al (30) and Pei et al (31) have recently studied the spread of MRSA in hospital networks. They developed a large-scale data-driven contact network model that included the dynamics of patient referrals. This includes the dynamics of patient referrals within and between wards and hospitals. As their model captures the interaction patterns emerging from time-varying contact data in the real world, they do not take the type and structure of hospital departments into account explicitly. Here, the sectors which are likely to play a key role in the spread of MDR-E was identified based on hospital-specific and network-centric data. In the absence of interventions, the simulations showed a strong positive correlation between the average incidence of each department and network characteristics (e.g., degree and weighted degree).

2.2. Study implication and strengths

HAIs threaten not only individual patient's health as well as global healthcare systems. Each year 3.1-4.6 million people acquire a HAI in acute care hospitals in European Union (EU) countries, Iceland, Norway, and the United Kingdom (1). HAIs increase morbidity and mortality and even trigger psychological problems for patients as well as a financial burden for the healthcare system. Each year, more than 90 000 people die in EU countries, including Iceland, Norway, and the United Kingdom due to the six most common HAIs (32). By studying the patient transfer patterns and modelling the HAI spread based on them, insights into patient transfer management, hospital infection prevention and control can be provided. This kind of studies requires a complete patient discharge data, for example researchers from England, Netherland as well as France (2, 7, 8) have used national medical registration datasets. However, this requirement cannot always be easily satisfied. Since intra-hospital transfers and the HAI spread based on them influences the prevalence of discharged patients, it is also necessary to study this topic systematically. However, most studies which studying spread of HAIs inside a hospital are limited to intensive care units (ICUs), or to some medical specialty departments.

This thesis has strength in terms of providing a further study of the above topics and its methodological aspects. Germany was taken as an example to study the impact of data incompleteness and a systematic method was delivered for scaling up the data. Situations and boarders which fit to this scale-up methods were also be clearly described. The study of intra-hospital patient transfers is not limited in single hospitals or only inside some special hospital departments. Considering the diversity of hospitals. It took data from different hospitals and even from various countries: University Medical Center Utrecht (UMCU), Utrecht, The Netherlands; Hospital Universitario Virgen Macarena (HUVM), Seville, Spain; Charité Universitätsmedizin (CUM), Berlin, Germany; Beilinson Hospital (BH), Rabin Medical Center, Petah Tikva, Israel; and Universitätsklinikum Halle (UKH), Halle, Germany. It investigated not only the patient transfer patterns but also provided a patient stratification methods according to the high and low risk of colonization based on ICD-10 codes.

For the future work, the methods provided in this thesis can help researchers alleviate the dependence of data completeness and design a more sophisticated model. Hospitals could be divided into different departments and patients can be align different colonization rate and recovery rate by using the risk stratification. Based on these information, the transfer probability matrix in previous work (33) can be refined and thus, a better deterministic SIS model which includes not only the inter-hospital transfer but also the intra-hospital patient movements can be delivered. Obviously, the patient-movement based epidemic model in the first publication can also be improved.

2.3. Study limitations

For investigating inter-hospital transfers, data from only three insurance companies from selected regions are used, therefore, the findings might not be generalizable to other German regions or to other countries. There could also be differences between insurance companies with respect to characteristics of the insured patients, which could again affect generalization of the findings even within the studied regions. For example, a dataset from an insurance company with younger insured patients who have less hospitalizations could tolerate lower fraction of the included population before too many transfer links are lost and epidemic model results are biased. In addition, the transmission model is relatively simple. More complex infections could bring additional problems which were not addressed in the study. The assumption is that that all susceptible patients had the same probability of becoming infected within each hospital although it is known that the physical and organizational structures of a hospital affect these numbers.

For the study of intra-hospital transfers, disparities in the department structure exist between hospitals, e.g., two departments in one hospital might be merged as single department in another hospital. The structure in every hospital is different and a generalization of the hospital structure is therefore not possible. In this study, the German hospitals did include data from emergency departments, but other hospitals did not include emergency departments in their datasets, such that it is difficult to compare the hospitals with each other.

For both studies, with respect to a successful transmission, a patient is assumed to become immediately infectious or colonized, and can transmit the pathogen to others, which may not be the case in reality. In future work, a latent period can be included to allow some time delay for a patient to acquire enough bacterial load before transmitting to others. We do not expect strong influence of a latent period on the correlation results, but a latent period might affect prevalence levels. Moreover, it can be considered to include a direct pathway from susceptible to an infected disease state.

2.4. Conclusions

Despite the bias in hospital network measures at higher levels of incompleteness, networks based on incomplete data still maintain similar patterns of patient transfer. Furthermore, even with high levels of incompleteness, the prevalence simulated in the patient-movement based model show only small deviations. At the upper limit of the incompleteness level, the scale-up method improves the robustness of the model. While incompleteness of patient transfer data remains a challenge, robust estimates can be achieved despite incomplete data across a wide range of assumptions. The scale-up methods can be easily expanded to approximate the federal datasets by using sex and age groups.

The intra-hospital study highlights the importance of patient flow within hospitals and their impact on the spread of pathogens. By targeting interventions at hubs, i.e., departments with a higher degree and weighted degree of centrality, it may help to control the spread of MDR-E. Furthermore, when the colonisation status of patients from different departments is unknown, a departmental ranking system based on network centrality may be used to improve the efficiency of interventions.

3. References

1. Suetens C, Latour K, Karki T, Ricchizzi E, Kinross P, Moro ML, et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. Euro Surveill. 2018;23(46).
2. Donker T, Wallinga J, Grundmann H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. PLoS Comput Biol. 2010;6(3):e1000715.
3. Lee BY, McGlone SM, Song Y, Avery TR, Eubank S, Chang CC, et al. Social network analysis of patient sharing among hospitals in Orange County, California. Am J Public Health. 2011;101(4):707-13.
4. Donker T, Wallinga J, Slack R, Grundmann H. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. PLoS One. 2012;7(4):e35002.
5. Belik V, Hövel P, Mikolajczyk R. Control of Epidemics on Hospital Networks. 2016. p. 431-40.
6. Belik V, Karch A, Hövel P, Mikolajczyk R. Leveraging Topological and Temporal Structure of Hospital Referral Networks for Epidemic Control. 2017. p. 199-214.
7. Donker T, Smieszek T, Henderson KL, Johnson AP, Walker AS, Robotham JV. Measuring distance through dense weighted networks: The case of hospital-associated pathogens. PLoS Comput Biol. 2017;13(8):e1005622.
8. Nekkab N, Astagneau P, Temime L, Crepey P. Spread of hospital-acquired infections: A comparison of healthcare networks. PLoS Comput Biol. 2017;13(8):e1005666.
9. Die gesetzlichen Krankenkassen: Anzahl der Krankenkassen im Zeitablauf-Konzentrationsprozess durch Fusionen: GKV-Spitzenverband; [cited 2023/03/04]. Available from: https://www.gkv-spitzenverband.de/krankenversicherung/kv_grundprinzipien/alle_gesetzlichen_krankenkassen/alle_gesetzlichen_krankenkassen.jsp#:~:text=Waren%20es%201970%20noch%201815,Januar%202023.
10. Harris AD, Kotetishvili M, Shurland S, Johnson JA, Morris JG, Nemoy LL, et al. How important is patient-to-patient transmission in extended-spectrum beta-lactamase Escherichia coli acquisition. Am J Infect Control. 2007;35(2):97-101.
11. Harris AD, Perencevich EN, Johnson JK, Paterson DL, Morris JG, Strauss SM, et al. Patient-to-patient transmission is important in extended-spectrum beta-lactamase-producing Klebsiella pneumoniae acquisition. Clin Infect Dis. 2007;45(10):1347-50.
12. Ajao AO, Johnson JK, Harris AD, Zhan M, McGregor JC, Thom KA, et al. Risk of acquiring extended-spectrum beta-lactamase-producing Klebsiella species and Escherichia coli from prior room occupants in the intensive care unit. Infect Control Hosp Epidemiol. 2013;34(5):453-8.
13. Gurieva T, Dautzenberg MJD, Gniadkowski M, Derde LPG, Bonten MJM, Bootsma MCJ. The Transmissibility of Antibiotic-Resistant Enterobacteriaceae in Intensive Care Units. Clin Infect Dis. 2018;66(4):489-93.
14. Smit PW, Stoesser N, Pol S, van Kleef E, Oonsivilai M, Tan P, et al. Transmission Dynamics of Hyper-Endemic Multi-Drug Resistant Klebsiella pneumoniae in a Southeast Asian Neonatal Unit: A Longitudinal Study With Whole Genome Sequencing. Front Microbiol. 2018;9:1197.
15. Gesundheitssystem Deutschlands: WIKIPEDIA; [cited 2023/03/04]. Available from: https://de.wikipedia.org/wiki/Gesundheitssystem_Deutschlands.
16. Gesetzliche Krankenversicherung: WIKIPEDIA; [cited 2023/03/04]. Available from: https://de.wikipedia.org/wiki/Gesetzliche_Krankenversicherung.
17. Gesundheit Bf. Daten des Gesundheitswesens 2019. Bundesgesundheitsministerium für Gesundheit; 2019.

18. GmbH MBV. „dfg-GKV-Ranking – Liste der deutschen Krankenkassen“ nach Mitgliederzahlen. 2015.
19. Anzahl der Versicherten in der AOK nach Altersgruppe und Geschlecht im Jahr 2015 [Internet]. Statistica. [cited 2016/10/01]. Available from: https://de.statista.com/statistik/daten/studie/407958/umfrage/anzahl-der-versicherten-in-der-aok-nach-altersgruppe-und-geschlecht/.
20. Newman MEJ. Analysis of weighted networks. Phys Rev E. 2004;70(5):056131.
21. Tore Opsahl FA, John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010;32(3):245-51.
22. Brass DJ. Being in the Right Place: A Structural Analysis of Individual Influence in an Organization. Administrative Science Quarterly. 1984;29(4):518-39.
23. Borgatti SP. Centrality and network flow. Social Networks. 2005;27(1):55-71.
24. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E. 2001;64(1):016132.
25. U. B. A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology. 2001;25(2):163-77.
26. Dijkstra EW. A note on two problems in connexion with graphs. Numerische Mathematik. 1959;1(1):269-71.
27. Deeny SR, Cooper BS, Cookson B, Hopkins S, Robotham JV. Targeted versus universal screening and decolonization to reduce healthcare-associated meticillin-resistant Staphylococcus aureus infection. J Hosp Infect. 2013;85(1):33-44.
28. Sadsad R, Sintchenko V, McDonnell GD, Gilbert GL. Effectiveness of hospital-wide methicillin-resistant Staphylococcus aureus (MRSA) infection control policies differs by ward specialty. PLoS One. 2013;8(12):e83099.
29. Vilches TN, Bonesso MF, Guerra HM, Fortaleza C, Park AW, Ferreira CP. The role of intra and inter-hospital patient transfer in the dissemination of heathcare-associated multidrug-resistant pathogens. Epidemics. 2019;26:104-15.
30. Rocha LEC, Singh V, Esch M, Lenaerts T, Liljeros F, Thorson A. Dynamic contact networks of patients and MRSA spread in hospitals. Scientific Reports. 2020;10(1):9336.
31. Pei S, Morone F, Liljeros F, Makse H, Shaman JL. Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. Elife. 2018;7.
32. Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank HP, Ducomble T, et al. Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study. PLoS Med. 2016;13(10):e1002150.
33. Piotrowska MJ, Sakowski K, Karch A, Tahir H, Horn J, Kretzschmar ME, et al. Modelling pathogen spread in a healthcare network: Indirect patient movements. PLoS Comput Biol. 2020;16(11):e1008442.

4. Theses

1) The current study investigates how varying degrees of data incompleteness affect the estimation of network structure and compares the associated epidemiological simulations in German inter-hospital networks.
2) Network measures (degree, intensity and closeness) are quite sensitive to data incompleteness, while prevalence, as a result of applying the susceptible-infected-susceptible (SIS) model, is quite robust to incompleteness.
3) By adopting the scale-up method proposed in this study, a coverage of less than 10% of the local population of the Federal State is sufficient to bring the relative deviation from the prevalence of the various transmission parameters encountered in clinical practice to less than 10%.
4) The current study emphasizes the importance of patient movements inside hospitals and their influence on pathogen spread.
5) Risk stratification of patients according to the ICD-10 codes revealed that length of hospital stay, patient age and average number of movements per admission were higher in the high-risk group.
6) When the colonization status of patients from different units is unknown, a ranking system based on unit centrality can be used to design more effective interventions to control the spread of pathogens.

Publications

List of publications on which the cumulative thesis is built

1. Xia H, Horn J, Piotrowska MJ, Sakowski K, Karch A, Tahir H, et al. Effects of incomplete inter-hospital network data on the assessment of transmission dynamics of hospital-acquired infections. PLoS Comput Biol. 2021;17(5): e1008941.
2. Tahir H, Lopez-Cortes LE, Kola A, Yahav D, Karch A, Xia H, et al. Relevance of intra-hospital patient movements for the spread of healthcare-associated infections within hospitals - a mathematical modeling study. PLoS Comput Biol. 2021;17(2):e1008600.

Publication 1

Contribution as an author:

I have contributed to the conception and design of the work, analysis of the data, method development and implementation, interpretation of results, writing and editing the manuscript. In addition, I was responsible for the whole submission process until the manuscript was published.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Effects of incomplete inter-hospital network data on the assessment of transmission dynamics of hospital-acquired infections

Hanjue Xia [1], Johannes Horn [1], Monika J. Piotrowska [2], Konrad Sakowski [2,3], André Karch [4], Hannan Tahir [5], Mirjam Kretzschmar [5], Rafael Mikolajczyk [1]*

1 Institute for Medical Epidemiology, Biometrics and Informatics (IMEBI), Interdisciplinary Center for Health Sciences, Medical School of the Martin-Luther University Halle-Wittenberg, Halle, Saxony-Anhalt, Germany, **2** Institute of Applied Mathematics and Mechanics, University of Warsaw, Warsaw, Poland, **3** Institute of High Pressure Physics, Polish Academy of Sciences, Warsaw, Poland, **4** Institute for Epidemiology and Social Medicine, University of Münster, Münster, North Rhine-Westphalia, Germany, **5** Julius Center for Health Sciences & Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

* rafael.mikolajczyk@uk-halle.de

Citation: Xia H, Horn J, Piotrowska MJ, Sakowski K, Karch A, Tahir H, et al. (2021) Effects of incomplete inter-hospital network data on the assessment of transmission dynamics of hospital-acquired infections. PLoS Comput Biol 17(5): e1008941. https://doi.org/10.1371/journal.pcbi.1008941

Editor: Benjamin Muir Althouse, Institute for Disease Modeling, UNITED STATES

Received: October 5, 2020

Accepted: April 6, 2021

Published: May 6, 2021

Data Availability Statement: The anonymized insurance data are owned by a third party (AOK Lower Saxony) and authors do not have permission to share them. These data may be requested from: AOK Bavaria, Carl-Wery- Straße 28, 81739 München; https://www.aok.de/pk/bayern/ AOK Lower Saxony: AOK Niedersachsen: Hildesheimer Straße 273, 30519 Hannover; https://niedersachsen.aok.de/ AOK PLUS, Sternplatz 7, 01067 Dresden; https://www.aok.de/pk/plus/.

Abstract

In the year 2020, there were 105 different statutory insurance companies in Germany with heterogeneous regional coverage. Obtaining data from all insurance companies is challenging, so that it is likely that projects will have to rely on data not covering the whole population. Consequently, the study of epidemic spread in hospital referral networks using data-driven models may be biased. We studied this bias using data from three German regional insurance companies covering four federal states: AOK (historically "general local health insurance company", but currently only the abbreviation is used) Lower Saxony (in Federal State of Lower Saxony), AOK Bavaria (in Bavaria), and AOK PLUS (in Thuringia and Saxony). To understand how incomplete data influence network characteristics and related epidemic simulations, we created sampled datasets by randomly dropping a proportion of patients from the full datasets and replacing them with random copies of the remaining patients to obtain scale-up datasets to the original size. For the sampled and scale-up datasets, we calculated several commonly used network measures, and compared them to those derived from the original data. We found that the network measures (degree, strength and closeness) were rather sensitive to incompleteness. Infection prevalence as an outcome from the applied susceptible-infectious-susceptible (SIS) model was fairly robust against incompleteness. At incompleteness levels as high as 90% of the original datasets the prevalence estimation bias was below 5% in scale-up datasets. Consequently, a coverage as low as 10% of the local population of the federal state population was sufficient to maintain the relative bias in prevalence below 10% for a wide range of transmission parameters as encountered in clinical settings. Our findings are reassuring that despite incomplete coverage of the population, German health insurance data can be used to study effects of patient traffic between institutions on the spread of pathogens within healthcare networks.

Funding: This work was supported by grants from the following national funding agencies: Bundesministerium für Bildung und Forschung (BMBF), Germany, 01KI1704C (Martin-Luther-University Halle-Wittenberg, Medical Faculty, Institute of medical epidemiology, biostatistics and informatics) to RM and National Science Centre, Poland, Unisono: 2016/22/Z/ST1/00690 (University of Warsaw, Faculty of Mathematics, Informatics and Mechanics, Institute of Applied Mathematics and Mechanics) to MJP and KS and the Netherlands ZonMw grant number 547001005 (Julius Centre, University Medical Centre Utrecht) to MK within the 3rd JPI AMR framework (Joint Programming Initiative on Antimicrobial Resistance) cofound grant no 681055 for the consortium EMerGE-Net (Effectiveness of infection control strategies against intra- and inter-hospital transmission of MultidruG-resistant Enterobacteriaceae). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Patterns of patients' transfer between different hospitals contribute crucially to the risk of hospital-acquired infections (HAIs) in the health care system. To quantify this risk, network models can be applied. The estimated risk can be inaccurate in the case of incomplete data on hospital admissions, which can be a consequence of the multiplicity of insurance companies as it is the case in Germany. To develop a better understanding of how incompleteness of data affects network measures and the simulated spread of HAI, we compared those measures derived from sampled, scale-up and original data, based on hospitalization data from three AOK insurance companies. We found that common network measures were affected by incompleteness, but the simulated prevalence as a measure of epidemic spread in the network was robust over a large range of incompleteness proportions. Epidemics and the transition of the infectious diseases may be modelled on hospital data with a coverage as low as 10% of the local population, whilst maintaining accuracy to within 10% of the true population prevalence.

Introduction

Transfers of patients between hospitals have an important impact on transmission pathways of hospital-acquired infections (HAIs) [1–11]. In recent years, hospital discharge databases based on English, Dutch as well as French national medical registration datasets have been used to construct "healthcare networks" to provide insights into patient transfer management, hospital infection prevention and control [2, 3, 5, 7, 8, 11]. In these networks, nodes represent hospitals and edges between pairs of nodes represent patient transfers between the linked pairs of hospitals. Based on these data, network measures like degree, closeness, and also network density were calculated [4, 5, 8]. The networks were used for simulating the spread of HAIs, evaluating epidemic risk [1–3, 5–7, 9–11], and recommending control strategies [1, 5, 11, 12].

In Germany, a central discharge database does not exist. In 2020, there were 105 statutory and several private health insurance companies in Germany. More than 90% of the population are insured by any one of the statutory health insurance companies and every insurant can freely choose any of these at any time, so that changes between companies occur frequently. It is very challenging to obtain data from all insurance companies, so that any projects based on empirical data will suffer from incomplete data access. Statistical properties of the network and the modelling predictions based on these incomplete data can be biased. The incomplete data may for example lead to missing edges in a network graph representing hospital connections. For person-to-person contact networks, incompleteness is a common problem; thus, several studies have focused on inferring network statistics from incomplete contact data in various contexts [13–22]. Nevertheless, the effects of incomplete patient transfer data have not been studied previously.

Thus, in this paper we aimed to assess how different levels of data incompleteness affect estimates of network measures and modelled infection spread.

Materials and methods

Data processing

We used anonymized data from hospitalization databases, provided by three different German insurance companies, specific to each region: AOK (historically "general local health insurance company", but currently only the abbreviation is used) Lower Saxony, AOK Bavaria, and AOK

PLUS (merger of AOK Saxony and AOK Thuringia). AOKs are insurance companies, which historically exclusively insured persons from the federal states where they were founded. As a consequence, they have high coverage of the population in their own federal state and low coverage outside, and can thus be used to study regional networks. In each dataset, the following are available: anonymized patient ID, anonymized hospital facility ID, the federal state the hospital is located in, admission and discharge date, main diagnosis (ICD 10 GM code) as well as year of birth and sex of the patient.

The three datasets covered different time spans. To ensure comparability, we studied data from each across the interval of same length of six-years. Some hospitals were not active during the whole time and based on observed distribution of hospitalizations, we excluded hospitals with less than 100 hospitalizations during this six-year time period from further analysis (see S1 Fig). We also excluded all hospitals located outside the respective federal state as they did not have the same local coverage of the studied population in their catchment area. The datasets contained various types of overlaps, which suggested that a patient stayed in multiple hospitals on a given day [23, 24]. Overlaps can happen due to several factors. If, for example, a hospitalization is completely included in another hospitalization at a different hospital, the patient may have been moved from one hospital to another before returning to the originating facility. In this case, for reimbursement purposes, the second stay had to be considered as part of the first, resulting in the record of continuous hospitalization in the first hospital. However, not all overlaps could be explained following this logic. If patients appeared in multiple hospitals simultaneously at a one-day overlap (indicating possibly a coding error for a direct transfer), we randomly chose one of hospitals. The filtering procedures are in detail in Fig 1.

Surrogate data construction

Construction of sampled and scale-up data. We considered the original data provided by the companies as a starting point (Table 1). By randomly removing a fraction of patients from these datasets, we obtained a subset with remaining patients, defined as sampled dataset.

In the next step, we randomly resampled patients from the sampled datasets to compensate for the excluded patients. For each resampling, once the patient was chosen, their information was cloned and labelled as a new patient. The progress was repeated until there was the same number of patients in the resampled dataset as there was in the original dataset.

Network construction

Two distinct types of networks were defined to represent the patient transfer data. One type was a time-independent (static) network, defined as a hospital network as by Iwashyna et al. [27], where patient transfers between the hospitals within a certain period were described. This network had N nodes and e directed weighted edges. In this network, nodes represented the hospitals and edges starting from node i towards node j with weight indicating the total number of patient transfers from hospital i to hospital j during the whole time period [28]. From this static network, we further extracted two different kinds of subnetworks: The first one consisted only of patient transfers from one hospital to another with no stay at home in between; we referred to this as the "direct hospital network". The other network consisted solely of indirect transfers, i.e. transfers where patients stayed at home for at least one day between discharge from one hospital and admission to the next one. The resulting transfer network, we termed the "indirect hospital network".

The second type of network was a time-dependent (transient) two-mode network [29], named patient-movement network (also known as affiliation or bipartite network). In this approach, there were two distinct sets of nodes: one was the set of all individuals, and the other

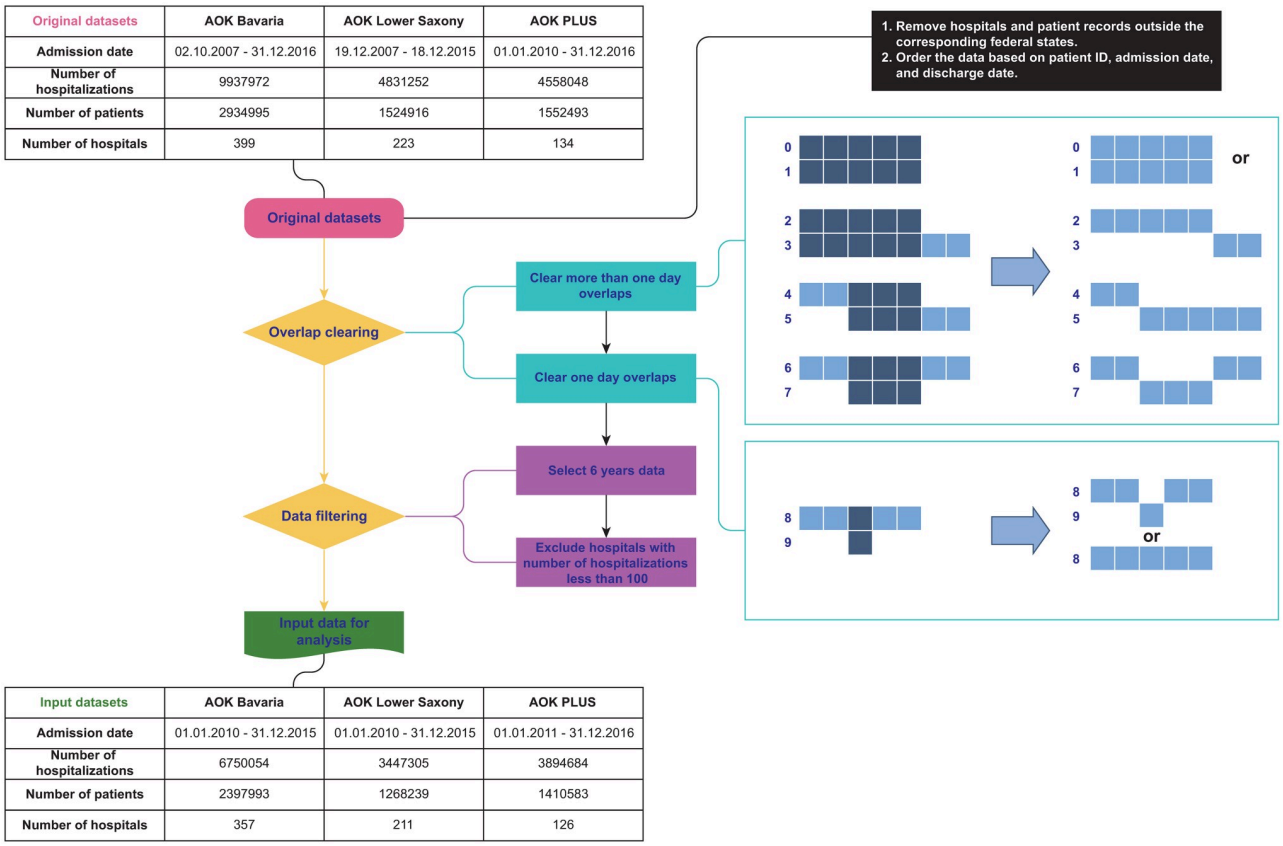

Fig 1. Data filtering process.

https://doi.org/10.1371/journal.pcbi.1008941.g001

one was the set of all hospitals (as well as one community node for individuals currently not hospitalized). In this type of network, links existed only between nodes belonging to different sets. This network was composed of M patient nodes and $N + 1$ location nodes. All nodes in this network were preserved during the covered period.

In our study, we used the static network for calculating the network measures, and the transient network for running the epidemic model, which obeyed the actual patient locations.

Description of network measures and epidemic models

Since network measures are widely utilized for describing the importance of hospitals in hospital networks, we adopted some commonly used ones for describing hospital network structures.

Table 1. Description of health insurance datasets.

Dataset	Time span	Population coverage*	Hospitalization coverage*
AOK Bavaria	2010–2015	33.9%	38.3%
AOK Lower Saxony	2010–2015	30.7%	34.2%
AOK PLUS	2011–2016	44.0%	40.3%

* Percentage of population and all hospitalizations in the federal state based on the year 2015 [25, 26].

https://doi.org/10.1371/journal.pcbi.1008941.t001

Hospital network measures. Each static weighted directed network $G(E, V)$ with E the edge set and V the vertex set can be represented by a $N \times N$ adjacency matrix A, defined as

$$A_{ij} = \begin{cases} w_{ij}, & \text{if } \{i,j\} \in E \text{ and } i \neq j; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Here, network weight w_{ij} denotes the total number of patient transfers during 6 years from node i to j and N is the number of hospitals. To depict the network cohesion, we calculated the network density, defined as $\rho = \frac{e}{N^2}$ where e denotes the number of network edges. Moreover, we considered four additional network measures with further subdivision regarding to directions: in- and out-degree, in- and out-strength, the shortest path, as well as in- and out-closeness.

Degree and strength were used for describing the transfer activity of a node. For node i, its in- and out-degrees are defined by $k_i^{in} = \sum_{j=1}^{N} sgn(w_{ji})$ and $k_i^{out} = \sum_{j=1}^{N} sgn(w_{ij})$, respectively, which indicate the number of neighbour hospitals of hospital i. Analogously, its in- and out-strengths are given by $s_i^{in} = \sum_{j=1}^{N} w_{ji}$ and $s_i^{out} = \sum_{j=1}^{N} w_{ij}$, which stand for the number of ingoing and outgoing patients of hospital i, respectively. sgn is the sign function with

$$sgn(x) = \begin{cases} -1 & , \text{ if } x < 0; \\ 0 & , \text{ if } x = 0; \\ 1 & , \text{ if } x > 0. \end{cases} \quad (2)$$

With respect to pathogen propagation, it was essential to determine the distances or shortest paths between hospitals [8]. The shortest path through the static weighted directed network from node i to node j is defined as $d_{ij} = \min_h \left(\frac{1}{w_{ih}} + \cdots + \frac{1}{w_{hj}} \right), h \in V \backslash \{i, j\}$. It was used for calculating the measure "closeness" [8, 30].

In addition, we calculated the in- and out-closeness centralities of node i: $C_i^{in} = \sum_{j=1}^{N} \frac{1}{d_{ji}}$ also $C_i^{out} = \sum_{j=1}^{N} \frac{1}{d_{ij}}$ [8, 30]. The closeness determines the risk posed by an outbreak in any of the other hospitals [8]. In-closeness C_i^{in} indicates that the outbreak risk in hospital i was triggered by receiving patients from other hospitals. Out-closeness C_i^{out} shows the risk caused by hospital i in its neighbour hospitals.

Patient-movement based epidemic models. We used a susceptible-infectious-susceptible (SIS) model and applied parameters for hospital-acquired methicillin-resistant Staphylococcus aureus (MRSA). Patients colonized or infected with MRSA were defined as "Infectious". In the following, we defined "become infected" not exclusively as infection, but also as colonization by the pathogen. On modelling day t, we first identified the location of every patient based on the transient network. Then, by keeping their epidemic statuses on day $t - 1$, we calculated the total number of patients N_L^t as well as the number of infectious patients $N_{L,I}^t$ in location L. Finally, we ran the following epidemic models and calculated the prevalence of infection in hospitals and the community node, defined as the proportion of infectious patients according to the patient status on day t.

In our model, a susceptible patient became infected with probability β_L^t on day t through contacts with infectious patients at the same location L on day t. The probability of becoming

infected at location L on day t was

$$\beta_L^t = \begin{cases} \frac{N_{L,I}^t}{N_L^t - 1} \cdot \beta, & N_{L,I}^t \geq 1 \text{ and } N_L^t \geq 2; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Here, $N_{L,I}^t$ denoted the number of infectious patients at location L on day t and N_L^t was the total number of patients with $t = T_1, T_1 + 1, \cdots, T_2$. T_1 and T_2 were the start and end days of the simulation. Here, we assumed that the population mixed homogeneously within each location.

For the initial state of every simulation, we randomly selected 4% of the patients to be infectious [31–33]. We assumed that no transmission occurred in the community node, and that the discharged patients remained infectious until they recovered. Infectious patients recovered spontaneously with probability γ per day. Following Scanvic et al. [34] and Donker et al. [11], we set $\gamma = \frac{1}{365} days^{-1}$ based on the assumption that the mean time of MRSA colonization was 365 days. We assumed that the recovery probabilities in different hospitals and the community were identical. Following the previous work of Donker et al. [11], we assumed the reference transmission probability per patient contact to be $\beta = 0.085$ in hospital nodes and $\beta = 0$ in the community node. Since we used stochastic approach to simulate pathogen spread within the network, we performed 100 independent simulations for each set of values (e.g. different levels of data incompleteness).

Measures used for comparison of network characteristics and spread of pathogens.

- **Absolute value of relative bias (ARB)**: As different network measures, applied on the hospital network, yielded deviations in varying degrees, we computed the ARB for sampled and scale-up datasets in comparison with the original data at node i:

$$ARB_M(Net_O, Net_E)_i = \left| \frac{M_i^{(Net_O)} - M_i^{(Net_E)}}{M_i^{(Net_O)}} \right|,$$

where M_i denoted the weighted directed static network measure (degree, strength, or Closeness) at node i. Net_O and Net_E represented the network based on the original dataset and sampled or scale-up dataset, respectively. As the nodal ARB did not take into account that different nodes can correspond to very different number of patient hospitalizations and transfers, we also considered the weighted ARB, defined as:

$$ARB_M^w(Net_O, Net_E) = \sum_{i=1}^{N} H_i^W \cdot ARB_M(Net_O, Net_E)_i \quad ,$$

where $H_i^W = \frac{H_i(Net_O)}{\sum_{j=1}^{N} H_j(Net_O)}$, H_i was the total number of patient admissions at node i and N was the number of hospitals.

The values for ARB and weighted ARB ranged from 0 to 1. Values close to 0 indicated that the estimation based on the surrogate data was close to the one based on the original data.

- **Cosine similarity (CS)**: We further applied CS as a similarity measure defined for vectors. The value of CS was bounded between −1 and + 1, with a value of 1 for parallel vectors with the same orientation, a value of −1 for parallel vectors with opposite orientation, and 0 for

perpendicular vectors [35]. The local CSs of node i were defined as:

$$CS^{in}(Net_O, Net_E)_i = \frac{\sum_j w_{ji}^{Net_O} \cdot w_{ji}^{Net_E}}{\sqrt{\sum_j (w_{ji}^{Net_O})^2}\sqrt{\sum_j (w_{ji}^{Net_E})^2}} \quad (4)$$

$$CS^{out}(Net_O, Net_E)_i = \frac{\sum_j w_{ij}^{Net_O} \cdot w_{ij}^{Net_E}}{\sqrt{\sum_j (w_{ij}^{Net_O})^2}\sqrt{\sum_j (w_{ij}^{Net_E})^2}} \quad (5)$$

$w_{ij}^{Net_O}$ and $w_{ij}^{Net_E}$ were the numbers of total patient movements from hospital i to its neighbours j in Net_O and Net_E during considered years, respectively. In addition, to study the global similarity, we calculated the weighted CSs:

$$CS^{W,in}(Net_O, Net_E) = \sum_{i=1}^{N} H_i^W \cdot CS^{in}(Net_O, Net_E)_i \quad (6)$$

$$CS^{W,out}(Net_O, Net_E) = \sum_{i=1}^{N} H_i^W \cdot CS^{in}(Net_O, Net_E)_i \quad (7)$$

H_i^W was defined as in the above section on ARB.

- **Prevalence relative bias (PRB) and final PRB**: To evaluate the estimation of daily prevalence in the SIS model, we then computed PRB as follow:

$$PRB(Net_O, Net_E)_t = \frac{Pr_t^{Net_O} - Pr_t^{Net_E}}{Pr_t^{Net_O}} \quad (8)$$

Pr_t was the prevalence on day t with $t = T_1, T_1 + 1, \cdots, T_2$, which had been defined in the section "Patient-movement based epidemic models". The time-dependent PRB reflected the underestimated proportion of the daily prevalence if $PRB > 0$, while $PRB < 0$ indicated the overestimated proportion. Since PRB depended on time, we additionally defined a measure "final PRB" by averaging the PRB values in their steady state.

Threshold for preservation of transmission networks regarding incompleteness. Subsequently, we investigated at which threshold of incompleteness the network characteristics were still reasonably preserved and estimates of epidemic spread are maintained. We considered error levels of 5% and 10%, measured by final PRB as acceptable. In addition to the original transmission parameters, we studied the error levels for varying parameters for β and γ in the patient-movement based epidemic models.

Software. R package "tnet" [36] was used for creating and analysing the hospital networks. Own code was written for the stochastic SIS simulation. To visualize the data we used the R package "ggplot2" [37]. For all analyses, we used R version 4.0.3 [38].

Results

Description of the AOK networks

The data from AOK PLUS included more than 40%, whereas AOK Bavaria as well as AOK Lower Saxony included around 30% of the population of the corresponding federal states (Table 1 in section Materials and Methods).

Table 2. Basic description of AOK hospital networks.

Dataset	No. of hospitals (N)	ρ^{DT}	ρ^{in-DT}	$\bar{w}^{DT}$	$\bar{w}^{in-DT}$
AOK Bavaria	357	0.10	0.37	3.29	30.86
AOK Lower Saxony	211	0.13	0.41	4.47	44.47
AOK PLUS	126	0.28	0.61	12.25	144.23

The static networks had N nodes; ρ represented the network density with the subscripts "DT" and "in-DT" denoting the direct and indirect hospital networks. The "$\bar{w}$" columns gave the average number of patient transfers per day between different hospitals.

https://doi.org/10.1371/journal.pcbi.1008941.t002

The hospital network and patient characteristics of the original AOK datasets were presented in Tables 2 and 3. Based on the network densities ρ^{DT} and ρ^{in-DT} (see Table 2), we inferred that the indirect hospital networks had much more connections than the direct networks. The average number of hospitalizations and average length of stay (LOS) from different AOK datasets were close to each other, whereas the average LOS in AOK PLUS was substantially higher than in the other two datasets (Table 3).

To provide further information about the data, we showed also the distribution of LOS in S2–S4 Figs, of hospitalizations in S5 Fig, of network weights in S6 Fig, of in- and out-degrees of each hospital in S7 and S8 Figs, and of in- and out-strengths in S9 and S10 Figs. More detailed information about the AOK data can be also accessed in our previous work [23, 24].

Effects of incompleteness on estimated network characteristics

The static network measures: in- and out-degree, in- and out-strength, and in- and out-closeness were calculated for the sampled and scale-up datasets. The results were compared with the ones evolving from the original datasets by using the evaluation measure ARB (Fig 2).

Despite the fact that there were differences in basic network properties in the various insurance datasets, the ARBs of different network measures at different incompleteness levels were similar across these datasets (see also Supporting information S11 Fig). There was also no strong difference in ARBs between direct and indirect networks in each dataset. Among all network measures, only strength showed a difference between sampled and scale-up datasets.

To further assess the impacts of incompleteness on network characteristics, we used cosine similarity (CS). The findings are presented in S12–S15 Figs. The CS showed more stable results in the dataset of the AOK PLUS, which had less area of low CS values at the same incompleteness level as the other two datasets. When incompleteness reached 90%, several middle and small hospitals displayed CS values lower than 0.6 and 0.7 in AOK Bavaria and AOK Lower Saxony, while there were very few low CS values in AOK PLUS and only in some small hospitals. The indirect hospital networks were less affected by incompleteness than the direct hospitals. The scale-up of the sampled dataset to the original size provided little improvement.

Table 3. Basic description of AOK patient characteristics.

Dataset	No. of patients (M)	Avg. No. of hospitalizations per person	Avg. No. of persons per day per hospital	Avg. LOS
AOK Bavaria	2,397,993	2.81	76.80	8.90
AOK Lower Saxony	1,268,239	2.72	66.23	8.88
AOK PLUS	1,410,588	2.76	136.74	9.70

The transient network included M individuals who were insured by the AOKs and were hospitalized at least once during the 6 years period. The abbreviation "Avg." was short for "average", "No." for "number", and "LOS" for the "length of stay".

https://doi.org/10.1371/journal.pcbi.1008941.t003

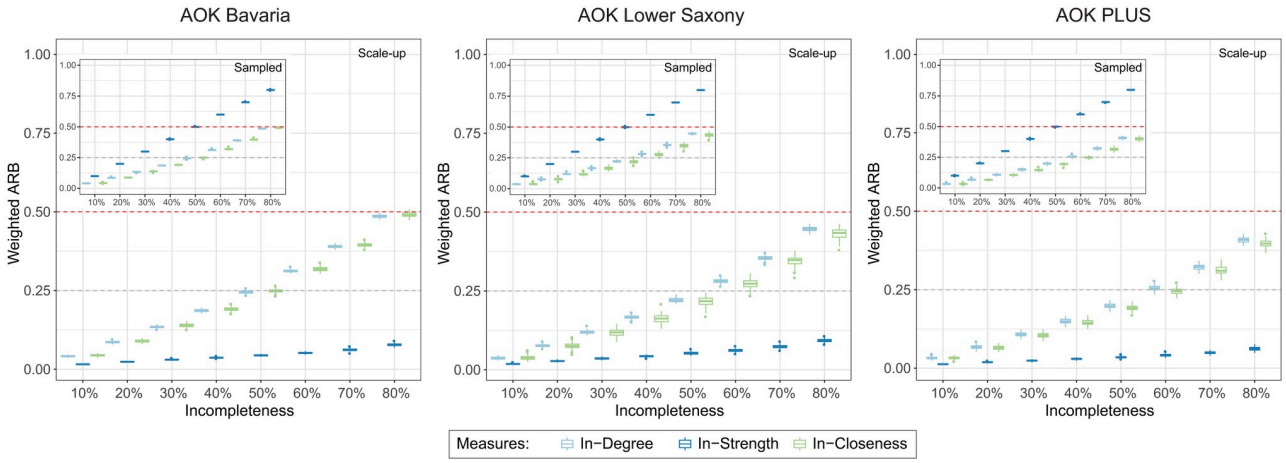

Fig 2. Impact of sampling and scale-up on direct hospital network measures (visualized in boxplots). The weighted absolute value of relative biases (ARBs) across different network measures on sampled (inset) and scale-up (main plot) datasets for various incompleteness levels. Incompleteness was defined as the percentage of removed patients from the original datasets.

https://doi.org/10.1371/journal.pcbi.1008941.g002

To investigate the overall effect of incompleteness on patient transfer patterns, we calculated the weighted CS. Direct networks were affected by lower degrees of incompleteness than indirect networks, which displayed CS above 0.98 up to incompleteness of 97–98% (Fig 3).

Effects of incompleteness on simulated spread of infections

The daily prevalence was hardly affected by incompleteness as high as 90% of the original datasets, when scale-up was applied (S16–S18 Figs). Even at higher incompleteness levels the scale-

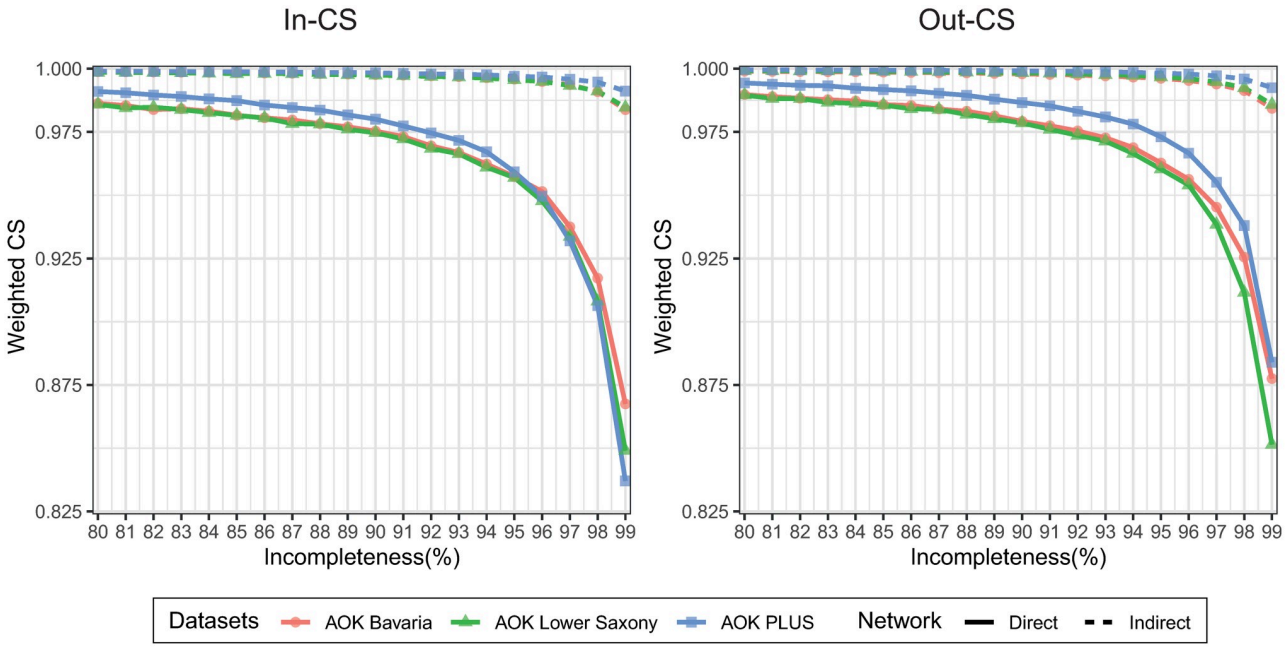

Fig 3. Weighted cosine similarities (CSs). Weighted CSs in directions "in" and "out" between the sampled networks and the whole hospital networks as a function of the incompleteness levels for three insurance datasets. The solid and dashed lines represented the weighted CSs on direct and indirect hospital networks based on these AOK datasets, respectively.

https://doi.org/10.1371/journal.pcbi.1008941.g003

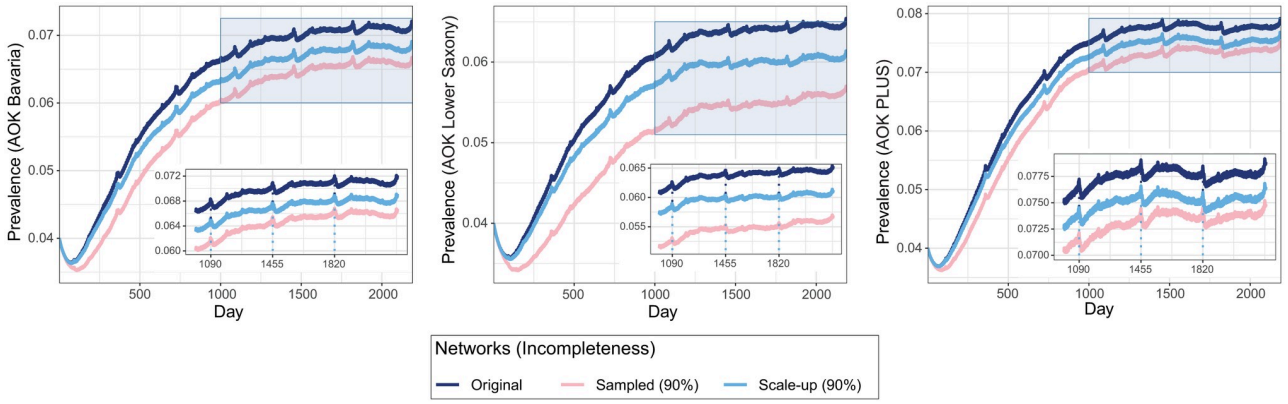

Fig 4. Prevalence in community nodes obtained from transmission model simulations for different AOK datasets. The dashed lines in each graph point the x-axis of prevalence peaks.

https://doi.org/10.1371/journal.pcbi.1008941.g004

up of the data markedly reduced the bias in estimation of prevalence, which slightly decreased for less complete data (see S16–S18 Figs). There were fluctuations in prevalence related to holidays and weekends when there were fewer patients in the hospitals (see Fig 4 and S19 Fig). This was especially true for the time between Christmas and New Year. These seasonal patterns were consistent across years and AOK datasets. Overall, incompleteness had less impact on prevalence in hospitals than on the prevalence in the community node.

Incompleteness threshold for preservation of transmission characteristics

After determining that the effects of incompleteness became visible, when incompleteness levels surpassed 90%, we further focussed on this area and studied the higher incompleteness more in detail (Fig 5). At this high incompleteness level, the scale-up datasets performed better than sampled datasets. In the sampled datasets, AOK Lower Saxony appeared less robust to higher incompleteness levels compared to the other two datasets. 10% PRB was reached already at 90% incompleteness in the AOK Lower Saxony and at 94–95% incompleteness in the AOK Bavaria and AOK PLUS datasets (for the community node prevalence). When additional, scale-up is applied, even higher incompleteness levels can be tolerated.

To further investigate the effects of incompleteness on the final PRBs based on the SIS model, we used the AOK PLUS data with varying β and γ values. We could show that even across a wide range of values (reflecting various pathogen with realistic characteristics) final PRBs were preserved for high incompleteness levels of 96% (see Fig 6 and S20 Fig). For even higher incompleteness levels, there was some indication that for some combination of parameters final PRBs became substantially larger.

Discussion

We demonstrated that while incomplete coverage of the population affected the studied network measures, this did not greatly bias the prevalence as a measure of epidemic spread until the level of incompleteness exceeded 90% (in relation to the original dataset). Scale-up by "cloning" the patients provided little improvement, unless for very high incompleteness levels.

Previous research demonstrated that the epidemic risk of single hospitals in the absence of transmission models can be estimated by utilizing their network measures [31]. Unfortunately, incomplete data would affect such estimates.

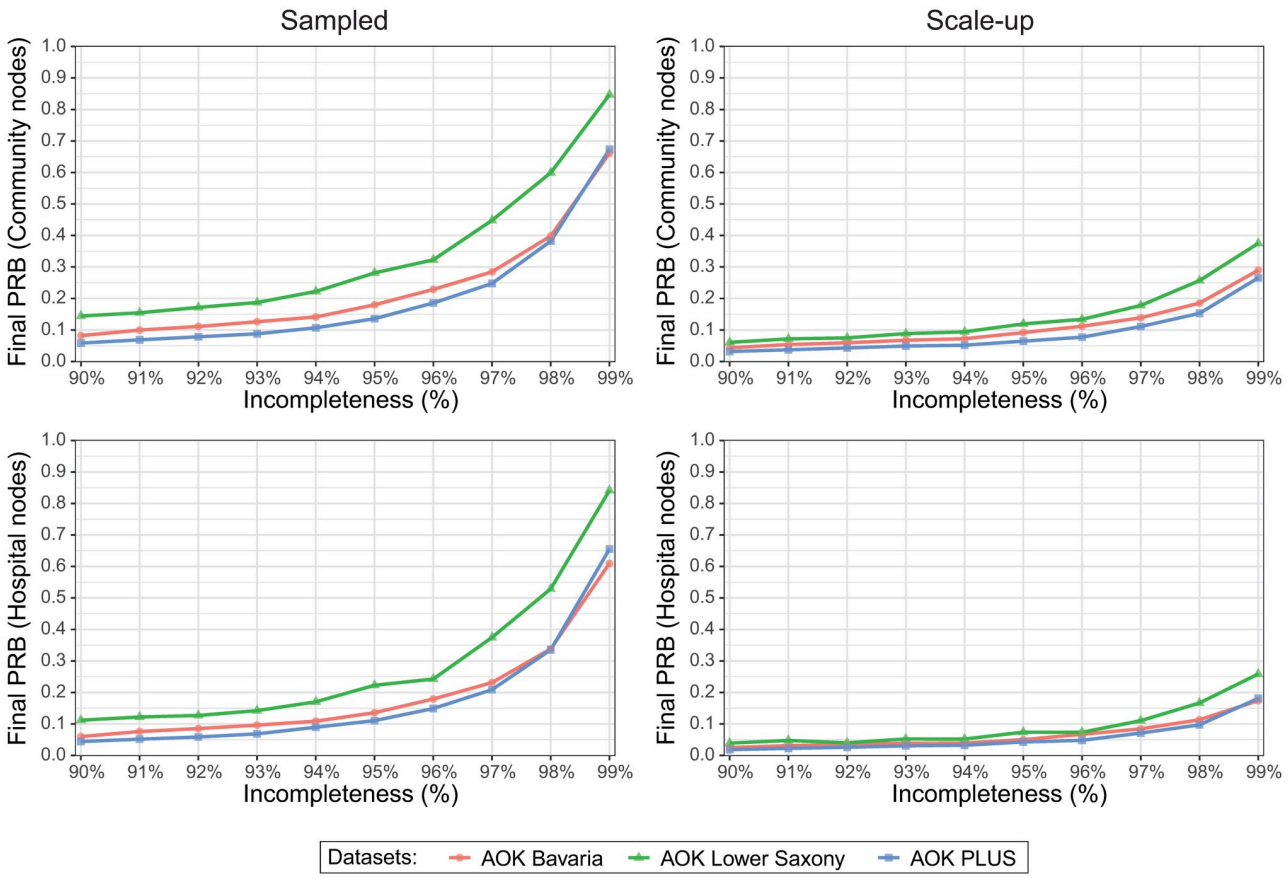

Fig 5. Final PRBs of simulated networks for each AOK plotted against percentage of incompleteness.

https://doi.org/10.1371/journal.pcbi.1008941.g005

Degree and closeness displayed strong bias of around 25% at incompleteness levels of 50%–60% and nearly 50% at an incompleteness level of 80%. This effect was mainly due to the removal of weaker links, caused by missing patient records with transfers between particular hospitals, which would underestimate the risk of pathogen transmission between those hospitals. In case of the network measure strength, this bias was reduced to 10% or less by adopting the scale-up method based on "cloning" of patients. However, since the scale-up could not impute lost edges, it provided little benefit with respect to the other measures.

The CS as measures of affinity was relatively high even at high incompleteness levels, likely due to the fact that removal of patients occurred at random, as demonstrated in previous research [14]. German health insurance system is scattered in multiple companies, with some of them being more similar the others in terms of insurants and some known for more specific characteristics. Nevertheless, there is one system of hospitals where the patients are admitted. So, the insurance datasets are likely to differ in average hospitalization rates, but they should not generate disparate or differently weighted hospital networks.

For modelling HAI spread in the healthcare system, we used transmission parameters based on previous research [2, 3, 7, 8, 11] and varied those across a range of clinically relevant parameters. For example, an important factor in modelling multidrug-resistant pathogens acquired in healthcare institutions is that patients can become carriers without displaying symptoms of infection. Consequently, they might remain colonized for a long time, i.e. either undetected and, therefore, not treated or because some of the pathogens are overall difficult to

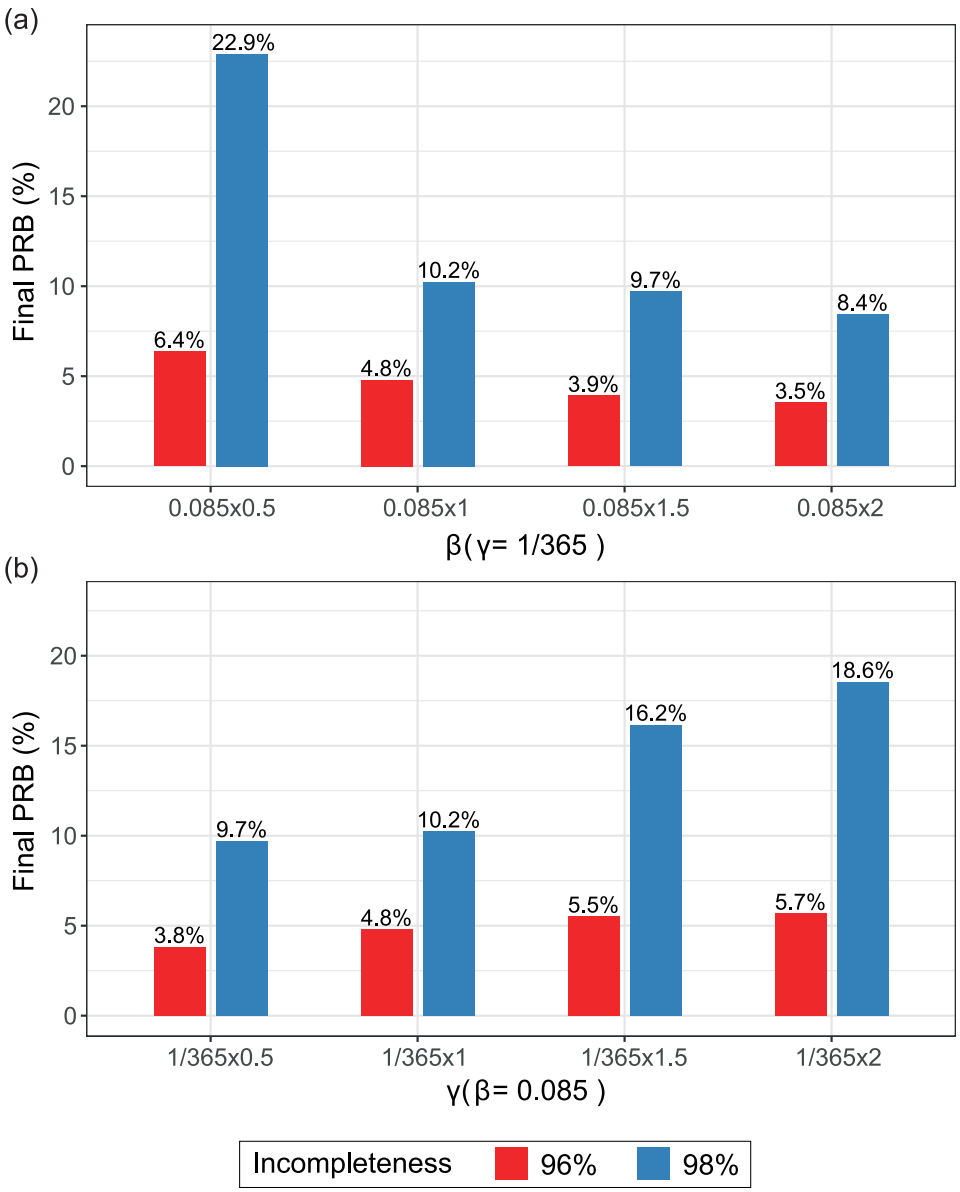

Fig 6. Final prevalence relative biases (PRBs) in hospital nodes for varying transmission parameters. The final PRBs in hospital nodes were calculated based on the simulations on AOK PLUS patient-movement networks, built by scale-up at incompleteness levels of 96% and 98%. In (a) we presented the final PRBs in hospital nodes for different values of β. In (b) we presented the final PRBs in hospital nodes for different values of γ.

https://doi.org/10.1371/journal.pcbi.1008941.g006

remove, despite decolonization efforts [39]. Some estimates of clearance rate are available. However, clearance might not be homogeneous, and some patients can carry the pathogen much longer. If such patients are readmitted to the same hospital, or possibly to another hospital, for the same or unrelated reason, they can reintroduce the pathogen. This is particularly the case, if such patients receive antibiotic treatment, which can lead to selection of resistant strains and uncover a partially hidden colonization.

Prevalence estimates as a measure of epidemic spread in the network were robust until high level of incompleteness of 90% and more. Accounting for the fact that the datasets were themselves subsamples of the populations of the federal states, one could assume that 0.1 times

30–40%, i.e. a minimum coverage of 3–4% of the local population is required to represent the patient-movement network structure in the region relevant for pathogen transmission. In each region, typically several insurance companies have a larger share of 5% of coverage, although above 10% there usually only few, with AOKs insuring 30% or more of the local population in most federal states, followed by possibly other more local insurance companies and any of the three largest countrywide companies. In this sense for each federal state in Germany, there are 3–5 potential data sources for such analyses, which fulfil the requirement of the coverage of more than 10% of the local population. They would provide basis for robust modelling of pathogen spread in regional hospital networks. Conversely, for the whole of Germany there 1–2 health insurance companies which fulfil this criterion in most of the federal states, in addition to the composite of all local AOKs taken together.

Limitations

We used data from only three insurance companies from selected regions, so that our findings might not be generalizable to other regions of Germany or to other countries. Also, there could be differences between insurance companies with respect to characteristics of the insurants, which could again affect generalization of our findings even within the studied regions. For example, a dataset from an insurance company with younger insurants who have less hospitalizations could tolerate lower fraction of the included population before too many transfer links are lost and epidemic model results are biased. In addition, we used only a relatively simple transmission model, and more complex infections could bring additional problems not addressed in our analyses. We assumed that all susceptible patients had the same probability of becoming infected within each hospital although it is known that the physical and organizational structures of a hospital affect these numbers [40, 41]. In our model, we also neglected the existence of smaller subunits as hospital wards. Patients are likely to transmit the pathogen to patients in the same ward and not to all patients in the hospital. Finally, we investigated only a range of assumptions regarding transmission parameters in the epidemiological model. While our focus was on parameters observed for MRSA and other multidrug-resistant pathogens transmitted in the hospitals, it could be that some specific pathogens were not covered in the chosen range or that new pathogens will emerge with different characteristics.

Conclusions

Although hospital network measures were biased when incompleteness levels were high, networks based on incomplete data still maintained similar patient transfer patterns. In addition, even up to high levels of incompleteness, simulated infection prevalence in SIS transmission models displayed only small bias. At the upper boundary of incompleteness levels, scale-up improved the robustness of patient-movement based transmission model. While incompleteness of patient transfer data remains a challenge, our findings are reassuring that across a broad range of assumptions, robust estimates can be reached despite incomplete data.

Supporting information

S1 Fig. Distribution of number of hospitalizations per hospital in the three AOK datasets including hospitals with less than 100 hospitalizations during six-year time period. The x-axis represents hospitals, ranked from largest to smallest (left to right) based on the number of hospitalizations that occurred there. The dash lines indicate the bounds of 100 hospitalizations.
(EPS)

S2 Fig. Density distribution of length of stay based on age and sex categories in AOK Bavaria. The upper and lower panels show the distributions for female and male patents, respectively.
(EPS)

S3 Fig. Density distribution of length of stay based on age and sex categories in AOK Lower Saxony. The upper and lower panels show the distributions for female and male patents, respectively.
(EPS)

S4 Fig. Density distribution of length of stay based on age and sex categories in AOK PLUS. The upper and lower panels show the distributions for female and male patents, respectively.
(EPS)

S5 Fig. Distribution of number of hospitalizations per hospital in the three AOK datasets. The x-axis represents hospitals, ranked from largest to smallest (left to right) based on the number of hospitalizations that occurred there.
(EPS)

S6 Fig. Distribution of weights for hospital edges. Here, the "Direct" and "Indirect" indicates the direct and indirect hospital networks, respectively.
(EPS)

S7 Fig. Distribution of degrees on direct hospital networks. The x-axis represents hospitals, ranked from large to small (left to right) based on the number of hospitalizations that occurred there.
(EPS)

S8 Fig. Distribution of degrees on indirect hospital networks. The x-axis represents hospitals, ranked from large to small (left to right) based on the number of hospitalizations that occurred there.
(EPS)

S9 Fig. Distribution of strengths on direct hospital networks. The x-axis represents hospitals, ranked from large to small (left to right) based on the number of hospitalizations that occurred there.
(EPS)

S10 Fig. Distribution of strengths on indirect hospital networks. The x-axis represents hospitals, ranked from large to small (left to right) based on the number of hospitalizations that occurred there.
(EPS)

S11 Fig. Impact of sampling and scale-up on static direct hospital subnetwork measures in boxplots. We plotted the weighted absolute value of relative biases (weighted ARBs) across different network measures on sampled (inset) and scale-up (main plot) static hospital networks for various incompleteness levels of patient removed. The boxplots showed the progressive disparities of ARBs of diverse network measures as incompleteness increases. Each row specified the type of hospital subnetwork and the legend in the right side showed the corresponding network measures. Each column involved the graphs from the same AOK data.
(EPS)

S12 Fig. Cosine similarities (CSs) in direction "In" between sampled and original hospital networks according to 9 incompleteness levels. The right y-axis displays incompleteness levels. The bar plots on the top of the heat maps indicate hospital sizes, ranking from highest to lowest (left to right) based on the number of hospitalizations that occurred there, which also defines the x-axis.
(EPS)

S13 Fig. Cosine similarities (CSs) in direction "Out" between sampled and original hospital networks according to 9 incompleteness levels. The right y-axis displays incompleteness levels. The bar plots on the top of the heat maps indicate hospital sizes, ranking from highest to lowest (left to right) based on the number of hospitalizations that occurred there, which also defines the x-axis.
(EPS)

S14 Fig. Cosine similarities (CSs) in direction "In" between scale-up and original hospital networks according to 9 incompleteness levels. The right y-axis displays incompleteness levels. The bar plots on the top of the heat maps indicate hospital sizes, ranking from highest to lowest (left to right) based on the number of hospitalizations that occurred there, which also defines the x-axis.
(EPS)

S15 Fig. Cosine similarities (CSs) in direction "Out" between scale-up and original hospital networks according to 9 incompleteness levels. The right y-axis displays incompleteness levels. The bar plots on the top of the heat maps indicate hospital sizes, ranking from highest to lowest (left to right) based on the number of hospitalizations that occurred there, which also defines the x-axis.
(EPS)

S16 Fig. Daily prevalence PRBs in AOK Bavaria according to incompleteness levels equal or above 90%. The red and dark blue dashed lines indicate the bounds of 5% and 10% estimation biases.
(EPS)

S17 Fig. Daily prevalence PRBs in AOK Lower Saxony according to incompleteness levels equal or above 90%. The red and dark blue dashed lines indicate the bounds of 5% and 10% estimation biases.
(EPS)

S18 Fig. Daily prevalence PRBs in AOK PLUS according to incompleteness levels equal or above 90%. The red and dark blue dashed lines indicate the bounds of 5% and 10% estimation biases.
(EPS)

S19 Fig. Prevalence in hospital nodes obtained from transmission model simulations for different AOK datasets. The dashed lines in each graph point towards the x-axis of prevalence peaks.
(EPS)

S20 Fig. Final prevalence relative biases (PRBs) in community nodes for varying transmission parameters. The final PRBs in community nodes were calculated based on the simulations on AOK PLUS patient-movement networks, built by scale-up at incompleteness levels of

96% and 98%. In (a) we presented the final PRBs in community nodes for different values of β. In (b) we presented the final PRBs in community nodes for different values of γ.
(EPS)

Author Contributions

Conceptualization: Hanjue Xia, Johannes Horn, Monika J. Piotrowska, Konrad Sakowski, André Karch, Hannan Tahir, Mirjam Kretzschmar, Rafael Mikolajczyk.

Formal analysis: Hanjue Xia, Johannes Horn.

Funding acquisition: Rafael Mikolajczyk.

Investigation: Hanjue Xia.

Methodology: Hanjue Xia, Johannes Horn.

Software: Hanjue Xia.

Supervision: Rafael Mikolajczyk.

Validation: Hanjue Xia.

Visualization: Hanjue Xia.

Writing – original draft: Hanjue Xia.

Writing – review & editing: Hanjue Xia, Johannes Horn, Monika J. Piotrowska, Konrad Sakowski, André Karch, Hannan Tahir, Mirjam Kretzschmar, Rafael Mikolajczyk.

References

1. Belik V, Hövel P, Mikolajczyk R. In: Control of epidemics on hospital networks; 2016. p. 431–440.
2. Donker T, Wallinga J, Grundmann H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. PLoS Comput Biol. 2010; 6(3):e1000715. https://doi.org/10.1371/journal.pcbi.1000715 PMID: 20333236
3. Donker T, Wallinga J, Slack R, Grundmann H. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. PLoS One. 2012; 7(4):e35002. https://doi.org/10.1371/journal.pone.0035002 PMID: 22558106
4. Lee BY, McGlone SM, Song Y, Avery TR, Eubank S, Chang CC, et al. Social network analysis of patient sharing among hospitals in Orange County, California. Am J Public Health. 2011; 101(4):707–13. https://doi.org/10.2105/AJPH.2010.202754 PMID: 21330578
5. Nekkab N, Astagneau P, Temime L, Crepey P. Spread of hospital-acquired infections: A comparison of healthcare networks. PLoS Comput Biol. 2017; 13(8):e1005666. https://doi.org/10.1371/journal.pcbi.1005666 PMID: 28837555
6. Piotrowska MJ, Sakowski K, Karch A, Tahir H, Horn J, Kretzschmar ME, et al. Modelling pathogen spread in a healthcare network: Indirect patient movements. PLOS Computational Biology. 2020; 16(11):1–22. https://doi.org/10.1371/journal.pcbi.1008442
7. Ciccolini M, Donker T, Grundmann H, Bonten MJ, Woolhouse ME. Efficient surveillance for healthcare-associated infections spreading between hospitals. Proc Natl Acad Sci U S A. 2014; 111(6):2271–6. https://doi.org/10.1073/pnas.1308062111 PMID: 24469791
8. Donker T, Smieszek T, Henderson KL, Johnson AP, Walker AS, Robotham JV. Measuring distance through dense weighted networks: The case of hospital-associated pathogens. PLoS Comput Biol. 2017; 13(8):e1005622. https://doi.org/10.1371/journal.pcbi.1005622 PMID: 28771581
9. Huang SS, Avery TR, Song Y, Elkins KR, Nguyen CC, Nutter SK, et al. Quantifying interhospital patient sharing as a mechanism for infectious disease spread. Infect Control Hosp Epidemiol. 2010; 31(11):1160–9. https://doi.org/10.1086/656747 PMID: 20874503
10. Lee BY, McGlone SM, Wong KF, Yilmaz SL, Avery TR, Song Y, et al. Modeling the spread of methicillin-resistant Staphylococcus aureus (MRSA) outbreaks throughout the hospitals in Orange County, California. Infect Control Hosp Epidemiol. 2011; 32(6):562–72. https://doi.org/10.1086/660014 PMID: 21558768

11. Donker T, Wallinga J, Grundmann H. Dispersal of antibiotic-resistant high-risk clones by hospital networks: changing the patient direction can make all the difference. J Hosp Infect. 2014; 86(1):34–41. https://doi.org/10.1016/j.jhin.2013.06.021 PMID: 24075292

12. Belik V, Karch A, Hövel P, Mikolajczyk R. In: Leveraging Topological and Temporal Structure of Hospital Referral Networks for Epidemic Control; 2017. p. 199–214.

13. Fournet J, Barrat A. Estimating the epidemic risk using non-uniformly sampled contact data. Sci Rep. 2017; 7(1):9975. https://doi.org/10.1038/s41598-017-10340-y PMID: 28855718

14. Genois M, Vestergaard CL, Cattuto C, Barrat A. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. Nat Commun. 2015; 6:8860. https://doi.org/10.1038/ncomms9860 PMID: 26563418

15. Sapienza A, Barrat A, Cattuto C, Gauvin L. Estimating the outcome of spreading processes on networks with incomplete information: A dimensionality reduction approach. Phys Rev E. 2018; 98(1-1):012317. https://doi.org/10.1103/PhysRevE.98.012317 PMID: 30110805

16. Vestergaard C, Valdano E, Génois M, Poletto C, Colizza V, Barrat A. Impact of spatially constrained sampling of temporal contact networks on the evaluation of the epidemic risk. European Journal of Applied Mathematics. 2016; 27(6):941–957. https://doi.org/10.1017/S0956792516000309

17. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. Clin Microbiol Infect. 2014; 20(1):10–6. https://doi.org/10.1111/1469-0691.12472 PMID: 24267942

18. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton JF, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS One. 2010; 5(7):e11596. https://doi.org/10.1371/journal.pone.0011596 PMID: 20657651

19. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. J Infect Dis. 2012; 206(10):1549–57. https://doi.org/10.1093/infdis/jis542 PMID: 23045621

20. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One. 2011; 6(2):e17144. https://doi.org/10.1371/journal.pone.0017144 PMID: 21386902

21. Tizzoni M, Sun K, Benusiglio D, Karsai M, Perra N. The Scaling of Human Contacts and Epidemic Processes in Metapopulation Networks. Sci Rep. 2015; 5:15111. https://doi.org/10.1038/srep15111 PMID: 26478209

22. Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Regis C, et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS One. 2013; 8(9):e73970. https://doi.org/10.1371/journal.pone.0073970 PMID: 24040129

23. Piotrowska MJ, Sakowski K. Analysis of the AOK Lower Saxony hospitalisation records data (years 2008–2015); 2019.

24. Lonc A, Piotrowska MJ, Sakowski K. Analysis of the AOK Plus data and derived hospital network. Mathematica Applicanda. 2019; 47(1):127–139. https://doi.org/10.14708/ma.v47i1.6497

25. Mitglieder und Versicherte der Gesetzlichen Krankenversicherung (GKV);. https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte.html.

26. Mitglieder und Versicherte der Gesetzlichen Krankenversicherung (GKV)Bevölkerung mit Angaben zur Krankenversicherung nach Krankenkasse/Krankenversicherung in 1000. Gliederungsmerkmale: Jahre, Region, Krankenkasse/Krankenversicherung, Art des Versicherungsverhältnisses;. https://www.gbe-bund.de/gbe/!pkg_olap_tables.prc_set_orientation?p_uid=gast&p_aid=63268712&p_sprache=D&p_help=2&p_indnr=866&p_ansnr=57379115&p_version=2&D.000=3&D.001=3&D.437=1&D.438=2.

27. Iwashyna TJ, Christie JD, Moody J, Kahn JM, Asch DA. The structure of critical care transfer networks. Med Care. 2009; 47(7):787–93. https://doi.org/10.1097/MLR.0b013e318197b1f5 PMID: 19536030

28. Lonc A, Piotrowska MJ, Sakowski K. Analysis of the AOK Plus data and derived hospital network. MATHEMATICA APPLICANDA. 2019; 47:127–139. https://doi.org/10.14708/ma.v47i1.6497

29. Latapy M, Magnien C, Vecchio N. Basic notions for the analysis of large two-mode networks. Social Networks. 2008; 30(1):31–48. https://doi.org/10.1016/j.socnet.2007.04.006

30. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010; 32(3):245–251. https://doi.org/10.1016/j.socnet.2010.03.006

31. Marschall J, Muhlemann K. Duration of methicillin-resistant Staphylococcus aureus carriage, according to risk factors for acquisition. Infect Control Hosp Epidemiol. 2006; 27(11):1206–12. https://doi.org/10.1086/507917 PMID: 17080378

32. Tübbicke A, Hübner C, Kramer A, Hübner NO, Fleßa S. Transmission rates, screening methods and costs of MRSA—a systematic literature review related to the prevalence in Germany. European Journal of Clinical Microbiology & Infectious Diseases. 2012; 31(10):2497–2511. https://doi.org/10.1007/s10096-012-1632-8

33. Williamson K, Bisaga A, Paquette K, Lovell E. The prevalence of methicillin-resistant Staphylococcus aureus colonization in emergency department fast track patients. World J Emerg Med. 2013; 4(4):278–9. https://doi.org/10.5847/wjem.j.issn.1920-8642.2013.04.006 PMID: 25215132

34. Scanvic A, Denic L, Gaillon S, Giry P, Andremont A, Lucet JC. Duration of colonization by methicillin-resistant Staphylococcus aureus after hospital discharge and risk factors for prolonged carriage. Clin Infect Dis. 2001; 32(10):1393–8. https://doi.org/10.1086/320151 PMID: 11317238

35. Dehak N, Dehak R, Glass JR, Reynolds DA, Kenny P. Cosine similarity scoring without score normalization techniques. In: Odyssey;. p. 15.

36. Opsahl T. Structure and Evolution of Weighted Networks. University of London (Queen Mary College), London, UK; 2009. Available from: http://toreopsahl.com/publications/thesis/.

37. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org.

38. R Core Team. R: A Language and Environment for Statistical Computing; 2020. Available from: https://www.R-project.org/.

39. Tacconelli E, Mazzaferri F, de Smet A, Bragantini D, Eggimann P, Huttner B, et al. ESCMID-EUCIC clinical guidelines on decolonization of multidrug-resistant Gram-negative bacteria carriers. Clin Microbiol Infect. 2019; 25(7):807–817. https://doi.org/10.1016/j.cmi.2019.01.005 PMID: 30708122

40. De Cao E, Zagheni E, Manfredi P, Melegaro A. The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections. Biostatistics. 2014; 15(3):470–83. https://doi.org/10.1093/biostatistics/kxu008 PMID: 24705143

41. Smieszek T. A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. Theor Biol Med Model. 2009; 6:25. https://doi.org/10.1186/1742-4682-6-25 PMID: 19919678

Publication 2

Contribution as an author:

I have contributed to the conception and design of the work, method development, editing the manuscript.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Relevance of intra-hospital patient movements for the spread of healthcare-associated infections within hospitals - a mathematical modeling study

Hannan Tahir[1]*, Luis Eduardo López-Cortés[2], Axel Kola[3], Dafna Yahav[4], André Karch[5], Hanjue Xia[6], Johannes Horn[6], Konrad Sakowski[7,8], Monika J. Piotrowska[7], Leonard Leibovici[9], Rafael T. Mikolajczyk[6], Mirjam E. Kretzschmar[1]

1 Julius Center for Health Sciences & Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands, **2** Unidad Clínica de Enfermedades Infecciosas, Microbiología y Medicina Preventiva, Hospital Universitario Virgen Macarena, Sevilla, Spain, **3** Institute of Institute of Hygiene and Environmental Medicine, Charité- University Medicine Berlin, Berlin, Germany, **4** Infectious Diseases Unit, Rabin Medical Center, Beilinson Hospital, Petah-Tiqva, Israel, **5** Institute for Epidemiology and Social Medicine, University of Münster, Münster, Germany, **6** Institute for Medical Epidemiology, Biometry, and Informatics (IMEBI), Interdisciplinary Center for Health Sciences, Medical Faculty of the Martin Luther University Halle-Wittenberg, Halle, Germany, **7** Institute of Applied Mathematics and Mechanics, University of Warsaw, Warsaw, Poland, **8** Institute of High Pressure Physics, Polish Academy of Sciences, Warsaw, Poland, **9** Department of Medicine E; Rabin Medical Center, Beilinson Hospital, Petah-Tiqva, Israel

* h.tahir@umcutrecht.nl

Citation: Tahir H, López-Cortés LE, Kola A, Yahav D, Karch A, Xia H, et al. (2021) Relevance of intra-hospital patient movements for the spread of healthcare-associated infections within hospitals - a mathematical modeling study. PLoS Comput Biol 17(2): e1008600. https://doi.org/10.1371/journal.pcbi.1008600

Editor: Roger Dimitri Kouyos, University of Zurich, SWITZERLAND

Received: June 8, 2020

Accepted: December 2, 2020

Published: February 3, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This publication was made possible by grants from following national funding agencies: National Science Centre, Poland, Unisono: 2016/22/Z/ST1/00690 (University of Warsaw, Faculty of Mathematics, Informatics and Mechanics, Institute of Applied Mathematics and Mechanics) and 01KI1704C (Martin-Luther-University Halle-

Abstract

The aim of this study is to analyze patient movement patterns between hospital departments to derive the underlying intra-hospital movement network, and to assess if movement patterns differ between patients at high or low risk of colonization. For that purpose, we analyzed patient electronic medical record data from five hospitals to extract information on risk stratification and patient intra-hospital movements. Movement patterns were visualized as networks, and network centrality measures were calculated. Next, using an agent-based model where agents represent patients and intra-hospital patient movements were explicitly modeled, we simulated the spread of multidrug resistant enterobacteriacae (MDR-E) inside a hospital. Risk stratification of patients according to certain ICD-10 codes revealed that length of stay, patient age, and mean number of movements per admission were higher in the high-risk groups. Movement networks in all hospitals displayed a high variability among departments concerning their network centrality and connectedness with a few highly connected departments and many weakly connected peripheral departments. Simulating the spread of a pathogen in one hospital network showed positive correlation between department prevalence and network centrality measures. This study highlights the importance of intra-hospital patient movements and their possible impact on pathogen spread. Targeting interventions to departments of higher (weighted) degree may help to control the spread of MDR-E. Moreover, when the colonization status of patients coming from different departments is unknown, a ranking system based on department centralities may be used to design more effective interventions that mitigate pathogen spread.

Wittenberg, Medical Faculty, Institute of medical epidemiology, biostatistics and informatics) and the Netherlands ZonMw grant number 547001005 (Julius Centre, University Medical Centre Utrecht) within the 3rd JPI AMR framework (Joint Programming Initiative on Antimicrobial Resistance) cofound grant no 681055 for the consortium EMerGE-Net (Effectiveness of infection control strategies against intra- and inter-hospital transmission of MultidruG-resistant Enterobacteriaceae). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Pathogens including multidrug resistant enterobacteriacae (MDR-E) inside hospital settings are associated with higher morbidity, mortality, and healthcare costs. Better understanding of the transmission routes of these pathogens is required to develop targeted and efficient measures to contain the spread of MDR-E in a hospital. We analyzed datasets from five hospitals in different countries to explore patient movement patterns between departments of these hospitals (intra-hospital movements). We assessed whether movement patterns differ between patients at high or low risk of colonization. Our results show that in every intra-hospital network, there exist a few departments which are strongly connected and many peripheral departments which are loosely connected. High-risk patients stay on average longer in the hospital, and move more frequently between departments than low-risk patients. Targeting interventions to strongly connected departments may help to reduce pathogen spread inside the hospital. To explore this, we simulated the spread of MDR-E inside one hospital using an agent-based model that includes patient movements. In the simulations, we found positive correlations between departments' prevalence and network characteristics such as degree and weighted degree, thus highlighting the importance of targeting interventions to departments of higher (weighted) degree to control the spread of MDR-E.

Introduction

Multidrug resistant enterobacteriacae (MDR-E) are a common cause of hospital-acquired infections (HAIs) [1–4] and are considered a major public health threat. HAIs due to MDR-E are associated with higher morbidity, mortality, and healthcare costs [5,6]. A better understanding of the transmission routes of MDR-E pathogens may provide valuable insight to develop more effective and targeted infection control measures. When dealing with the spread of MDR-E in a single hospital, several factors such as contact precautions, inadequate hygiene protocols, and prolonged hospital stays play an important role. However, in recent years, inter-hospital patient movements between healthcare facilities have been recognized as an important route of transmission of pathogens between healthcare facilities. Various studies have used data on inter-hospital transfers of patients to construct healthcare networks. Based on those networks, various innovative infection control measures were proposed to contain the spread of HAIs [7–10]. Moreover, the burden of HAIs in a healthcare system has been proposed to be dependent on the structure of the inter-hospital network [9–12].

Similar to the role of inter-hospital patient movements, patient movements between the departments of a single hospital (intra-hospital movements) may contribute to spread of pathogens within a hospital. The effect of intra-hospital movements of patients on pathogen spread in a single hospital, however, is not fully understood. Various studies have reported the spread of MDR-E pathogens inside a hospital [13–17]. However, such studies were either limited to intensive care units (ICUs), or to some specific medical specialty departments, and patient movements between the departments were not assessed. Few modeling studies related to the spread of Methicillin-resistant *Staphylococcus aureus* (MRSA) [18,19] and Carbapenem-resistant *Klebsiella pneumoniae* [20] have evaluated the role of patient movements inside a hospital, but either very few departments were considered or departments other than ICUs were considered equal in terms of department structures (e.g. similar number of beds per department, similar movement rates between the departments etc.). These studies did not account for the hospital

department network when investigating the role of intra-hospital patient movements. Certain departments such as ICUs, emergency and surgery departments tend to have more patient movements than others. Applying infection control strategies to those departments may, therefore, help to limit the spread of the pathogen. A detailed understanding of how patient movements inside a hospital contribute to the spread and prevalence of pathogens in departments and in the entire hospital may help to develop more effective infection control strategies.

Several risk factors for acquiring HAIs due to MDR-E pathogens have been reported in the literature such as prolonged ICU or hospital stays [21,22], prior antibiotic usage [21,23–25], older age [26], renal dysfunction [26,27], mechanical ventilation [27], and recent invasive surgical procedures [22,25]. However, a patient's risk of acquiring colonization varies between patients, and not all patients are equally likely to become colonized with MDR-E pathogens. Multiple illnesses such as cancer, diabetes mellitus, dialysis, chronic renal disease, chronic alcoholism, chronic liver disease, and solid-organ transplantation have also been identified as risk factors for infection with MDR-E pathogens, as they weaken host defenses and thus increase host susceptibility to developing an infection [27–31]. It is, however, not yet clear whether such illnesses are associated with acquisition and colonization, or only with infection [27]. It is also worth noting here that a patient's susceptibility to acquiring colonization may not differ between patients with or without chronic diseases, but that observed differences in colonization rates between patients might be due to difference in exposure. Patients with severe disorders or chronic diseases are more likely to be in need of repeated hospital admission, and require more intense care from healthcare workers. Frequent contacts with healthcare workers may put such patients at high risk for acquiring colonization during their hospital stay.

The aim of this study is to assess and understand patient movement patterns in hospitals from different countries based on electronic hospital information systems data. Our analysis includes a stratification of movement patterns by risk level based on ICD-10 codes at discharge. To study the impact of intra-hospital patient movements on pathogen spread, we performed simulations using an agent-based transmission model including patient movements between departments. We analyzed the association between departments' prevalence and various network centrality measures obtained from the agent-based simulations. Finally, we discuss implications for targeted intervention measures to reduce pathogen spread in hospitals.

Material and methods

Data

We obtained routine hospital admission data of five hospitals from Spain, The Netherlands, Germany, and Israel. The pseudonymized admission data were extracted from electronic hospital information systems and do not include any sensitive information. Participating hospitals were: University Medical Center Utrecht (UMCU), Utrecht, The Netherlands; Hospital Universitario Virgen Macarena (HUVM), Seville, Spain; Charité Universitätsmedizin (CUM), Berlin, Germany; Beilinson Hospital (BH), Rabin Medical Center, Petah Tikva, Israel; and Universitätsklinikum Halle (UKH), Halle, Germany. Basic details of every dataset are given in *Table 1*.

The provided data included: patient ID, hospital ID, patient birth year, admission and discharge dates, and ICD-10 diagnosis codes. Additionally, the names of all departments in which patients stayed during their hospital stay were provided for the respective time periods. Data extraction periods were different for every hospital. Since information on patients who were transferred to other hospitals was not available from every hospital, we considered those movements as discharges.

We have, further, excluded admissions with a hospital length of stay (LOS) of less than a day (outpatient admissions) and only considered inpatient admissions (hospital LOS $>$ 1 day)

Table 1. Descriptive data of participating hospitals.

Hospitals	University Medical Center Utrecht, The Netherlands (UMCU)	Hospital Universitario Virgen Macarena, Spain (HUVM)	Charité Universitäts-medizin, Germany (CUM)	Beilinson Hospital, Israel (BH)	Universitäts klinikum Halle, Germany (UKH)
Data Period	01.01.2014–31.12.2017	01.01.2016–30.01.2017	01.01.2016–31.12.2016	01.01.2012–31.12.2017	01.01.2017–31.12.2017
Hospital Size (number of beds)	1042	950	3011	800	950
Number of Departments	61	36	58	26	58
Number of patients per year *	30,823	26,724	416,751	17,410	32,863
Number of admissions per year **	81,516	34,364	841,221	33,754	47,279
Total number of admissions (N) **	326,064	37,227	841,221	202,524	47,279
Number of admissions used in the analysis (N (%))	117,758 (36.12)	28,191 (75.73)	124,946 (14.85)	169,541 (83.71)	37,977 (80.33)

* A patient can be admitted more than once. Every patient gets a unique patient ID and it remains unchanged for future admissions.

** Admissions include both in-patient and out-patient admissions. For every (re)admission, a unique admission ID is assigned to a patient.

https://doi.org/10.1371/journal.pcbi.1008600.t001

in the study. The reason behind this exclusion is that outpatients may not be exposed to the parts of the hospital where inpatients stay, and therefore do not contribute to spread of pathogens. Moreover, admissions to psychiatric departments were excluded because patients in those departments tend to have completely different (often very long) LOS and are usually not bedridden. We also excluded admissions in obstetrics departments because we observed that newly born children do not immediately get their own patient and admission ID, but those of their mothers instead. Thus, it is difficult to differentiate between newborn children from their mothers in our data. Moreover, in obstetrics departments, transfers to other departments are rare and the duration of stay is short. *Table 1* reports the total number of admissions included in the current study from every hospital after the above exclusions.

Patient risk stratification

Patients having severe disorders or chronic diseases and immunocompromised patients may have a higher risk of acquiring colonization and subsequent infection. A risk stratification into low-risk and high-risk patients [32], was implemented based on certain diagnoses (ICD10 codes), which are known to be associated with patient disabilities (*Table 2*). This risk stratification was applied consecutively to any further hospitalization of the patient, i.e. once a patient is defined as a high-risk person, the patient will automatically be defined as high-risk for any consecutive hospitalization independent of the respective diagnoses.

Intra-hospital movements

When a patient is transferred from one department to another, we counted this transfer as a single movement. From department-level data, we extracted such intra-hospital movements for every hospital stay. Patient movements within a single department were not considered in this study.

We did not put any constraints on a minimum time between movements. However, if a patient stays in another department only for a short time, the probability that transmission

Table 2. Patient risk stratification. A patient having any of the mentioned ICD-10 codes was considered as a high-risk patient.

Disease	ICD-10 Code(s)
Cancer	C00-C96
Diabetes mellitus	E10-E14
Heart failure	I50
Chronic kidney disease (moderate or severe)	N18.3-N18.6
Immune system disease	D80-D89
Systemic sclerosis and other systemic involvement of connective tissue	M34-M35
Psoriasis (chronic skin disease)	L40
Abnormal immunological findings in serum	R76

https://doi.org/10.1371/journal.pcbi.1008600.t002

occurs is very small, so short stays do not contribute much to overall transmissions in the hospital. Using intra-hospital movements data, we created a directed movement network for each risk group and for each hospital in order to visualize patient movements patterns. In such a network, nodes represent departments and links represent patient transfers between the departments. We further computed network statistics such as degree, weighted degree, graph density, average path length, average clustering coefficient and network diameter.

Data analysis tools

Python Pandas [33] was used for data cleaning, filtering, stratification, and analysis. From *Python Pandas*, patients' intra-hospital movements for each hospital were exported as weighted edge-lists, where weights represent numbers of patient movements in each direction. These weighted edge-lists were later imported into *Gephi* software [34] for network visualization and computation of network statistics. In this study, we used the mean and standard deviation for most of the descriptive quantities, but for some other variables we used median and show interquartile range (IQR).

Agent-based model

In order to evaluate the role of patient intra-hospital movements and their implications towards MDR-E spread, we developed a discrete-event agent-based model (ABM) to simulate the spread of a pathogen inside a hospital. To demonstrate our methods, we performed simulations for the Spanish hospital HUVM, for which we had the most complete data. The model was built using *Python* library *Mesa* which is an open source ABM framework [35]. In the HUVM hospital, there were 34 departments present in total. Number of beds in each department was estimated from the HUVM dataset using mean number of patients present every day in each department (*S10 Fig*).

In the model, we simulated patients explicitly as individual agents and every agent had several descriptive attributes, namely unique id, risk score, length of stay (LOS), disease state, department number, and bed number. In the ABM, agents' attributes were updated in discrete time steps of 5 minutes. During the simulation, the new patient's arrivals process is modeled as a Poisson process with estimated daily arrival rate. Moreover, for all the simulations, patients were uniformly distributed at admission to available beds in the departments. At the time of admission, a LOS in discrete time units was assigned to every individual agent from an exponential probability density function (estimated from the HUVM dataset). At every time step, patient's LOS was decremented by 1, and once the LOS for a patient reached zero, the patient was discharged from the hospital. Our model does not account for changes in the patient population due to death.

In the data, for patients who moved at least twice during their stay in the HUVM hospital, we observed that in 81% of such movements, patients were moved back to the previous ward. Based on this observation, we implemented in the ABM that there is an 81% chance to return to the previous ward for patients with two or more movements. For the remaining 19% of the movements, patients follow the department selection algorithm through a preference matrix explained below. Daily patient movement rates to other departments were estimated for every department of the HUVM hospital (*S11 Fig*). These movement rates were then divided by the size of the departments to obtain department-specific daily movement probabilities. Given those daily movement probabilities, numbers of patients to be moved to other departments were calculated for every department every simulated day, and added to a department specific counter. This counter kept track of the number of patients to be moved from every department.

For each patient movement, we determined a future department using a preference matrix composed of preference probabilities [36]. For this preference matrix, a pivot table was first computed based on every department's weighted in-degree and weighted out-degree from the HUVM hospital. This pivot table was then normalized row-wise to obtain preference probabilities. Given those preference probabilities [36], a patient's next department was selected and the patient is moved to the new department. It is worthwhile to note, that there are fixed numbers of beds per department, which can all be occupied at a certain moment. If all beds in potential new departments were occupied at a certain moment, the patient stayed in the current location and the above explained procedure is repeated in the next time step.

Each patient also has a disease state: *susceptible*, *colonized*, or *infected*. A patient in a *susceptible* state can immediately become colonized after being exposed to MDR-E pathogens. A patient in a *colonized* state remains asymptomatically colonized with MDR-E and can transmit the pathogen to others upon contact. A patient in an *infected* state is symptomatically colonized showing disease symptoms. Infected patients can still spread the pathogen to others. In the model, we assumed that the transition from *susceptible* to *infected* requires passing through *colonized* and, therefore, neglect a direct pathway from *susceptible* to *infected*. More details on the disease progression can be found in *S12 Fig*. When a patient becomes infected in the model, no further movement of that patient will be allowed. This assumption is based on discussions with clinicians from the participating hospitals, where the majority of the patients after infection diagnosis are only moved to other departments in case of emergencies.

Intra-hospital models usually assume that transmission occurs via the cross-transmission route, which involves effective contacts between patients and healthcare workers (HCWs). Since we did not model HCWs explicitly, the transmission was implemented as a force of infection (FOI), which gives the probability per unit of time t for a susceptible patient to acquire the pathogen and to become colonized. The FOI was dependent on the transmission parameter β, the number of colonized and infected patients, and the total number of patients present in a given department. Martin et al. [27] reported that approximately 5% of the patients with gastrointestinal carriage of Carbapenem-resistant *Klebsiella pneumoniae* developed an infection. We calibrated our model to achieve a similar cumulative percentage of infected patients using a probability of 0.012/day in Bernoulli trials. Once the disease state of a patient was changed to *infected*, the LOS of that patient was increased by three days. After the end of this extended LOS period, the patient returned back to a *colonized* state. Since infected patients return back to a *colonized* state, they may become infected again in the model. The increase in the LOS by three days for infected patients was a parameter in our model; thus, we also checked the effects of varying this parameter on the department prevalence (*S13*–*S15 Figs*). In our model, daily transmission rates for both high-risk and low-risk patients were the same. However, it was assumed that there was a difference in the mean LOS, with a slightly

longer LOS for high-risk patients when compared to the low-risk group. Further details on the model and parameters are given in *S1 and S1A Text*, respectively.

To evaluate the role of patient movements for pathogen spread, we tested the four different scenarios using the ABM described below. We apply these scenarios starting from day 30 so that a stable hospital population of susceptible patients can be assumed.

- *Scenario 1*: On day 30 (simulation time) one colonized patient was admitted to the highest weighted degree centrality department (ICU–department 8). The motivation for this scenario is to highlight the impact of patient movements on pathogen spread in case of a single imported case (no continuous inflow of colonized patients).
- *Scenario 2*: continuous inflow of 1% colonized patients from day 30 onwards. We used a probability of 0.01 for a patient to arrive in a colonized state into the hospital. These colonized patients were then randomly distributed to all departments of the hospital.
- *Scenario 3*: continuous inflow of 5% colonized patients from day 30 onwards, distributed randomly to all departments of the hospital.
- *Scenario 4*: continuous inflow of 15% colonized patients from day 30 onwards, distributed randomly to all departments of the hospital.

In all scenarios, we tested different values of the transmission parameter β (range 0.0005–0.30 per day). Spearman's rank correlation coefficients between departments'network characteristics and departments'prevalence were computed to evaluate the association between these two quantities. For scenarios with continuous inflow of colonized patients, we used stable prevalence states for the Spearman's rank correlation test.

Results

Descriptive statistics of hospital data

Descriptive statistics for each risk group for every hospital are shown in *Table 3*. In all hospitals, proportions of male patients and proportions of male admissions in the high-risk groups were always higher when compared to proportions of low-risk male patients and admissions. Proportions of high-risk admissions vary between 30.20%– 36.69% between the hospitals with CUM being the lowest and HUVM having the highest proportion.

Patients in the high-risk groups were on average older than low-risk patients (*Fig 1A*). *Fig 1A* shows that BH hospital had the highest mean age in both risk groups. *Fig 1B* displays LOS distributions where we observed longer mean LOS in the high-risk groups when compared to

Table 3. Descriptive statistics for data included in the analysis for participating hospitals.

Hospitals		UMCU		HUVM		CUM		BH		UKH	
		Low-risk	High-risk	Low-risk	High-risk	Low-risk	High-risk	Low-risk	High-risk	Low-risk	High-risk
Patients	Total (N)	52590	16870	15368	6910	68755	23008	70369	26480	19491	8487
	Males (%)	53.16	57.27	50.78	55.18	48.24	56.10	52.98	54.56	47.60	56.00
Admissions	Total (N (%))	75147 (63.82)	42611 (36.18)	17848 (63.31)	10343 (36.69)	87219 (69.80)	37727 (30.20)	109985 (64.87)	59556 (35.13)	24071 (63.38)	13906 (36.62)
	Males (%)	52.91	57.67	51.70	56.39	48.26	57.30	52.89	55.25	48.6	57.92
Admissions per day (SD)		51.43 (21.02)	29.16 (14.03)	44.49 (13.57)	25.67 (8.45)	238.3 (133.8)	103.07 (85.31)	50.17 (17.15)	27.16 (11.37)	65.94 (28.77)	38.09 (20.64)
Mean LOS (Days (SD))		5.52 (9.88)	7.19 (11.46)	6.09 (8.84)	8.49 (9.59)	5.31 (8.06)	8.74 (13.69)	4.85 (7.16)	6.61 (9.55)	5.52 (7.94)	10.09 (12.7)
Patient Age (Median (IQR))		38 (6–62)	64 (50–72)	57 (34–74)	72 (61–81)	52 (31–69)	66 (54–76)	65 (47–78)	72 (62–81)	52 (27–67)	67 (57–77)

https://doi.org/10.1371/journal.pcbi.1008600.t003

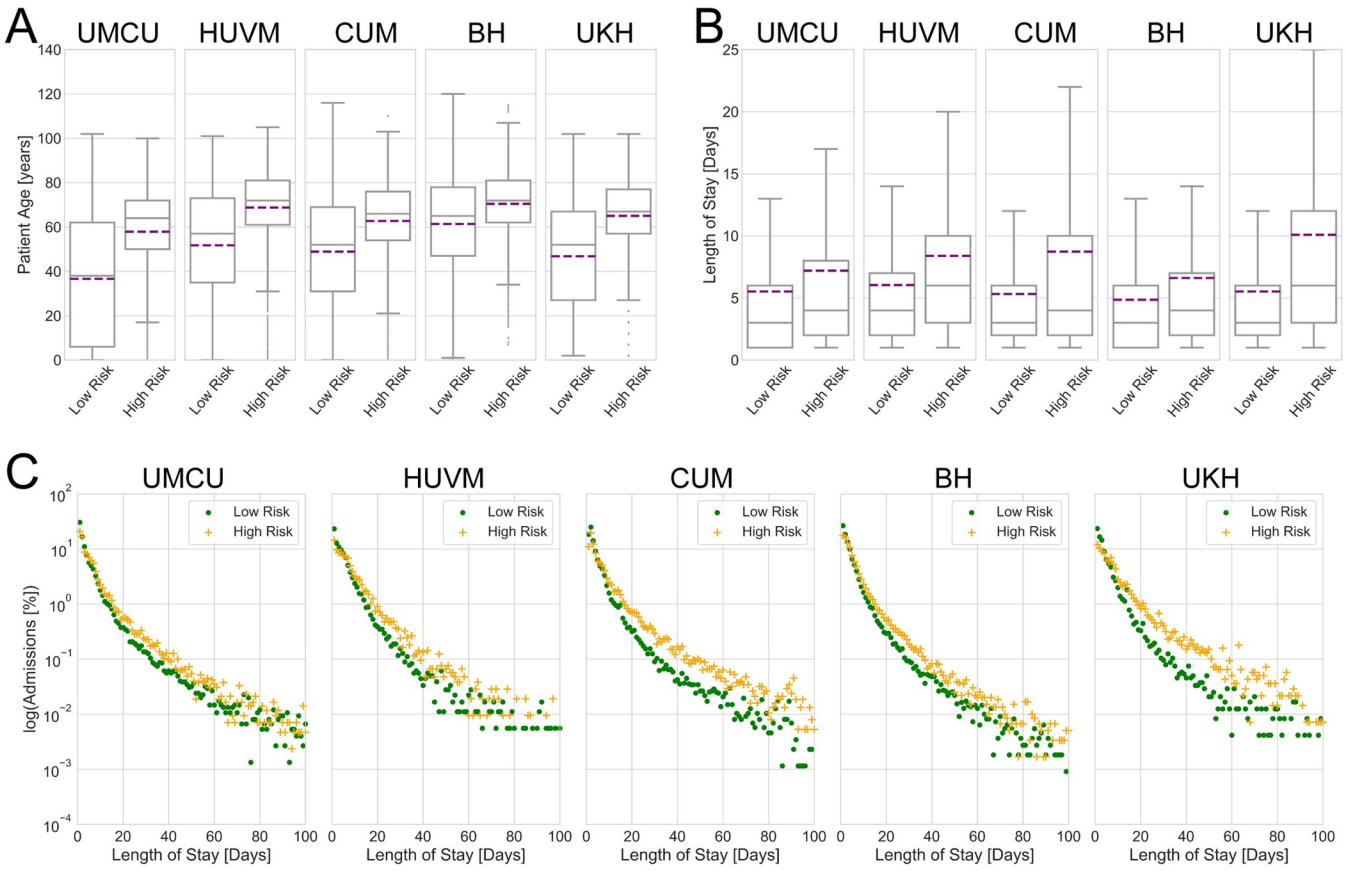

Fig 1. (A) Box plots showing patient age distribution for each hospital and risk group, (B) Box plots for admission LOS distribution for each hospital and risk group. In A and B, purple lines in the boxes show mean value whereas the grey line in the boxes show median of the data. (C) Proportion of admissions versus LOS on log-linear scale for each risk group in every hospital. To better visualize the trends between the risk groups at smaller LOS, proportion of admissions with LOS over 100 days are not shown in C. There are however, few data points above 100 days LOS (see S1 Numerical Data for complete data).

https://doi.org/10.1371/journal.pcbi.1008600.g001

low-risk groups in each hospital. High-risk groups in CUM, UKH and HUVM had large variability in the LOS as shown by the box plots (*Fig 1B*). We further plotted the proportion of admissions versus LOS in all the hospitals and risk groups on a log-linear scale to display the distributions of LOS. *Fig 1C* shows that in all hospitals, a large proportion of admissions in both risk groups had short LOS. Moreover, low-risk groups have a higher proportion of patients with short LOS (between 1–4 days). For both risk groups in each hospital, their LOS was approximately exponentially distributed.

Patient movements

Patient intra-hospital movements were estimated from patient transfers between the departments in every admission. *Fig 2A* shows mean movements per day normalized by hospital size, number of departments, and number of admissions per year in each risk group for every hospital. It is clear from *Fig 2A* that UKH had the highest mean intra-hospital movements per day in the high-risk group. Moreover, high-risk groups in all hospitals had a higher mean number of movements per day except for the UMCU. Mean number of intra-hospital movements per hospital admission are shown in *Fig 2B*. High-risk groups had notably higher means when compared to low-risk groups, except for the UMCU where there was little difference between the risk groups. BH had the lowest mean number of movements per admission.

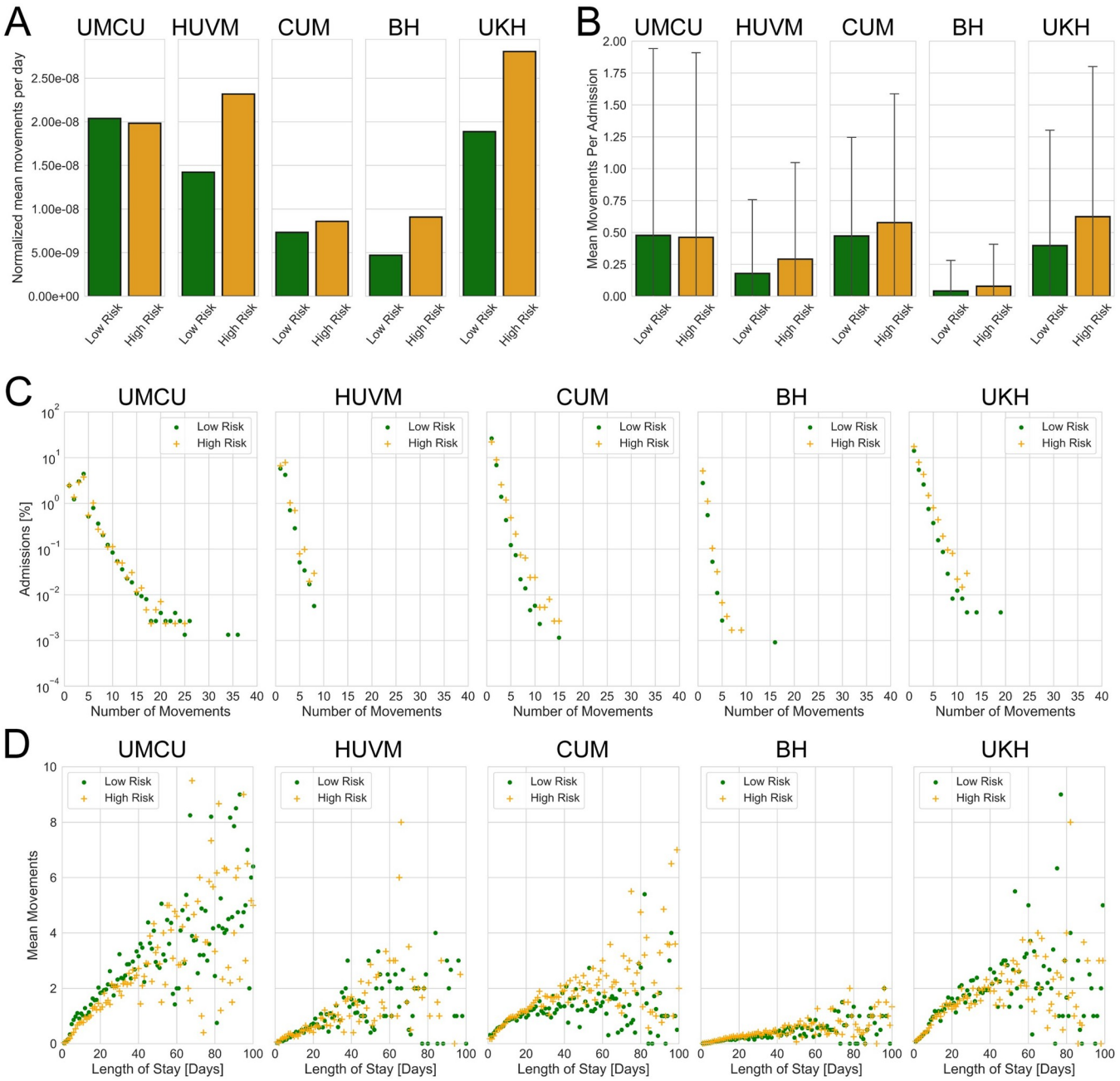

Fig 2. Patients movement results. (A) Mean number of movements per day normalized by hospital size, number of departments and the number of admissions per year in each risk group for every hospital. (B) Mean number of movements per hospital admission for each risk group and for each hospital. Error bars show standard deviation. (C) Proportion of admissions versus number of movements in each risk group for every hospital on log-linear scale. It is worth noting here that major proportion of admissions in each hospital has zero movements (UMCU (Low-risk 86.62%, High-risk 87.06%), HUVM (Low-risk 89.05%, High-risk 83.79%), CUM (Low-risk 64.81%, High-risk 64.43%), BH (Low-risk 96.6%, High-risk 93.63%), and UKH (Low-risk 76.45%, High-risk 66.01%)). (D) Mean number of movements versus admission LOS for each risk group and every hospital. To better visualize the trends between the risk groups at smaller LOS, data over 100 days LOS are not shown in D. There are however, few data points above 100 days LOS (see S2 Numerical Data for complete data).

https://doi.org/10.1371/journal.pcbi.1008600.g002

The majority of patients in each hospital did not move between the departments during their hospital stay. For UMCU, HUVM and BH, more than 83% of admissions in both risk groups had no movements. This effect was even more pronounced in the BH where patients in 93% of

the admissions in the high-risk group, and 96% in the low-risk group, did not move between the departments. The German hospitals CUM and UHK had smaller proportions of admissions with zero movements (low-risk 64.8% versus high-risk 64.4% in CUM, low-risk 76.5% versus high-risk 66% in UKH). In *Fig 2C*, we plotted the proportion of admissions with a given number of movements for all admissions with at least one movement on a log-linear scale.

When plotting admission LOS against mean number of movements per admission (*Fig 2D*), we observed a positive correlation in all participating hospitals, indicating that a patient with longer LOS is more likely to move more frequently between the departments during the hospital stay. To identify differences between the risk groups for the data shown in *Fig 2D*, we plotted number of movements against LOS for every hospital admission in each risk group for every hospital. There seems to be no clear difference between the risk groups for every hospital.

Characteristics of intra-hospital movement networks

The intra-hospital movements were visualized as a weighted directed network, where nodes represent departments in a hospital and links are defined by patient flows. Directed weights of the links are based on the number of patient movements in each direction. Degree of a department is defined as the number of its connections to other departments. Weighted degree indicates the number of patients moving in and out from one department to other departments. *Fig 3* shows the visualization of the HUVM intra-hospital movement networks for both risk groups as well as for the complete data without stratification. Nodes were ordered alphabetically based on node names in a counter clockwise direction from the top node. The color of the node is based on node degree (sum of in-degree and out-degree), whereas the size of the node is based on node weighted degree (sum of weighted in-degree and weighted out-degree). The width of the link is based on weights (number of patient movements), where the thickness of a link is based on the number of patients moving in that direction. A clustering layout of the HUVM network is also shown in *S1 Fig*. Moreover, the number of patient movements between the departments are visualized as a heatmap (*S1 Fig*). The hospital movement networks for other considered hospitals are shown in *S2*–*S9 Figs*. Networks displayed in *Figs 3* and *S2*–*S9* clearly show that these networks had several central hubs with much incoming and outgoing patient movements, and many peripheral nodes that were only loosely connected to the network. The most common departments to act as hubs for both risk groups include ICU, emergency department, internal medicine, anesthesiology, cardiology and cardiothoracic surgery, neurology, cardiovascular surgery, and nephrology. However, pediatric departments tended to appear in the top few nodes for all networks in the low-risk groups except for BH, where no pediatric patients were admitted due to the absence of a pediatrics specialty.

To better compare the networks, we computed network statistics as shown in *Table 4* (see supplementary *S1 Text* for methods of calculating the reported network statistics). Number of edges and average degree were higher in the high-risk groups except HUVM and CUM where an opposite trend was observed. When compared to low-risk groups, slightly higher maximum degrees (degree range) were observed for high-risk groups for all hospitals. There were no clear relationships observed in the weighted degree, network diameter, and average path length between the risk groups and between the hospitals as these characteristics largely depend on the number of movements in each risk group of every hospital. *Table 4* also shows that average clustering coefficients in the high-risk groups were slightly higher except for HUVM. Among hospitals, the largest average clustering coefficients in both risk groups were observed in the BH. Similarly, graph density, showing the completeness and connectedness of a network, was higher in the high-risk groups for all hospitals except HUVM and CUM, where it was slightly

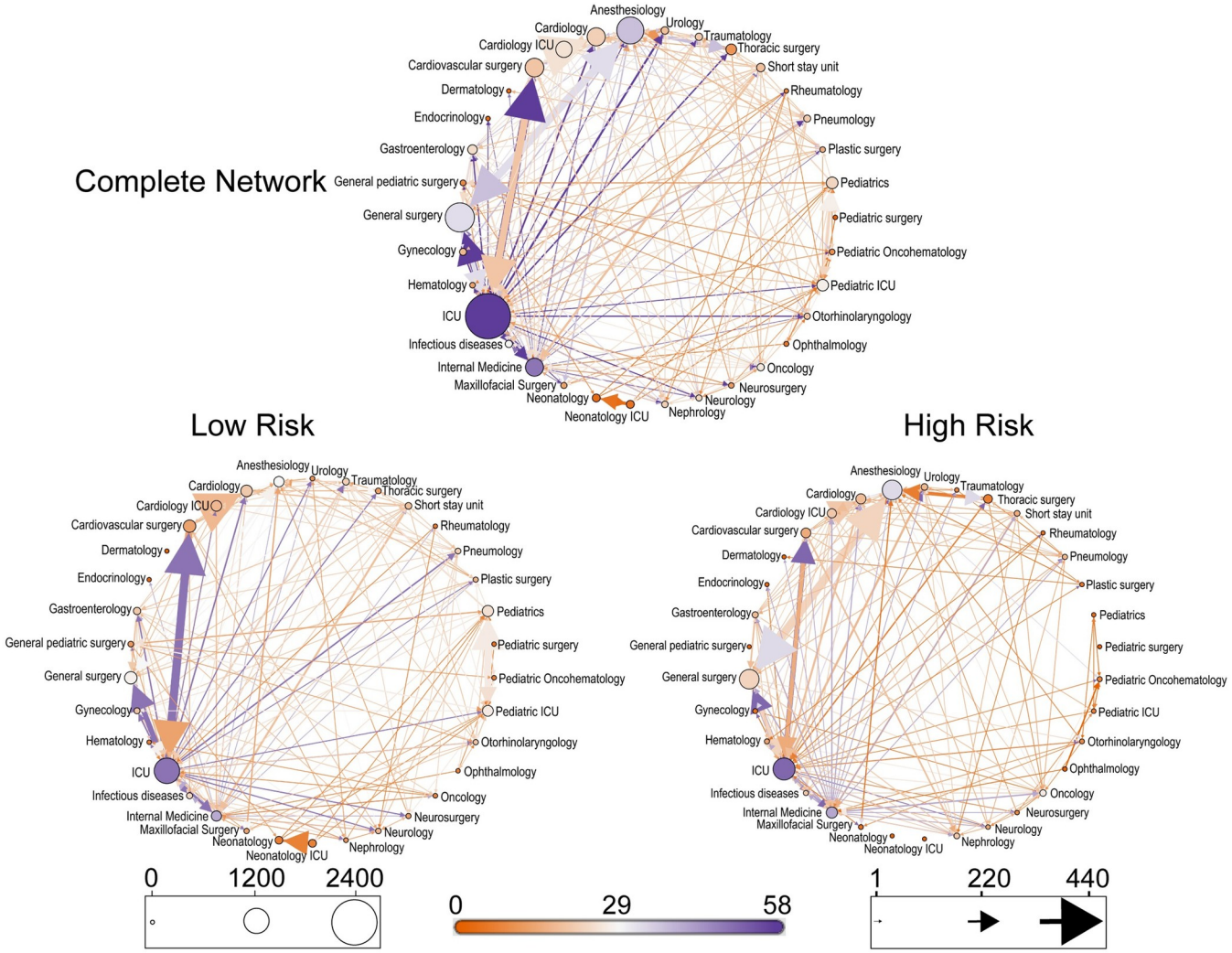

Fig 3. Intra-hospital hospital networks of HUVM showing patient directed movements from one department to another. Nodes represent departments and arrows represent patient movements between these departments. Color of the nodes was based on nodes degree whereas size of the nodes was based on nodes weighted degree. Arrows thickness is based on the directed number of transfers (weights) between the departments and color of an arrow is assigned similar to the node color from where the arrow is originating. For visualization of the clustering, a clustering layout of the complete HUVM network is also shown in the supplementary S1 Fig.

https://doi.org/10.1371/journal.pcbi.1008600.g003

lower than the density of the low-risk group network. BH showed relatively high graph density in both risk groups when compared to other hospitals (*Table 4*).

Simulation results

Based on patient movement patterns, we investigated the impact of movement network structure on pathogen spread. To do that, we used an agent-based model parameterized with data from the HUVM hospital. *Fig 4A* represents a network of patient movements averaged over 50 simulations. Visually, this network matches well with the un-stratified network of the HUVM (*Fig 4B*). *Fig 4A* shows the large variation in weighted degree of the nodes (departments), where the ICU acts as the biggest hub with highest weighted degree compared to other departments.

For each scenario, distributions of department specific daily MDR-E prevalence (calculated as sum of colonized and infected patients divided by the total number of patients in a

Table 4. Intra-hospital hospital networks statistics for each risk group. Complete data correspond to unstratified data including low-risk and high-risk groups.

Hospitals		UMCU			HUVM			CUM			BH			UKH		
		Complete	Low-risk	High-risk	Complete	Low-risk	High-risk	Complete	Low-risk	High-risk	Complete	Low-risk	High-risk	Complete	Low-risk	High-risk
Nodes		59			34			56			24			56		
Edges		1026	703	729	335	252	218	1219	1004	904	435	370	375	1076	720	871
Graph Density		0.3	0.205	0.213	0.298	0.225	0.194	0.396	0.326	0.294	0.788	0.67	0.679	0.349	0.234	0.283
Degree	Mean	17.39	11.92	12.356	9.85	7.41	6.41	21.77	17.93	16.14	18.13	15.42	15.62	19.21	12.86	15.55
	Range (min—max)	1–86	0–76	0–79	2–58	0–48	0–50	8–74	7–62	0–66	16–46	11–42	10–46	1–82	1–65	0–78
Avg. Weighted Degree per year		235.35	150.6	83.14	158.54	85.63	81.12	1118.6	723.14	366.48	63.83	31.40	32.43	340.80	172.73	168.07
Network Diameter		3	5	4	4	4	4	3	3	4	2	2	2	4	5	5
Avg. Clustering Coefficient		0.641	0.564	0.588	0.595	0.507	0.494	0.566	0.52	0.554	0.841	0.767	0.789	0.592	0.479	0.53
Avg. Path Length		1.759	1.968	2.03	1.852	1.962	1.855	1.644	1.757	1.857	1.212	1.332	1.321	1.769	2.03	1.884

https://doi.org/10.1371/journal.pcbi.1008600.t004

department) are plotted as box plots in *Fig 5*. Although we ran simulations with different values of β (*S1A Text*), we only showed results from $\beta = 0.25$ in *Fig 5*.

It is clear from *Fig 5A* that when MDR-E pathogen transmissions occurred in a highly-connected ICU department (*scenario 1*), it affected all the major departments of the hospital connected via patient movements. Although *scenario 1* was a hypothetical scenario with only a single admission of a colonized patient, it highlights the impact of patient movements on pathogen spread (*Fig 5A*). In reality, hospitals continuously receive

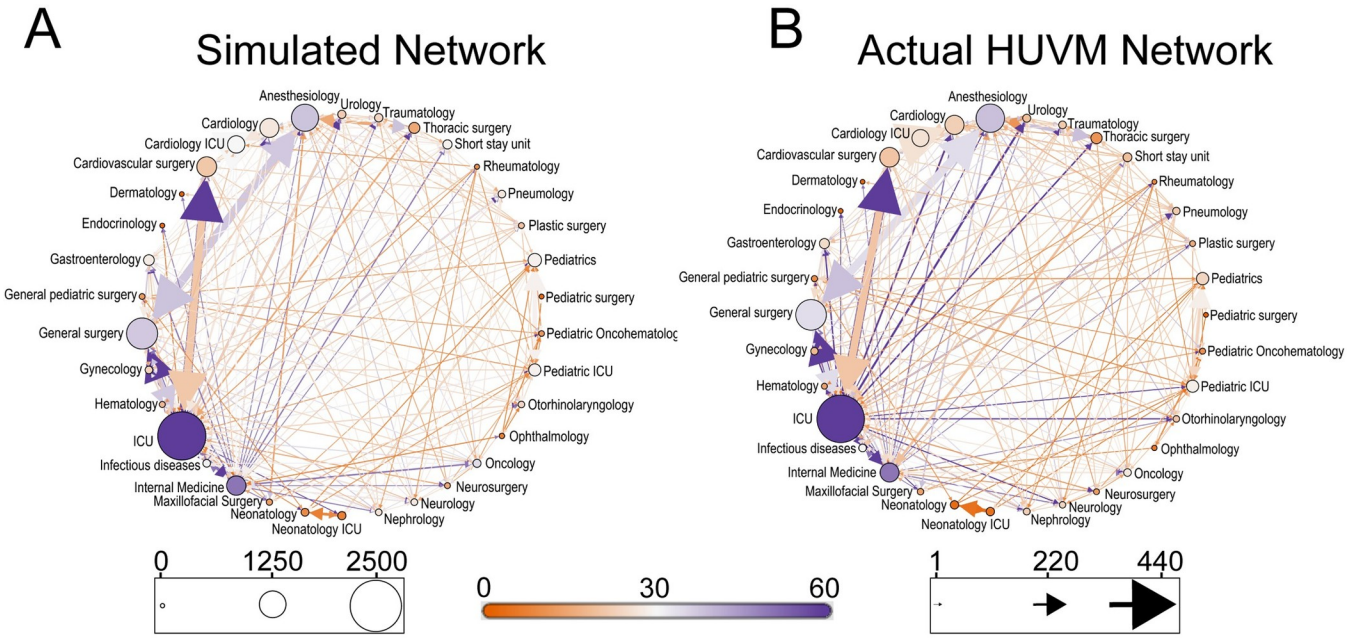

Fig 4. (A) Patient movements network generated from the ABM simulation, (B) Actual HUVM patient movement network without stratification of patients into low-risk and high-risk. Nodes represent departments and arrows represent patient movements between these departments. Color of the nodes is based on nodes degree and dark orange color refers to low values of degree whereas purple color refers to high degree. Size of the nodes is based on nodes weighted degree.

https://doi.org/10.1371/journal.pcbi.1008600.g004

colonized patients at admission. *Fig 5B* shows a steady state department specific daily prevalence of MDR-E pathogen for *scenarios 2*, *3*, *and 4*, which consider 1%, 5% and 15% of daily arrivals being colonized at admission, respectively. *Fig 5B* clearly illustrates that an increase in the percentage of colonized patients at admission has a direct impact on the departments' MDR-E prevalence. Scenarios shown in *Fig 5B* show quite high prevalences

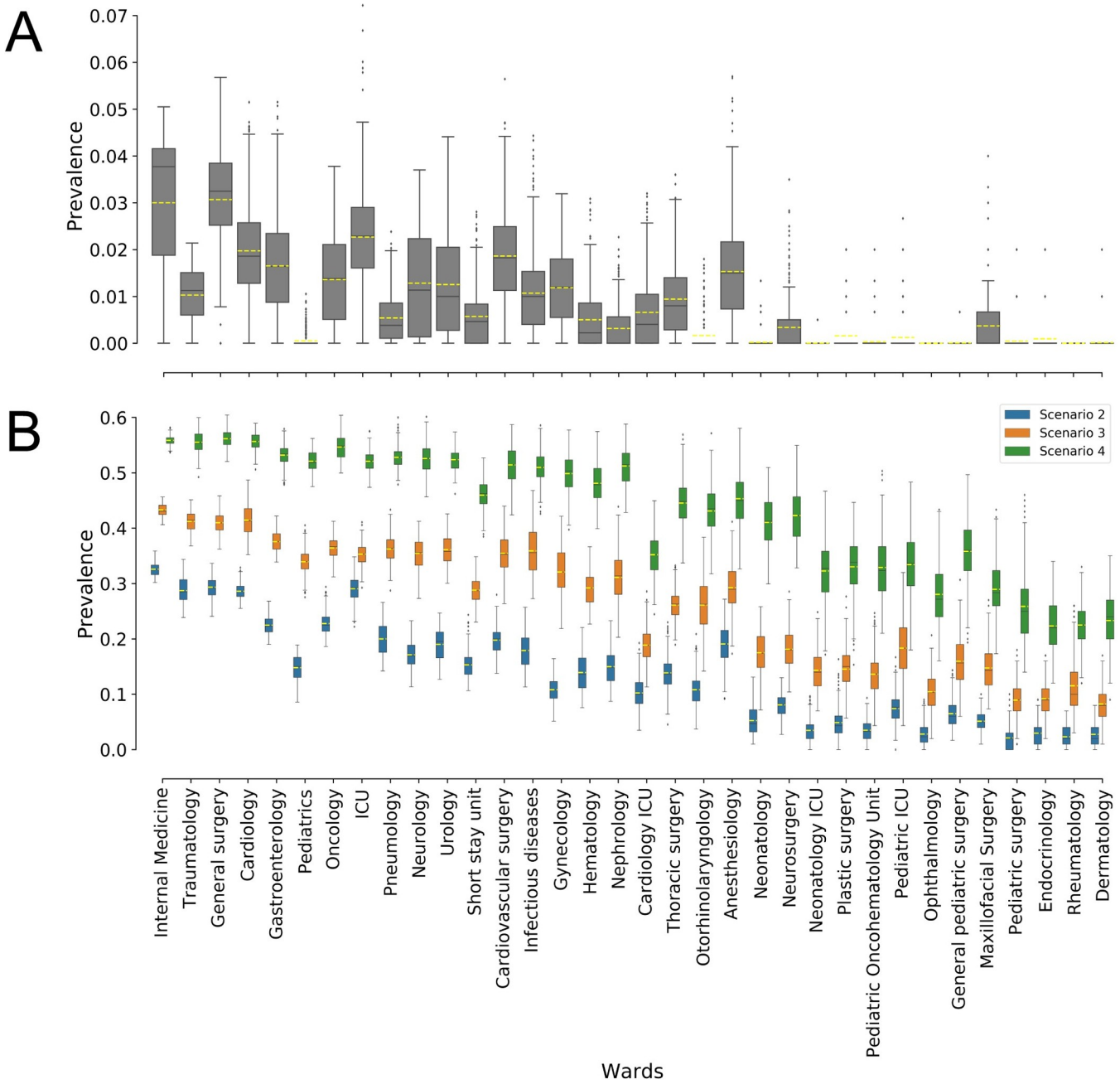

Fig 5. (A) Distribution of daily departments' prevalence of MDR-E for a period of 395 days for scenario 1 where a single colonized patient was admitted to the ICU on day 30. (B) Distribution of steady state daily departments' prevalence of MDR-E (from 200 days onwards) for scenario 2, 3 & 4 which include 1%, 5% and 15% daily arrivals of colonized patients respectively from day 30 onwards. Yellow lines represent the mean prevalence per department for every scenario. For all results shown in Fig 5, transmission parameter $\beta = 0.25$ was used for all departments. Departments are ordered by size with internal medicine department being the largest department. Results shown in Fig 5 are based on 50 simulations for every scenario.

https://doi.org/10.1371/journal.pcbi.1008600.g005

in the larger and high centrality departments such as ICU, internal medicine, general surgery, cardiology and cardiovascular surgery. When the percentage of incoming colonized patients is low (1%, *scenario 2*), stochastic variations in departments' mean prevalence are large for the large departments. However, with the increase in daily colonized arrivals (15% *scenario 4*), the majority of the large departments show similar levels of MDR-E mean prevalence.

In order to quantify the impact of network characteristics on department specific MDR-E mean prevalence, we calculated Spearman's rank correlation coefficients between departments' MDR-E mean prevalence and network characteristics such as nodes degree (*Fig 6A*) and nodes weighted degree (*Fig 6B*) for different transmission rates (β) and for every scenario presented in *Fig 5*. *Fig 6A* and *6B* show that when a colonized patient is admitted to a high-centrality ICU department (*scenario 1*), departments' prevalences showed a positive correlation with degree and weighted degree centralities respectively for all values of β. The correlations were much stronger for high β values ($\beta > 0.10$). For scenarios with continuous inflow of colonized patients (*scenarios 2–4*), strong associations between departments' prevalences and both network degree and weighted degree were observed for $\beta > 0.05$ (*Fig 6A and 6B*). At extremely low β value ($\beta = 0.0005$), no association between departments' prevalences and network centralities was observed for *scenarios 2–4*. At low β values ($0.001 \leq \beta \leq 0.05$), *scenarios 2 and 3* showed weak correlations, however, *scenario 4* did show an increasing trend between β values ($0.001 \leq \beta \leq 0.05$) and correlation coefficients.

Considering our modeling results and assuming no interventions in place, Table 5 indicates the top three departments with highest mean prevalence expected in the event of MDR-E spread for the considered hospitals, based on the above-mentioned network characteristics. Controlling patient movements out of these departments or applying other infection control interventions targeted to those high centrality departments may prevent MDR-E from spreading to remaining departments of the hospital.

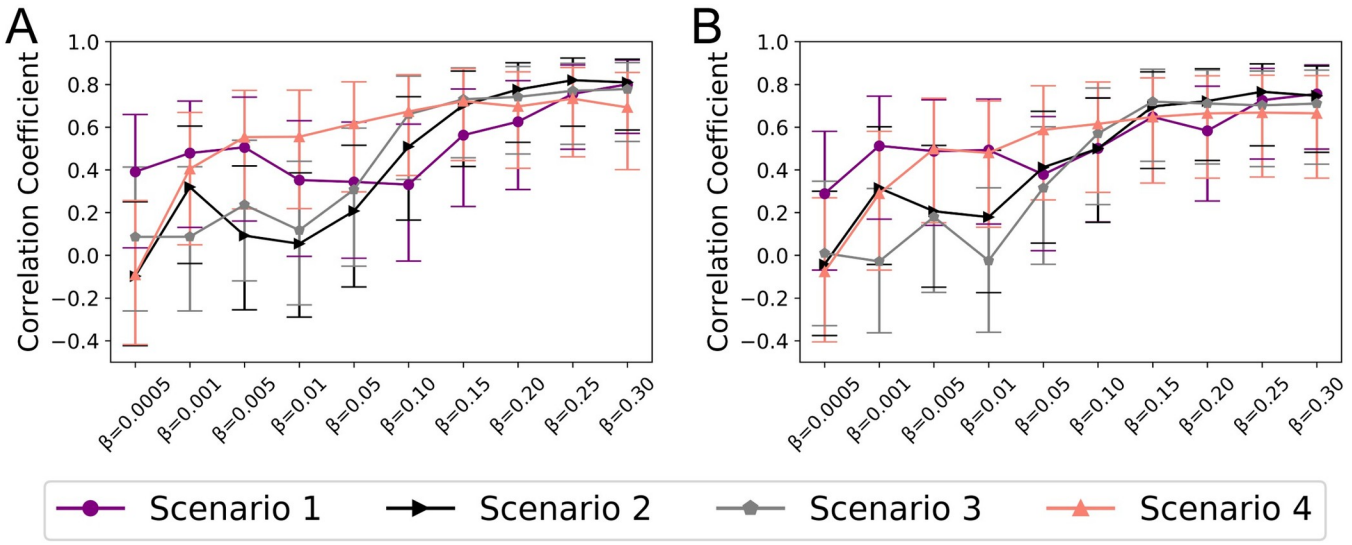

Fig 6. Spearman's rank correlation coefficient results between averaged network statistics from 50 simulations and departments' prevalence of MDR-E. (A) Correlation between departments' degree and departments' mean prevalence. (B) Correlation between departments' weighted degree and departments' mean prevalence. In all scenarios, different transmission parameter β values were tested. For Scenario 1, mean prevalence over a period of 395 days was used. For Scenario 2–4, steady state mean prevalence from 200 days onwards was used. Joining datapoints with lines was only done to improve readability but it does not show a functional relationship as x-axis is a categorical axis. For both A and B, errorbars represent 95% confidence intervals.

https://doi.org/10.1371/journal.pcbi.1008600.g006

Table 5. Top three departments of each participating hospital with simulated high mean prevalence based on network characteristics with respect to a hospital-wide MDR-E spread in the absence of interventions.

Hospital	Degree	Weighted Degree
UMCU	ICU, Neurology, Internal Medicine	ICU, Cardiothoracic surgery, Pediatrics ICU
HUVM	ICU, Internal Medicine, Anesthesiology	ICU, General surgery, Anesthesiology
CUM	CVK Anesthesiology, CVK Nephrology/Internal Intensive Care, CCM Anesthesiology	CBF Emergency Department, CVK Internal Emergency Department, CVK Anesthesiology
BH	Internal medicine D, Internal medicine A, Internal medicine C	Internal medicine D, Internal medicine E, Internal medicine B
UKH	ICU, Anesthesiology 1, Anesthesiology 2	Interdisciplinary emergency, ICU, Cardiac surgery 1

https://doi.org/10.1371/journal.pcbi.1008600.t005

Impact of intervention

We used the model to assess the impact of two interventions: (i) contact isolation of infected patients where we assumed that contact isolated patients were placed in separate rooms and their contacts with HCWs were reduced. We also assumed that HCWs were required to wear gloves and gowns and to follow strict hand hygiene protocols when entering contact isolated patients' rooms. Depending on the effectivenss of these measures, contact isolated patients contributed less towards transmissions in the hospital. For this intervention, we tested different contact isolation effectiveness scenarios (30%, 70% and 100%), (ii) a network intervention where patients moving in and out from the highest (weighted) degree department (ICU) were screened (see Table 5). If a patient was detected as positive, the patient was put on contact isolation (assuming 90% effectiveness of contact isolation) for the remaining hospital stay. We calculated the percent reduction in the number of transmissions from the baseline scenario, where no intervention was applied. Fig 7 shows the impact of both interventions as percent reduction in the number of transmissions in the hospital over a period of 395 days for different transmission rates (β). We observed a clear impact of the contact isolation effectiveness where 100% effectiveness resulted in larger reductions in number of transmissions as compared to 30% and 70% contact isolation effectiveness. Depending on the transmission rate (β), the network intervention applied to just one department with the highest degree and weighted degree resulted in 8–11% reduction in the number of transmissions. Results for extremely low β values ($\beta < 0.01$) are not plotted in Fig 7 because the number of transmissions were very low and percent reduction results did not make any sense. Fig 7 also shows that a contact isolation of

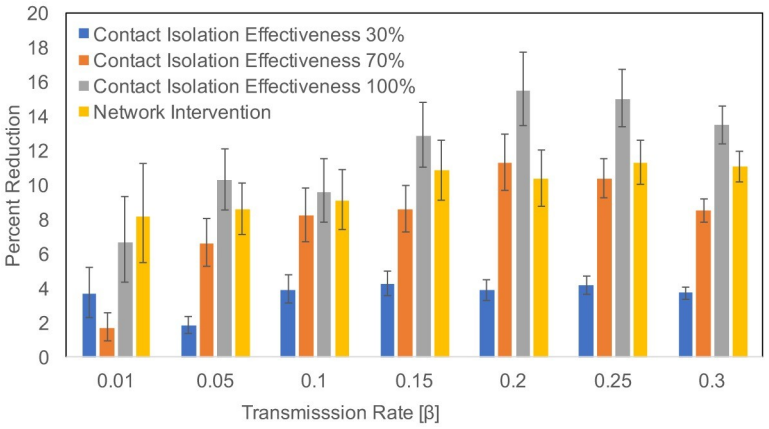

Fig 7. Impact of interventions on transmissions percent reductions for different β values. Error bars represent 95% confidence intervals.

https://doi.org/10.1371/journal.pcbi.1008600.g007

infected patients which is 100% effective is best, but if contact isolation is not perfect, a network based approach may be better. Instead of targeting just one highest degree and weighted degree department, extending the network intervention to include several departments may result in more reductions in the number of transmissions in the hospital.

Discussion

Our study provides a detailed description of patterns of patients' movements between departments in a hospital based on data from several hospitals, and an analysis of how these movements may impact the transmission of bacterial pathogens within hospitals.

Our analysis includes a stratification by patients' risk levels and its implications for patients' movements and risk of becoming colonized. Risk of acquiring colonization may differ between patients and may be determined by several factors such as functional status, immune responses, chronic or severe diseases. Taking such factors into account, we stratified patients into low-risk and high-risk groups based on certain ICD-10 codes. Results indicate that high-risk patients were on average slightly older and stay longer in the hospital than low-risk patients. High-risk patients moved between departments slightly more often per admission than low-risk patients; however, this higher number of movements may be due to their longer hospital stay. We further visualized patient intra-hospital movement patterns from participating hospitals as networks, and used an ABM to further assess the impact of intra-hospital movements on pathogen spread. Our modeling results clearly show that a MDR-E spread in one of the high centrality departments can spread out to all the departments in a hospital. Positive correlations between departments' centralities and departments' mean prevalence show that departments with high centralities will eventually have high prevalence at higher transmission rates. Therefore, consideration of departments' centralities and patient movements in a given hospital could improve the efficiency of interventions.

Patient movements inside the hospital have been included in few modeling studies for MRSA and MDR-E pathogens [18–20] but those studies have considered only a few specialty departments and the departments' structure (e.g. different department sizes) is often neglected. Rocha et al. [37] and Pei et al. [38] have recently studied MRSA spread in a hospital network where they developed a large scale data-driven contact network model including the dynamics of patient referrals within and between wards and hospitals. Since their model captures the interaction patterns that were formed from time varying real-world contact data, they did not explicitly consider the department types and structure of hospitals. Here we identified, based on data from specific hospitals and data on network centrality, which departments may play a crucial role in the spread of MDR-E in the entire hospital. In the absence of interventions, simulation results showed a strong positive correlation between departments' mean prevalence and network characteristics such as degree and weighted degree.

Study limitations

Disparities in the department structure exist between hospitals, e.g., a hospital could have two departments that might be a single department in another hospital. Every hospital structure is different and a generalization of the hospital structure is therefore not possible. For the hospitals included in our study, the German hospitals (CUM and UKH) did include data from emergency departments, but other hospitals did not include emergency departments in their data set. Such differences in the hospital structures and data sets made it much harder to compare these hospitals with each other. Moreover, the CUM hospital is a very large hospital in Berlin which has three separate campuses across Berlin. Although it would also be possible to consider these CUM campuses as separate hospitals, we did

observe substantial patient movements between these campuses. Furthermore, the BH did not have a pediatrics department, and ICU data were stored in a separate database, which was not provided for the analysis.

In the current work, movement networks were created from routine patient data. It could be that some patient movements in the datasets were not physical movements, but represented administrative events, when a patient received treatment from another department, while staying in the same place. We are aware of this issue, however we did not have sufficient information to filter out such movement records from the datasets. A more accurate way of tracing patient movements between the departments of a hospital would be to use wireless wearable sensors as used in the close proximity interaction studies [39–41], however, gathering data with such methods is often limited to shorter time periods given the high costs and privacy protection issues.

In the presented model, upon a successful transmission, a patient is assumed to become immediately colonized, and starts transmitting the pathogen to others, which may not be the case in reality. In future work, a latent period may be included to allow some time delay for a patient to acquire enough bacterial load before transmitting to others. We do not expect strong impact of a latent period on the correlation results, but a latent period might affect prevalence levels. Moreover, it can be considered to include a direct pathway from *susceptible* to an *infected* disease state.

Implications for clinical practice

In the current study, patient movements between the departments of a single hospital were extracted from anonymized hospital admission data. Such data is stored in almost every hospital information system. This work highlights the potential of using such data to evaluate patient movement patterns and their implications for pathogens spread inside the hospital.

When an infectious disease is severe, large changes in the patient movement patterns may become necessary, and will be implemented as witnessed during the COVID-19 pandemic. However, changing patient movements in a hospital to prevent spread of pathogens such as MDR-E may not be a viable option. In view of our results about correlation between network characteristics and prevalence, one could consider establishing a ranking system based on departments' network characteristics, such as degree and weighted degree. In such a ranking system, departments are sorted by their degree and weighted degree. When the colonization status of patients coming from different departments is unknown, a risk assessment may be based upon the rank of the department the patient is coming from. If the respective department has a high rank, the patient may be placed in isolation or increased hygienic measures could be taken as a precautionary infection control measure. The advised intervention should only be applied when there are available resources, in terms of free beds available in the department. Moreover, priority should be given to patients in need of urgent medical care and after that, if resources are still available, patients coming from different departments can be handled based on the movement ranking system. This may prevent spread of pathogens among departments that are connected via patient movements.

To conclude, our study emphasizes the importance of intra-hospital patient movements and their impact on pathogen spread. Applying interventions by targeting hubs, i.e. departments of higher degree and weighted degree centrality may help to control the spread of MDR-E. Moreover, when the colonization status of patients coming from different departments is unknown, a department ranking system based on centrality measures could be used to improve the efficiency of the interventions.

Supporting information

S1 Text. Agent Based Intra-hospital model.
(DOCX)

S1 Fig. (A) Inter-department complete HUVM hospital network showing clustering of the departments. Clustering is computed based on the modularity algorithm in the Gephi software which detects nodes that are more densely connected together than to the rest of the network. Node colors show the cluster to which a node belongs. The color of the arrow is based on the color of the node from where the arrow is originating. The thickness of the arrow is based on the number of patient's transfers (weight). The size of the node is based on the weighted degree. (B) Heat map showing the number of transfers from one department to another department for the complete HUVM network. A patient is transferred from the source to the target department.
(PDF)

S2 Fig. Inter-department hospital networks of the UMCU hospital showing patient directed movements from one department to another. (A) Complete UMCU network without stratification, (B) Low-risk UMCU network, (C) High-risk UMCU network. Nodes represent departments and arrows represent patient movements between these departments. The color of the nodes was based on nodes degree whereas size of the nodes was based on the nodes' weighted degree.
(PDF)

S3 Fig. (A) Inter-department complete UMCU hospital network showing clustering of the departments. Clustering is computed based on the modularity algorithm in the Gephi software which detects nodes that are more densely connected together than to the rest of the network. Node colors show the cluster to which a node belongs. The color of the arrow is based on the color of the node from where the arrow is originating. The thickness of the arrow is based on the number of patient's transfers (weight). The size of the node is based on the weighted degree. (B) Heat map showing the number of transfers from one department to another department for the complete UMCU network. A patient is transferred from the source to the target department.
(PDF)

S4 Fig. Inter-department hospital networks of the CUM hospital showing patient directed movements from one department to another. (A) Complete CUM network without stratification, (B) Low-risk CUM network, (C) High-risk CUM network. Nodes represent departments and arrows represent patient movements between these departments. The color of the nodes was based on nodes degree whereas size of the nodes was based on nodes weighted degree. CBF, CCM and CVK are different campuses of the CUM hospital.
(PDF)

S5 Fig. (A) Inter-department complete CUM hospital network showing clustering of the departments. Clustering is computed based on the modularity algorithm in the Gephi software which detects nodes that are more densely connected together than to the rest of the network. Node colors show the cluster to which a node belongs. The color of the arrow is based on the color of the node from where the arrow is originating. The thickness of the arrow is based on the number of patient's transfers (weight). The size of the node is based on the weighted degree. (B) Heat map showing the number of transfers from one department to another department for the complete CUM network. A patient is transferred from the source to the target department.
(PDF)

S6 Fig. Inter-department hospital networks of the BH hospital showing patient directed movements from one department to another. (A) Complete BH network without stratification, (B) Low-risk BH network, (C) High-risk BH network. Nodes represent departments and arrows represent patient movements between these departments. The color of the nodes was based on nodes degree whereas size of the nodes was based on the nodes' weighted degree.
(PDF)

S7 Fig. (A) Inter-department complete BH hospital network showing clustering of the departments. Clustering is computed based on the modularity algorithm in the Gephi software which detects nodes that are more densely connected together than to the rest of the network. Node colors show the cluster to which a node belongs. The color of the arrow is based on the color of the node from where the arrow is originating. The thickness of the arrow is based on the number of patient's transfers (weight). The size of the node is based on the weighted degree. (B) Heat map showing the number of transfers from one department to another department for the complete BH network. A patient is transferred from the source to the target department.
(PDF)

S8 Fig. Inter-department hospital networks of the UKH hospital showing patient directed movements from one department to another. (A) Complete UKH network without stratification, (B) Low-risk UKH network, (C) High-risk UKH network. Nodes represent departments and arrows represent patient movements between these departments. The color of the nodes was based on nodes degree whereas size of the nodes was based on the nodes weighted degree.
(PDF)

S9 Fig. (A) Inter-department complete UKH hospital network showing clustering of the departments. Clustering is computed based on the modularity algorithm in the Gephi software which detects nodes that are more densely connected together than to the rest of the network. Node colors show the cluster to which a node belongs. The color of the arrow is based on the color of the node from where the arrow is originating. The thickness of the arrow is based on the number of patient's transfers (weight). The size of the node is based on the weighted degree. (B) Heat map showing the number of transfers from one department to another department for the complete UKH network. A patient is transferred from the source to the target department.
(PDF)

S10 Fig. Mean number of patients present per day in every department of the HUVM hospital. This data was used to define department size in terms of beds per department.
(PDF)

S11 Fig. Daily inter-department discharge rates from every department in the HUVM hospital. Departments are ordered by department size as shown in S10 Fig.
(PDF)

S12 Fig. Patient disease state flow chart. S refers to Susceptible, C refers to Colonized, and I refers to symptomatic infected patients. FOI is the force of infection.
(PDF)

S13 Fig. Impact of additional LOS for infected patients on steady state prevalence using Scenario 3 (5% continuous arrival of colonized patients). Transmission parameter $\beta = 0.25$

was used in each department.
(PDF)

S14 Fig. Spearman's rank correlation coefficients between departments prevalence and network characteristics (degree and weighted degree). Different values for increase in the LOS for infected patients were used to identify the impact of this parameter on the correlation between steady state departments prevalence and network characteristics. Scenario 3 (5% continuous arrival of colonized patients) with transmission parameter $\beta = 0.25$ for every department was used.
(PDF)

S15 Fig. Impact of additional LOS for infected patients on the overall hospital LOS distributions in both risk groups (Low-risk and High-risk) and its comparison with the actual HUVM data for both risk groups. Grey lines in the boxes show median of the data.
(PDF)

S1 Numerical Data. Numerical data of Fig 1C.
(XLSX)

S2 Numerical Data. Numerical data of Fig 2D.
(XLSX)

Acknowledgments

We would like to thank Dr. M.C.J. Bootsma and Thi Mui Pham for useful suggestions and discussions.

Author Contributions

Conceptualization: Hannan Tahir, Luis Eduardo López-Cortés, Axel Kola, Dafna Yahav, André Karch, Hanjue Xia, Johannes Horn, Konrad Sakowski, Monika J. Piotrowska, Leonard Leibovici, Rafael T. Mikolajczyk, Mirjam E. Kretzschmar.

Formal analysis: Hannan Tahir.

Funding acquisition: Rafael T. Mikolajczyk.

Investigation: Hannan Tahir, Mirjam E. Kretzschmar.

Methodology: Hannan Tahir, Monika J. Piotrowska, Rafael T. Mikolajczyk, Mirjam E. Kretzschmar.

Project administration: Hannan Tahir, Mirjam E. Kretzschmar.

Software: Hannan Tahir.

Supervision: Mirjam E. Kretzschmar.

Validation: Hannan Tahir.

Visualization: Hannan Tahir.

Writing – original draft: Hannan Tahir.

Writing – review & editing: Hannan Tahir, Luis Eduardo López-Cortés, Axel Kola, Dafna Yahav, André Karch, Hanjue Xia, Johannes Horn, Konrad Sakowski, Monika J. Piotrowska, Leonard Leibovici, Rafael T. Mikolajczyk, Mirjam E. Kretzschmar.

References

1. Pitout JD, Laupland KB. Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis. 2008; 8 (3):159–66. https://doi.org/10.1016/S1473-3099(08)70041-0 PMID: 18291338
2. Van Duin D, Paterson DL. Multidrug-resistant bacteria in the community: trends and lessons learned. Infect Dis Clin. 2016; 30 (2):377–90. https://doi.org/10.1016/j.idc.2016.02.004 PMID: 27208764
3. O'neill J. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Rev Antimicrob Resist. 2014; 20:1–16.
4. Suetens C, Hopkins S, Kolman J, Diaz Högberg L. European centre for disease prevention and control. Point prevalence survey of healthcare associated infections and antimicrobial use in European acute care hospitals Sweden: ECDC. 2013.
5. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis. 2019; 19 (1):56–66. https://doi.org/10.1016/S1473-3099(18)30605-4 PMID: 30409683
6. Cassini A, Plachouras D, Eckmanns T, Sin MA, Blank H-P, Ducomble T, et al. Burden of six healthcare-associated infections on European population health: estimating incidence-based disability-adjusted life years through a population prevalence-based modelling study. PLoS Med. 2016; 13 (10):e1002150. https://doi.org/10.1371/journal.pmed.1002150 PMID: 27755545
7. Donker T, Wallinga J, Slack R, Grundmann H. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. PLoS One. 2012; 7 (4):e35002. https://doi.org/10.1371/journal.pone.0035002 PMID: 22558106
8. Donker T, Wallinga J, Grundmann H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. PLoS Comput Biol. 2010; 6 (3):e1000715. https://doi.org/10.1371/journal.pcbi.1000715 PMID: 20333236
9. Nekkab N, Astagneau P, Temime L, Crépey P. Spread of hospital-acquired infections: A comparison of healthcare networks. PLoS Comput Biol. 2017; 13 (8):e1005666. https://doi.org/10.1371/journal.pcbi.1005666 PMID: 28837555
10. Piotrowska MJ, Sakowski, K., Lonc, A., Tahir, H., Kretzschmar, M.E. Impact of inter-hospital transfers on the prevalence of resistant pathogens in a hospital-community system. Epidemics (accepted for publication). 2020.
11. Ciccolini M, Donker T, Köck R, Mielke M, Hendrix R, Jurke A, et al. Infection prevention in a connected world: The case for a regional approach. Int J Med Microbiol. 2013; 303 (6–7):380–7. https://doi.org/10.1016/j.ijmm.2013.02.003 PMID: 23499307
12. Piotrowska MJ, Sakowski K, Karch A, Tahir H, Horn J, Kretzschmar M, et al. Modelling pathogen spread in a healthcare network: indirect patient movements. Plos Computational Biology (Revision submitted). 2020;arXiv preprint arXiv. 2001:05875.
13. Harris A, Perencevich E, Johnson J, Paterson D, Morris J, Strauss S, et al. Patient-to-patient transmission is important in extended-spectrum β-lactamase–producing Klebsiella pneumoniae acquisition. Clin Infect Dis. 2007; 45 (10):1347–50. https://doi.org/10.1086/522657 PMID: 17968833
14. Harris AD, Kotetishvili M, Shurland S, Johnson JA, Morris JG, Nemoy LL, et al. How important is patient-to-patient transmission in extended-spectrum β-lactamase Escherichia coli acquisition. Am J Infect Control. 2007; 35 (2):97–101. https://doi.org/10.1016/j.ajic.2006.09.011 PMID: 17327188
15. Gurieva T, Dautzenberg MJ, Gniadkowski M, Derde LP, Bonten MJ, Bootsma MC. The transmissibility of antibiotic-resistant Enterobacteriaceae in intensive care units. Clin Infect Dis. 2018; 66 (4):489–93. https://doi.org/10.1093/cid/cix825 PMID: 29020273
16. Smit PW, Stoesser N, Pol S, van Kleef E, Oonsivilai M, Tan P, et al. Transmission dynamics of hyperendemic multi-drug resistant Klebsiella pneumoniae in a Southeast Asian neonatal unit: a longitudinal study with whole genome sequencing. Front Microbiol. 2018; 9:1197. https://doi.org/10.3389/fmicb.2018.01197 PMID: 29951041
17. Ajao AO, Johnson JK, Harris AD, Zhan M, McGregor JC, Thom KA, et al. Risk of acquiring extended-spectrum β-lactamase–producing Klebsiella species and Escherichia coli from prior room occupants in the intensive care unit. Infection Control & Hospital Epidemiology. 2013; 34 (5):453–8. https://doi.org/10.1086/670216 PMID: 23571360
18. Deeny S, Cooper B, Cookson B, Hopkins S, Robotham J. Targeted versus universal screening and decolonization to reduce healthcare-associated meticillin-resistant Staphylococcus aureus infection. J Hosp Infect. 2013; 85 (1):33–44. https://doi.org/10.1016/j.jhin.2013.03.011 PMID: 23911111

19. Sadsad R, Sintchenko V, McDonnell GD, Gilbert GL. Effectiveness of hospital-wide methicillin-resistant Staphylococcus aureus (MRSA) infection control policies differs by ward specialty. PLoS One. 2013; 8 (12):e83099. https://doi.org/10.1371/journal.pone.0083099 PMID: 24340085

20. Vilches T, Bonesso M, Guerra H, Fortaleza C, Park A, Ferreira C. The role of intra and inter-hospital patient transfer in the dissemination of heathcare-associated multidrug-resistant pathogens. Epidemics. 2019; 26:104–15. https://doi.org/10.1016/j.epidem.2018.11.001 PMID: 30583920

21. Skippen I, Shemko M, Turton J, Kaufmann M, Palmer C, Shetty N. Epidemiology of infections caused by extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella spp.: a nested case–control study from a tertiary hospital in London. J Hosp Infect. 2006; 64 (2):115–23. https://doi.org/10.1016/j.jhin.2006.05.010 PMID: 16859810

22. Lucet J-C, Chevret S, Decré D, Vanjak D, Macrez A, Bédos J-P, et al. Outbreak of multiply resistant Enterobacteriaceae in an intensive care unit: epidemiology and risk factors for acquisition. Clin Infect Dis. 1996; 22 (3):430–6. https://doi.org/10.1093/clinids/22.3.430 PMID: 8852958

23. Harris AD, McGregor JC, Johnson JA, Strauss SM, Moore AC, Standiford HC, et al. Risk factors for colonization with extended-spectrum β-lactamase–producing bacteria and intensive care unit admission. Emerg Infect Dis. 2007; 13 (8):1144. https://doi.org/10.3201/eid1308.070071 PMID: 17953083

24. Asensio A, Oliver A, González-Diego P, Baquero F, Perez-Diaz JC, Ros P, et al. Outbreak of a multiresistant Klebsiella pneumoniae strain in an intensive care unit: antibiotic use as risk factor for colonization and infection. Clin Infect Dis. 2000; 30 (1):55–60. https://doi.org/10.1086/313590 PMID: 10619733

25. Safdar N, Maki DG. The commonality of risk factors for nosocomial colonization and infection with antimicrobial-resistant Staphylococcus aureus, enterococcus, gram-negative bacilli, Clostridium difficile, and Candida. Ann Intern Med. 2002; 136 (11):834–44. https://doi.org/10.7326/0003-4819-136-11-200206040-00013 PMID: 12044132

26. Kofteridis DP, Valachis A, Dimopoulou D, Maraki S, Christidou A, Mantadakis E, et al. Risk factors for carbapenem-resistant Klebsiella pneumoniae infection/colonization: A case–case-control study. J Infect Chemother. 2014; 20 (5):293–7. https://doi.org/10.1016/j.jiac.2013.11.007 PMID: 24703709

27. Martin RM, Bachman MA. Colonization, infection, and the accessory genome of Klebsiella pneumoniae. Front Cell Infect Microbiol. 2018; 8:4. https://doi.org/10.3389/fcimb.2018.00004 PMID: 29404282

28. Menéndez R, Méndez R, Polverino E, Rosales-Mayor E, Amara-Elori I, Reyes S, et al. Risk factors for multidrug-resistant pathogens in bronchiectasis exacerbations. BMC Infect Dis. 2017; 17 (1):659. https://doi.org/10.1186/s12879-017-2754-5 PMID: 28964261

29. Paczosa MK, Mecsas J. Klebsiella pneumoniae: going on the offense with a strong defense. Microbiol Mol Biol Rev. 2016; 80 (3):629–61. https://doi.org/10.1128/MMBR.00078-15 PMID: 27307579

30. Tsai S-S, Huang J-C, Chen S-T, Sun J-H, Wang C-C, Lin S-F, et al. Characteristics of Klebsiella pneumoniae bacteremia in community-acquired and nosocomial infections in diabetic patients. Chang Gung Med J. 2010; 33 (5):532–9. PMID: 20979704

31. Biehl LM, Schmidt-Hieber M, Liss B, Cornely OA, Vehreschild MJ. Colonization and infection with extended spectrum beta-lactamase producing Enterobacteriaceae in high-risk patients–Review of the literature from a clinical perspective. Crit Rev Microbiol. 2016; 42 (1):1–16. https://doi.org/10.3109/1040841X.2013.875515 PMID: 24495097

32. Lim SMS, Wong PL, Sulaiman H, Atiya N, Shunmugam RH, Liew SM. Clinical prediction models for ESBL-Enterobacteriaceae colonization or infection: a systematic review. J Hosp Infect. 2019.

33. McKinney W, editor Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference; 2010: Austin, TX.

34. Bastian M., Heymann S., Jacomy M. Eds. Gephi: an open source software for exploring and manipulating networks. Third international AAAI conference on weblogs and social. Media. 2009.

35. Masad D, Kazil J, editors. MESA: an agent-based modeling framework. 14th PYTHON in Science Conference; 2015.

36. HUVM. Patient movement preference matrix, https://www.dropbox.com/s/l9edlqwus2g0hmm/HUVM_preference_matrix.csv?dl=0.

37. Rocha LE, Singh V, Esch M, Lenaerts T, Liljeros F, Thorson A. Dynamic contact networks of patients and MRSA spread in hospitals. Sci Rep. 2020; 10 (1):1–10. https://doi.org/10.1038/s41598-019-56847-4 PMID: 31913322

38. Pei S, Morone F, Liljeros F, Makse H, Shaman JL. Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. elife. 2018; 7:e40977. https://doi.org/10.7554/eLife.40977 PMID: 30560786

39. Duval A, Obadia T, Boëlle P-Y, Fleury E, Herrmann J-L, Guillemot D, et al. Close proximity interactions support transmission of ESBL-K. pneumoniae but not ESBL-E. coli in healthcare settings. PLoS Comput Biol. 2019; 15 (5):e1006496. https://doi.org/10.1371/journal.pcbi.1006496 PMID: 31145725

40. Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS One. 2013; 8 (9): e73970. https://doi.org/10.1371/journal.pone.0073970 PMID: 24040129

41. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. Clin Microbiol Infect. 2014; 20 (1):10–6. https://doi.org/10.1111/1469-0691.12472 PMID: 24267942

Declarations of independence

(1) I declare that I have not completed or initiated a doctorate procedure at any other universities.

(2) I declare that all information given is accurate and complete. The thesis has not been used previously at this or any other university in order to achieve an academic degree.

(3) I declare under oath that this thesis is my own work entirely and has been written without any help from other people. I complied with all regulation of good scientific practice, and I used only the sources mentioned and included all the citations correctly both in word and content.

Halle (Saale), Hanjue Xia

Declarations of previous dissertation attempt

I declare that this work is the first attempt of writing a dissertation. Also, I declare that this work is exclusively submitted as a dissertation for the Medical Faculty of Martin Luther University Halle-Wittenberg.

Halle (Saale), Hanjue Xia

Acknowledgments

I would like to thank the following individuals for supporting this doctoral research:

- Prof. Dr. Rafael Mikolajczyk, for his supervision, guidance, and support throughout this project.
- Johannes Horn, PhD, for his insightful discussion and valuable support during the doctoral research project.
- apl. Prof. Dr. Andreas Wienke, for his valuable assistance.
- Very special gratitude goes out to all Institute for Medical Epidemiology, Biometrics and Informatics (IMEBI) doctoral candidates for the warm friendships and knowledge sharing.
- I wish to acknowledge the help provided by the technical and support staff at IMEBI.
- My beloved parents Zhidong Xia and Jinfeng Chen, husband Thorsten Goebel for their unconditional support and understanding.
- I thank grant from the national funding agency Bundesministerium für Bildung und Forschung (BMBF), Germany, 01KI1704C (Martin-Luther-University Halle-Wittenberg, Medical Faculty, Institute of medical epidemiology, biostatistics and informatics) to Rafael Mikolajczyk.