

## RESEARCH ARTICLE

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

Hanjue Xia <sup>1</sup>, Johannes Horn <sup>1</sup>, Monika J. Piotrowska <sup>2</sup>, Konrad Sakowski <sup>2,3</sup>, André Karch <sup>4</sup>, Hannan Tahir <sup>5</sup>, Mirjam Kretzschmar <sup>5</sup>, Rafael Mikolajczyk <sup>1\*</sup>

**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



## OPEN ACCESS

**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

**Copyright:** © 2021 Xia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**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, Unicono: 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

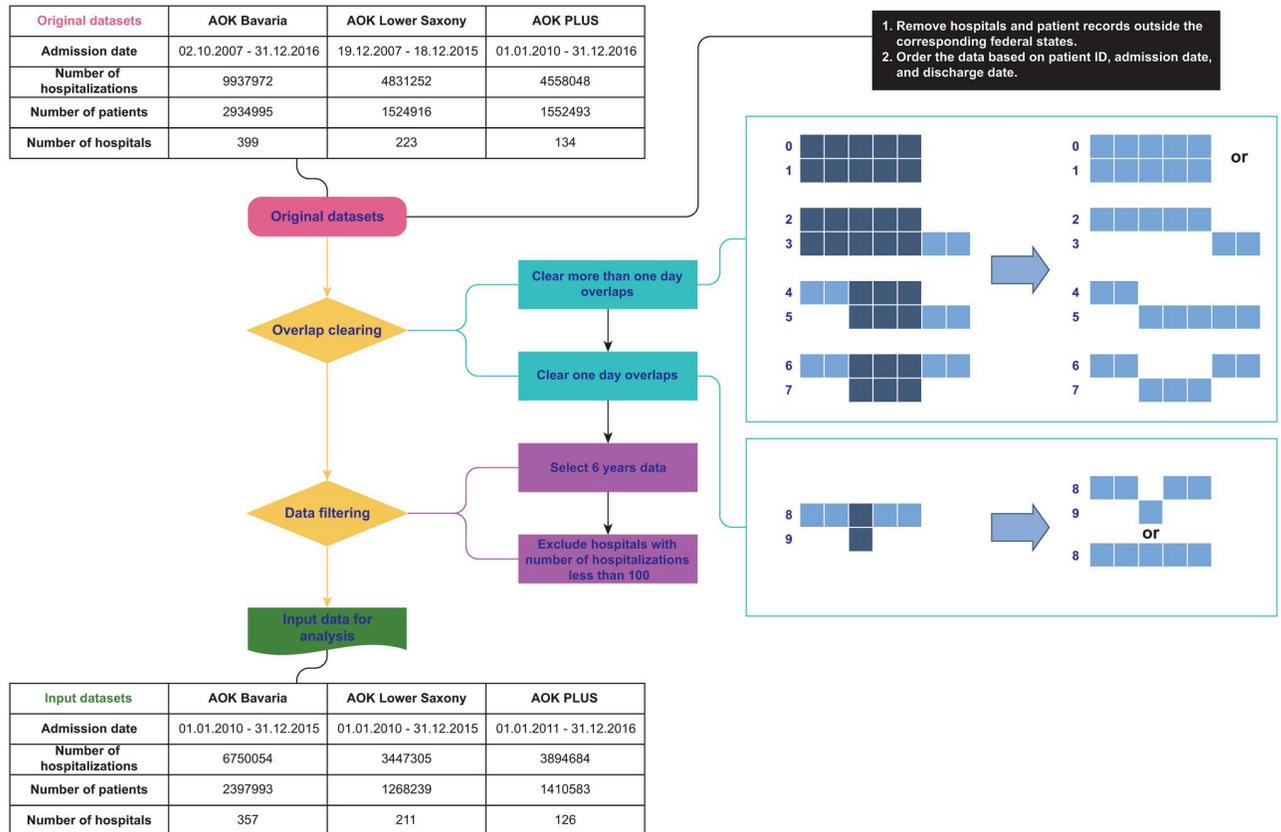


Fig 1. Data filtering process.

<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

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

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  $k_i^{in} = \sum_{j=1}^N \text{sgn}(w_{ji})$  and  $k_i^{out} = \sum_{j=1}^N \text{sgn}(w_{ij})$ , 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.  $\text{sgn}$  is the sign function with

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

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  $d_{ij} = \min_h \left( \frac{1}{w_{ih}} + \dots + \frac{1}{w_{hj}} \right)$ ,  $h \in V \setminus \{i, j\}$ . 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  $\beta_L^t$  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

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

Here,  $N_{L,L}^t$  denoted the number of infectious patients at location  $L$  on day  $t$  and  $N_L^t$  was the total number of patients with  $t = T_1, T_1 + 1, \dots, T_2$ .  $T_1$  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  $\gamma$  per day. Following Scanvic et al. [34] and Donker et al. [11], we set  $\gamma = \frac{1}{365} \text{ days}^{-1}$  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:

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

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:

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

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

$w_{ij}^{Net_O}$  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:

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

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

$H_i^W$  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:

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

$Pr_t$  was the prevalence on day  $t$  with  $t = T_1, T_1 + 1, \dots, T_2$ , 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  $\beta$  and  $\gamma$  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 ( <i>N</i> )	$\rho^{DT}$	$\rho^{in-DT}$	$\bar{w}^{DT}$	$\bar{w}^{in-DT}$
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;  $\rho$  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  $\rho^{DT}$  and  $\rho^{in-DT}$  (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.

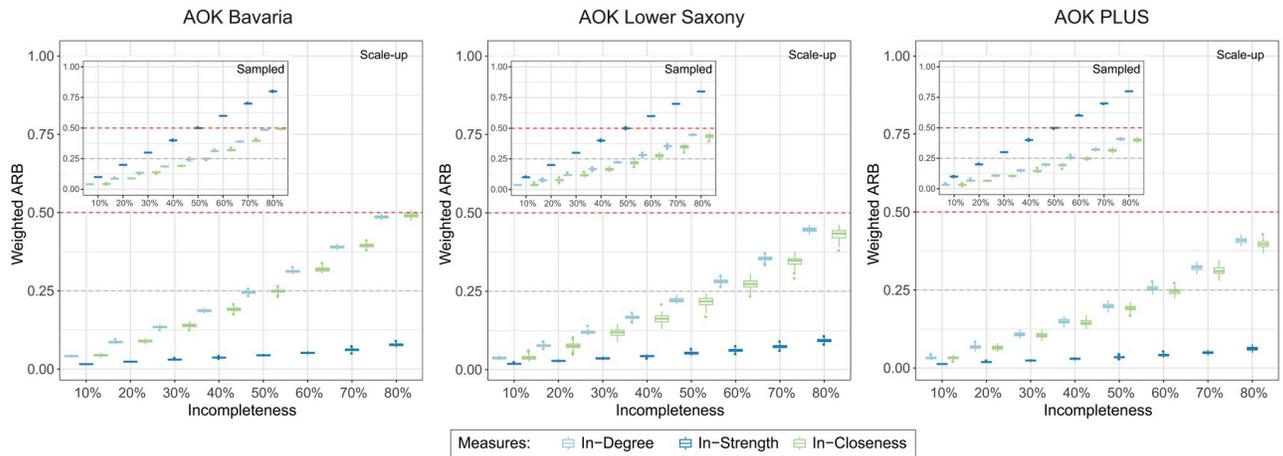
To further assess the impacts of incompleteness on network characteristics, we used cosine similarity (CS). The findings are presented in S12–S15 Figs. The CS showed more stable results in the dataset of the AOK PLUS, which had less area of low CS values at the same incompleteness level as the other 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 ( <i>M</i> )	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>



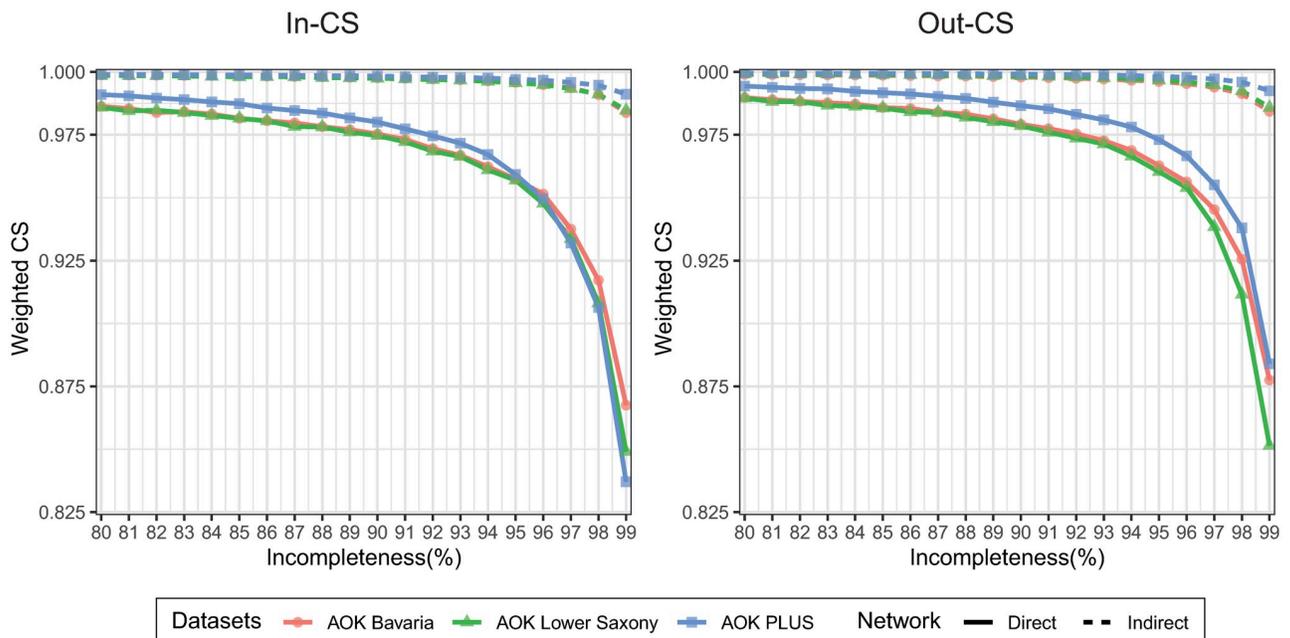
**Fig 2. Impact of sampling and scale-up on direct hospital network measures (visualized in boxplots).** The weighted absolute value of relative biases (ARBs) across different network measures on sampled (inset) and scale-up (main plot) datasets for various incompleteness levels. Incompleteness was defined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

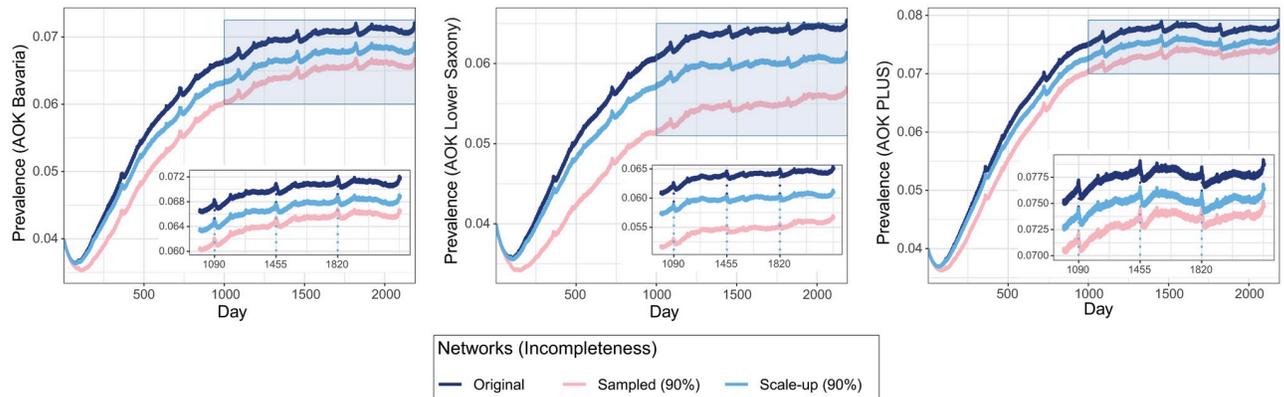
### Effects of incompleteness on simulated spread of infections

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

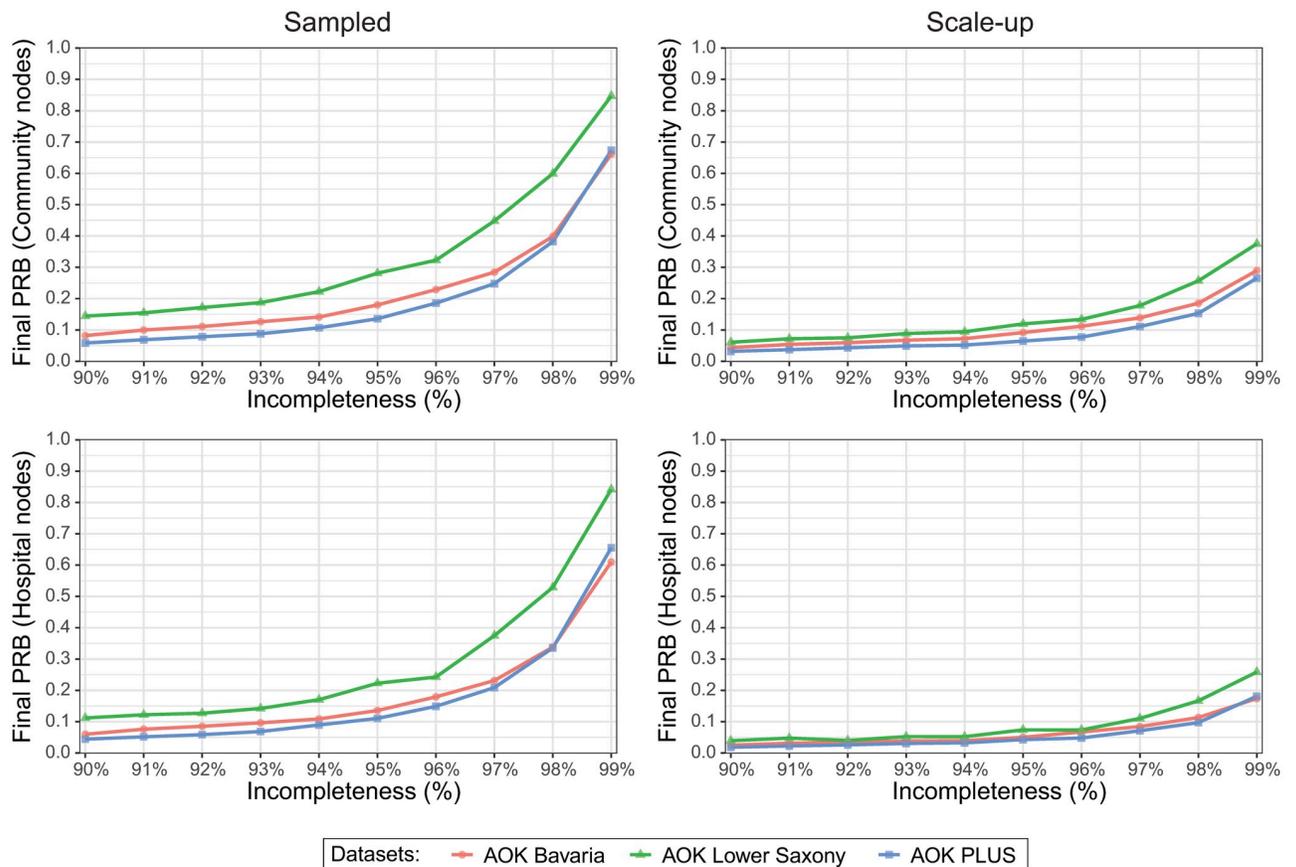
After determining that the effects of incompleteness became visible, when incompleteness levels surpassed 90%, we further focussed on this area and studied the higher incompleteness more in detail (Fig 5). At this high incompleteness level, the scale-up datasets performed better than sampled datasets. 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  $\beta$  and  $\gamma$  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.



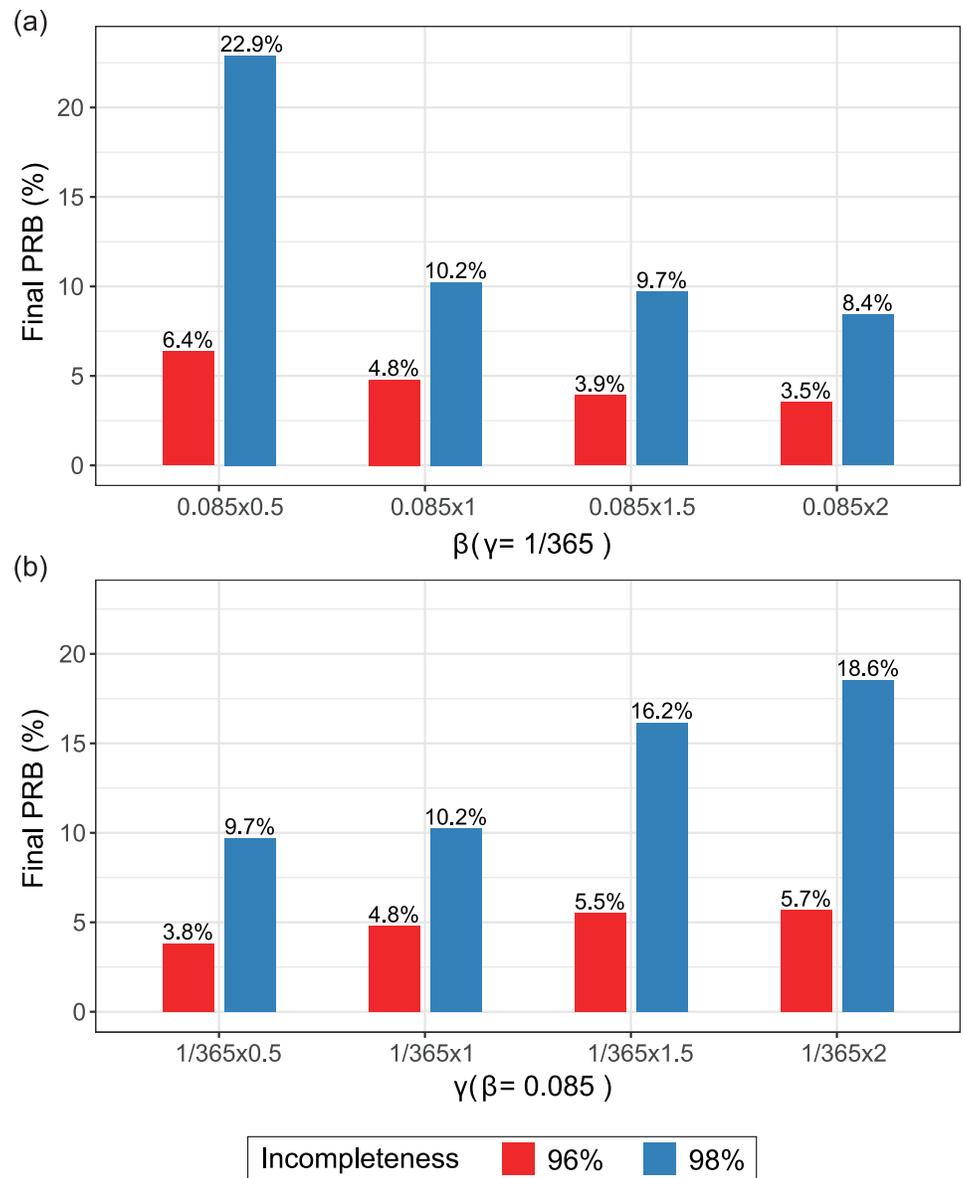
**Fig 5. Final PRBs of simulated networks for each AOK plotted against percentage of incompleteness.**

<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  $\beta$ . In (b) we presented the final PRBs in hospital nodes for different values of  $\gamma$ .

<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 insureds, which could again affect generalization of our findings even within the studied regions. For example, a dataset from an insurance company with younger insureds 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  $\beta$ . In (b) we presented the final PRBs in community nodes for different values of  $\gamma$ . (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. Genoï 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, Génoï 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](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