Transmission Visualizations of Healthcare Infection Clusters: A Pilot Scoping Review

by

Mya Brady

BS in Psychology, University of Pittsburgh, 2020

Submitted to the Graduate Faculty of the

Department of Epidemiology

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

Mya Brady

on

December 10, 2021

and approved by

Essay Advisor: Nancy Glynn, PhD, Associate Professor, Department of Epidemiology,
Graduate School of Public Health, University of Pittsburgh

Reader: Graham Snyder, MD SM, Associate Professor, Medical Director of Infection Prevention
and Hospital Epidemiology, University of Pittsburgh Medical Center

Reader: Elise Martin, MD MS, Assistant Professor, Department of Medicine, University of
Pittsburgh School of Medicine

Reader: Lee Harrison, MD, Professor, Department of Epidemiology, Graduate School of Public
Health, University of Pittsburgh

Reader: Ravi Sharma, PhD, Adjunct Assistant Professor, Department of Epidemiology,
Graduate School of Public Health, University of Pittsburgh

Transmission Visualizations of Healthcare Infection Clusters: A Pilot Scoping Review

Mya Brady, MPH

University of Pittsburgh, 2021

Abstract

Background: Understanding transmission of COVID-19 in healthcare settings is critical to infection prevention and control efforts to mitigate its spread. Implementing interventions to interrupt transmission requires deriving hypothesized transmission pathways. Visualizations of transmission pathways can aid in hypothesis generation. SARS-CoV-2 provides a unique opportunity to determine ways in which data visualizations can be improved to aid in outbreaks due to presymptomatic/asymptomatic transmission, highly variable incubation period and transmission that can involve multiple individuals in healthcare facilities. The objective of this review was to conduct a scoping review of the current literature of transmission visualizations in the healthcare setting to describe the types and frequency of data elements used in these types of visualizations.

Methods: Medline (Ovid) was searched using a combination of MeSH terms, title, abstract and keywords developed in tandem with a University of Pittsburgh Graduate School of Public Health Health Sciences Librarian. Terms were cross-referenced with a set of known studies to ensure that the search would capture relevant articles. Article eligibility criteria was determined a priori. Inclusion criteria contained the following: published after 1985, written in English, peer-reviewed, healthcare facility infectious disease transmission, an infectious disease with ≥ 1 transmission event or infectious diseases with a National Healthcare safety Network (NHSN) definition, ≥ 1 data visualizations of transmission using data observable by an infection

preventionist showing temporal and/or spatial relationships using patient health data. The articles were screened and selected using DistillerSR (Evidence Partners) reviewing software and the protocol was published on Open Science Forum (OSF) for transparency purposes.

Results: The initial search yielded 1,958 articles; 15% (299 articles) were used for this pilot review based on alphabetical order of author's last name. Eleven articles were eligible for full review and 21 data visualizations were analyzed. Of the 21 data visualizations, all described either bacterial or viral transmission, almost all visualizations contained spatial data elements and patient data elements. None of the visualizations contained contagious periods.

Conclusion: The findings from this review support the need for a standardized data visualization tool to implement public health and infection prevention interventions sooner to interrupt transmission.

Table of Contents

List of Tables

List of Figures

Preface

I would like to thank everyone that has helped and guided me through writing this masters essay and throughout my graduate school career. First, I would like to thank the faculty at the Graduate School of Public Health for providing me with wisdom, knowledge and critical thinking skills during my semesters at the University of Pittsburgh. I would also like to thank Dr. Nancy Glynn for providing me with support, guidance, and courage through various points of my MPH journey. I would like to thank Julie Slaughter for her exceptional time management skills and for providing me with all of the support and kindness needed to keep me on track for this essay. To Dr. Graham Snyder for always being available to have important conversations, for constant guidance throughout this essay process, and for providing an abundance of confidence in the project. Additionally, I would like to thank Dr. Elise Martin for always being willing to help and for being available to provide guidance. Thank you to Dr. Ravi Sharma and Dr. Lee Harrison for agreeing to be part of this essay committee and for your helpful insight on this research area. A big thank you to the UPMC Presbyterian-Shadyside Infection Prevention and Control team for allowing me to shadow the important work conducted on a regular basis to ensure that the hospital is the safest place for patients during an unprecedented time. I would also like to thank Helena VonVille for her confidence in this project and her scoping review guidance and expertise. Furthermore, I am beyond grateful for the support I have received from the community I was able to establish at Pitt Public Health during the pandemic. I would like to thank Joseph White, Nathan Raabe, Amina Chtourou, Rosie Benford, and Kelly Friday for providing immense moral support.

1.0 Introduction

1.1 Infection Preventionists and Outbreak Investigations

Infection preventionists (IPs) perform surveillance for infections within healthcare settings and work to prevent nosocomial infections. Surveillance data are used to identify and prevent transmission of serious pathogens between patients, healthcare workers, and other individuals in the healthcare setting. Infection rates are used to determine whether a certain infectious disease or pathogen is appearing in a higher frequency than usual. If the usual threshold is exceeded, an investigation is typically initiated.[1] Cluster investigations can also be initiated in response to individual events of infections when the pathogen is unusual or is appearing more frequently than expected. Conducting cluster investigations is essential for preventing spread of infections.[2]

Descriptive epidemiology techniques used during a cluster investigation include performing surveillance for additional case-finding, identifying potential contacts, extracting information from electronic medical records, and merging this data with transmission information (such as time of exposure to the infectious agent, disease onset, and symptomatic period), and identifying potential risk factors for patients.[3] Transmission events hypothesized by clustering patterns of a common organism are often confirmed with a method to determine genetic relatedness, such as whole genome sequencing.[4]

Transmission visualization tools are data visualizations showing networks of transmission (individuals and interactions/connections between them) comprising an outbreak.[5] They can be used to aid in the outbreak investigation process and may useful in a dynamic setting such as healthcare facilities due to the complexity of the conditions of this environment.[5] Additionally,

transmission visualization can lead to hypothesis generation that can then be tested using methods to establish genetic relatedness.[5] Therefore, visualizing transmission pathways can be a highly informative step for IPs to investigate infectious disease clusters and respond with implementation of transmission-interrupting interventions. [5]

Currently, transmission visualizations of transmission pathways are produced by incorporating different elements of transmission events and different software tools. These types of visualizations may display time, space, and/or other elements; however, there is no established best practice.

1.2 COVID-19 Epidemiology

Thus far, the COVID-19 pandemic has resulted in over 700,000 deaths in the United States.[6] SARS-CoV-2 transmits from person-to-person through small and large respiratory droplets. Individuals become infected when the viral particles from the respiratory droplets are inhaled or land on areas with mucosal membranes (eyes, mouth, nose) through close contact with an infected individual.[7] Some ways to mitigate COVID-19 spread are through identification of sources of outbreaks, using surveillance techniques to track spread, studying COVID-19 through surveillance data, and to create guidance for action that can reduce the spread and reduce the burden of disease.[8] However, more effective measures to reduce the spread of COVID-19 are social distancing, masking in indoor spaces, and vaccination.[9]

1.3 COVID-19 Transmission

In the early months of the pandemic, healthcare systems experienced overwhelming influxes of COVID-19 cases due to lack of precautions such as masking, distancing, COVID-19 restrictions, and the absence of a vaccine.[10] Nosocomial transmission of SARS-COV-2 rates vary depending on the population affected.[11] In a SARS-CoV-2 outbreak in a tertiary care center in London, in March-April 2020, 15% of positive cases were linked to nosocomial transmission.[12] However, in a Boston Women's Hospital during an outbreak that occurred from March 7th and May 30th 2020, only one positive case was associated with nosocomial transmission.[11] Determining exact pathways of infection can be difficult for every COVID-19 positive case, especially within a healthcare setting where patient turnover rate is high, visitors come and go, and healthcare workers sometimes work on different floors and units within a hospital.

An observational cohort study from January 2020 to October 2020 conducted in acute care facilities in the United Kingdom looked at associations between transmission pathways in healthcare workers and patients using linear, logistic, additive and mixed regression models.[13] This paper found strong associations of nosocomial transmission between healthcare workers with COVID-19 infections and patients. Additionally, they found that healthcare workers who were nurses had highest risks of infection with COVID-19. Finally, risk of transmission to healthcare workers was strongly associated with healthcare workers in general medical units, and when healthcare workers and patients were infectious on the same unit.[13]

A prospective cohort study conducted from March 24, 2020, to April 23rd, 2020 compared the risk of SARS-COV-2 infections among healthcare workers to the general public in the United Kingdom and the United States to determine prevalence of SARS-CoV-2 infection. Based on the study population enrolled from both the U.S. and the U.K., the investigators determined that the

prevalence of COVID-19 in this particular population during this time in the pandemic was 2,724 cases per 100,00 healthcare workers compared to the prevalence of COVID-19 in the community, which was found to be 242 per 100,00 people.[14] Ultimately, this study concluded that in the U.K. and the U.S. front-line healthcare workers had approximately a threefold increased risk of COVID-19 infection compared to the general population when adjusting for additional risk factors.[14]

To understand healthcare worker to understand the risk of healthcare worker to patient transmission of SARS-CoV-2, an observational prospective study was conducted from March 1st, 2020 to June 10, 2020 based on contact tracing efforts for positive healthcare workers and patients at an acute care facility.[15] Among 253 patients exposed to healthcare workers infected with SARS-CoV-2 in a multitude of settings (inpatient, outpatient clinics, emergency departments), transmission was <1%.[15] Even though transmission risk is lower for healthcare workers to patients, compared to patients to healthcare workers, exposure and transmission is still possible between these groups of individuals and is something to consider for when investigating potential nosocomial SARS-CoV-2 cases.

In an acute care facility in Israel in July 2021, there was nosocomial spread of SARS-CoV-2.[16] Even though this population was highly vaccinated, during this outbreak investigation cases were linked to a single patient case with an attack rate (AR) of 23.3% among patients, and an AR of 10.3% among healthcare workers.[16] This outbreak might be attributed to waning immunity from vaccination, however, it still shows that the risk of SARS-CoV-2 remains an issue for nosocomial transmission even within a vaccinated population.

From April 4, 2021 through July 17, 2021 in the United States, the age standardized rate of hospitalizations among unvaccinated and vaccinated individuals was 9.1 per 100,000 persons and 0.7 per 100,000, respectively.[17] As of November 15, 2021 in the United States, only 56% of

white individuals, 49% of Black individuals, 54% of Hispanic individuals, and 72% of Asian individuals have received a COVID-19 vaccine.[18] There is still a significant percent of the population in the United States that remains unvaccinated against COVID-19. This difference in hospitalizations between unvaccinated individuals compared to vaccinated individuals, the percentage of individuals that remain unvaccinated, and the ability for nosocomial spread to still occur even in a country with a high vaccination rate, provides strong evidence that SARS-CoV-2 infection has and will continue to result in hospitalizations. Influxes of hospitalized COVID-19 patients results in high prevalence within healthcare facilities leading to a potential higher risk of nosocomial transmission events if infection prevention and control measures are not initiated or are not followed. Especially in the case of SARS-CoV-2 transmission, vaccination status is important as it may affect whether the individual becomes infected with the virus and the length of time/procedures for quarantine/isolation. Therefore, this should be considered an important variable within a data visualization that is not necessarily accounted for depending on the type of pathogen transmission depicted.

Infection control efforts by IPs for COVID-19 can be particularly challenging due to presymptomatic or asymptomatic transmission.[19] To find presymptomatic and asymptomatic COVID-19 cases, regular Polymerase Chain Reaction (PCR) screening is needed[20] along with contact tracing and this needs to be initiated directly after potential exposures. Transmission in presymptomatic and asymptomatic cases can also be halted by quarantining all individuals exposed for 14 days or testing 100% of the population daily without contact tracing but these measures are extensive and may not be cost effective or feasible when dealing with a large healthcare facility. However, even these efforts may not be enough to completely halt SARS-CoV-2 transmission if not initiated quickly and efficiently. Healthcare settings are dynamic

environments for transmission events to occur. There can be variability in patient and healthcare adherence to masking, that can increase risk of exposure and transmission to other healthcare workers or patients.[14] There is also constant movement of healthcare , visitors, and patients within healthcare facilities that can complicate transmission and transmission pathways. The wide range incubation period of SARS-CoV-2 of 2-14 days also makes it challenging to determine the exact source of acquisition.[21] This emphasizes the importance of understanding infectious disease transmission within healthcare settings to reduce the risk of patient harm.

SARS-CoV-2 makes for the ideal pathogen to ensure that transmission visualizations can be the most informative and effective for generating transmission hypotheses. SARS-CoV-2 transmission is complex as it can affect healthcare workers, visitors, and patients. Additionally, this virus has the possibility of asymptomatic or presymptomatic contagiousness which can create difficulties when attempting to contain the spread, break the chain of transmission, and determine the exact point of exposure. The highly variable incubation period of SARS-CoV-2 can be modulated through proper PPE use and vaccination status; however, improper PPE use and unvaccinated individuals can contribute to the spread of the virus especially within healthcare settings. Therefore, this virus provides a unique opportunity to understand how data visualization tools can be helpful in the event of an outbreak in a healthcare setting.

1.4 Important Elements for Transmission Visualizations

To determine the types of important features for inclusion in transmission visualizations, two separate focus groups were held in June 2021 prior to the start of the article search. The focus groups were conducted by the author with current IPs in a University of Pittsburgh Medical Center (UPMC) acute care facility, an IP at a UPMC long-term care facility, and an IP at a UPMC behavioral health facility on Microsoft Teams. IPs were recruited through known contacts of the research mentor (G.S.). The first group contained the acute care IP, the long-term care IP, the behavioral health IP, and the author, and the research mentor. The second focus group contained two UPMC acute care facility IPs and the author. This preliminary work was conducted for two reasons: first, to determine the scope of the problem, and second, to determine data elements for data extraction from the eligible articles for the scoping review. These discussions were structured by first introducing the problem, asking whether this tool would be useful for their facility, what elements they thought were necessary to have in a data visualization, and to discuss concerns or unique aspects of their facility when thinking about transmission in their healthcare facilities.

 In these conversations, essential features for a visualization tool were determined to be: having an element of space (e.g., unit, floor, room), identification of role of individuals (e.g., healthcare workers, patients, visitors), unique identifiers for infected individuals, test positive date, and symptom onset date. Features that would be helpful, but are not necessary to include for data visualizations: designating individuals as visitors, displaying vaccination status, patient room transfers (i.e., moving from ICU to a general medicine floor), and whether support persons/patient visitors complied with hand hygiene and any necessary personal protective equipment. Procedures in other departments are also important to consider for when investigating a cluster. This is supported by results from The Enhanced Detection System for Healthcare-Associated

Transmission (EDS-HAT), a system that uses genome sequencing technology and computer algorithms that utilize patient electronic medical record data to detect transmission routes and sources of outbreaks.[22] The frequency of these elements were evaluated in this scoping review, an initial comprehensive evaluation of the existing literature surrounding a topic that has location, time, and source,[23] to determine which of these elements were the most pertinent based on aggregate data and should therefore be included in a standardized transmission visualization.

1.5 Software to Visualize Transmission Pathways

A paper published in 2020 described the overall history and theory behind data visualizations in order to provide some recommendations for the infection prevention field.[24] In this paper, Salinas et al. described the limitation that comes with epidemiological surveillance and creating data visualizations. The researchers highlighted the difficulty of having a data visualization that is a "one-size fits all" approach due to differing needs of healthcare facilities. Additionally, they specified best practices for creating a data visualization such as the use of color, zero as the baseline for displaying information, using graphs and ways to visualize information such as case counts, rates, ratios, SIRs, and surgical site infection compliance data. However, this paper did not specifically discuss using data visualizations as a tool to aid in transmission hypothesis generation or best practices for using data visualizations to display this type of data.

There was a systematic review conducted in 2013 that aimed to provide a broad overview of infectious disease data visualizations. The review found that usability and utility were considered to be important functions when determining whether a data visualization tool is helpful to the user. The article also highlighted the limited success of widespread implementation of data

visualization techniques and tools. However, the researchers concluded that organizations focusing on public health practice need to find infectious disease data visualizations that incorporate input from public health professionals throughout the development process to determine the best practice for these types of software.[5] This paper mostly analyzed the reasons and limitations for infectious disease data visualization use and did not discuss types of data visualizations that should be made to display transmission events or best practices for a standardized transmission visualization tool.

Another systematic review conducted in 2014 analyzing spatial and temporal analyses of disease transmission in healthcare settings aimed to identify the use of these types of analyses in preventing Healthcare Associated Infections (HAIs) and to propose expanding the use of these methods for investigations. The researchers found that only 15% of studies included in the systematic review included spatiotemporal analyses.[24] The researchers argued that this was indicative of a need for implementing the use of spatiotemporal analyses in the prevention of transmission of HAIs.[25] This article only described whether analyses included spatiotemporal analyses, it did not look at the types of elements used within data visualizations or describe the types of common elements that should be included in a standardized data visualization tool.

Finding an ideal software to visualize infection clusters in a healthcare setting may aid in the efficiency and effectiveness of outbreak investigations. Examples of current transmission visualizations used to describe infectious diseases clusters include using heat maps, dot maps, Gantt charts, commercial products like TheraDoc (Premier Inc), line lists and timelines (**Figure 1**). Based on focus group conversations, Excel (Microsoft Corp) is also used as a data visualization tool used to show time on the x-axis and cases as rows with specific indictors such as contagious windows, test dates, and locations. However, a few programs have been developed by researchers to try and make transmission visualizations more innovative.

9

Dotmapper was developed by researchers to be an easily accessible interactive mapping tool to visualize infection clusters using Geographic Information Systems (GIS) (Esri).[26] However, this tool is limited, as it only visualizes clusters over large geographic regions.[26] It is important to understand transmission within a large indoor setting with constant movement of individuals, yet GIS is non-inclusive of these settings such as hospitals. Knox statistical testing has also been used to understand spatial relationships of case clusters since the 1960s.[25] The Knox test only determines whether pairs of cases are clustered in time and space and does not consider other susceptible individuals, or controls. Therefore, this statistical method does not provide a full visual of the transmission event.[25]

A visualization application like Dotmapper that was developed for nosocomial transmission is HospMapper (developed by: Smith, University College London, UK).[26] This tool can show data for patients who have a positive test for a pathogen. Its output includes an epidemic curve, a timeline. and a schematic ward plan. However, this software does not have support for troubleshooting and would need someone with expertise in the software program/programing language R (Developed by: Smith & Hayward, University College London, UK).[26] It also has disadvantages regarding integration in current data from healthcare systems, and would need training for formatting purposes. This software also does not look at healthcare workers or other individuals that may pass through the hospital.[27] This is important, as patients are not the only individuals that need to be investigated when there is an outbreak in a healthcare setting.

Figure 1 Types of data visualizations commonly used and the information displayed

Types of Data Visualizations
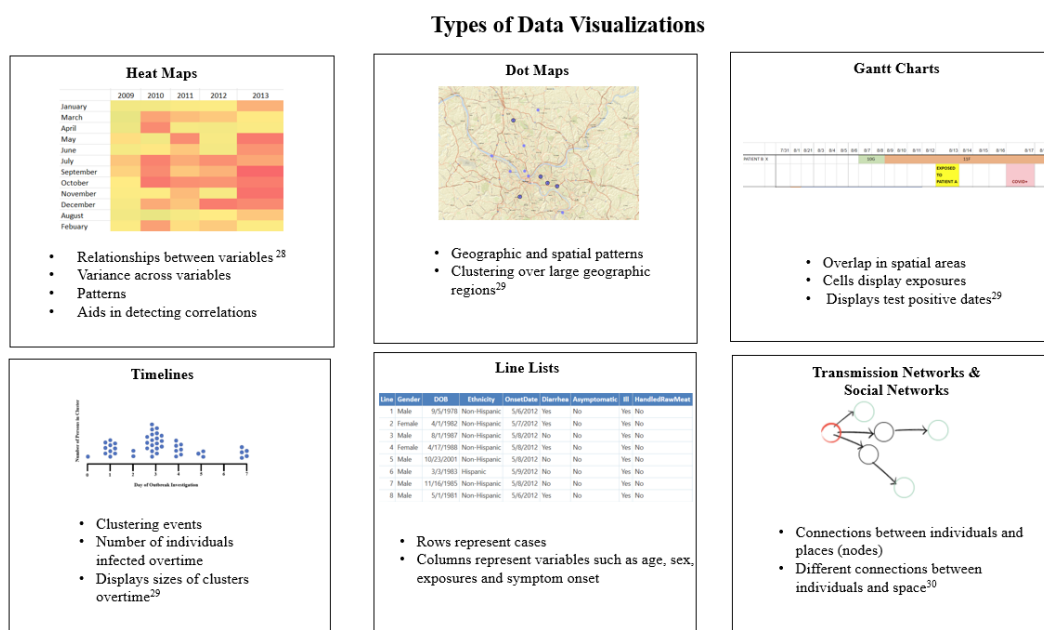
Note: citations for information for the data visualizations are as follows: Heat Map[28], Dot Maps[29], Timelines[29], Gantt Chart[29], Transmission networks & social networks.[30] All visualizations are hypothetical examples created by the author this year.

1.6 Gaps in Knowledge

To my knowledge, there is no peer-reviewed English-language literature describing the differences in elements contained within data visualizations to show transmission pathways of clusters in healthcare settings. Additionally, the different types of spatiotemporal elements that can be incorporated within these visualizations have not been systematically characterized. Since there is no standardized practice or criteria for creating transmission visualizations, describing published methods of visualizing transmission pathways in healthcare settings will aid IPs in understanding available resources, and allow for the description of an ideal software for creating these types of visualizations that programmers and IPs may jointly create. Therefore, this scoping review

describes the need for best practice criteria based on essential visualization elements to understand nosocomial transmission and generate transmission hypotheses.

1.7 Public Health Significance

This review will aid in IP's understanding of the types of visualizations that can be used to generate transmission hypotheses. This will allow for earlier initiation of infection prevention interventions, which may lead to earlier termination of nosocomial outbreaks.

2.0 Objectives

The objectives of this study were to:1) evaluate the infectious disease transmission visualizations used within healthcare settings; and 2) determine which methods of visualizing transmission of infectious diseases in a healthcare setting are most effective in interrupting transmission.

3.0 Methods

3.1 Eligibility Criteria

For articles to be considered eligible for this scoping review, the articles needed to be peer-reviewed, written in English, and published after 1985, when IPs first documented using electronic surveillance, and using electronic medical records and information technology tools that allow for data collection methods that are automatic).[31,32] All eligibility criteria were selected in the search algorithm. For this pilot study, 15% of articles from the original search were evaluated,, selected alphabetically by author's last name.

Articles were excluded if they did not contain patient data with a methodology appropriate for evaluating a healthcare associated infectious disease or infectious pathogen cluster, such as observational cluster investigation methods.[2] Therefore, cost-benefit analyses, meta-analyses, randomized control trials, community outbreaks, and other non-healthcare studies were not included in the review. Articles were also excluded if they did not meet any of the below inclusion criteria.

Full article review used the following inclusion criteria to select articles for data abstraction:

- Infectious diseases needed to meet Healthcare Associated Infection (HAI) definitions per the National Healthcare Safety Network (NHSN)[33] but could be an infectious disease without an NHSN HAI classification if ≥1 transmission event occurred within a healthcare facility.
- All sources needed to describe an infectious disease cluster.

14

- Articles had to contain one or more data visualizations demonstrating pathogen transmission pathways with data observable by an IP showing temporal and/or spatial relationships using patient health data and epidemiological data. For example, data visualizations cannot contain only a phylogenetic tree, which alone would not include various person type data (i.e., patients, healthcare workers) or test positive dates; however, a phylogenetic tree can serve as a data visualization if it incorporates other non-genetic epidemiologically relevant data.

- The facilities where transmission occurred must be considered a healthcare facility. Healthcare facilities were defined as structures whose primary purpose is designed to provide care to individuals where person-to-person transmission can occur and where healthcare personnel are present. Examples of healthcare facilities include long-term care facilities, acute care facilities, rehabilitation facilities, and behavioral health facilities.

- Multiple clusters in an outbreak or publication may be described if each cluster is contained in one healthcare facility.

Articles which failed to meet any of the above criteria, were excluded from further analysis.

3.2 Information Sources and Search Strategy

Medline (Ovid) was searched by a health sciences librarian, Helena VonVille, MLS, MPH, University of Pittsburgh Graduate School of Public Health, with systematic review experience. The date of the last search was August 24, 2021. Concepts that made up the search were healthcare acquired infections and cluster analysis or geographic mapping. A combination of MeSH terms

and title, abstract, and keywords was used to develop the Medline search that was checked against a known set of studies.

Articles were also considered for review if they are cited as references within the papers found through the original Medline search and appear to have relevance for this review. These articles needed to meet inclusion criteria to be reviewed.

3.3 Search Strategy

The following table (**Table 1**) details the initial Medline (Ovid) search:

Table 1 Medline search terms used for initial article selection in Ovid-Medline

Line #	Search Terms
1.	Cross Infection/ or "Healthcare-Associated Pneumonia"/ or "Infectious Disease Transmission, Professional-to-Patient"/ or "Infectious Disease Transmission, Patient-to-Professional"/
2.	((Cross adj1 Infection*) or (((Health adj1 Care) or Healthcare or nurse* or physician*) adj3 (acquired or associated or infected or Infection* or (tested adj3 positive))) or (hospital adj3 outbreak*) or nosocomial).ti,ab,kf.
3.	1 or 2
4.	Cluster Analysis/ or Geographic Mapping/ or Spatio-Temporal Analysis/
5.	((chain or chains or cluster or clustering or clusters or geographic* or heat or indirect or spatial or (spatio adj1 temporal) or spatiotemporal or temporal or time or visual) adj4 (analysis or distribution or distributions or identification or map or mapped or mapping or representation or transmission)).ti,ab,kf.
6.	dotmapper.ti,ab,kw.
7.	4 or 5 or 6
8.	3 and 7
9.	limit 8 to (english language and yr="1985 - 2022")
10.	9 not (animals/ not humans/)

3.4 Selection of Sources of Evidence

For this review, one primary reviewer and author, Mya Brady (MB), BS, University of Pittsburgh Graduate School of Public Health, was used to conduct the title and abstract screen, as well as, the data abstraction from the final selected eligible articles. A second reviewer, Joseph White (JW), BS, University of Pittsburgh Graduate School of Public Health, helped the primary reviewer screen articles eligible for full-text review. The research mentor, Dr. Graham Snyder (GS), MD, SM, University of Pittsburgh Medical Center, aided in the creation of inclusion and exclusion criteria, and the data abstraction form.

Citations were uploaded from EndNote to DistillerSR (Evidence Partners) for the study selection process. All study selection decisions were stored in DistillerSR.

Two reviewers (MB, JW) took part in a Cohen's kappa interrater reliability test prior to screening all full-text articles. For the full-text review, the reviewers needed to reach agreement on at least 80% of their decisions. If not, eligibility questions were evaluated for clarity. When at least 80% agreement was achieved, the reviewers independently screened all full-text articles following the *a priori* eligibility criteria developed by the research mentor and one reviewer. The two reviewers (MB, JW) discussed items in which they disagreed and attempted to reach a consensus. When a consensus could not be reached by the author and a reviewer(MB, JW), the research mentor (GS) independently reviewed the citation and abstract to determine eligibility. The data charting was completed independently by one reviewer (MB) for articles that passed the full-text screening process.

3.5 Selection of Sources

A total of 1,958 articles were found through the Ovid Medline database search. Of these 1,958 articles, 15% of the total were used for this pilot study and 299 articles were screened for title and abstract eligibility (**Figure 2**). In total, 246 (82%) articles failed to meet inclusion criteria and were removed from the first screening. A total of 53 (22%) articles moved to the second screening process. Of the 53 articles in the full-text review second screen, 42 (79%) articles were removed. These articles were removed because three articles did not contain clusters of outbreaks, 14 articles did not have a data visualization, 15 articles did not contain a cluster or a transmission visualization, eight articles did not contain a cluster or a healthcare facility, and two articles did not contain a data visualization or a healthcare facility. Eleven (21%) articles fulfilled the eligibility criteria and were selected for further data abstraction for the pilot scoping review.

3.6 Data Items

Data extracted from the articles were placed into DistillerSR for collection purposes using a data charting abstraction form developed by the research mentor and the author. The following data were abstracted (**Table 2**):

Table 2 Data extraction items with the type of data element along with descriptions and examples

Data Element	Description
Basic Article Citation	Author(s), article title, journal title, and year of publication
Type of Healthcare Facility	Acute care hospital, long-term care facility, behavioral health clinic, etc.
Spatial Elements	Emergency department or other unit of space if used (i.e. unit, floor, facility)
Temporal Elements	Hours, days, months
Number of Cases	Ex. Cluster size: 356 infected individuals
Types of Surveillance	Active surveillance (asymptomatic testing), retrospective surveillance, prospective surveillance
Status of Infected Individuals	Healthcare worker (including type of worker) or patient
Infectious Pathogen	SARS-CoV-2, MERS, influenza, *Clostridioides difficile*, needs to be described by healthcare transmission
Infectious Pathogen Type	Virus, bacteria, fungus
Methods to Determine Genetic Relatedness	Genus or genus & species, antimicrobial resistance phenotype, intermediate genetic relatedness testing (Pulsed-field gel electrophoresis (PFGE), Multilocus sequence typing (MLST), spa typing) etc.
Type of Transmission	Direct, Indirect, Probable, Suspected
Scope of Transmission	Multiple floors, multiple buildings, multiple units, single unit/department
Number and Type of Cases Defined/Identified	Ex. 3 healthcare workers, 5 patients
Software Used for Creating Visualization	GIS, R, Excel, DotMapper, etc.
Type of Visualization	Flow chart, line list, Gantt chart, heat map, cluster map, plotted cases in a floor plan, transmission network
Other Features	Incubation period, specific symptoms, laboratory test details such as specimen location or Ct value, nature of interaction

4.0 Results

The 11 eligible studies were conducted in internationally-represented acute care facilities,[30,34–43] and 36% were from the United States.[36,37,40,41] Other countries included Portugal[34], Brazil,[35] Korea,[38] Australia,[39] Canada,[30] Singapore,[43] and Spain (**Table 3**).[42] These 11 studies were predominantly observational retrospective studies (n=5, 45%).[35,37–39,41] The 11 studies included described clusters with a mean cluster size of 61 individuals (range, 14 to 790 individuals)[39,41] lasting a median of 122 days (range, 4 days – 2,190 days).[38,41] Of the clusters described, 88% of articles discussed patients involved in the transmission clusters (**Table 4**).[30,34–37,39–43] The composition of the infected individuals within the clusters described were mostly patients (N=10, 91%),[30,34–37,39–43] nurses (N=5, 45%),[34,36–38,43] and advanced practice practitioners and physicians (N=5, 45%) (**Table 4**).[34,36–38,43] The causative pathogens described in clusters either viral (N=6, 55%) [34,36–38,42,43] or bacterial clusters (N=5, 45%).[30,35,39–41]

There were 21 data visualizations found in the 11 eligible review articles, with a range of one to three visualizations per study. Almost all data visualizations included spatial elements (n=19, 90%),[30,34–43] about half of the data visualizations contained the patient type person (n=11, 52%).[30,34–38,40,42] None of the data visualizations contained contagious periods (**Table 5**). Most studies did not specify the type of software they used (n=6, 55%),[30,34,37–39,41–43] if any, to create their visualizations. The following software programs were reported to create data visualizations in the eligible articles; (n=5, 45%) AutoCAD,[35] QGis,[35] Cytoscape,[36] PathoSPOT,[27] Pajek 4.09[30] and ArcGIS[40] (**Table 6**).

Less frequently utilized elements were non-patient person types (n=9, 43%), [30,34,36–38,40,42] case type (confirmed versus suspected)[35,37–40,42] (n=6, 29%), time elements (n=9, 43%),[30,34,36,38,39,41,42] and test positive dates (n=5, 24%)[34,36–38,42] across data visualizations. Across Timelines,[36,38] spatial and time elements were always included. Gantt Charts[34] contained patients and spatial elements. Among the three cluster maps,[37,38,43] spatial elements were always utilized. Transmission visualizations[38,39,42] had a varying degree of included elements, with confirmed versus hypothesized pathways (n=2, 67%),[38,42] spatial elements (n=2, 67%)[38,39] and time elements (n=2, 67%)[39,42] utilized most frequently. Social network analyses[30,40] always utilized patients and spatial elements. Spatial proximity visualizations[30,40,41] most frequently contained spatial elements (n=3, 75%).[40,41] Across interaction visualizations[36] and heat maps,[35] both utilized patients, healthcare workers and spatial elements.

5.0 Discussion

In this pilot scoping review of 299 eligible articles, 11 studies with 21 data visualizations were identified that described clusters of infections in healthcare settings. Across the 21 visualizations, almost all included spatial elements, half included patients as a person type, and none of the data visualizations contained contagious periods. This pilot scoping review showed the range of diversity across data visualizations. While none of the data visualizations analyzed in this review incorporated all data elements including test positive date, symptom onset date, time elements, transmission pathways, and case type, several types of visualizations incorporated multiple elements. This review will assist IPs in selecting data visualizations that are most relevant for specific clusters being evaluated in their facilities.

Significant variability was seen in the data elements included across visualization types. With the wide range of data elements included in each of the visualizations analyzed for this review, an IP should explore different data visualization types for the specific circumstances to fit their cluster to analyze it in the most effective way. This variability also suggests that there could be a better data visualization tool developed that is able to incorporate all the elements from the 21 visualizations analyzed since these types of data elements can be important for investigations of clusters.

Across all data visualizations, the most commonly used elements were spatial elements (n=19, 90%)[30,34–43] and patients (n=11, 52%).[30,34–38,40,42] The high inclusion frequency of these types of data elements may be related to with accessibility and availability of information that healthcare facilities typically have. However, it is important to note that to make these types of data visualizations, data needs to be readily available and accessible, which can prove to be a

limitation if not necessarily collected or specified, such as symptom onset date. This may be a reason to integrate electronic medical records with data visualization software to improve accessibility and efficiency when attempting to visualize clusters of infections. Especially within cluster investigations, it is important to be able to have access to and use spatial and patient data to understand how and when a transmission event may have occurred. In the cases of respiratory viral transmission and bacterial transmission, understanding where an infected patient is in space will determine the types of infection prevention and control measures that should be taken to prevent further transmission events from occurring. These types of data, patient information, as well as locations of patients/healthcare workers in space, fit into the way in which standard outbreak investigations are conducted by IPs. IPs are usually notified of a positive patient, or a suspected positive patient, and then can sometimes use a software such as TheraDoc's bed trace to see all the places that the patient has visited in the hospital during their stay and the duration in these locations. IPs can also determine symptom onset using electronic medical records and through discussions with the patient's provider. This is an indication that a standardized data visualization tool would need to be aware of the data IPs have readily available to them.

Within the 21 data visualizations, spatial elements were used 90% of the time and time elements were used in 43% of visualizations (n=9, 43%).[30,34,36,38,39,41,42] These inclusion frequencies have increased from the 15% inclusion of spatiotemporal elements seen in data visualizations from 2014.[25] The rise in spatiotemporal elements may be due to an increased interest in evaluating spatial and temporal relationships between infection clusters in healthcare settings. In the case of a SARS-CoV-2 cluster of infections, dates of positive PCR test are used to determine when to isolate a patient, how long a patient should be placed in isolation, in order to ensure that all healthcare workers are taking proper airborne/droplet precautions when performing

care on the patient and when the patient can safely come out of isolation and standard precautions can be utilized again. Symptom onset and test positive dates are used when patients are symptomatic, and during contact tracing investigations to identify potential exposures to other patients or healthcare workers and break the chain of transmission by determining appropriate recommendations for mitigation, i.e. self-monitoring for symptoms or quarantine. Being able to understand these types of spatiotemporal relationships are especially important with SARS-CoV-2 transmission to initiate infection prevention and control measures as soon as possible to interrupt further transmission.

One observed pattern that is worth further analysis is the difference between the types of visualizations used to analyze droplet versus contact transmission. Bacterial clusters use spatial proximity maps and social network analyses more frequently. In contrast, viral clusters used timelines, cluster maps, Gantt charts, and transmission maps most often. In addition, many of the viral cluster visualizations included patients and healthcare workers while bacterial cluster visualizations mostly included patients. Inclusion of both healthcare workers and patients in viral cluster visualizations may be due to respiratory pathogens such as SARS-CoV-2 and Influenza A having droplet transmission. Especially in the case of SARS-CoV-2, exposures due to poor masking place both healthcare workers and their patients at higher risk of respiratory viral infection. Test positive dates were depicted more often in SARS-CoV-2 visualizations, as well as, in the visualization showing Influenza A transmission which are both respiratory viral infections that can spread through droplet transmission. In visualizations depicting SARS-COV-2 transmission, including test positive dates and symptom onset dates is essential for determining exposure times when the case is nosocomial. *C. difficile* transmission can occur through contact, and one of the main concerns is the spread from patient-to-patient rather than patient-to-healthcare

worker if contact precautions are not adhered to by healthcare workers. *C. difficile* and some other bacterial infections can be spread through contact with the environment which means that temporal associations of these cluster events may be more spaced out compared to temporal associations with viral cluster events. Data visualizations used for different infectious pathogens and the data elements used to describe the clusters might be based on the mechanisms of transmission of the pathogen under investigation. For example, respiratory virus infections cluster events can typically be linked to the same time and space as an infected individual due to exposure time and droplet transmission, and certain bacterial infection cluster events can be linked to the same environmental surfaces or through healthcare worker contact transmission.

The less frequently used elements in data visualizations were contagious periods and symptom onset dates. Symptom onset dates were only used for one SARS-CoV-2 cluster map, whereas contagious period was not contained in any of the data visualizations evaluated. Despite the additional burden that collection of these elements place on healthcare facilities, symptom onset dates and contagious periods should be collected and considered important features in cluster transmission visualizations. For contact tracing and the development of hypotheses of transmission, these types of elements are essential to fully understand the extent of the transmission event.

For all the visualizations evaluated, only three articles mentioned the type of software used. This review provided an overview of the types of elements that data visualizations can display; however, due to the lack of software mentioned, it is difficult to fully comprehend how to create them. Even though some data visualizations mentioned the software used, not all are accessible or intuitive to use to develop a successful data visualization program if an individual lacks training in programs such as R and GIS. Some of the software that was created by researchers to aid in

cluster visualization such as DotMapper and HospMapper require specific software knowledge and technological limitations such as needing to have the data formatted in a specific way before running it through the program. Ideally, software would need to be easily integrated into current infection prevention workflow to be an efficient and helpful tool. This software would also need to include all necessary elements to understand a cluster without being difficult to interpret or read.

All of the outbreaks in the articles were described after the outbreak occurred within the healthcare facility and infection prevention measures were initiated. The data visualizations contained within these articles illustrated the transmission events and pathways that were already known. However, it would be ideal to find a software that is able to generate hypotheses about transmission events at the beginning of an outbreak, so that it can aid in an outbreak investigation and be used as a resource to prevent further transmission events from occurring.

5.1 Strengths and Limitations

A key strength of this paper was the range of bacterial and viral transmission, as well as contact and droplet routes of transmission in the visualizations included. The inclusion of articles with a variety of both viral and bacterial types of pathogens allow for increased generalizability and applicability of this paper to different types of clusters that might be evaluated by IPs in their facilities. Additionally, the articles were from a range of countries with varying cluster sizes. Finally, to our knowledge this review is the first of its kind evaluating specific data visualization elements in the published literature from an infection prevention perspective.

This review also has limitations. Although not a requirement for inclusion, all studies in the initial pilot were from acute care facilities and this initial pilot has a small sample size with only 11 articles and 21 data visualizations included. All articles were chosen based on alphabetical selection, which was non-random. Therefore, the common elements found in these articles may not be generalizable to other healthcare facilities looking to visualize transmission pathways such as long-term care facilities, behavioral health facilities, or psychiatric facilities. These types of healthcare settings have some distinct differences from acute care facilities and transmission can occur differently due to shared common spaces by patients. Although multiple bacteria and viruses were included in this pilot, many of the studies depicted SARS-CoV-2 transmission and it is possible that the findings of this study are skewed towards features that are important in SARS-CoV-2 transmission compared to other pathogens that may not be adequately represented in this study. Finally, publication bias may have affected the methods of this paper. Healthcare facilities may be less likely to publish articles about clusters due to potential concern for negative coverage that may impact a facilities' reputation. Therefore, the selection of articles published describing clusters may be limited to larger healthcare facilities compared to smaller ones, or may only publish smaller clusters that were well controlled. There may also be publication bias towards large academic acute care facilities that usually have more healthcare personnel, more patients and more of an ability to publish studies based on the large amount of data that the facility has access to. None of the studies included in this review were smaller institutions and so this review has underrepresented these types of healthcare facilities and the scopes of outbreaks that can occur under a smaller setting.

5.2 Conclusions

When looking through the 11 articles describing cluster investigations and analyzing 21 data visualizations across articles, there was significant variability in the types of data elements included in visualizations, and almost all visualizations included spatial elements and most contained patients. Overall, the results from this scoping review suggest that there are still many inconsistencies and variations across transmission visualizations in the types of data elements that are displayed. Reports of cluster investigations should be included in the methods along with a clear description of the software and elements included in the data visualization. It would be advantageous to conduct a larger scoping review that can analyze more articles to fully comprehend the types of elements that should be included in these types of data visualizations since this review only analyzed 15% of the original articles returned by the initial search. In addition, a data visualization tool should be created to aid in the standardization of visualizations to reduce variability and increase efficiency. Finally, the data gathered from these articles will be used to inform the creation of a standardized visualization tool that can aid in current SARS-CoV-2 cluster investigations, with the overall goal of allowing for visualization of other infectious pathogen clusters in healthcare settings to reduce nosocomial transmission. By understanding the types of common data visualization elements utilized in transmission visualizations, this gives IPs a generalized understanding of the types of possible data visualizations and data elements that can be included for different infectious pathogens. This pilot study also identifies the need for a standardized data visualization that can utilize electronic medical records to allow for the generation of hypotheses regarding transmission, and therefore, improve public health and patient safety through faster infection prevention and control interventions to interrupt transmission.

Appendix A

Figure 2 PRISMA flowchart of number of articles included in the initial search, articles removed after each screening level, and reasons for exclusion
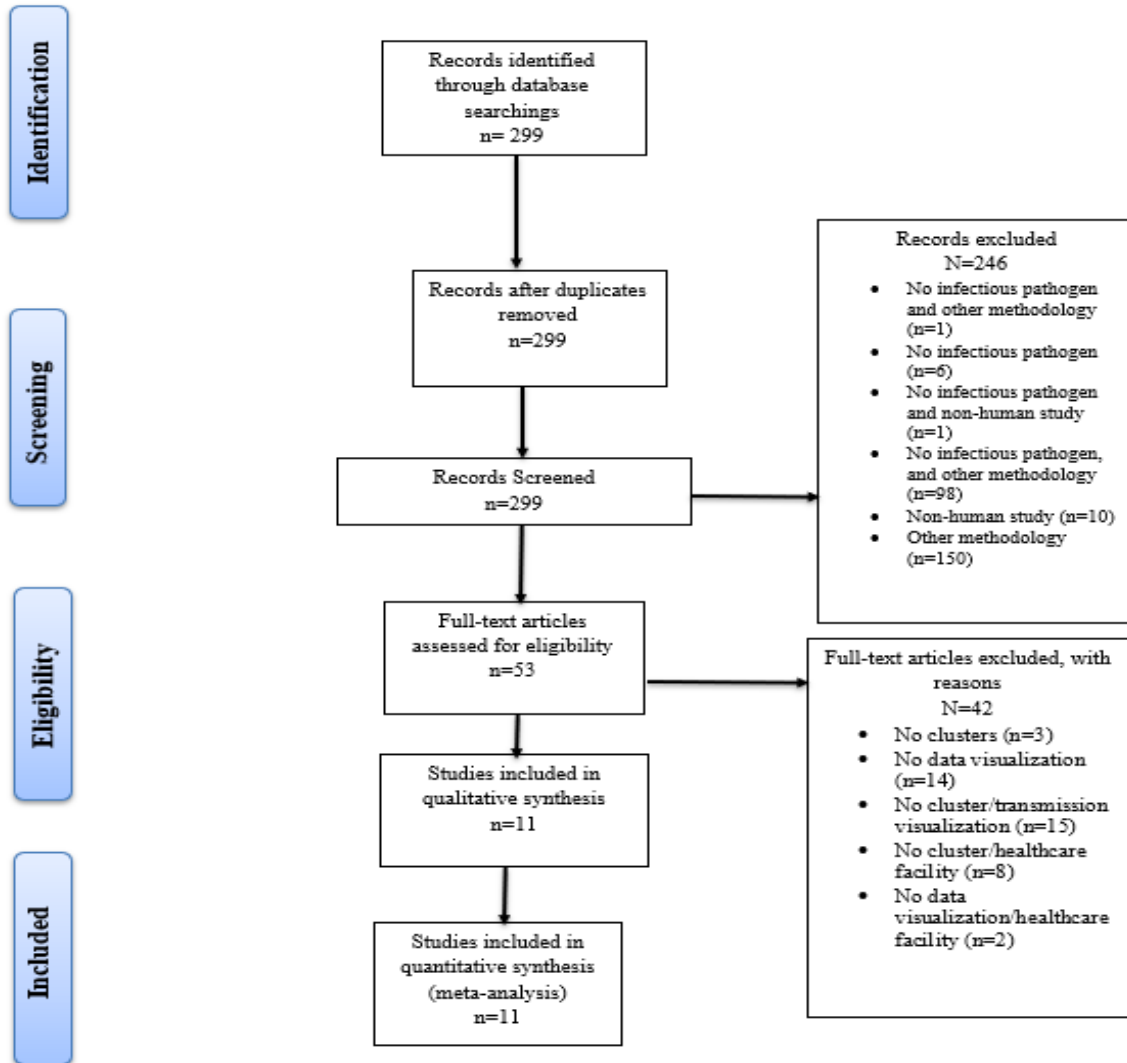
Table 3 Characteristics of studies examining clusters of infectious pathogens with data visualizations in healthcare settings

Study	Country	Study Type	Setting	Scope of Outbreak	Infectious Pathogen	Cluster Size	Cluster Duration
Borges et al., 2021	Portugal	Observational Prospective	Acute Care	Internal Medicine, Hemodialysis Units	*SARS-CoV-2*	48	122 days
De Silva et al., 2021	Brazil	Observational Retrospective	Acute Care	Intensive and non-Intensive Units	*Klebsiella pneumoniae*	24	365 days
Javaid et al., 2020	USA	Observational Prospective	Acute Care	Psychiatric Unit, Oncology Unit, Emergency Department, Medical/Surgical Inpatient Unit, Cardiac Interventional Observation, Inpatient Telemetry Unit, Cardiac Care Unit, ICU, Medical Unit, Cardiac Catheter Lab	*Influenza A*	107	12-days
Klompas et al., 2020	USA	Observational Retrospective	Acute Care	Four Inpatient Units	*SARS-COV-2*	53	120 days
Lee et al., 2021	Korea	Observational Retrospective	Acute Care	Neurosurgery ICU, Neurosurgery ward, Radiology, Bank (in hospital), Intervention room, Echo lab, general internal medicine wards,	*SARS-CoV-2*	85	4 days
Marmor et al., 2019	Australia	Observational Retrospective	Acute Care	Renal Unit, Surgical Unit, ICU, Hematology Unit	*Enterobacteriaceae*	14	1,460 days
McHaney-Lindstrom et al., 2020	USA	Case-Control	Acute Care	Intra-hospital transfers across inpatient units	*Clostridioides difficile*	502	883 days
Moldovan et al., 2019	Canada	Observational Case-Control	Acute Care	Three campuses	*Staphylococcus aureus*	547	335 days
Pai et al., 2020	USA	Observational Retrospective	Acute Care	Across Inpatient Units	*Clostridioides difficile*	790	2,190 days
Pérez-Lago et al., 2021	Spain	Observational Retrospective	Acute Care	Gastroenterology Ward	*SARS-CoV-2*	18	28 days
Wee et al., 2020	Singapore	Prospective Cohort	Acute Care	Campus-wide	*SARS-CoV-2*	61	112 days

Table 4 Person type data elements included within each transmission visualization with the total percentage of person type present across studies

Study	Infectious Pathogen	Person Type							
		Patients	Nurses	APP/ Physicians	Trainees	Patient Care Technicians	Transporters	Other HCW	Total
Borges et al. 2021	*SARS-CoV-2*	21	10	3	12	–	–	3	48
de Silva et al. 2021	*Klebsiella pneumoniae*	24	–	–	–	–	–	–	24
Javaid et al.2020	*Influenza A*	18	16	6	24	8	–	35	107
Klompas et al. 2020	*SARS-CoV-2*	15	12	10	–	5	1	10	53
Lee et al. 2021	*SARS-CoV-2*	–	46	14	–	–	–	25	85
Marmor et al. 2019	*Entero-bacteriaceae*	14	–	–	–	–	–	–	14
McHaney-Lindstrom et al. 2020	*Clostridioides difficile*	502	–	–	–	–	–	–	502
Moldovan et al. 2019	*Staphylococc us aureus*	547	–	–	–	–	–	–	547
Pai et al. 2020	*Clostridioides difficile*	790	–	–	–	–	–	–	790
Pérez-Lago et al. 2021	*SARS-CoV-2*	15	–	–	–	–	–	3	18
Wee et al. 2020	*SARS-CoV-2*	23	17	21	–	–	–	–	61

Total Percentage of Person Type N=2,249	88%	4%	2%	2%	0.6%	0.04%	3%

Note: APP/Physicians = Advanced Practice Practitioners and Physicians (MDs, DOs); HCW= healthcare workers

Table 5 Types of data elements included across 21 transmission visualizations grouped by data visualization type in the 11 eligible articles reviewed

Study	Infectious Pathogen	Data Visualization Type	Person Type	Patients	Healthcare workers	Other	Case Type	Pathway	Spatial Element	Time Element	Symptom Onset Date	Test Positive Date	Contagious Period
												Data Visualizations Included	
Javaid et al. 2020	Influenza A	Timeline 1	—	—	—	—	—	—	✓	✓	—	—	—
Javaid et al. 2020	Influenza A	Timeline 2	✓	✓	✓	—	—	✓	✓	✓	—	✓	—
Lee et al. 2021	SARS-CoV-2	Timeline	✓	✓	✓	—	✓	—	✓	✓	—	✓	—
Timeline Subtotal N (%)			2(67)	2(67)	2(67)	—	1(33)	1(33)	3(100)	3(100)	—	2(67)	—
Borges et al. 2021	SARS-CoV-2	Gantt Chart	✓	✓	✓	—	—	—	✓	✓	—	✓	—
Borges et al. 2021	SARS-CoV-2	Gantt Chart	✓	✓	✓	✓	—	—	✓	—	—	—	—
Gantt Chart Subtotal N (%)			2(100)	2(100)	2(100)	1(50)	—	—	2(100)	1(50)	—	1(50)	—
de Silva et al. 2021	*Klebsiella pneumoniae*	Floor Plan*	—	—	—	—	✓	—	✓	—	—	—	—
Pérez-Lago et al. 2021	SARS-CoV-2	Floor Plan Map	✓	✓	—	—	✓	✓	✓	✓	—	✓	—
Floor Plan Map Subtotal N (%)			1(50)	1(50)	—	—	2(100)	1(50)	2(100)	1(50)	—	1(50)	—
Kloempas et al. 2020	SARS-CoV-2	Cluster Map	✓	✓	✓	—	✓	✓	✓	—	✓	✓	—
Lee et al. 2021	SARS-CoV-2	Cluster Map	—	—	—	—	—	—	✓	—	—	—	—
Wee et al. 2020	SARS-CoV-2	Cluster Map	—	—	—	—	—	—	✓	—	—	—	—
Cluster Map Subtotal N (%)			1(33)	1(33)	1(33)	—	1(33)	1(33)	3(100)	—	1(33)	1(33)	—
Lee et al. 2021	SARS-CoV-2	Transmission Map	✓	✓	✓	✓	—	✓	✓	—	—	—	—
Marmor et al. 2019	Enterobacteriaceae	Transmission Diagram	—	—	—	—	✓	—	✓	✓	—	—	—
Pérez-Lago et al.2021	SARS-CoV-2	Transmission Network	—	—	—	—	—	✓	—	✓	—	—	—
Transmission Map Subtotal N (%)			1(33)	1(33)	1(33)	1(33)	1(33)	2(67)	2(67)	2(67)	—	—	—
Moldovan et al. 2019	*Staphylococcus aureus*	Social Network*	✓	✓	✓	—	—	—	✓	✓	—	—	—
McHaney-Lindstrom et al. 2020	*Clostridioides difficile*	Social network analysis	✓	✓	—	—	—	—	✓	—	—	—	—
Social Network Analysis Subtotal N (%)			2(100)	2(100)	1(50)	—	—	—	2(100)	1(50)	—	—	—
McHaney-Lindstrom et al. 2020	*Clostridioides difficile*	Onset' Risk Rings	—	—	—	—	—	—	✓	—	—	—	—
McHaney-Lindstrom et al. 2020	*Clostridioides difficile*	RPROS Eigenvector and LISA Map*	—	—	—	—	✓	—	✓	—	—	—	—
Pai et al. 2020	*Clostridioides difficile*	Spatial Graph	—	—	—	—	—	✓	—	—	—	—	—
Pai et al. 2020	*Clostridioides difficile*	Case Proximity Graph	—	—	—	—	—	—	✓	✓	—	—	—
Spatial Proximity Map Subtotal N (%)			—	—	—	—	1(25)	1(25)	3(75)	1(25)	—	—	—
de Silva et al. 2021	*Klebsiella pneumoniae*	Heat Map*	✓	✓	✓	—	—	—	✓	—	—	—	—
Heat Map Subtotal N (%)			1(100)	1(100)	1(100)	—	—	—	1(100)	—	—	—	—
Javaid et al. 2020	Influenza A	Other (Interaction Map) *	✓	✓	✓	—	—	—	✓	—	—	—	—
Other Subtotal N (%)			1(100)	1(100)	1(100)	—	—	—	1(100)	—	—	—	—
Totals N=21, N(%)			11(52%)	11(52%)	9(43)	2(10%)	6(29)	6(29)	19(90)	9(43)	1(5)	5(24)	—

Note: *= Data visualizations with specific software stated, see **Table 6**.

Table 6 Types of Software used for specified data visualizations from Tables 4 and 5

Study	Infectious Pathogen	Data Visualization	Software Used
de Silva et al. 2021	*Klebsiella pneumoniae*	Floor Plan	AutoCAD, Qgis
Moldovan et al. 2019	*Staphylococcus aureus*	Social Network	Pajek 4.09
McHaney-Lindstrom et al. 2020	*Clostridioides difficile*	RPROS Eigenvector and LISA Map (Spatial proximity map)	GIS
de Silva et al. 2021	*Klebsiella pneumoniae*	Heat Map	AutoCAD, QGis
Javaid et al. 2021	*Influenza A*	Interaction Map	PathoSPOT, Cytoscape

Bibliography

1. Edmond MB, Wenzel RP. Infection prevention in the health care setting. In: *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Elsevier; 2015:3286-3293.e1. doi:10.1016/B978-1-4557-4801-3.00300-3

2. WHO. Dealing with Outbreaks. In: Prevention of Hospital-Acquired Infections. ; 2002:26-28. Accessed November 24, 2021. https://www.who.int/csr/resources/publications/whocdscsreph200212.pdf

3. Reintjes R, Zanuzdana A. Outbreak Investigations. In: Krämer A, Kretzschmar M, Krickeberg K, eds. *Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health*. Statistics for biology and health. Springer New York; 2010:159-176. doi:10.1007/978-0-387-93835-6_9

4. Nutman A, Marchaim D. How to: molecular investigation of a hospital outbreak. *Clin Microbiol Infect*. 2019;25(6):688-695. doi:10.1016/j.cmi.2018.09.017

5. Carroll LN, Au AP, Detwiler LT, Fu T-C, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *J Biomed Inform*. 2014;51:287-298. doi:10.1016/j.jbi.2014.04.006

6. Home - Johns Hopkins Coronavirus Resource Center. Accessed October 7, 2021. https://coronavirus.jhu.edu/

7. Coronavirus (COVID-19) frequently asked questions | CDC. Accessed October 23, 2021. https://www.cdc.gov/coronavirus/2019-ncov/faq.html

8. Centers for Disease Control. About COVID-19 Epidemiology. Centers for Disease Control and Prevention. July 1, 2021. Accessed November 7, 2021. https://www.cdc.gov/coronavirus/2019-ncov/science/about-epidemiology/index.html

9. Framework for Implementation of COVID-19 Community Mitigation Measures for Lower-Resource Countries | CDC. Accessed November 24, 2021. https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/community-mitigation-measures.html

10. Coronavirus Second Wave, Third Wave and Beyond: What Causes a COVID Surge | Johns Hopkins Medicine. Accessed November 25, 2021. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus

11. Rhee C, Baker M, Vaidya V, et al. Incidence of Nosocomial COVID-19 in Patients Hospitalized at a Large US Academic Medical Center. *JAMA Netw Open*. 2020;3(9):e2020498. doi:10.1001/jamanetworkopen.2020.20498

12. Rickman HM, Rampling T, Shaw K, et al. Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. *Clin Infect Dis*. 2021;72(4):690-693. doi:10.1093/cid/ciaa816

13. Mo Y, Eyre DW, Lumley SF, et al. Transmission of community- and hospital-acquired SARS-CoV-2 in hospital settings in the UK: A cohort study. *PLoS Med*. 2021;18(10):e1003816. doi:10.1371/journal.pmed.1003816

14. Nguyen LH, Drew DA, Graham MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health*. 2020;5(9):e475-e483. doi:10.1016/S2468-2667(20)30164-X

15. Baker MA, Fiumara K, Rhee C, et al. Low Risk of Coronavirus Disease 2019 (COVID-19) Among Patients Exposed to Infected Healthcare Workers. *Clin Infect Dis*. 2021;73(7):e1878-e1880. doi:10.1093/cid/ciaa1269

16. Shitrit P, Zuckerman NS, Mor O, Gottesman B-S, Chowers M. Nosocomial outbreak caused by the SARS-CoV-2 Delta variant in a highly vaccinated population, Israel, July 2021. *Euro Surveill*. 2021;26(39). doi:10.2807/1560-7917.ES.2021.26.39.2100822

17. Scobie HM, Johnson AG, Suthar AB, et al. Monitoring Incidence of COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Status - 13 U.S. Jurisdictions, April 4-July 17, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(37):1284-1290. doi:10.15585/mmwr.mm7037e1

18. Latest Data on COVID-19 Vaccinations by Race/Ethnicity | KFF. Accessed November 24, 2021. https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/

19. Bender JK, Brandl M, Höhle M, Buchholz U, Zeitlmann N. Analysis of Asymptomatic and Presymptomatic Transmission in SARS-CoV-2 Outbreak, Germany, 2020. *Emerging Infect Dis*. 2021;27(4). doi:10.3201/eid2704.204576

20. Hellewell J, Russell TW, SAFER Investigators and Field Study Team, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Med*. 2021;19(1):106. doi:10.1186/s12916-021-01982-x

21. Clinical Questions about COVID-19: Questions and Answers | CDC. Accessed November 1, 2021. https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html

22. Sundermann AJ, Chen J, Kumar P, et al. Whole genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. *Clin Infect Dis*. Published online November 17, 2021. doi:10.1093/cid/ciab946

23. What is a Scoping Review? - Systematic Reviews & Other Review Types - Research Guides at Temple University. Accessed November 23, 2021. https://guides.temple.edu/c.php?g=78618&p=4156607

24. Salinas JL, Kritzman J, Kobayashi T, Edmond MB, Ince D, Diekema DJ. A primer on data visualization in infection prevention and antimicrobial stewardship. *Infect Control Hosp Epidemiol*. 2020;41(8):948-957. doi:10.1017/ice.2020.142

25. Davis GS, Sevdalis N, Drumright LN. Spatial and temporal analyses to investigate infectious disease transmission within healthcare settings. *J Hosp Infect*. 2014;86(4):227-243. doi:10.1016/j.jhin.2014.01.010

26. Smith CM, Hayward AC. DotMapper: an open source tool for creating interactive disease point maps. *BMC Infect Dis*. 2016;16:145. doi:10.1186/s12879-016-1475-5

27. Smith CM, Allen DJ, Nawaz S, et al. An interactive data visualisation application to investigate nosocomial transmission of infections. [version 2; peer review: 2 approved]. *Wellcome Open Res*. 2019;4:100. doi:10.12688/wellcomeopenres.15240.2

28. Ohannessian R, Bénet T, Argaud L, et al. Heat map for data visualization in infection control epidemiology: An application describing the relationship between hospital-acquired infections, Simplified Acute Physiological Score II, and length of stay in adult intensive care units. *Am J Infect Control*. 2017;45(7):746-749. doi:10.1016/j.ajic.2017.02.012

29. Visualization Types - Data Visualization - LibGuides at Duke University. Accessed November 8, 2021. https://guides.library.duke.edu/datavis/vis_types

30. Moldovan ID, Suh K, Liu EY, Jolly A. Network analysis of cases with methicillin-resistant Staphylococcus aureus and controls in a large tertiary care facility. *Am J Infect Control*. 2019;47(12):1420-1425. doi:10.1016/j.ajic.2019.05.026

31. Schifman RB, Palmer RA. Surveillance of nosocomial infections by computer analysis of positive culture rates. *J Clin Microbiol*. 1985;21(4):493-495. doi:10.1128/jcm.21.4.493-495.1985

32. Grota PG, Stone PW, Jordan S, Pogorzelska M, Larson E. Electronic surveillance systems in infection prevention: organizational support, program characteristics, and user satisfaction. *Am J Infect Control*. 2010;38(7):509-514. doi:10.1016/j.ajic.2009.10.007

33. Centers for Disease Control. NHSN Organism List. Centers for. 2019. Accessed October 13, 2021. https://www.cdc.gov/nhsn/pdfs/validation/2019/2019-NHSN-Organisms-List-Validation.xlsx

34. Borges V, Isidro J, Macedo F, et al. Nosocomial Outbreak of SARS-CoV-2 in a "Non-COVID-19" Hospital Ward: Virus Genome Sequencing as a Key Tool to Understand Cryptic Transmission. *Viruses*. 2021;13(4). doi:10.3390/v13040604

35. da Silva PP, da Silva FA, Rodrigues CAS, et al. Geographical information system and spatial-temporal statistics for monitoring infectious agents in hospital: a model using Klebsiella pneumoniae complex. *Antimicrob Resist Infect Control*. 2021;10(1):92. doi:10.1186/s13756-021-00944-5

36. Javaid W, Ehni J, Gonzalez-Reiche AS, et al. Real-Time Investigation of a Large Nosocomial Influenza A Outbreak Informed by Genomic Epidemiology. *Clin Infect Dis*. Published online November 30, 2020. doi:10.1093/cid/ciaa1781

37. Klompas M, Baker MA, Rhee C, et al. A SARS-CoV-2 Cluster in an Acute Care Hospital. *Ann Intern Med*. 2021;174(6):794-802. doi:10.7326/M20-7567

38. Lee U, Kim SE, Lee SY, et al. Source Analysis and Effective Control of a COVID-19 Outbreak in a University Teaching Hospital during a Period of Increasing Community Prevalence of COVID-19. *J Korean Med Sci*. 2021;36(24):e179. doi:10.3346/jkms.2021.36.e179

39. Marmor A, Daveson K, Harley D, Coatsworth N, Kennedy K. Two carbapenemase-producing Enterobacteriaceae outbreaks detected retrospectively by whole-genome sequencing at an Australian tertiary hospital. *Infection, Disease & Health*. 2020;25(1):30-33. doi:10.1016/j.idh.2019.08.005

40. McHaney-Lindstrom M, Hebert C, Miller H, Moffatt-Bruce S, Root E. Network analysis of intra-hospital transfers and hospital onset clostridium difficile infection. *Health Info Libr J*. 2020;37(1):26-34. doi:10.1111/hir.12274

41. Pai S, Polgreen PM, Segre AM, Sewell DK, Pemmaraju SV, CDC MInD-Healthcare group. Spatiotemporal clustering of in-hospital Clostridioides difficile infection. *Infect Control Hosp Epidemiol*. 2020;41(4):418-424. doi:10.1017/ice.2019.350

42. Pérez-Lago L, Martínez-Lozano H, Pajares-Díaz JA, et al. Overlapping of Independent SARS-CoV-2 Nosocomial Transmissions in a Complex Outbreak. *mSphere*. 2021;6(4):e0038921. doi:10.1128/mSphere.00389-21

43. Wee LE, Venkatachalam I, Sim XYJ, et al. Containment of COVID-19 and reduction in healthcare-associated respiratory viral infections through a multi-tiered infection control strategy. *Infection, Disease & Health*. 2021;26(2):123-131. doi:10.1016/j.idh.2020.11.004