SHEA

Review

Whole-genome sequencing surveillance and machine learning for healthcare outbreak detection and investigation: A systematic review and summary

Alexander J. Sundermann DrPH[1,2,3], Jieshi Chen MS[4] (iD), James K. Miller PhD[4] (iD), Elise M. Martin MD[2,5] (iD),
Graham M. Snyder MD[2,5] (iD), Daria Van Tyne PhD[2] (iD), Jane W. Marsh PhD[1,2], Artur Dubrawski PhD[4] (iD) and
Lee H. Harrison MD[1,2,3]

[1]Microbial Genomic Epidemiology Laboratory, Center for Genomic Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, [2]Division of Infectious
Diseases, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, [3]Department of Epidemiology, Graduate School of Public Health, University of
Pittsburgh, Pittsburgh, Pennsylvania, [4]Auton Lab, Carnegie Mellon University, Pittsburgh, Pennsylvania and [5]Department of Infection Prevention and Hospital
Epidemiology, UPMC Presbyterian, Pittsburgh, Pennsylvania

Abstract

Background: Whole-genome sequencing (WGS) has traditionally been used in infection prevention to confirm or refute the presence of an outbreak after it has occurred. Due to decreasing costs of WGS, an increasing number of institutions have been utilizing WGS-based surveillance. Additionally, machine learning or statistical modeling to supplement infection prevention practice have also been used. We systematically reviewed the use of WGS surveillance and machine learning to detect and investigate outbreaks in healthcare settings.

Methods: We performed a PubMed search using separate terms for WGS surveillance and/or machine-learning technologies for infection prevention through March 15, 2021.

Results: Of 767 studies returned using the WGS search terms, 42 articles were included for review. Only 2 studies (4.8%) were performed in real time, and 39 (92.9%) studied only 1 pathogen. Nearly all studies (n = 41, 97.6%) found genetic relatedness between some isolates collected. Across all studies, 525 outbreaks were detected among 2,837 related isolates (average, 5.4 isolates per outbreak). Also, 35 studies (83.3%) only utilized geotemporal clustering to identify outbreak transmission routes. Of 21 studies identified using the machine-learning search terms, 4 were included for review. In each study, machine learning aided outbreak investigations by complementing methods to gather epidemiologic data and automating identification of transmission pathways.

Conclusions: WGS surveillance is an emerging method that can enhance outbreak detection. Machine learning has the potential to identify novel routes of pathogen transmission. Broader incorporation of WGS surveillance into infection prevention practice has the potential to transform the detection and control of healthcare outbreaks.

(Received 28 October 2021; accepted 4 November 2021)

Whole-genome sequencing (WGS) for infection prevention has traditionally been used in reaction to a suspected outbreak, usually at the end of an investigation to confirm or to refute the presence of an outbreak. In contrast, WGS surveillance of selected healthcare-associated pathogens regardless of whether an outbreak is suspected can be used to identify outbreaks that are not detected by traditional hospital epidemiologic methods. High costs and needed infrastructure for implementation have been historic barriers to widespread use of WGS surveillance. However, the cost of WGS has fallen, and the expansion of genomic surveillance due to coronavirus disease 2019 (COVID-19) may enable healthcare institutions to establish WGS surveillance programs for other pathogens. Additionally, our work and studies from Australia have identified cost benefits to implementing a WGS surveillance program with effective intervention.[1]

Although WGS surveillance is effective in identifying transmission, it does not provide information on the responsible transmission route, which is crucial for interrupting an outbreak. Traditional epidemiologic methods for identifying where transmission occurs have relied on geotemporal clustering within the hospital, which is inadequate for identifying more complex patterns of transmission.[2,3] Automated analysis of electronic health records (EHRs) creates an opportunity to use machine learning

Genomics Search Terms

(whole genome sequenc*) AND (surveillance OR routine) AND (healthcare OR hospital) AND transmission
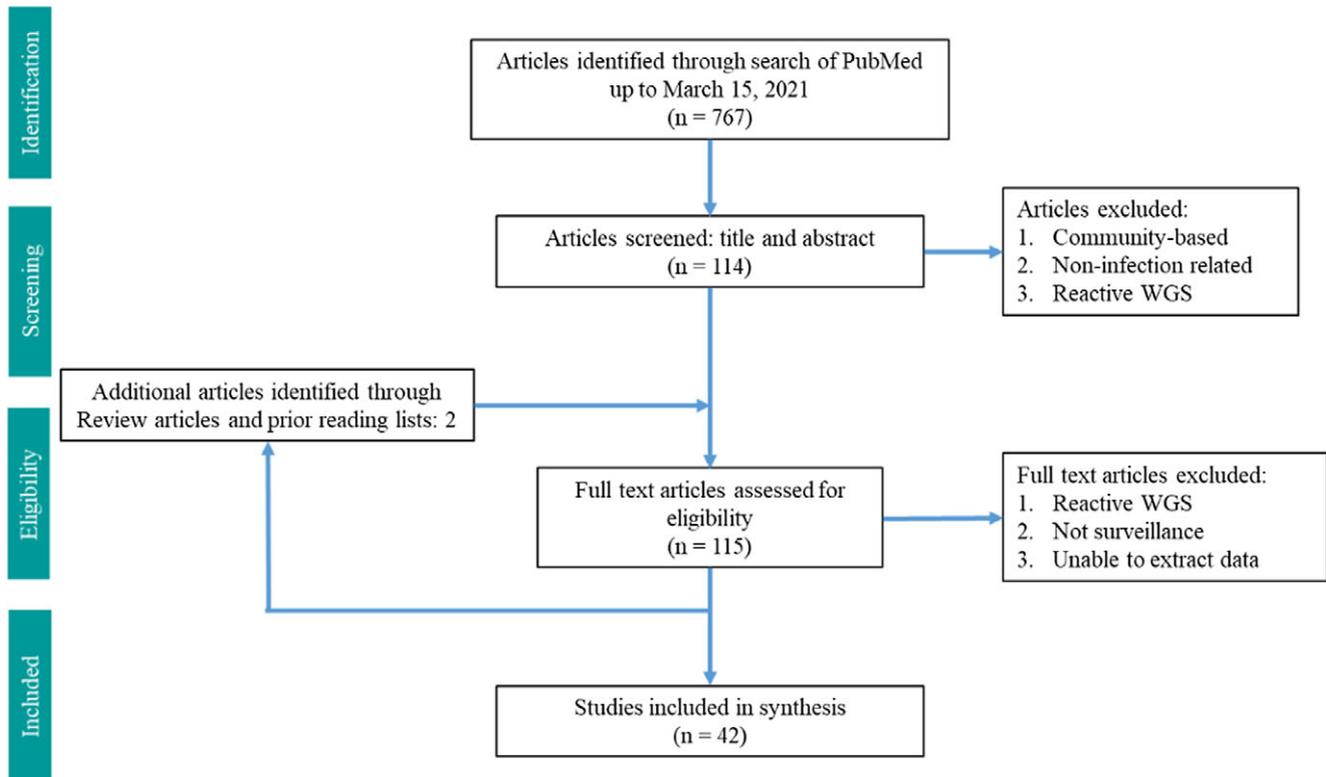

Fig. 1. Search terms in PubMed for whole-genome sequencing surveillance.

or statistical modeling approaches for determining the outbreak transmission routes identified by WGS surveillance.[4–8] These automated approaches would assist hospital infection prevention departments by providing systematic methods to investigate outbreaks and identify transmission routes.

In this systematic review, we provide a summary of prior studies utilizing WGS surveillance in healthcare settings for outbreak detection, as well as the use of machine learning and statistical modeling technologies to identify transmission routes. The purposes of this review were to summarize the current literature in this field, to identify barriers to widespread implementation, and to synthesize current knowledge on this topic to help guide decision making about implementation of WGS surveillance.

Methods

Two search terms were utilized in PubMed from inception until March 15, 2021 (Figs. 1 and 2). The WGS surveillance terms "(whole genome sequenc*) AND (surveillance OR routine) AND (healthcare OR hospital) AND transmission" returned 767 results. Article abstracts were screened to exclude studies that were solely community based, non–infection related, utilized non-WGS methods (eg, older molecular subtyping methods such as pulsed-field gel electrophoresis), or only utilized reactive WGS in response to suspected outbreaks. Genomic and epidemiologic data on organisms, number of isolates sequenced, percentage of isolates that were related, number of outbreaks, and epidemiological links

were extracted and summarized. Articles were excluded if the data were not sufficiently detailed for extraction.

The machine-learning search terms utilized were "("electronic health record" OR "electronic medical record" OR "artificial intelligence" OR "AI" OR "ML" OR "model") AND (outbreak OR transmission) AND ("data mining" OR "machine learning") AND (infection OR infectious) AND ("healthcare-associated" OR "hospital-associated" OR "healthcare-acquired" OR "hospital-acquired")" and returned 21 results. Article abstracts were screened to exclude infection prediction and outcome studies. Data on the methodology and findings of outbreak and transmission detection models were extracted and summarized.

Results

In total, 42 articles on WGS surveillance were included in the final review.[3–5,9–47] Among these studies, only 2 employed machine learning or statistical modeling to investigate transmission, which were also captured in the machine learning search. From 2013 to 2016, there was only one article per year, with a substantial increase thereafter (Fig. 3). Moreover, 12 studies were from the United States; 10 were from the United Kingdom; 5 were from Australia; 4 were from Germany; 2 were from Japan; and 1 study came per country came from China, Denmark, Finland, France, India, Italy, Netherlands, Spain, Sweden, and Thailand.

The duration of WGS surveillance varied substantially by study, with a median of 12 months and a range of 1–73 months

EHR/ML/Modeling Search Terms

("electronic health record" OR "electronic medical record" OR "artificial intelligence" OR "AI" OR "ML" OR "model") AND (outbreak OR transmission) AND ("data mining" OR "machine learning") AND (infection OR infectious) AND ("healthcare-associated" OR "hospital-associated" OR "healthcare-acquired" OR "hospital-acquired")
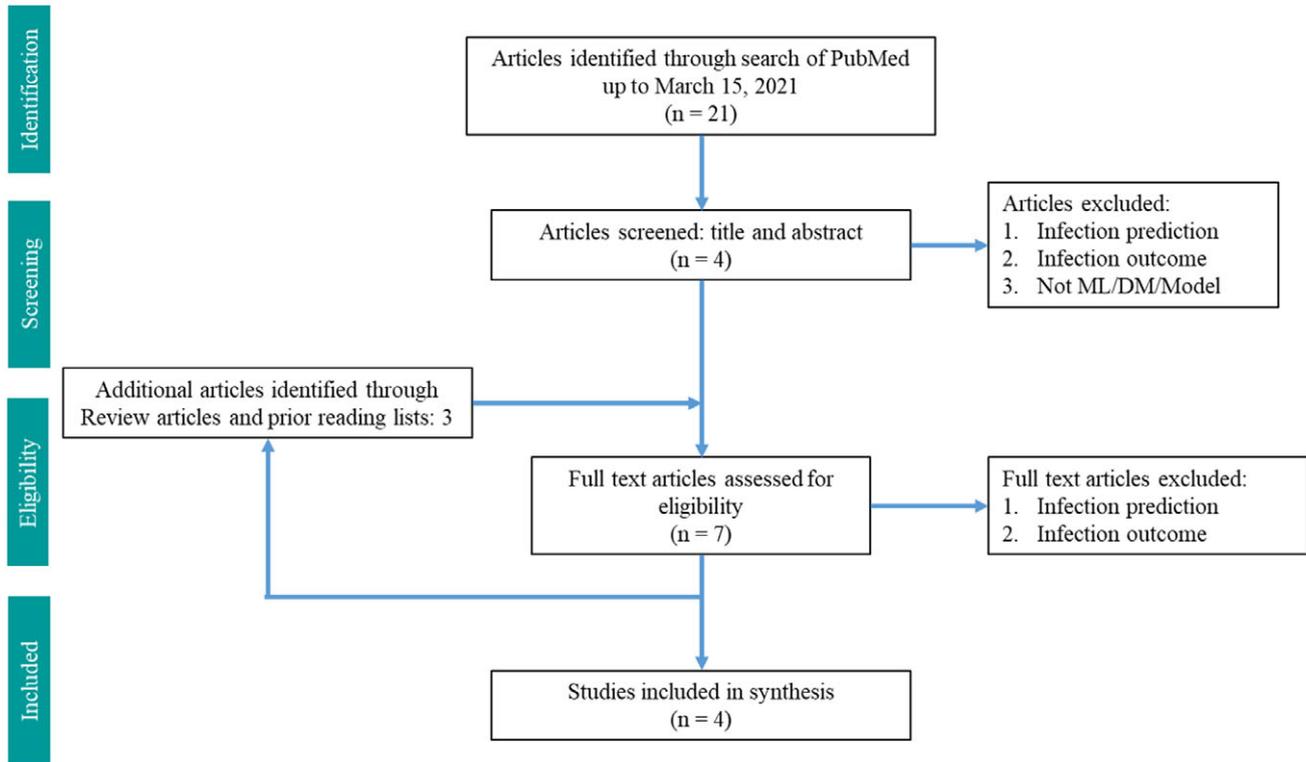
Fig. 2. Search terms in PubMed for machine learning and modeling.
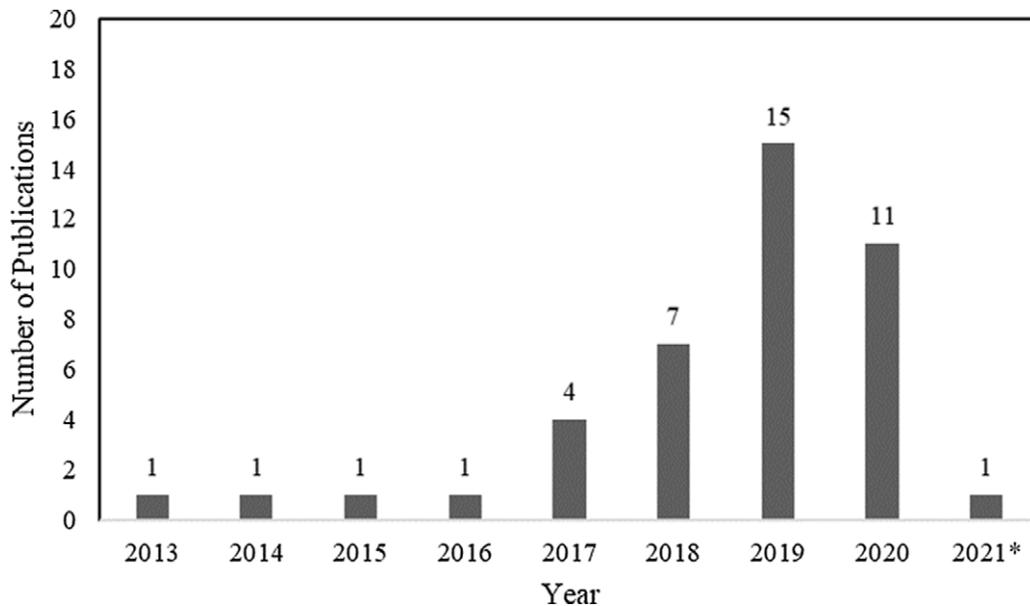
Fig. 3. Summary by year of 42 whole-genome sequencing (WGS) surveillance studies in PubMed through March 15, 2021. *Through March 15, 2021.

(Supplementary Table S1). Only 2 (4.8%) studies were performed in real time; all other studies were performed retrospectively. Moreover, 39 studies included a single pathogen and 3 studies included multiple pathogens (Table 1). *Staphylococcus aureus* was the most commonly studied organism (12 studies, 28.6%),

with 4 additional organisms present in >2 studies: 2 *Klebsiella pneumoniae*, 7 *Clostridioides difficile*, 6 *Enterococcus faecium*, and 3 *Pseudomonas aeruginosa*. Organisms selected for sequencing (eg, by anatomic site of infection, multilocus sequence type, and antibiotic resistance phenotype) were diverse across studies.

Table 1. Studies by Date, Organism, and Outbreaks Detected Utilizing WGS Surveillance

Year	First Author	Organism(s)	Type	Unique Isolates	Related, No. (%)	Outbreaks Detected	Epi Link, No. (%)
2021	Meredith[34]	SARS-CoV-2	...	299	159 (53.2)	35	35 (22)
2021	Miles-Jay[35]	*Escherichia coli*	ST131, H30	126	17 (13.5)	8	9 (52.9)
2021	Rose[39]	*Staphylococcus aureus*	Methicillin-resistant	56	15 (26.8)	7	7 (46.7)
2020	Berbel Caban[10]	*S. aureus*	Methicillin-resistant	224	33 (14.7)	8	...
2020	Cremers[12]	*S. aureus*	Methicillin-sensitive	84	40 (47.6)	14	0 (0)
2020	Gona[20]	*Klebsiella pneumoniae*	...	80	39 (48.8)	10	14 (35.9)
2020	Hammerum[23]	*K. pneumoniae*	...	103	36 (35)	13	11 (30.6)
2020	Marmor[31]	*Enterobacter cloacae*	...	63	7 (11.1)	1	0 (0)
		Citrobacter freundii	...		10 (15.9)	1	0 (0)
2020	Neumann[36]	*E. faecium*	Vancomycin-resistant	111	...	...	...
2020	Sundermann[5]	*P. aeruginosa*	ST27	882	31 (3.5)	10	1 (3.2)
2020	Sundermann[4]	*Enterococcus faecium*	Vancomycin-resistant, ST1471	439	10 (2.3)	...	1 (10)
2020	Tsujiwaki[44]	*S. aureus*	Methicillin-resistant	57	19 (33.3)	5	0 (0)
2019	Eigenbrod[14]	*Acinetobacter baumannii*	...	39	15 (38.5)	4	5 (33.3)
2019	Eyre[18]	*Clostridioides difficile*	...	299	43 (14.4)	6	20 (46.5)
2019	García-Fernández[19]	*C. difficile*	...	367	41 (11.2)	6	34 (82.9)
2019	Hall[22]	*S. aureus*	Methicillin-resistant	55	27 (49.1)	12	8 (29.6)
2019	Harada[24]	*K. pneumoniae*	Bloodstream infections	140	2 (1.4)	1	2 (100)
2019	Jakharia[26]	*C. difficile*	.	45	4 (8.9)	2	4 (100)
2019	Kossow[27]	*S. aureus*	Methicillin-resistant	.	8	1	0 (0)
2019	Mathur[33]	*K. pneumoniae*	Colistin-resistant	21	8 (38.1)	4	0 (0)
2019	Roy[40]	Influenza	A H1N1	36	5 (13.9)	2	2 (40)
2019	Sherry[41]	Enterobacteriaceae	Carbapenemase-producing	291	53 (18.2)	12	8 (15.1)
2019	Stenmark[42]	*S. capitis*	Bloodstream infections	46	12 (26.1)	6	12 (100)
2019	Sullivan[43]	*S. aureus*	Methicillin-resistant	141	28 (19.9)	4	2 (7.1)
2019	van Beek[45]	*K. pneumoniae*	Carbapenemase-producing, ST512	...	20	2	4 (20)
2019	Wang[46]	*Corynebacterium striatum*	...	91	18 (19.8)	6	3 (16.7)
2019	Ward[3]	*S. aureus*	...	953	85 (8.9)	28	65 (76.5)
		E. faecium	...	86	13 (15.1)	5	9 (69.2)
		Psuedomonas aeruginosa	...	118	2 (1.7)	1	2 (100)
		K. pneumoniae	...	100	0 (0)	0	0 (0)
2018	Auguet[9]	*S. aureus*	Methicillin-resistant	610	261 (42.8)	90	13 (5)
2018	Donskey[13]	*C. difficile*	...	66	12 (18.2)	4	4 (33.3)
2018	Houldcroft[25]	Adenovirus	...	43	6 (14)	2	0 (0)
2018	Kwong[28]	*K. pneumoniae*	Carbapenemase-producing	86	53 (61.6)	4	10 (18.9)
2018	Leong[29]	*E. faecium*	Vancomycin-resistant	80	10 (12.5)	2	3 (30)
2018	Martin[32]	*C. difficile*	...	640	227 (35.5)	.	69 (30.4)
2018	Wendel[47]	*A. baumannii*	...	36	20 (55.6)	2	2 (10)
2017	Coll[11]	*S. aureus*	Methicillin-resistant	1465	785 (53.6)	173	187 (23.8)
2017	Eyre[17]	*C. difficile*	...	652	128 (19.6)	...	...
2017	Gorrie[21]	*K. pneumoniae*	...	106	17 (16)	5	0 (0)
2017	Raven[37]	*E. faecium*	Bloodstream infections	293	93 (31.7)	6	.
2016	Elbadawi[15]	*K. pneumoniae*	Carbapenemase-producing	46	4 (8.7)	1	0 (0)

(Continued)

Table 1. (Continued)

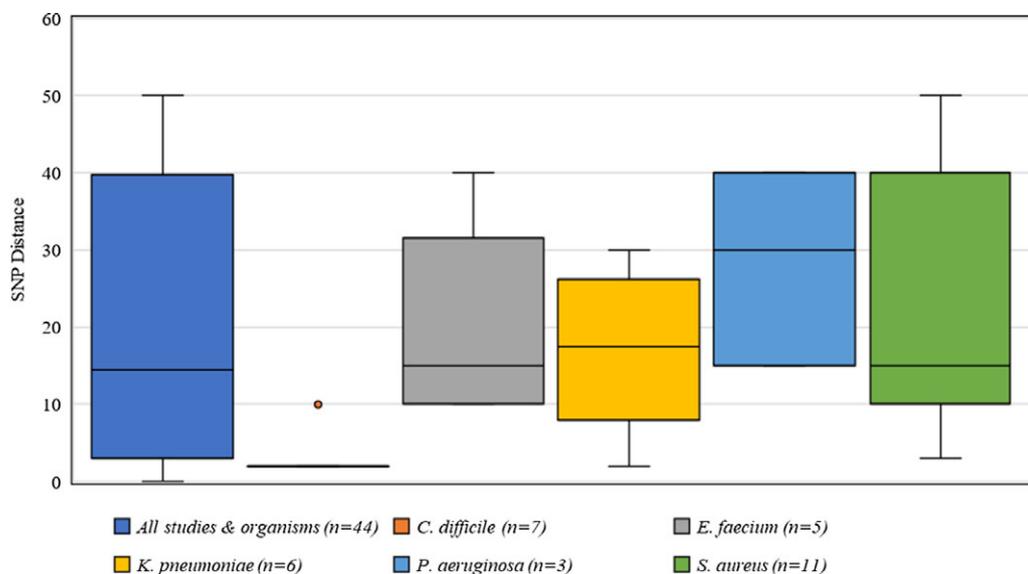| Year	First Author	Organism(s)	Type	Unique Isolates	Related, No. (%)	Outbreaks Detected	Epi Link, No. (%)
2015	Roach[38]	Staphylococcus epidermidis	. . .	178	56 (31.5)	10	. . .
		P. aeruginosa	. . .	44	7 (15.9)	3	. . .
		E. faecium	. . .	36	13 (36.1)	3	. . .
		S. aureus	. . .	118	4 (3.4)	2	. . .
		E. faecalis	. . .	72	6 (8.3)	3	. . .
		Stenotrophomonas maltophilia	. . .	58	2 (3.4)	1	. . .
2014	Long[30]	S. aureus	. . .	305	0 (0)	0	. . .
2013	Eyre[16]	C. difficile	. . .	957	333 (34.8)	.	152 (45.6)

Fig. 4. Distribution of single nucleotide polymorphisms (SNPs) for defining genetic relatedness from 42 studies.

Criteria for defining genetic relatedness were also highly variable between studies and were generally based on the number of single-nucleotide polymorphism (SNP) differences between genomes (Supplementary Table S1 online). Among organisms present in >2 studies, C. difficile had the most consistent SNP cutoff at 2 SNPs, and 1 study used 10 SNPs to identify related isolates (Fig. 4). Furthermore, S. aureus had the widest distribution of SNP cutoffs, ranging from 7 to 50 SNPs.

An analysis of the proportion of sequenced isolates that were determined to be genetically related to one another in each study revealed an average of 23.8% of isolates (range, 0%–61%). There were 525 outbreaks detected among 2,837 related isolates (average 5.4 isolates per outbreak). In addition, 41 studies (97.6%) found some level of genetic relatedness between the sequenced isolates.

We examined the methods employed to identify the responsible transmission routes for outbreaks that were detected by WGS. Overall, 35 studies (83.3%) restricted attempts to identify transmission routes to the same hospital unit during a defined period.[9,11–27,29–38,40–42,44–47] Only 7 studies (16.7%) examined other possible routes such as medical procedures or healthcare workers.[3–5,10,28,39,43]

Several studies were notable for uncovering otherwise unidentified transmissions, which is the main goal of WGS surveillance.

Sullivan et al[43] were prompted by an outbreak in a neonatal intensive care unit (NICU) to retrospectively investigate all MRSA bloodstream infections for 16 months. Their investigation uncovered isolates related to the NICU outbreak from adult patients in a separate tower. Further investigation revealed shared ventilators between the adult unit and the NICU, which was believed to have caused transmission. Separately, Roy et al[40] performed sequencing of influenza A H1N1for 6 months and found that traditional infection prevention practice falsely identified outbreaks, and WGS surveillance data were able to connect cases that were previously not believed to be epidemiologically related. Lastly, Berbel Caban et al[10] utilized WGS surveillance of MRSA over 2 years and found multiple undetected outbreaks within 2 New York City hospitals. One cluster of 24 isolates from 16 patients spanned 21 months and 9 different hospital wards with patterns of shared healthcare workers. In this study, the authors emphasized the limitations of investigating only geotemporal clustering in outbreak detection and investigation.

Furthermore, 4 articles identified using the machine-learning search terms included in the final synthesis, 2 of which overlapped in the WGS surveillance search terms.[5,28,48,49] Table 2 lists the methods and limitations of each study. One study utilized imputation of cultures to model transmission dynamics from

Table 2. Studies Utilizing Machine Learning or Modeling to Detect outbreaks Or Transmission

Year	First Author	Machine Learning or Model Method	Utility and Findings	Limitations
2018	Lensing[48]	Imputation of clinical and environmental cultures to model transmission dynamics	Aided in targeted environmental cleaning to decolonize plumbing systems and reduce the risk of transmission of carbapenem-resistant Enterobacteriaceae based upon learned positivity	Requires expert knowledge of plumbing system and prior data on colonization
2018	Kwong[28]	Bayesian transmission modeling using Markov chain Monte Carlo	Assisted in transmission modeling of KPC-producing *K. pneumoniae* to determine if spread resulted from interfacility or intrafacility transmission	Does not provide specific details in sequence of transmission within a complex outbreak
2020	Sundermann[5]	Bayesian inference with case-control methodology to describe transmission sources	Scanned electronic health record and provided statistical output for possible transmission routes beyond but including geotemporal clustering	Requires robust mapping of electronic health record charge codes
2017	Stachel[49]	WHONET-SaTScan	Utilizes space-time permutation scan statistics to identify potential outbreaks	Low positive predictive value creating a high number of "false alarms"

environmental sink contamination, 2 studies used Bayesian methods to model transmission, and 1 study combined WHONET and SaTScan tools to detect outbreaks. In all of these studies, tools were implemented to supplement outbreak detection or investigation, yet the importance of manual or expert input to further investigate the transmissions or outbreaks detected was noted in each of these studies. In the study by Satchel et al,[49] 45 outbreaks were identified, but only 6 were confirmed by an IP investigation. However, these researchers stated that the tool helped to streamline investigation efforts, which reduced time spent by the IP team.

Discussion

In this systematic review, we synthesized studies that demonstrate the utility of WGS surveillance in finding cryptic outbreaks in healthcare settings. Nearly all studies (97.6%) identified outbreaks, but few (4.8%) utilized machine learning or statistical modeling methods to investigate transmission routes. WGS surveillance, while uncommon but increasingly utilized, aided infection prevention practice in these studies by uncovering outbreaks and enabling intervention.

Studies utilizing WGS surveillance have primarily relied on geotemporal linkage to identify transmission routes. Restricting investigations to geotemporal linkage fails to identify potential transmission by procedures that are performed in areas of the hospital other than patient nursing units or healthcare workers, as shown in some of the studies in this review. Some studies stated the limitations of relying solely on geotemporal parameters for identifying the transmission route for related isolates. Regardless, WGS surveillance enabled many of these studies to uncover substantial and significant previously undetected outbreaks that likely affected patient outcomes and associated healthcare costs.

Almost all of the studies reviewed were retrospective in nature, which limits the potential impact of WGS surveillance on healthcare epidemiology and infection prevention. If performed in real time, IP teams have an opportunity to perform an investigation (eg, audit practices, collect environmental cultures, and interview staff), which is not possible in retrospective studies. Furthermore, many studies focused on 1 pathogen, which is less sensitive for detecting outbreaks than WGS surveillance of multiple pathogens. For example, a single transmission route can lead to the spread of multiple pathogens.

Substantial investment and infrastructure are needed to establish real-time WGS surveillance. Healthcare institutions must have appropriate laboratory capacity, bioinformaticians, and genomic epidemiologists to interpret the data. A recent paper by Parcell et al[50] discussed barriers to instituting a WGS surveillance program for outbreak detection from an economic and systemwide perspective. Indeed, it is often difficult to prove estimates of cost-effectiveness when considering prevention interventions, but 2 studies have demonstrated the cost-effectiveness of WGS surveillance programs.[1,7]

We identified very few studies on the utility of machine learning or statistical modeling methods for identification of outbreak transmission routes by WGS surveillance. In our experience, machine learning adds value in detecting transmission routes that do not involve geotemporal clustering such as invasive procedures, healthcare workers, outbreaks separated by unit, and outbreaks of longer duration.[5,6,8] The use of machine learning in combination with WGS surveillance is clearly an understudied area of healthcare epidemiology and infection prevention. Barriers such as interoperability of electronic health records and adoption of WGS surveillance prevent the implementation of such programs. However, adoption of public health WGS surveillance for COVID-19 may expedite the use of this technology by healthcare institutions.

The combination of prospective WGS surveillance, EHR data, and machine learning has the potential to dramatically transform the paradigm of outbreak detection and investigation for infection prevention and control by identifying outbreaks quicker and enabling early intervention to halt transmission. This approach will both improve patient safety and reduce healthcare costs. However, healthcare institutional investment into establishing WGS surveillance programs will be key to expansion and implementation of this approach.

References

1. Gordon LG, Elliott TM, Forde B, *et al*. Budget impact analysis of routinely using whole-genomic sequencing of six multidrug-resistant bacterial pathogens in Queensland, Australia. *BMJ Open* 2021;11:e041968.

2. Sherry NL, Lee RS, Gorrie CL, *et al.* Pilot study of a combined genomic and epidemiologic surveillance program for hospital-acquired multidrug-resistant pathogens across multiple hospital networks in Australia. *Infect Control Hosp Epidemiol* 2021;42:573–581.

3. Ward DV, Hoss AG, Kolde R, *et al.* Integration of genomic and clinical data augments surveillance of healthcare-acquired infections. *Infect Control Hosp Epidemiol* 2019;40:649–655.

4. Sundermann AJ, Babiker A, Marsh JW, *et al.* Outbreak of vancomycin-resistant *Enterococcus faecium* in interventional radiology: detection through whole-genome sequencing-based surveillance. *Clin Infect Dis* 2020;70:2336–2343.

5. Sundermann AJ, Chen J, Miller JK, *et al.* Outbreak of *Pseudomonas aeruginosa* infections from a contaminated gastroscope detected by whole-genome sequencing surveillance. *Clin Infect Dis* 2021;73:e638–e642.

6. Sundermann AJ, Miller JK, Marsh JW, *et al.* Automated data mining of the electronic health record for investigation of healthcare-associated outbreaks. *Infect Control Hosp Epidemiol* 2019;40:314–319.

7. Kumar P, Sundermann AJ, Martin EM, *et al.* Method for economic evaluation of bacterial whole genome sequencing surveillance compared to standard of care in detecting hospital outbreaks. *Clin Infect Dis* 2021;73:e9–e18.

8. Miller JK, Chen J, Sundermann A, *et al.* Statistical outbreak detection by joining medical records and pathogen similarity. *J Biomed Inform* 2019;91: 103126.

9. Tosas Auguet O, Stabler RA, Betley J, *et al.* Frequent undetected ward-based methicillin-resistant *Staphylococcus aureus* transmission linked to patient sharing between hospitals. *Clin Infect Dis* 2018;66:840–848.

10. Berbel Caban A, Pak TR, Obla A, *et al.* PathoSPOT genomic epidemiology reveals under-the-radar nosocomial outbreaks. *Genome Med* 2020;12:96.

11. Coll F, Harrison EM, Toleman MS, *et al.* Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci Transl Med* 2017;9(413):eaak9745.

12. Cremers AJH, Coolen JPM, Bleeker-Rovers CP, *et al.* Surveillance-embedded genomic outbreak resolution of methicillin-susceptible *Staphylococcus aureus* in a neonatal intensive care unit. *Sci Rep* 2020;10:2619.

13. Donskey CJ, Sunkesula VCK, Stone ND, *et al.* Transmission of *Clostridium difficile* from asymptomatically colonized or infected long-term care facility residents. *Infect Control Hosp Epidemiol* 2018;39:909–916.

14. Eigenbrod T, Reuter S, Gross A, *et al.* Molecular characterization of carbapenem-resistant *Acinetobacter baumannii* using WGS revealed missed transmission events in Germany from 2012-15. *J Antimicrob Chemother* 2019;74:3473–3480.

15. Elbadawi LI, Borlaug G, Gundlach KM, *et al.* Carbapenem-resistant *Enterobacteriaceae* transmission in healthcare facilities—Wisconsin, February–May 2015. *Morb Mortal Wkly Rep* 2016;65:906–909.

16. Eyre DW, Cule ML, Wilson DJ, *et al.* Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013;369:1195–1205.

17. Eyre DW, Fawley WN, Rajgopal A, *et al.* Comparison of control of *Clostridium difficile* Infection in six english hospitals using whole-genome sequencing. *Clin Infect Dis* 2017;65:433–441.

18. Eyre DW, Shaw R, Adams H, *et al.* WGS to determine the extent of *Clostridioides difficile* transmission in a high incidence setting in North Wales in 2015. *J Antimicrob Chemother* 2019;74:1092–1100.

19. García-Fernández S, Frentrup M, Steglich M, *et al.* Whole-genome sequencing reveals nosocomial *Clostridioides difficile* transmission and a previously unsuspected epidemic scenario. *Sci Rep* 2019;9:6959.

20. Gona F, Comandatore F, Battaglia S, *et al.* Comparison of core-genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. *Microb Genom* 2020;6(4). doi: 10.1099/mgen.0.000347.

21. Gorrie CL, Mirceta M, Wick RR, *et al.* Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin Infect Dis* 2017;65:208–215.

22. Hall MD, Holden MT, Srisomang P, *et al.* Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *Elife* 2019;8:e46402.

23. Hammerum AM, Lauridsen CAS, Blem SL, *et al.* Investigation of possible clonal transmission of carbapenemase-producing *Klebsiella pneumoniae* complex member isolates in Denmark using core genome MLST and National Patient Registry Data. *Int J Antimicrob Agents* 2020;55:105931.

24. Harada S, Aoki K, Yamamoto S, *et al.* Clinical and molecular characteristics of *Klebsiella pneumoniae* isolates causing bloodstream infections in Japan: occurrence of hypervirulent infections in health care. *J Clin Microbiol* 2019;57:e01206–19.

25. Houldcroft CJ, Roy S, Morfopoulou S, *et al.* Use of whole-genome sequencing of adenovirus in immunocompromised pediatric patients to identify nosocomial transmission and mixed-genotype infection. *J Infect Dis* 2018;218:1261–1271.

26. Jakharia KK, Ilaiwy G, Moose SS, *et al.* Use of whole-genome sequencing to guide a *Clostridioides difficile* diagnostic stewardship program. *Infect Control Hosp Epidemiol* 2019;40:804–806.

27. Kossow A, Kampmeier S, Schaumburg F, Knaack D, Moellers M, Mellmann A. Whole-genome sequencing reveals a prolonged and spatially spread nosocomial outbreak of Panton-Valentine leucocidin-positive meticillin-resistant *Staphylococcus aureus* (USA300). *J Hosp Infect* 2019;101:327–332.

28. Kwong JC, Lane CR, Romanes F, *et al.* Translating genomics into practice for real-time surveillance and response to carbapenemase-producing *Enterobacteriaceae*: evidence from a complex multi-institutional KPC outbreak. *Peer J* 2018;6:e4210.

29. Leong KWC, Cooley LA, Anderson TL, *et al.* Emergence of vancomycin-resistant *Enterococcus faecium* at an Australian hospital: a whole-genome sequencing analysis. *Sci Rep* 2018;8:6274.

30. Long SW, Beres SB, Olsen RJ, Musser JM. Absence of patient-to-patient intrahospital transmission of *Staphylococcus aureus* as determined by whole-genome sequencing. *mBio* 2014;5(5):e01692–01614.

31. 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. *Infect Dis Health* 2020;25:30–33.

32. Martin JSH, Eyre DW, Fawley WN, *et al.* Patient and strain characteristics associated with *Clostridium difficile* transmission and adverse outcomes. *Clin Infect Dis* 2018;67:1379–1387.

33. Mathur P, Khurana S, de Man TJB, *et al.* Multiple importations and transmission of colistin-resistant *Klebsiella pneumoniae* in a hospital in northern India. *Infect Control Hosp Epidemiol* 2019;40:1387–1393.

34. Meredith LW, Hamilton WL, Warne B, *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* 2020;20:1263–1271.

35. Miles-Jay A, Weissman SJ, Adler AL, Baseman JG, Zerr DM. Whole-genome sequencing detects minimal clustering among *Escherichia coli* sequence type 131-H30 isolates collected from United States children's hospitals. *J Pediatric Infect Dis Soc* 2021;10:183–187.

36. Neumann B, Bender JK, Maier BF, *et al.* Comprehensive integrated NGS-based surveillance and contact-network modeling unravels transmission dynamics of vancomycin-resistant enterococci in a high-risk population within a tertiary-care hospital. *PLoS One* 2020;15:e0235160.

37. Raven KE, Gouliouris T, Brodrick H, *et al.* Complex routes of nosocomial vancomycin-resistant *Enterococcus faecium* transmission revealed by genome sequencing. *Clin Infect Dis* 2017;64:886–893.

38. Roach DJ, Burton JN, Lee C, *et al.* A year of infection in the intensive care unit: prospective whole-genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet* 2015;11: e1005413.

39. Rose R, Nolan DJ, Moot S, *et al.* Molecular surveillance of methicillin-resistant *Staphylococcus aureus* genomes in hospital unexpectedly reveals discordance between temporal and genetic clustering. *Am J Infect Control* 2021;49:59–64.

40. Roy S, Hartley J, Dunn H, Williams R, Williams CA, Breuer J. Whole-genome sequencing provides data for stratifying infection prevention and control management of nosocomial influenza A. *Clin Infect Dis* 2019;69:1649–1656.

41. Sherry NL, Lane CR, Kwong JC, *et al.* Genomics for molecular epidemiology and detecting transmission of carbapenemase-producing *Enterobacterales* in Victoria, Australia, 2012 to 2016. *J Clin Microbiol* 2019;57:e00573–19.

42. Stenmark B, Hellmark B, Söderquist B. Genomic analysis of *Staphylococcus capitis* isolated from blood cultures in neonates at a neonatal intensive care unit in Sweden. *Eur J Clin Microbiol Infect Dis* 2019;38:2069–2075.

43. Sullivan MJ, Altman DR, Chacko KI, *et al.* A complete genome screening program of clinical methicillin-resistant *Staphylococcus aureus* isolates identifies the origin and progression of a neonatal intensive care unit outbreak. *J Clin Microbiol* 2019;57:e01261–19.

44. Tsujiwaki A, Hisata K, Tohyama Y, *et al.* Epidemiology of methicillin-resistant *Staphylococcus aureus* in a Japanese neonatal intensive care unit. *Pediatr Int* 2020;62:911–919.

45. van Beek J, Räisänen K, Broas M, *et al.* Tracing local and regional clusters of carbapenemase-producing *Klebsiella pneumoniae* ST512 with whole-genome sequencing, Finland, 2013 to 2018. *Euro Surveill* 2019;24(38). doi: 10.2807/1560-7917.ES.2019.24.38.1800522

46. Wang X, Zhou H, Chen D, *et al.* Whole-genome sequencing reveals a prolonged and persistent intrahospital transmission of *Corynebacterium striatum*, an emerging multidrug-resistant pathogen. *J Clin Microbiol* 2019;57: e00683–19.

47. Wendel AF, Malecki M, Otchwemah R, Tellez-Castillo CJ, Sakka SG, Mattner F. One-year molecular surveillance of carbapenem-susceptible *A. baumannii* on a German intensive care unit: diversity or clonality. *Antimicrob Resist Infect Control* 2018;7:145.

48. Julia L, Vilankar K, Kang H, Brown DE, Mathers A, Barnes LE. Environmental reservoirs of nosocomial infection: imputation methods for linking clinical and environmental microbiological data to understand infection transmission. *AMIA Annu Symp Proc* 2017;2017:1120–1129.

49. Stachel A, Pinto G, Stelling J, *et al.* Implementation and evaluation of an automated surveillance system to detect hospital outbreak. *Am J Infect Control* 2017;45:1372–1377.

50. Parcell BJ, Gillespie SH, Pettigrew KA, Holden MTG. Clinical perspectives in integrating whole-genome sequencing into the investigation of healthcare and public health outbreaks—hype or help? *J Hosp Infect* 2021;109:1–9.