Title

Statistical detection of geographic clusters of resistant Escherichia coli in a regional network with WHONET and SaTScan

Journal

Expert Review of Anti-infective Therapy, 14(11)

ISSN

1478-7210

Authors

Park, Rachel
O'Brien, Thomas F
Huang, Susan S
et al.

Publication Date

2016-11-01

DOI

10.1080/14787210.2016.1220303

Peer reviewed

Statistical detection of geographic clusters of resistant *Escherichia coli* in a regional network with WHONET and SaTScan

Rachel Park[a,b], Thomas F. O'Brien[a,c], Susan S. Huang[d], Meghan A. Baker[a,e], Deborah S. Yokoe[a], Martin Kulldorff[a,c,e], Craig Barrett[f], Jamie Swift[g] and John Stelling[a,c] for the Centers for Disease Control and Prevention Epicenters Program

[a]Brigham and Women's Hospital, Department of Medicine, Division of Infectious Diseases, Boston, MA, USA; [b]London School of Hygiene and Tropical Medicine, Faculty of Infectious and Tropical Diseases and Faculty of Public Health and Policy, London, UK; [c]Harvard Medical School, Department of Medicine, Boston, MA, USA; [d]University of California Irvine School of Medicine, Division of Infectious Diseases and Health Policy Research Institute, Orange, CA, USA; [e]Harvard Pilgrim Health Care and Harvard Medical School, Department of Population Medicine, Boston, MA, USA; [f]Premier, Inc., Safety Solutions, Charlotte, NC, USA; [g]Mountain States Health Alliance, Department of Infection Prevention, Johnson City, TN, USA

ABSTRACT

Background: While antimicrobial resistance threatens the prevention, treatment, and control of infectious diseases, systematic analysis of routine microbiology laboratory test results worldwide can alert new threats and promote timely response. This study explores statistical algorithms for recognizing geographic clustering of multi-resistant microbes within a healthcare network and monitoring the dissemination of new strains over time.

Methods: Escherichia coli antimicrobial susceptibility data from a three-year period stored in WHONET were analyzed across ten facilities in a healthcare network utilizing SaTScan's spatial multinomial model with two models for defining geographic proximity. We explored geographic clustering of multiresistance phenotypes within the network and changes in clustering over time.

Results: Geographic clustering identified from both latitude/longitude and non-parametric facility groupings geographic models were similar, while the latter was offers greater flexibility and generalizability. Iterative application of the clustering algorithms suggested the possible recognition of the initial appearance of invasive E. coli ST131 in the clinical database of a single hospital and subsequent dissemination to others.

Conclusion: Systematic analysis of routine antimicrobial resistance susceptibility test results supports the recognition of geographic clustering of microbial phenotypic subpopulations with WHONET and SaTScan, and iterative application of these algorithms can detect the initial appearance in and dissemination across a region prompting early investigation, response, and containment measures.

1. Introduction

1.1 Antimicrobial resistance

The emergence of antimicrobial resistance hinders our ability to treat infectious diseases with profound impacts on suffering, disability, death, and healthcare costs, and high rates of

resistance are seen in bacteria causing common healthcare-associated and community-acquired infections in all World Health Organization (WHO) regions [1]. Dr. Keiji Fukuda, WHO Assistant Director-General of Health Security, has warned that a 'postantibiotic era – in which common infections and minor injuries can kill. . . is a very real possibility for the 21st century' [2].

To confront this urgent threat, a global action plan was endorsed by the sixty-eighth World Health Assembly in May 2015 to confront antimicrobial resistance [2]. The global action plan outlined five strategic objectives to 'ensure, for as long as possible, continuity of successful treatment and prevention of infectious diseases with effective and safe medicines that are quality assured, used in a responsible way, and accessible to all who need them' [1]. These goals include improving awareness and understanding of antimicrobial resistance, strengthening surveillance and research, reducing the incidence of infection, optimizing the use of antimicrobial agents, and developing sustainable economic investments in new medicines, diagnostic tools, and other interventions for the needs of all countries [2].

Two key drivers impacting the epidemiology of microbial populations are: (1) selection pressure due to antimicrobial use and especially misuse in healthcare, agricultural, and industrial settings; and (2) transmission of resistant strains between individuals, communities, and nations due to local deficiencies in community and healthcare facility hygiene, and to domestic and international trade and travel [3]. In this work, we have explored algorithms for recognizing pathogen transmission and dissemination in order to support early response and containment [4–6].

1.2 Cluster detection with routine microbiology laboratory data

WHO highlighted in the global action plan the importance of surveillance in tracking and confronting antimicrobial resistance [1], while the WHO Global Strategy for Containment of Antimicrobial Resistance recognizes the role of national reference laboratories and surveillance networks as a 'fundamental priority' for containment efforts [7].

The WHO Collaborating Centre for Surveillance of Antimicrobial Resistance, located at Brigham and Women's Hospital in Boston, Massachusetts USA, has developed, disseminated, and supported WHONET as a surveillance software for the management of microbiological laboratory data since 1989 – based on earlier work with mainframe computers initiated in 1964 – with a special focus on the analysis and interpretation of antimicrobial susceptibility test results [8–10]. WHONET supports surveillance of infectious diseases and antimicrobial resistance in over 2,000 clinical, public health, veterinary, and food laboratories in over 120 countries worldwide. BacLink, the data import module of WHONET, is used to reformat and standardize data from diverse laboratory information systems into WHONET format. Data to be entered or uploaded into WHONET is configurable by the user, but typically includes patient identifiers and demographics (age, gender), medical encounter details (healthcare facility, medical service or ward, date of admission), specimen identifier and details (date, anatomical sample site), and microbiological findings (organism identification and antimicrobial susceptibility test results).

In addition to core descriptive statistics on organism frequency, distribution, and resistance

data and automated alerts on findings of public health importance, WHONET offers a number of more sophisticated algorithms for the statistical detection of case clusters suggestive of disease outbreaks. A number of such algorithms are already implemented in WHONET through an integration with the free SaTScan software [11,12], a free cluster detection tool using purely temporal, purely spatial, and spatiotemporal scan statistics in either a prospective or retrospective setting, for either count or continuous data. Using Monte Carlo simulations, SaTScan utilizes temporal and/or spatial scanning windows to search for clusters in which the number of observed cases within a given time period and/or geographic entity exceeds to a statistically significant degree the expected number of cases for that time period and location, adjusting for the multiple testing inherent in the many cluster locations and sizes evaluated [13].

We have explored the detection of clusters using a number of routinely available data elements: organism, serotype, phage type, multi-resistance phenotype, biochemical phenotype, medical ward (or groups of wards), medical service (or groups of services), latitude and longitude of the testing laboratory or healthcare facility, as well as combinations of these variables, for example resistance phenotype + ward [14– 18]. Of all of these variables, we have repeatedly demonstrated the value of multi-resistance phenotypes (i.e. the set of antimicrobials to which a microbial isolate is non-susceptible) in improving specificity of alerts (as detected cases have homogeneous characteristics), sensitivity (through decreases in background noise), and timeliness (through improved specificity and sensitivity) [16].

In previous work, we have explored use of the WHONETSaTScan integrated tool for the detection of possible outbreaks in the community [14,15], within individual healthcare and long-term care facilities [16,17], and across facilities in healthcare networks [18]. Although outbreak detection is limited when relying on routinely collected laboratory data, as patients may not present to medical attention and there may be biases in clinical sampling and test practices, microbiology laboratories around the world demand that we examine them for events of public health importance with subsequent epidemiological and/or microbiological investigation as warranted.

While the focus of our earlier work has been the detection of possible disease outbreaks meriting prompt recognition and response, in this study we explored the use of purely spatial, time-independent algorithms with alternate objectives in mind. The first was to explore the ability of spatial algorithms to recognize static geographic clustering of microorganisms and strain phenotypes using routine microbiology laboratory data. Such geographic clustering would be expected in some clinical settings, such as greater predominance of multi-resistant strains in intensive care units or tertiary care facilities, but unexpected in others, for example in general medical wards or certain long-term care facilities, but not others. The second objective was to explore whether the iterative application of these spatial algorithms over time could recognize the initial appearance and gradual dissemination of emerging threats that could be missed by analysis strategies optimized for the detection of acute changes in case incidence.

We explored the algorithms using the distribution of *Escherichia coli (E. coli)* across a multicenter network of healthcare facilities, in part because of *E. coli's* importance in both community and healthcare-associated infections and in part because of important therapeutic challenges posed by resistant strains. However, the approach developed is generalizable to

other pathogens and geographic settings, for example across wards within a healthcare facility [19]. High proportions of resistance in *E. coli* to third-generation cephalosporins, fluoroquinolones, and a number of other antimicrobial classes has required an increased reliance on reserve agents such as carbapenems, to which resistance is rapidly rising, and colistin. There is increased awareness of the threat posed by *E. coli* multi-locus sequence type 131 (ST131), recognized as a major cause of invasive multidrug-resistant infections in the USA and elsewhere [20].

2. Methods

2.1. Data sources and preparation

Ten facilities affiliated with one healthcare system and subscribing to the SafetySurveillor infection prevention module of Premier, Inc. participated in this project. The facilities included both hospitals and long-term care facilities. Ethical approval was received from the Partners Healthcare Institutional Review Board, as well as from the relevant Institutional Review Boards of each facility. Microbiological data covering a 3-year period included patient (study identifier), encounter (clinical service and ward, date of admission), sample (identifier, collection date, type), and microbiological findings (organism identification, susceptibility test results). Data files included both quantitative antimicrobial susceptibility test results, notably disk diffusion zone diameters and minimal inhibitory concentrations (MICs), but within this study we focused on the qualitative interpretations of the test as 'susceptible' versus 'nonsusceptible' using the interpretation provided by the submitting laboratory, consistent with recent recommendations of the Clinical and Laboratory Standards Institute (CLSI) guidelines [21]. This study was approved by the Partners Human Research Committee.

Results from all clinical isolates were available in the source analysis database, but only the first isolate of each species per patient in a 365-day period was included for study. Clinical isolates studied represented both community and healthcareassociated infections. The database included *E. coli* results for only diagnostic samples, no screening isolates. Consequently, the statistical clusters identified in this work reflect the patient populations (both inpatient and outpatient) served by particular hospital laboratories, and not necessarily patients hospitalized within those facilities. We have explored the application of the algorithms described in this application to presumptive isolates from healthcare-associated infections (based on the difference between the hospital admission date and the specimen collection date) [19], but the results are not presented here.

In past work, we have consistently demonstrated the value of multi-drug resistance phenotypes to be meaningful and specific proxy indicators of strain relatedness. As the susceptibility test practices in each facility were not identical, we explored the availability of data for each antimicrobial tested from each facility in order to identify a subset of all antimicrobials for which results were consistently available for most isolates. To accomplish this, we utilized the 'Number tested' column of WHONET's '%RIS and test measurements' analysis (%RIS = %Resistant, %Intermediate, %Susceptible). Our goal was to identify antimicrobials tested throughout the network, for which at least 90% of relevant *E. coli* isolates were tested within each facility. This core set of commonly tested antimicrobials were utilized to define a 'multi-resistance phenotype' for each microbial isolate.

2.2 Detection of geographic clusters

In our previous work, we have relied primarily on SaTScan's prospective space-time permutation model, integrated into WHONET, to identify outbreaks in time and space. In this study, we explored three probability models: Poisson (with hospital-based incidence of *E. coli* infection estimated with available 'patient-day' statistics), Bernoulli (comparing proportions of *E. coli* among all samples tested or organisms isolated between facilities), and multinomial (considering different *E. coli* resistance categories). Based on our early explorations of these three models, we considered the multinomial to be most promising and robust for further development and generalizable to a variety of scenarios as it does not require external denominators (patient catchment population, facilityor ward-level patient days, or laboratory test volumes) that may not always be available, reliable, or stable [22]. While the space-time permutation model and temporal Poisson models have been implemented within WHONET for several years, this is not yet the case with the multinomial model pending evaluation. Consequently, analyses were performed directly with SaTScan 9.4.2 [12,13].

SaTScan's 'spatial' algorithms require a 'spatial' variable. Most frequently, this is a latitude and longitude coordinate associated with some aspect of the clinical case, for example, patient home residence or place of work, or site of clinical care. In the previous work, we have utilized grid coordinate-based spatial variables such as the latitude and longitude of the testing laboratory, nonparametric spatial variables such as medical ward or service, or conceptual spatial variables such as microbial species, resistance phenotype, serotype, or phage type [15–18]. In this study, our focus was on the location of the patient's medical facility. The relationship between geographic locations can be defined in two ways in SaTScan: (1) grid coordinates, such as latitude and longitude; and (2) nonparametric groupings utilizing SaTScan's non-Euclidean 'neighbor' and 'meta-location' features. Both approaches were evaluated in this study.

For grid coordinates, the SaTScan 'coordinate location file' was populated with the longitude and latitude of the healthcare facility associated with each patient isolate. Latitudes and longitudes were determined from the batch geocoding utility found at www.spatialepidemiology.net, maintained by Imperial College London and the Wellcome Trust [23].

For nonparametric geographic relationships, we utilized the SaTScan 'meta-location' file feature to create meaningful location groupings not defined by grid coordinates. Meaningful groupings can be defined based on similar type of care, geographic proximity, and known patient referral patterns between facilities. An example of a 'meta' location is the grouping of a hospital with its affiliated long-term care facilities, such as nursing homes and rehabilitation centers, as in the following example: [CityA] = Hosp1, Hosp2, LTC1, LTC2, LTC3.

The multinomial model requires the identification of a 'categorical' variable, and explores whether cases within each category are randomly distributed or not across the geographical regions or entities. For our analyses, an isolate's multi-resistance phenotype was the categorical variable used – in other words, we explored: (1) whether the distribution of different types of E. coli across the healthcare network as defined by their observed multi-resistance phenotype was consistent with random variation or not; and (2) if not, which facility (or facility groupings) clusters were of greatest statistical significance and which

observed resistance phenotypes exceeded expectations to the greatest degree as estimated by relative risk.

SaTScan scan statistics use circular (or optionally elliptical) scan windows of varying radii in order to determine the most statistically significant clusters in geographic space involving one or more healthcare facility. The default SaTScan settings of 999 Monte Carlo replications and a geographic maximum scan window of 50% of cases were utilized. Clusters with the largest reported likelihood ratios indicate the most likely clusters, of either high or low incidence.

2. Results

3.1. Descriptive statistics

Institutional characteristics of the ten healthcare facilities studied are provided in Table 1, derived from the American Hospital Directory [24] and the institutional web page of each facility.

As indicated in Table 2, there were 54,651 unique patients with positive specimens in the entire network within the 3year study period, accounting for 73,705 microbial isolates in total. A total of 19,362 patients (35% of all patients) tested positive for *E. coli* with 28,377 total isolates. The distribution of *E. coli* isolates by healthcare facility is displayed in Figure 1 The largest facility was Hospital 3, the tertiary care teaching hospital for this network with 514 patient beds and its laboratory accounts for 42% of all isolates and 39% of all E. coli. Hospital 2 was the next largest facility with 155 patient beds accounting for 23% of all isolates and 24% of all E. coli. Samples collected in long-term care facilities contributed 1% of the total isolates and 0.7% of the E. coli.

The categorization of strains by multi-resistance phenotypes depends on the identification of a core set of antimicrobials that consistently have data available for analysis. Over 15 distinct antimicrobials were tested by the laboratories of the various facilities, but the specific antimicrobial tested varied from facility to facility, as did their testing practices (e.g. only testing certain antimicrobials in urine) and reporting practices (e.g. suppress results for third-line agents if an organism is found to be susceptible to a number of first-line agents). In examining the number of test results for each antimicrobial for each facility, we excluded from further consideration antimicrobials that were infrequently tested. If the microbiological activity and testing of two antimicrobials were substantially similar, for example ciprofloxacin and levofloxacin (LVX), only one of the two was included in the set of core antimicrobials. Given these considerations, we identified a core set of six antimicrobials suitable for assigning a multi-resistance phenotype to each strain: ampicillin (AMP), ceftriaxone (CRO), ceftazidime (CAZ), gentamicin (GEN), LVX, and trimethoprim/ sulfamethoxazole (SXT). For example, a strain with phenotype 'AMP CRO CAZ' is non-susceptible to the three antimicrobials indicated, but susceptible to the remaining three. This set of six antimicrobials worked well for purposes of strain characterization with the exception of Hospital 4 (Table 2), in which case the CAZ result was frequently missing. Consequently, this facility was excluded from further analyses. Excluding Hospital 4, complete results for these six antimicrobial were available for 93% of the E. coli studied. Because of this good level of data completeness and the lack of meaningful information provided by isolates missing test results, we utilized the WHONET feature 'exclude isolates missing results for one or more antimicrobials.' The distribution of all resistance phenotypes by facility is provided in Table 3.

Special public health concern about invasive, multi-resistant E. coli ST131 was noted above. Conveniently, there is a specific multi-resistance phenotype observed in a large proportion of E. coli ST131 isolates which can be used as a useful proxy phenotypic marker of this strain – specifically, these strains are often possess a CTX-M-15 or CTX-M-14 extended spectrum beta-lactamase, which confers resistance to CTX but

Table 1. Facility characteristics.

Facility	Facility type	Number of beds	Number of patient-days	Average length-of-stay (days to the nearest integer)
Hosp1	Short-term acute care	50–100	10,000–30,000	6
Hosp2	Short-term acute care	101–200	30,001–50,000	4
Hosp3	Short-term acute care	>200	>50,000	6
Hosp4	Short-term acute care	101–200	10,000–30,000	4
Hosp5	Short-term acute care	50–100	<10,000	5
Hosp6	Short-term acute care	101–200	10,000–30,000	4
LTC1	Nursing care	101–200	30,001–50,000	*
LTC2	Transitional care	<50	*	*
LTC3	Rehabilitation	<50	*	*
LTC4	Psychiatric	50–100	*	*

*Information not available from either the American Hospital Directory[11] or the individual facility's web site.

Table 2. Isolate distribution by facility.

Facility	Number (%) of isolates	Number (%) of *E. coli* isolates	Number (%) of *E. coli* tested against core antimicrobials
Hosp1	2,917 (5)	1,008 (5)	829 (82)
Hosp2	12,820 (23)	4,696 (24)	4,675 (99.6)
Hosp3	22,927 (42)	7,464 (39)	7,192 (96)
Hosp4	6,241 (11)	2,593 (13)	137 (5)
Hosp5	4,061 (7)	1,518 (8)	923 (61)
Hosp6	5,165 (9)	1,954 (10)	1,882 (96)
LTC1	16 (0.03)	9 (0.05)	2 (22)
LTC2	261 (0.5)	59 (0.3)	56 (95)
LTC3	191 (0.3)	42 (0.2)	41 (98)
LTC4	52 (0.1)	19 (0.1)	19 (100)
TOTAL	54,651	19,362	15,756 (81)

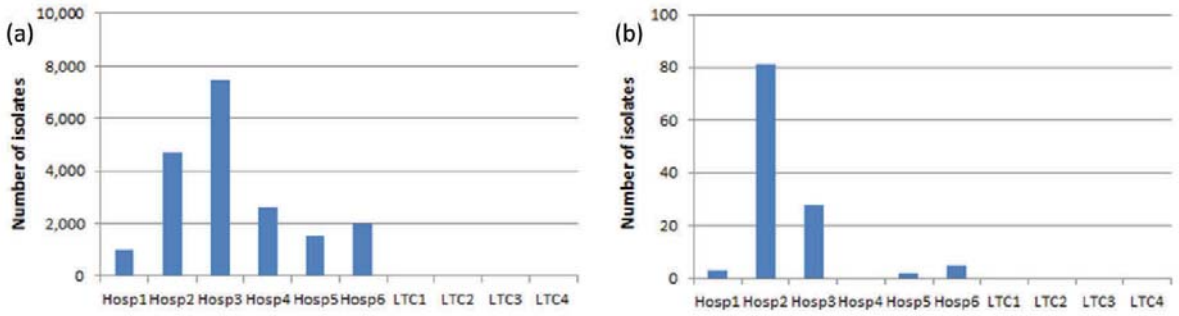

Figure 1. Distribution by of all *E. coli* isolates (1a) and *E. coli* with the typical *E. coli* ST131 multi-resistance phenotype – non-susceptible to AMP, CTX, and LVX, but susceptible to CAZ (1b).

Table 3. *E. coli* resistance phenotypes by facility.

Resistance profile					Possible *E. coli* ST131 phenotype	Hosp1	Hosp2	Hosp3	Hosp4	Hosp5	Hosp6	LTC1	LTC2	LTC3	LTC4
Susceptible to all						366	2084	3579	58	437	882	1	25	18	6
SXT						13	46	100	3	13	29		1		
LVX						15	106	142	4	23	42		1		
GEN							10	16		6	1				
AMP						136	601	1128	22	140	298		9	8	6
LVX	SXT					11	62	88	1	6	29			1	
GEN	SXT						1	2		1					
GEN	LVX						2	3		2	1				
CAZ	GEN							1							
AMP	SXT					51	276	449	9	64	120		1	2	1
AMP	LVX					57	326	401	8	50	101		10	6	
AMP	GEN					2	25	36	3	4	6				
AMP	CAZ					1	5	3							
AMP	CRO						2								
GEN	LVX	SXT				1	7	2							
AMP	LVX	SXT				47	471	401	13	60	126	1	3	3	2
AMP	GEN	SXT				28	114	139	4	19	42				
AMP	GEN	LVX				22	76	120	1	11	21			1	1
AMP	CAZ	LVX					2			1					
AMP	CAZ	GEN					1	1							
AMP	CRO	SXT				1	1								
AMP	CRO	LVX			Yes		5	3		1					
AMP	CRO	GEN					1	1							
AMP	CRO	CAZ				2	11	53		12	12				1
AMP	GEN	LVX	SXT			26	171	203	5	28	45		4	2	2
AMP	CAZ	LVX	SXT			1	2	2							
AMP	CAZ	GEN	LVX								1				
AMP	CRO	LVX	SXT		Yes		23	7		1	2				
AMP	CRO	GEN	SXT				3	1							
AMP	CRO	GEN	LVX		Yes		1	1							
AMP	CRO	CAZ	SXT			1	5	4		5					
AMP	CRO	CAZ	LVX			7	54	64	3	13	60				
AMP	CRO	CAZ	GEN				3	4		1	1				
AMP	CAZ	GEN	LVX	SXT			5	4			3				
AMP	CRO	GEN	LVX	SXT	Yes	3	15	5			1				
AMP	CRO	CAZ	LVX	SXT		19	94	102	1	9	39		1		
AMP	CRO	CAZ	GEN	SXT				8	2	13					
AMP	CRO	CAZ	GEN	LVX		5	13	38		2	4				
AMP	CRO	CAZ	GEN	LVX		14	51	81		1	16		1		

not to CAZ, as well as typical mutations to the gyrA and parC conferring resistance to fluoroquinolones, such as LVX [25]. Not all E. coli isolates with this phenotype will have the ST131 sequence type, and not all E. coli ST131 isolates will display this phenotype, but a number of studies have highlighted a useful association of this phenotype with this genotype [26,27], including a study in two US states demonstrating that 48% of CTX-M-15 producing E. coli and 66% of CTX-M-14producing E. coli were indeed E. coli ST131 [28]. Multi-resistance phenotypes that include this type of resistance ('AMP CRO LVX') are highlighted in Table 3 as consistent with the typical E. coli ST131 phenotype. As seen in Figure 1, only 28% of such isolates were from patients at the tertiary care center Hospital 3, whereas 81% were isolated from the much smaller Hospital 2.

3.2. Cross-sectional multinomial analysis

Table 4 presents the results of the multinomial analysis where relatedness of healthcare facilities was defined by geographic proximity through latitude and longitude. Three statistically significant clusters were found: Cluster 1 included two

Table 4. Spatial multinomial analysis cluster results utilizing latitude and longitude. Findings have been filtered to phenotypes with relative risk greater than 1.5 with at least two isolates. A relative risk of ∞ indicates that this resistance phenotype was seen only in this facility and no other.

Clusters	Facilities	Number of isolates	Cluster p-value	Resistance profiles					*E. coli* ST131 phenotype	Number of isolates	Relative risk
1	Hosp2	5601	0.001	AMP	LVX	SXT				524	1.58
	LTC2			AMP	CRO	LVX			Yes	5	2.27
	LTC3			AMP	CAZ	LVX	SXT			3	2.72
	Hosp1			AMP	CAZ					6	3.63
				AMP	CAZ	LVX				2	3.63
				AMP	CRO	LVX	SXT		Yes	23	4.17
				AMP	CRO	GEN	SXT			3	5.44
				AMP	CRO	GEN	LVX	SXT	Yes	18	5.44
				GEN	LVX	SXT				8	7.25
				AMP	CRO					2	∞
				AMP	CRO	SXT				2	∞
2	Hosp3	7211	0.001	AMP	CRO	CAZ				54	1.73
	LTC4			AMP	CRO	CAZ	GEN	LVX		38	1.88
3	Hosp6	1882	0.001	AMP	CAZ	GEN	LVX	SXT		3	2.46
				AMP	CRO	CAZ	LVX			60	3.14

hospitals and two long-term care facilities, and the relative risk associated with each resistance phenotype within these groups of facilities is provided. For example, a patient within these four facilities (Hospital 1, Hospital 2, LTC 2, and LTC 3) has a 5.44 greater risk of testing positive for an *E. coli* isolate with multi-resistance phenotype 'AMP CRO GEN LVX SXT' compared to a patient in the remaining six facilities in the network . Cluster 2 involves a single hospital and an associated long-term care facility providing psychiatric care, while Cluster 3 only involves a single hospital. A relative risk of ∞ indicates that this resistance phenotype was only seen in this facility/ facility grouping and no others, a finding which in itself should merit further investigation.

Table 5 presents the corresponding results for the multinomial analysis utilizing facilities and facility groupings. The overall findings in Table 5 have many similarities to those of Table 4. For example, Cluster 1 in the latitude-longitude analysis is comparable to Cluster 1 in the meta-location analysis; Cluster 2 in the former is comparable to Cluster 2 in the latter; and Cluster 3 in the former is comparable to Cluster 4 in the latter. The observed relative risks between the two analyses are also very similar, for example in reviewing the findings of Cluster 1 in the two analyses, the three resistance phenotypes consistent with the putative *E. coli* ST131 phenotype have relative risks of 2.27, 4.17, and 5.44 in the former and 2.96,
3.95, and 5.45 in the latter.

The meta-location analysis disclosed statistically significant findings in Hospital 5 (Cluster 3) that were not seen in the latitude-longitude analysis. A noteworthy finding from this cluster is that the relative risk associated with phenotype 'AMP CRO CAZ GEN SXT' is 25.94, which should prompt investigation into the epidemiology and risk factors associated with this phenotype in the patient population served by Hospital 5.

3.3. Iterative multinomial analysis

Further analyses explored whether there were any changes in geographic clustering of E. coli resistance phenotypes during the 3 years of the study. We compared location clusters identified in Year 1 with those identified in Year 3 with results displayed in Table 6. As the previous section demonstrated that the clusters derived from latitude and longitude-based analyses were broadly similar to those of the meta-location analyses, only the latter are presented here. Cluster 1 was clearly seen in both years with similar resistance phenotypes, but if one considers the relative risks associated with the three E. coli ST131 phenotypes, these dropped

from 4.89, 9.79, and 14.68 in Year 1 to 3.39, 3.14, and 5.03, respectively in Year 3 suggestive of either an overall decrease in the frequency of the phenotype (which was not the observation in this situation) or more likely a dissemination from Hospital 2 to Hospital 1 such that the statistical significance of the clustering within the Hospital 2 patient population was less in Year 3 as the organism has become more dispersed across the network.

The Hospital 5 cluster in Year 1 was associated with a high relative risk for phenotype 'AMP CRO CAZ SXT' (RR = 11.64) and 'AMP CRO CAZ GEN SXT' (RR = 27.17), but this statistical cluster disappeared entirely by Year 3. In contrast, no statistically significant findings were associated with Hospital 1 or Long-term Care Facility 3 in Year 1, but in Year 3, they appear in Cluster 2 along with Hospitals 3 and 6. Three of these four facilities are in the same city of approximately 65,000 inhabitants, while the fourth is a neighboring community 7 miles away suggesting possible dissemination of the indicated multi-resistant strain phenotypes between the population catchment areas of the four facilities during the 3 years of the study.

4. Discussion

The traditional statistical approach for detecting infectious disease outbreaks is the application of temporal [29,30] or spatiotemporal algorithms [15–18] to identify significant increases in case number or incidence case cluster suggestive of a public health event meriting real-time investigation and timely response. In this study, we explored the use of purely

Table 5. Multinomial cluster analysis results utilizing facilities and facility groupings. Findings have been filtered to show phenotypes with relative risk greater than 1.5 with at least two isolates. A relative risk of ∞ indicates that this resistance phenotype was seen only in this facility and no other.

Cluster ID	Facility or facilities	Number of *E. coli*	Cluster p-value	Resistance profiles	*E. coli* ST131 phenotype	Number of isolates	Relative risk
1	Hosp2	4,675	0.001	AMP CAZ LVX SXT		2	1.58
				AMP CAZ GEN LVX SXT		5	1.69
				AMP LVX SXT		471	1.7
				AMP CRO LVX	Yes	5	2.96
				AMP CAZ		5	2.96
				AMP CRO GEN LVX SXT	Yes	15	3.95
				AMP CAZ LVX		2	4.74
				AMP CRO LVX SXT	Yes	23	5.45
				GEN LVX SXT		7	5.53
				AMP CRO GEN SXT		3	7.11
				AMP CRO		2	∞
2	Hosp3	7,192	0.001	AMP CRO CAZ		53	1.66
				AMP CRO CAZ GEN LVX		38	1.89
3	Hosp5	923	0.001	AMP CRO CAZ		12	2.1
				GEN		6	3.07
				GEN LVX		2	4.61
				AMP CRO CAZ SXT		5	6.92
				AMP CRO CAZ GEN SXT		15	25.94
4	Hosp6	1,882	0.001	AMP CAZ GEN LVX SXT		3	2.46
				AMP CRO CAZ LVX		60	3.14

Table 6. Comparison of geographic clustering in Year 1 (6a) versus Year 3 (6b) filtered for relative risk >1.5. A relative risk of ∞ indicates that this resistance phenotype was seen only in this facility and no other.

Year 1

Cluster	Facilities	Number of isolates	Cluster p-value	Resistance profiles	*E. coli* ST131 phenotype	Number of isolates	Relative risk
1	Hosp2	1735	0.001	GEN		5	1.53
				AMP CRO CAZ GEN		2	1.63
				AMP GEN LVX SXT		80	1.65
				AMP LVX SXT		207	1.91
				AMP GEN		13	2.45
				AMP CAZ GEN LVX SXT		2	2.45
				GEN LVX SXT		3	3.67
				AMP CRO GEN LVX SXT	Yes	4	4.89
				AMP CRO LVX	Yes	4	9.79
				AMP CRO LVX SXT	Yes	6	14.68
2	Hosp3	2987	0.001	AMP CRO CAZ		24	1.6
				AMP CRO CAZ GEN LVX		16	2
3	Hosp5	473	0.001	AMP GEN		3	1.52
				LVX		16	1.62
				AMP GEN SXT		15	1.62
				GEN		3	3.49
				AMP CRO CAZ SXT		2	11.64
				AMP CRO CAZ GEN SXT		7	27.17
4	Hosp6	614	0.002	AMP CRO CAZ GEN LVX SXT		8	1.63
				LVX		22	1.76
				AMP CRO CAZ LVX SXT		21	2.38
				AMP CRO CAZ LVX		23	3.66

Year 3

Cluster	Facilities	Number of Isolates	Cluster p-value	Resistance profiles	*E. coli* ST131 phenotype	Number of isolates	Relative risk
1	Hosp2	1809	0.001	AMP CRO CAZ GEN LVX SXT		49	1.54
				AMP LVX SXT		210	1.84
				AMP CRO LVX	Yes	5	3.14
				AMP CRO GEN LVX SXT	Yes	9	3.39
				AMP CAZ GEN		2	3.77
				AMP CRO SXT		2	3.77
				AMP CAZ LVX SXT		2	3.77
				AMP CRO LVX SXT	Yes	16	5.03
				AMP CAZ		3	5.66
				GEN LVX SXT		3	∞
2	Hosp1 Hosp3 LTC3 Hosp6	2586	0.001	AMP GEN LVX		61	1.78
				AMP CRO CAZ GEN LVX		19	1.94
				AMP CRO CAZ		17	2.47
				AMP CRO CAZ GEN SXT		4	4.07
				GEN SXT		2	∞
3	Hosp6	824	0.001	AMP CRO CAZ LVX		36	3.62
				AMP CAZ GEN LVX SXT		3	5.33

spatial, time-independent algorithms with alternate objectives in mind. The first objective was to explore the ability of spatial algorithms to recognize purely geographic clustering of microorganisms and strain phenotypes using routine microbiology laboratory data. The second was to explore the iterative application of these spatial algorithms over time to recognize the initial appearance, subsequent movement and diffusion, and in some instance disappearance of strain phenotypes in a geographic region. Based solely on statistical findings, one cannot be certain whether the clusters identified are true or important, but we believe that many would suggest concerning issues and emerging threats that would merit further epidemiological and/or microbiological investigation and potentially response.

A key motivator to this work was the recognition that healthcare facilities often consider – incorrectly – that findings within their own facility are 'typical' of those in their own area. Infection control staff often believe that similar nearby facilities have the same types of microorganisms in a similar proportion and with similar epidemiology as their own. Prior to this study, we found many examples where this was not the case, so we sought to develop algorithms for the systematic detection of such clustering. In some instances, geographic clustering of multidrug-resistant bacteria is to be expected, for example within intensive care and

burn units or within tertiary care hospitals and long-term care facilities. However, in many cases, observed clustering does not have an obvious explanation. Contributory factors may include poor local hygiene measures, inappropriate antimicrobial use practices, chance historical events, or underlying population demographics of the community served by the laboratory. The recognition that the challenges faced by each facility are different is a first step toward targeted investigation, intervention, and control efforts. In this study, we selected E. coli as a pathogen for study and the testing laboratory as the geographic unit of study, but the methods are generalizable and can be applied to other species and location coordinates.

4.1. Comparison of SaTScan multinomial model to traditional chi-square

We applied spatial algorithms employing the multinomial probability model to compare the observed proportions of distinct multidrug-resistant phenotypes of E. coli isolates among the network facilities. The SaTScan spatial multinomial model algorithm is conceptually similar to traditional chisquare statistics – quantifying whether the 'observed' distribution of cases is statistically consistent with the 'expected' distribution of cases, but with two crucial advantages: (1) the SaTScan approach utilizes configurable scanning windows, so is not restricted to predefined geographic entities and groupings as traditional chi-square would be; and (2) critically, a traditional chi-square calculation provides only a simplistic 'yes' or 'no' response to the question of whether the observed distribution of cases in its entirety is or is not consistent with random variation – but without identifying particular rows or columns which deviate most from expectations. In contrast, the multinomial approach provides more granular details such as p-values for cluster significance and relative risk associated with each multi-resistance phenotypes which facilitate the recognition of the major phenotypic strains and locations associated with statistically significant clusters.

These characteristics are well highlighted in Table 3. A traditional chi-square based on the rows and columns of this table demonstrates a p-value of $p < 0.00001$, though the minimal data requirements for performing chi-square are not met and there is no obvious conclusion to be drawn as to which phenotypes and which locations deviate most significantly from expectations. Furthermore, this traditional approach cannot recognize clusters associated with facility groupings, for example Hosp1+Hosp2+LTC3, which can be explored with the SaTScan spatial multinomial model.

The strengths of the multinomial approach can be seen in Table 4. A p-value is associated with each cluster (which could represent a single facility or multiple), and for each resistance phenotype within this cluster an associated relative risk is presented. For example, patients associated with Hospital 5 have nearly 26 times the risk of having an E. coli with multi-resistant phenotype 'AMP CRO CAZ GEN SXT' than patients at other facilities, a finding probably unknown to staff members at Hospital 5, who perhaps consider this to be a 'usual finding' while staff members in other facilities are likely unaware of how frequent this particular strain is at one of the other facilities in their network. Table 4 highlights three rows with the resistance phenotype commonly associated with E. coli ST131, one row with the base phenotype of 'AMP CRO LVX' non-susceptibility, one row with additional SXT non-susceptibility, and one row with nonsusceptibility to both SXT and GEN. Of note, all three of these rows were seen in Hospital 2 patients, suggesting an important clustering of distinct strains of possible E. coli ST131 in the patients served by the laboratory of this facility. As this was a retrospective analysis of clinical isolates, molecular confirmation of the E. coli ST131

genotype was not possible, but in realtime monitoring, molecular confirmation and epidemiological investigation and response would be feasible.

Table 7 is an extract of Table 3, but including only rows with this putative E. coli ST131 phenotype. So the clustering of the putative E. coli ST131 in the Hospital 2 patient population (smaller than population served by Hospital 3) suggested by the multinomial analysis could potentially have been suspected from a thorough but tedious manual inspection of the original phenotype distributions and figures, but elucidated in a manner which is more comprehensive (addressing all locations and phenotypes, both high-priority concerns like

E. coli ST131 and non-priority), systematic (amenable to automation and configurable alerts), flexible (considering both individual facilities and groups of facilities), and generalizable to other settings and data sources.

4.2. Changes in geographical clustering over time

While we did not utilize the space-time version of the multinomial scan statistic, it is meaningful to look at changes in these static snapshots of geographical clustering over time, especially to explore important changes in the distribution of strains over time that are too gradual to be recognized by algorithms optimized for the detection of abrupt and timelimited outbreaks. Using the clinical databases of the network facilities, one can distinguish three priority scenarios of public health relevance:

- Stable geographic clustering: In this scenario, the underlying epidemiology of the involved microbial populations appears stable. The clustering of resistant strains may reflect higher risk populations, as in intensive care units or university hospitals, or unexpected findings in low-risk medical wards or certain long-term care facilities, but not others.
- Early geographic clustering that disappears in time: The disappearance of clustering over time may simply reflect a decrease or disappearance in the number of isolates with a given phenotype over time. This could be due to evolutionary fade-out or aggressive-targeted infection prevention measures. Of greater public health concern would be the progressive dissemination of a strain from an initial concentrated focus of infection (in which control efforts could be aggressive and locally targeted) across a geographic region so that the strain is increasingly common but with a more diffusion distribution (in which targeted control efforts are less likely to be successful).
- Geographic clustering that appears in time: Clustering of this type would suggest the incursion of a new threat into the geographic region studied (or alternatively a concentration of a strain previously broadly disseminated). The appearance of new strains in a defined geographic unit should prompt timely investigation, confirmation, and where appropriate response actions to limit spread to other geographic areas.

Table 7. Distribution by facility of isolates with the putative *E. coli* ST131 phenotype: non-susceptible to AMP, CTX, and LVX, but susceptible to CAZ, extracted from Table 3.

Resistance profile					Hosp1	Hosp2	Hosp3	Hosp4	Hosp5	Hosp6	LTC1	LTC2	LTC3	LTC4
AMP	CRO	LVX				5	3		1					
AMP	CRO	LVX	SXT			23	7		1	2				
AMP	CRO	GEN	LVX			1	1							
AMP	CRO	GEN	LVX	SXT	3	15	5			1				
Total (%)					3 (4%)	44 (65%)	16 (24%)		2 (3%)	3 (4%)				

Table 6 displays a comparison of the most prominent statistical clusters in Year 1 of the study to those of the Year 3, and the strong statistical association of isolates with the frequent

E. coli ST131 resistance phenotype in the Hospital 2 population appears in both years. However, one can note the great drop in relative risks associated with these phenotypes between the two time periods – RR of 9.79, 14.68, and 4.89 in Year 1 dropping 3.14, 5.03, and 3.39, respectively, in Year 3. This is consistent with a manual review of the facility-level statistics by year – notably 6 of 7 (86%) isolates with phenotype AMPCRO-LVX-SXT, were seen in the Hospital 2 population in Year 1, but only 16 of 22 (73%) in Year 3.

Despite the increased number of Hospital 2 patients in Year 3, there is also a greater number of Hospital 3 patients with this phenotype in Year 3: (1) suggesting a gradual movement of this phenotype across the region; and (2) explaining the decrease in statistical significance of the Hospital 2 clustering. The number of cases with this phenotype in the other hospitals is between 0 and 3 in the full 3year study suggesting that these are sporadic cases of infection, whereas by Year 3, the strain appears to be well established and endemic in both the Hospital 2 and Hospital 3 patient population.

Of note, there were no isolates with the typical E. coli ST131 phenotype during the first 4 months of 3-year data collection, so it is possible that we have observed the initial or early appearance of this strain in a clinical database in this network, beginning in the patient population covered by Hospital 2, but without data from earlier time periods, one cannot be certain of this. In fact, this type of observation suggests that there is value in the iterative monthly application of these algorithms to support prospective detection of strain emergence and gradual spread that might be missed by algorithms optimized for acute outbreak detection.

It is useful to contrast the results of the current study with the findings of parallel work [17] conducted by our group in which we applied the SaTScan space-time permutation model, an algorithm we have optimized for purposes of detecting acute outbreaks (over the course of several days to a few months), to the same database studied here. In that work, we identified 16 statistical clusters suggestive of possible outbreaks over 3 years in the 10 facilities, but only one was with E. coli, and that statistical cluster did not overlap with any of the clusters identified in this study. A useful conclusion is that different algorithms can serve different purposes – the space-time permutation model has been optimized for detecting acute/short-term clusters, often associated with unwarranted transmission of microbes within and between healthcare facilities, and the multinomial model, which we believe offers a new approach for identifying (1) stable geographic clustering; and (2) slow movements and changes in distributions of isolates.

4.3. Comparison of latitude-longitude coordinates versus meta-location groups

We explored two approaches for describing the geographic relatedness of healthcare facilities: (1) grid coordinates utilizing latitude and longitude; and (2) 'meta-location' meaningful

groupings of facilities based on geographic proximity, types of medical care, or patient referral patterns. The findings in this study indicate that statistical findings between the two approaches were similar.

Though there are merits in both approaches, the metalocation approach offers greater flexibility in defining the relationship between healthcare facilities utilizing a number of relevant dimensions beyond pure physical proximity. The approach is more robust with regard to the details of the geographic layout, for example if a healthcare network involves both urban and community facilities, mountainous or island communities (in which latitude and longitude do not adequately capture the practical distance between facilities). Most importantly, the meta-location approach is more generalizable to a broad variety of settings, for example permitting the application of these clustering algorithms to medical wards or groups of medical wards within a hospital, facilitating the detection and tracking of healthcare-associated outbreaks [19].

5. Conclusions

Previous work by us and others has demonstrated the value of statistical algorithms for the timely detection of emerging infectious disease threats and early response. In this study, we have aimed to expand the range of validated analytical approaches to incorporate nontemporal algorithms which identify unexpected spatial clustering in static snapshots. We found the SaTScan spatial multinomial model to be conceptually similar to traditional chi-square analyses, but with significant advantages in granularity of statistical detail and insights, flexibility with regard to location definitions and groupings either utilizing latitude and longitude grid coordinators or meaningful hierarchical groupings of medical wards or geographical entities, robustness with regard to chi-square assumptions which often do not hold in practice, and generalizability. Based on the encouraging results obtained in this study and other work, our intent is to proceed with full integration of the multinomial model as a standard feature within WHONET.

The iterative application of the geographical analyses permits the recognition of stable geographic clustering, both expected and non-expected, as well as the gradual appearance, disappearance, and dissemination of distinct microbial subpopulations that would be missed with algorithms optimized for the detection of acute, time-limited outbreaks. For example, in this study, we likely documented the early incursion of strains with the typical E. coli ST131 resistance phenotype into the patient population served by one of the hospitals of the network with later spread and endemic establishment in the referral center of the network, but only sporadic cases in the other centers.

While the algorithms that we studied have very generalizable application, a number of studies have found that the use of antimicrobial resistance phenotypes as meaningful proxy indicators for microbial subpopulations greatly improves the specificity of detected signals – since cases detected are phenotypically homogeneous – and sensitivity – through decrease in background noise related to unrelated strains. It was necessary to exclude one facility from several of the analyses because of lack of availability of one of the key antibiotics studied, emphasizing the value of coordination among facilities in providing consistently available data for at least a core set of antimicrobials.

An important aspect of this work is its reliance on the widely used WHONET software

to perform sophisticated statistical analyses for cluster recognition through its integration of a variety of statistical algorithms including those available within the free SaTScan software. WHONET is used to support surveillance of antimicrobial resistance in over 2,000 laboratories worldwide in over 110 countries. Thus, our aim is not only to evaluate and validate data management and alert strategies, such as those described in this paper, but also to support their real-time integration into routine practice by microbiologists, infection control practitioners, epidemiologists, and public health authorities at local, regional, and national levels worldwide to support early detection, response, and containment efforts.

Key issues

- Traditional statistical approaches for the detection of infectious clusters focus on the temporal component and deviations from a historical baseline.
- In this work, we applied the Monte Carlo-based spatial multinomial model available within SaTScan to explore geographic clustering of clinical isolates of multi-resistant E. coli in a healthcare network of six hospitals and four long-term care facilities.
- SaTScan's spatial multinomial model is conceptually similar to a classic chi-square RxC contingency table but with significant advantages in granularity of statistical feedback, flexibility in grouping geographic entities, and robustness with regards to underlying data assumptions.
- We analyzed space utilizing both latitude and longitude grid coordinates as well as meaningful facility groupings based on geographic proximity and patient referral patterns with similar findings observed for the two approaches. The meta-location approach offers greater robustness to assumptions on meaningful 'distances', flexibility in defining relationships between locations, and generalizability to a number of clinical scenarios.
- The iterative application of geographic cluster algorithms permits the recognition of stable clustering over time (both expected and unexpected) as well as the appearance, disappearance, and gradual dissemination of resistant populations that would be missed by statistical algorithms optimized for acute outbreak detection.
- Antimicrobial multi-resistance test results are valuable phenotypic strain markers which improves the specificity, sensitivity, and epidemiological relevance of cluster detection.
- WHONET and SaTScan software is utilized to support surveillance of infectious diseases and antimicrobial resistance in over 2000 microbiology laboratories in over 110 countries, so the strategies elaborated within this paper have broad applicability by a range of non-statistician healthcare professionals worldwide.

Acknowledgments

The work was carried out in satisfaction of Ms. Park's thesis requirement for the London School of Hygiene and Tropical Medicine's Master of Science degree in Control of Infectious Diseases. We thank Premier, Inc. and the participating network for their collaboration in this research.

Funding

This paper was funded by grants from the National Institutes of Health (grant number RR025040) and the Centers for Disease Control and Prevention (CDC) Epicenters Program (Cooperative Agreement Number U54CK000172). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. World Health Organization. WHO Antimicrobial resistance – Global report on surveillance. WHO/HSE/PED/AIP/2014.2, Geneva, 2015.
2. World Health Organization. WHO Global action plan on antimicrobial resistance. ISBN 978 92 4 150976 3, Geneva, 2015.
3. Huttner A, Harbarth S, Carlet J, et al. Antimicrobial resistance: a global view from the 2013 world healthcare-associated infections forum. Antimicrob Resist Infect Control. 2013;2(1):31.
4. Allegranzi B, Nejad S, Combescure C, et al. Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. The Lancet 2011;377(9761):228–241.
5. Zarb P, Coignard B, Griskeviciene J, et al. The European Centre for Disease Prevention and Control (ECDC) pilot point prevalence survey of healthcare-associated infections and antimicrobial use. Euro Survillance 2012;17(46).
6. Magill S, Edwards J, Bamberg W, et al. Multistate point-prevalence survey of health care-associated infections. New England J Med 2014;370(13):1198–1208.
7. World Health Organization. WHO global strategy for containment of antimicrobial resistance. Geneva, 2001, WHO/CDS/CSR/DRS/2001.2.
8. Stelling J, O'Brien TF. WHONET – software for surveillance of infecting microbes and their resistance to antimicrobial agents. In: Persing DH, Tenover FC, Hayden RT, et al., editors. Molecular microbiology: diagnostic principles and practice. 3rd ed. Washington (DC): ASM Press; 2016. doi:10.1128/9781555819071.ch48. ISBN-13: 978-1555819088, ISBN-10: 1555819087.
• This book chapter provides a historical and technical overview of WHONET's inception, development and use since 1964. Examples of WHONET's current capabilities and future directions are provided.
9. WHO Collaborating Centre for Surveillance of Antimicrobial Resistance. WHONET Home Page. Available from: www.whonet.org.
•• Home page from which the WHONET software, tutorials, and training materials can be download.

10. World Health Organization. WHO Home page for WHONET software, 2015. Available from: www.who.int/drugresistance/ whonetsoftware

11. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods 1997; 26:1481–1496.

12. Kulldorff M. Information Management Services (2015) SaTScan: Software for the spatial and space–time scan statistics, version

9.4.2. Available from: www.satscan.org

•• Home page from which the SaTScan software, tutorials, and training materials can be download.

13. Kulldorff M. SaTScanTM User Guide for Version 9.4, Boston, Feb 2015.

14. Stelling J, Yih WK, Galas M, et al. Automated use of WHONET and SaTScan to detect outbreaks of Shigella spp. Using Antimicrobial Resistance Phenotypes. Epidemiol Infect 2010 Jun;138(6):873–883.

15. Viñas M, Tuduri E, Galar A, et al. Working group MIDAS Argentina, Kulldorff M, Galas M. Laboratory-based prospective surveillance for community outbreaks of Shigella spp. In Argentina. PloS Negl Trop Dis 2013;7(12): e2521

• The authors applied the SaTScan space-time permutation model to the prospective detection of possible community outbreaks of shigellosis with molecular investigation and confirmation performed by pulsed-field gel electrophoresis and investigation by local epidemiologists responsible for outbreaks of foodborne pathogens.

16. Huang SS, Yokoe DS, Stelling J, et al. Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study. PloS Med 2010 Feb;7(2):e1000238.

17. Baker MA, Yokoe DS, Stelling J, et al. Automated outbreak detection of hospital-associated infections. Oral abstract presented in IDWeek 2015, San Diego: Infectious Diseases Society of America. October 8 2015.

• This abstract summarizes work in which the SaTScan space-time permutation model was applied to 45 separate healthcare facilities to identify possible outbreaks of healthcare-associated infections and the subsequent validation and assessment of statistical findings by local infection control practitioners and hospital epidemiologists. A full manuscript summarizing the study and findings is in preparation for submission for publication.

18. Lishman H Outbreak detection of antimicrobialresistant healthcare associated infections at the hospital network level. Thesis submitted for degree in Master of Science in Control of Infectious Diseases of the London School of Hygiene and Tropical Medicine, London, 2014.

19. Park R WHONET and SaTScan for hospital network and intra-hospital geographic cluster analysis of antimicrobial resistance. Thesis submitted for degree in Master of Science in Control of Infectious Diseases of the London School of Hygiene and Tropical Medicine, London, 2015.

•• The work described in the current manuscript is based on a more comprehensive literature review and data analysis plan included in thesis submitted as a partial requirement for a Master of Science degree. The full thesis also explores application of the multinomial manual to geographic cluster detection within a single medical facility and also discrimination of clusters associated with putative healthcare-associated and community-acquired infections based on date of admission.

20. Johnson J, Johnston B, Clabots C, et al. Escherichia coli sequence type ST131 as the

major cause of serious multidrug-Resistant E. coli infections in the United States. Clin Infect Dis 2010;51(3):286– 294.

21. Clinical and Laboratory Standards Institute. M100-S25 Performance standards for antimicrobial susceptibility testing; Twenty-fifth informational supplement, Wayne, Pennsylvania. ISBN 1-56238-989-0, 2015.

22. Jung I, Kulldorff M, Richard OJ A spatial scan statistic for multinomial data. Stat Med 2010, 29:1910–1918.

• This paper elaborates the theoretical foundation of the multinomial probability implemented within SaTScan described in this study.

23. Imperial College London, Wellcome Trust. Spatialepidemiology.net – Mapping infectious disease epidemiology Home Page. 2015. Available from: www.spatialepidemiology.net.

24. American Hospital Directory, Inc. American Hospital Directory Home Page. 2015. Available from: www.ahd.com.

25. Suzuki S, Shibata N, Yamane K, et al. Change in the prevalence of extended-spectrum-β-lactamase-producing Escherichia coli in Japan by clonal spread. J Antimicrob Chemother 2008;63 (1):72–79.

26. Peirano G, Richardson D, Nigrin J et al. High prevalence of ST131 isolates producing CTX-M-15 and CTX-M-14 among extendedspectrum-b-lactamase-producing Escherichia coli isolates from Canada. Antimicrob Agents Chemother 2010; 54: 1327–1330.

27. Petty NK, Zakour Ben NL, Stanton-Cook M, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A 2014. 111:5694–5699.

28. Chen LF, Freeman JT, Nicholson B, et al. Anderson DJ Widespread dissemination of CTX-M-15 genotype extended-spectrum-β-lactamase-producing Enterobacteriaceae among patients presenting to community hospitals in the southeastern United States. Antimicrob Agents Chemother 2014. 58:1200–1202.

29. Hutwagner L, Thompson W, Seeman GM, et al. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003 Jun;80(2 Suppl 1):i89–96.

30. Farrington CP, Beale AD Computer-aided detection of temporal clusters of organisms reported to the communicable disease surveillance centre. Commun Dis Rep CDR Rev 1993 May 21;3 (6):R78–82.

Present affiliation for Martin Kulldorff is Brigham and Women's Hospital, Department of Medicine, Division of Pharmacoepidemiology and Pharmacovigilance, Boston, MA, USA.