

RESEARCH ARTICLE

Open Access



# Geographic surveillance of community associated MRSA infections in children using electronic health record data

Lilly Cheng Immergluck<sup>1,2\*</sup> , Traci Leong<sup>3</sup>, Khusdeep Malhotra<sup>4</sup>, Trisha Chan Parker<sup>1</sup>, Fatima Ali<sup>1</sup>, Robert C. Jerris<sup>2,5</sup> and George S. Rust<sup>6</sup>

## Abstract

**Background:** Community-associated methicillin resistant *Staphylococcus aureus* (CA-MRSA) cause serious infections and rates continue to rise worldwide. Use of geocoded electronic health record (EHR) data to prevent spread of disease is limited in health service research. We demonstrate how geocoded EHR and spatial analyses can be used to identify risks for CA-MRSA in children, which are tied to place-based determinants and would not be uncovered using traditional EHR data analyses.

**Methods:** An epidemiology study was conducted on children from January 1, 2002 through December 31, 2010 who were treated for *Staphylococcus aureus* infections. A generalized estimated equations (GEE) model was developed and crude and adjusted odds ratios were based on *S. aureus* risks. We measured the risk of *S. aureus* as standardized incidence ratios (SIR) calculated within aggregated US 2010 Census tracts called spatially adaptive filters, and then created maps that differentiate the geographic patterns of antibiotic resistant and non-resistant forms of *S. aureus*.

**Results:** CA-MRSA rates increased at higher rates compared to non-resistant forms,  $p = 0.01$ . Children with no or public health insurance had higher odds of CA-MRSA infection. Black children were almost 1.5 times as likely as white children to have CA-MRSA infections (aOR 95% CI 1.44, 1.75,  $p < 0.0001$ ); this finding persisted at the block group level ( $p < 0.001$ ) along with household crowding ( $p < 0.001$ ). The youngest category of age ( $< 4$  years) also had increased risk for CA-MRSA (aOR 1.65, 95%CI 1.48, 1.83,  $p < 0.0001$ ). CA-MRSA encompasses larger areas with higher SIRs compared to non-resistant forms and were found in block groups with higher proportion of blacks ( $r = 0.517$ ,  $p < 0.001$ ), younger age ( $r = 0.137$ ,  $p < 0.001$ ), and crowding ( $r = 0.320$ ,  $p < 0.001$ ).

**Conclusions:** In the Atlanta MSA, the risk for CA-MRSA is associated with neighborhood-level measures of racial composition, household crowding, and age of children. Neighborhoods which have higher proportion of blacks, household crowding, and children  $< 4$  years of age are at greatest risk. Understanding spatial relationship at a community level and how it relates to risks for antibiotic resistant infections is important to combat the growing numbers and spread of such infections like CA-MRSA.

**Keywords:** Pediatric, Antibiotic resistant, Spatial analyses, *Staphylococcus aureus* infections, Methicillin resistant *Staphylococcus aureus*

\* Correspondence: [limmergluck@msm.edu](mailto:limmergluck@msm.edu)

<sup>1</sup>Department of Microbiology/Biochemistry/Immunology, Department of Pediatrics and Clinical Research Center, Morehouse School of Medicine, 720 Westview Drive, SW, Atlanta, GA 30310, USA

<sup>2</sup>Children's Healthcare of Atlanta, 1405 Clifton Road NE, Atlanta, GA 30322, USA

Full list of author information is available at the end of the article



## Background

The use of geocoded electronic health record (EHR) data for improving health outcomes is in its infancy [1] particularly for infectious diseases such as community-associated methicillin resistant *Staphylococcus aureus* (CA-MRSA) infections [2–4]. This is an important disease to study because it is preventable and the number of children hospitalized with CA-MRSA infections in the United States increased from 6.7 cases per 1000 admissions in 2002 to 21.2 cases in 2007 [5]. We demonstrated how geocoded EHR data in conjunction with a geographic information system (GIS) can be used to enhance the geographic surveillance of CA-MRSA. In doing so, we 1) characterized the communities where children infected by the bacteria *Staphylococcus aureus* (*S. aureus*) resided during the height of the CA-MRSA epidemic in the US [6] (2002–2010); and 2) identified areas at high risk for CA-MRSA where targeted interventions aimed at preventing disease transmission will have potentially the most impact.

Electronic health records are becoming a routine part of patient care. Their primary purpose is to follow the provision of care to individual patients. However, they are also a source of data for public health surveillance that cannot be obtained from reportable disease registries or vital statistics databases. EHR data typically includes basic demographic characteristics of the patient as well as their list of diagnoses, prescriptions, results from laboratory tests, referrals to specialists, and provider characteristics [7]. They also contain residential information about the patient which can be geocoded and linked to data about the patient's social, physical and built environment, community resources and other data about their community [8]. Use of geocoded EHR for public health is still emerging, but there are a wide range of recent application areas including hospital readmissions for pediatric asthma [9], preventable emergency department visits [10, 11], hospitalizations for common illnesses [12], obesity in children [12], and Type 2 diabetes [8, 13, 14].

Population based studies have shown a higher rate of invasive MRSA infections among blacks and other minorities [15–18] and in densely populated areas with high numbers of pre-school and school aged children [19]. The subset of MRSA infections emerging from the community, or CA-MRSA infections, have been endemic in many urban areas with significantly increased risk for invasive infection among blacks [6, 17, 20]. Risks associated with CA-MRSA disease may also be related to socioeconomic factors [4, 21], and are not well characterized for this community acquired infection [6, 16, 17]. For instance, household crowding has been cited as a risk for CA-MRSA infections, but the impact of neighborhood crowding has not previously been addressed [22, 23].

The goals of this study are to characterize the neighborhoods with the highest rates of CA-MRSA infection in the Atlanta–Sandy Springs–Roswell, GA Metropolitan Statistical Area (the Atlanta MSA), and to analyze the spatial distribution of CA-MRSA compared to the non-antibiotic resistant form of *S. aureus*, also known as community-associated methicillin sensitive *Staphylococcus aureus* (CA-MSSA). There are very few studies to characterize CA-MRSA infections in this region of the US, compared to other parts. This spatial characterization would enable us to identify specific areas at greatest risk for antibiotic resistance and thus, identify those communities most likely to benefit from targeted interventions. Comparing spatial patterns of the nonresistant forms of *S. aureus* to resistant forms would identify clonal spread of specific resistant strains which may be circulating in a community or neighborhood; primary and secondary intervention plans aimed at specific geographic areas could be tailored, based on the risks associated with a particular location. For example, CA-MRSA rates within certain communities [24] have decreased over the last 5 years when community-level prevention measures (e.g., increased hand sanitizers located throughout public areas) were implemented. The importance of achieving these goals would allow public health officials and area primary care providers to combat the spread of CA-MRSA infections through targeted interventions aimed at those communities at highest risk for acquisition and transmission. Such interventions would contribute to the development of health policy guidelines for infection control, effective home and school-based interventions, and identifying but not stigmatizing children who may be at increased risk for CA-MRSA based on individual and area level risk factors.

## Methods

### Overview

CA-MRSA infections in the Atlanta MSA from 2002 to 2010 were examined by applying exploratory spatial analysis in order to identify risk factors and characterize geographic variations of CA-MRSA rates compared to CA-MSSA rates.

### Study design

A retrospective epidemiology study was conducted of children from January 1, 2002 through December 31, 2010 who were treated for *S. aureus* infections at two pediatric hospitals in the Atlanta MSA (Scottish Rite Children's Hospital and Egleston Children's Hospital). Children diagnosed with *S. aureus* infection and who were managed in the emergency department (ED), or as an inpatient were included; no data was collected on *S. aureus* carriage or colonization. All had an International Classification of Diseases, Clinical Modification (ICD- 9-CM) code compatible with a staphylococcal infection, a positive *S. aureus*

culture, and lived within the 20-counties that comprise the Atlanta–Sandy Springs–Roswell, GA Metropolitan Statistical Area (Metro Atlanta). (The two hospitals consistently provide more than 80% of the pediatric hospitalizations for the Atlanta MSA [25].) This study was approved by Institutional Review Boards of hospitals and affiliated academic institutions.

Each unique child met the case definition if their EHR reported a positive MRSA infection on the first culture during a single hospitalization. MRSA carriage or colonization status is not routinely collected on patients with MRSA infection, so this was not collected for our study. Cases were then categorized as ‘community associated’ or ‘hospital acquired’ infections. Using CDC’s guidelines for case definitions, ‘community-associated’ infections were those which occurred within or at 48 h of hospital admission, otherwise, they were ‘hospital acquired’ [26]. Clinical isolates were identified using routine laboratory methods [27, 28]. Demographic information (race, age, gender and type of health insurance) and residential addresses were obtained from the EHR. Dates of hospital admission and discharge, patient’s body site(s) of infection, previous *S. aureus* related hospitalizations, and days to a positive culture from time of admission were also collected.

At the neighborhood level, the Census block group where a patient resided was identified [29–31]. We obtained American Community Survey [32] data to determine household crowding, proportion of black by block group, and proportion below poverty. Race/ethnicity, children living in poverty, and median household income were obtained from 2010 Decennial Census [32]. Boundary files for use in GIS for Georgia block groups and counties were downloaded from the National Historical Geographic Information System [33].

#### **Race/ethnicity**

Race/ethnicity was categorized as white, black, Hispanic and other (‘other’ included: Asian, Native American, multi-racial, Native Hawaiian, and other/declined) and proportions within each block were determined. Based on the proportions of race (black, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander) reported in the 2010 US Census, we determined which block groups were ‘concentrated’ area of blacks by using the overall percentage of blacks for the state as a reference point. [34] For example, the overall percentage of blacks in Georgia was 30.5% [34] in 2010, so any block group  $\geq 30.5\%$  blacks, was thereby defined as a ‘concentrated’ area of blacks.

#### **Poverty**

We used the ratio of income to poverty to control for variations in household size and to determine relative poverty by block group. It was determined by dividing

the median 2010 family income by the 2010 poverty threshold for a family of four (\$22,491) [35] and then assigning block groups as very low income ( $< 0.99$ ), low income (0.99–1.25), moderate income (1.25–2.50), high income ( $> 2.5$ ) [36, 37].

#### **Household crowding**

From US Housing and Urban Development definition, household crowding [38] was any housing unit with more than one person per room [4, 39]. Proportions of crowded households were calculated for each block group.

#### **Data analysis**

The basic unit of observation was each child diagnosed with *S. aureus* infection. Age, gender, and race entered the models as individual level covariates and all others were measured at the Census-block level. All risk factors were categorical except age. Since there were possible unknown correlations between outcomes of interest, and estimates of parameters using a generalized linear model, we applied a generalized estimated equations (GEE) model using an exchangeable covariance structure with model based standard errors; crude odds ratios (OR) were based on conditions determined a priori to be associated with risk of *S. aureus* and used as estimates of relative risks. Adjusted odds ratios (aORs) were calculated using a generalized linear mixed model, allowing for random effects and correlated errors for non-normal data.

Three separate models were applied for multiple regression analyses: Model 1- individual factors only, Model 2 - neighborhood factors, and Model 3- individual and neighborhood factors. Specific individual level and neighborhood level variables were identified, after comparing Model 1 to Model 3, and Model 2 to Model 3; this was done to ascertain which variables remained significant. Variables were included in the final adjusted model if, e.g., when comparing Model 1 to Model 3, the log likelihood statistics for the fitted model were significant for these variables. All tests for significance were two-tailed, and a  $p$ -value of  $\leq 0.05$  was considered significant. Statistical analysis was performed using SAS 9.4 (SAS Institute, Cary, NC).

#### **Spatial analysis**

For spatial analyses, we used data at the 2010 US Census block group because of the relatively homogeneous nature and close relationship of smaller census-derived units to health outcomes [40]. Data from all 20 counties of the Atlanta MSA were used for the multilevel and spatial analysis. To improve the geographic visualization of the disease patterns, the maps only show the eight counties that surround the Atlanta City limits; 78% of all cases were in these counties.

**Geographic distribution of community acquired- MRSA and -MSSA**

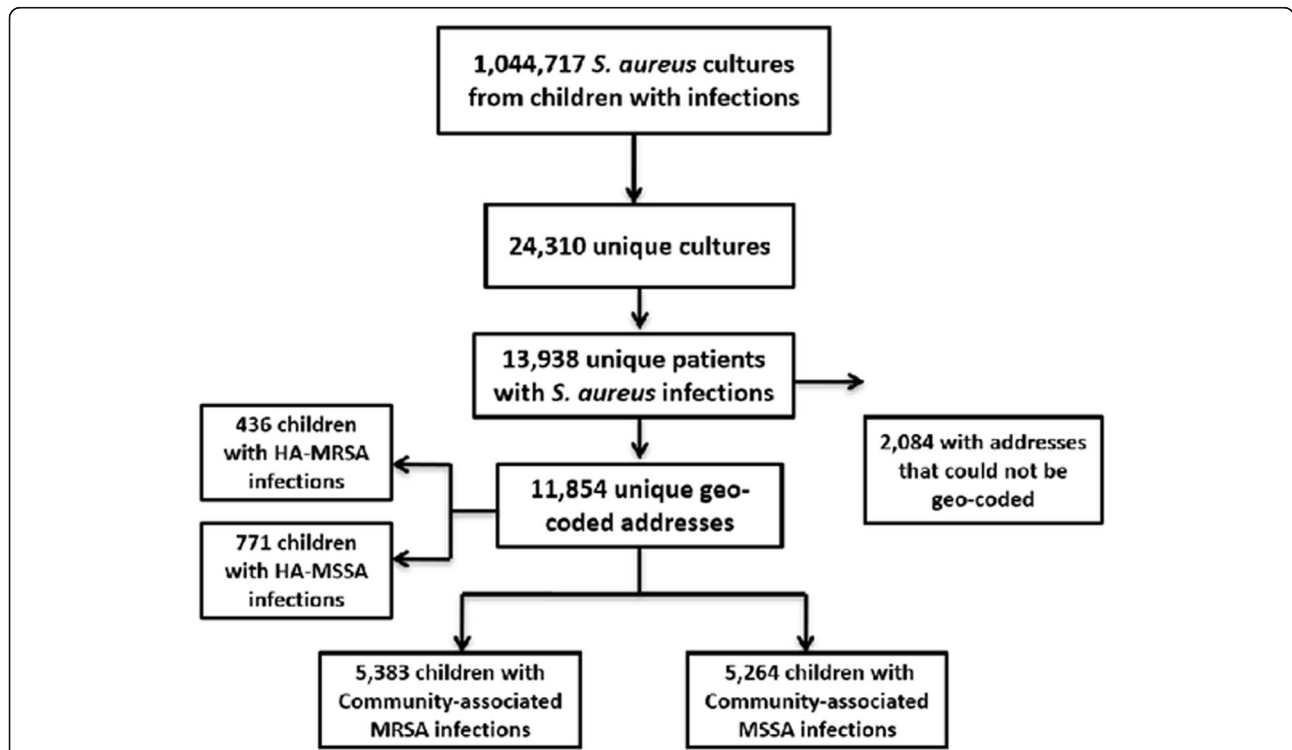
We measured the risk of CA-MRSA and CA-MSSA as standardized incidence ratios (SIR) which use the observed number of events in an area in the numerator and the expected number of events in the denominator. The number of expected MRSA events was calculated by applying the regional rate of CA-MRSA (14.2 per 100,000) to the number of children in each Census block group and the expected number of CA-MSSA events were calculated by applying the regional rate of CA-MSSA (13.2 per 100,000) to the number of children in each Census block group.

We calculated these SIRs in overlapping units of geography called spatially adaptive filter areas [41] (spatial filters). Spatial filters are aggregations of smaller geographic units, which, by themselves, do not contain sufficiently large enough populations to calculate a reliable disease rate. Spatial filter areas effectively address a common problem in the disease mapping literature called the small number problem [42], which occurs when the population in an administrative unit of geography (i.e., a county or a Census block group) is too small to calculate a reliable disease rate. The sizes of the spatial filter vary across the study area

and depend on the number of neighboring Census block groups needed for a threshold value; spatial filters are generally much larger in less populated areas than those in denser areas.

Each spatial filter is centered on a sampling location that we derived from the geometric center of each Census block group within the Atlanta MSA. Here, the smaller units that comprise each spatial filter are the set of Census block groups nearest to each sampling location, which we aggregated together until a threshold value of 30 expected CA-MRSA cases and 30 expected CA-MSSA cases were obtained. To accomplish this, we created a unique spatial filter for CA-MRSA and a unique spatial filter for CA-MSSA and then used whichever produced the largest filter area [43]. We chose 30 because the standard error of the SIR can never exceed 0.183 and any small area with an SIR less than 0.702 or greater than 1.425 will be statistically significant at  $p < 0.01$ .

The SIRs for CA-MRSA and CA-MSSA were then attributed to the sampling locations and spatially interpolated to reveal a continuous geographic pattern of statistically reliable CA-MRSA and CA-MSSA risk estimates. A by-product of the spatially adaptive filtering procedure is a map of smoothed disease rates, but it



**Fig. 1** Enrollment Schema. All patients admitted to Scottish Rite Children’s Hospital and Egleston Children’s Hospital from 2002 to 2010 with positive culture for *Staphylococcus aureus* (*S. aureus*) and had diagnosis of *S.aureus* infection were included. (Data collected did not include any *S. aureus* colonization or carriage information as this was not available.) All unique patients’ addresses were geo-coded and divided into those with community associated methicillin resistant *Staphylococcus aureus* (MRSA) and those with community associated methicillin sensitive *Staphylococcus aureus*

differs from other rate smoothing algorithms—such as headbanging [44–46] or localized empirical Bayesian estimation methods [47–49]—which use potentially unreliable disease rates pre-computed in administrative units. The rates calculated in spatially adaptive filters are not pre-computed; we first aggregated the raw number of *S. aureus* cases in neighboring geographic units until a large enough population to calculate a reliable disease rate is obtained and then calculate the disease rates. The effect of this procedure is that all SIRs are based on a minimum level of statistical reliability because the size of the geographic units is based on the principle of choosing an acceptable minimum level of reliability [50].

We also used the same spatial filter areas to calculate covariates. For a given spatial filter area, the numerator for *percent black* is the sum of the number of blacks in the Census block groups that comprise the spatial filter and the denominator is the sum of the people of any race in the same Census block groups. The numerator for the age covariate is the sum of the population younger than 5 years and the denominator is all children younger than 19 years. The numerator for the household crowding is the sum of the number of housing units with more than one person per room and the denominator is the sum of the housing units. We then used Pearson’s R correlation coefficients to measure the association between the percent of *S. aureus* that were

**Table 1** Patient level and neighborhood characteristics

Variable	Overall (n = 10,642)	MRSA (n = 5379)	MSSA (n = 5263)	p-value
<b>Patient Level Characteristics</b>				
Hospital				< 0.001
Egleston	4537 (42.6)	2331 (43.3)	2206 (41.9)	
Scottish Rite	6105 (57.4)	3048 (56.7)	3057 (58.1)	
Race, n (%)				< 0.001
Black	3553 (33.4)	2224 (41.3)	1329 (25.3)	
Hispanic	427 (4.0)	212 (3.9)	215 (4.1)	
Other	857 (8.1)	377 (7.0)	480 (9.1)	
Unknown	156 (1.5)	65 (1.2)	91 (1.7)	
White	5649 (53.1)	2501 (46.5)	3148 (59.8)	
Gender, n (%)				< 0.001
F	5102 (47.9)	2731 (50.8)	2371 (45.1)	
M	5540 (52.1)	2648 (49.2)	2892 (54.9)	
Age group, n (%)				< 0.001
0–3 yrs	4820 (45.3)	2888 (53.7)	1932 (36.7)	
4–12 yrs	3350 (31.5)	1399 (26.0)	1951 (37.1)	
> 12 yrs	2472 (23.2)	1092 (20.3)	1380 (26.2)	
Age, Mean (SD)	6.69 (5.65)	5.95 (5.55)	7.44 (5.66)	< 0.001
Insurance, n (%)				< 0.001
Multiple Types	20 (0.2)	10 (0.2)	10 (0.2)	
Private	4876 (45.8)	2101 (39.1)	2775 (52.7)	
Public	5282 (49.6)	3002 (55.8)	2280 (43.3)	
Self Pay	462 (4.3)	266 (4.9)	196 (3.7)	
<b>Neighborhood Level Characteristics</b>				
Household crowding, n (%)				< 0.001
≤ 1 person/room	5434 (51.1)	2639 (49.1)	2795 (53.1)	
> 1 person/room	5208 (48.9)	2740 (50.9)	2468 (46.9)	
Black population, n (%)				< 0.001
< 40.46%	7322 (68.8)	3361 (62.5)	3961 (75.3)	
≥ 40.46%	3320 (31.2)	2018 (37.5)	1302 (24.7)	
%Below poverty, Mean (SD)	17.1 (14.5)	18.3 (15.1)	15.9 (13.8)	< 0.001

CA-MRSA and the three covariates calculated in the spatial filters.

**Results**

**Sample size and population characteristics**

Figure 1 shows the selection criteria for the study. Of 13,938 unique children with a *S. aureus* infection, 2084 were excluded for invalid addresses (missing, postal office designations, invalid addresses); and 1207 children were excluded with hospital-acquired infections. In 2010, there were 5533 blocks in Georgia and 2097 block groups (38% of all block groups) had one or more observed cases of CA-MRSA. There were a total of 5379 CA-MRSA cases reported during this time period [51]. We determined there were 10,642 community-associated *S. aureus* infections with complete information, and 50.5% were the antibiotic resistant form, CA-MRSA (Table 1). We then stratified the infections based on the hospital location (Egleston Children’s Hospital or Scottish Rite Children’s Hospital). Differences were found between CA-MRSA and CA-MSSA infections with regard to race, age, gender and type of health insurance. Although white children had the highest proportion of CA-MRSA and CA-MSSA, black children with CA-MRSA was nearly double that of black children with CA-MSSA infections. Children with CA-MRSA also were younger and had higher rates of public health insurance compared to CA-MSSA ( $p < 0.01$ ).

**Prevalence and trend of antibiotic resistant *Staphylococcus aureus* (2002–2010)**

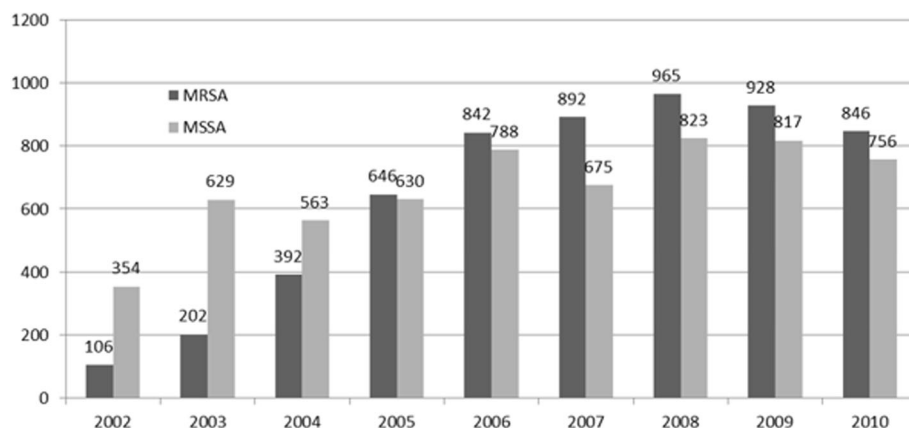
CA-MRSA cases increased per year, rising at higher rates compared to CA-MSSA,  $p = 0.01$ . CA-MRSA infections more than doubled from 23.2% in 2002 to 53.7% in 2006. (Fig. 2).

**Risk factors for community associated MRSA**

Children with no or public health insurance had higher odds of CA-MRSA infection compared to those with private insurance ( $p < 0.01$ ). Black children were one and a half times (aOR 1.58 (1.44-1.75,  $p < 0.0001$ )) as likely as whites to have CA-MRSA infection, after adjusting for gender, age, and health insurance. The youngest category of age (< 4 years) also had increased risk for CA-MRSA (aOR 1.65, 95% CI 1.48, 1.83,  $p < 0.0001$ ). Both crude and adjusted Odds Ratios from Model 1 as described in Methods are summarized in Table 2.

At the neighborhood level (Model 2), the unadjusted Model 2 showed significant differences between CA-MRSA and CA-MSSA, with risks for antibiotic resistance being higher in blocks with concentrated black populations ( $\geq 40.46\%$ ,  $p < 0.001$ ) and evidence of household crowding ( $p < 0.001$ ). However, household crowding no longer was a significant risk factor for CA-MRSA ( $p = 0.6633$ ), after adjusting for poverty and black concentration, while poverty and concentrated blacks living within blocks remained significant in the adjusted model ( $p < 0.0001$  and  $p = 0.0035$ , respectively). (Table 3) Blocks in the third and fourth quartiles of crowding had higher odds of poverty levels (3rd quartile’s OR 1.32; 95% CI 1.17–1.48;  $p$ -value  $< 0.0001$ ; 4th quartile’s OR 1.54, 95% CI 1.38–1.75,  $p < 0.0001$ ). This risk was only slightly higher for CA-MRSA in block groups living below the poverty level (aOR 1.01, 95% CI 1.002, 1.008), compared to aOR of 1.73 (95% CI 1.58, 1.89) if the neighborhood block was concentrated black ( $p < 0.0001$ ).

Multi-level or mixed model analysis (Model 3) showed blacks living in neighborhoods with  $\geq 40.6\%$  blacks, being female, public- or self-pay insurance, and younger age remained highly significant for risk of CA-MRSA infections compared to CA-MSSA, even after adjusting for



**Fig. 2** Prevalence of Antibiotic Resistant and Non Antibiotic Resistant *Staphylococcus aureus* Infections in Children, 2002–2010. The number of unique patients with antibiotic resistant (MRSA) and non antibiotic resistant (MSSA) infections from 2002 to 2010 is shown by year. (MRSA, methicillin resistant *Staphylococcus aureus*; MSSA, methicillin sensitive *Staphylococcus aureus*)

**Table 2** Individual-Level Factors for CA-MRSA Compared to CA-MSSA Infection for Crude and Adjusted Odds Ratios

Variable	Crude OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
<b>Race</b>				
White	Referent	–	Referent	–
Black	1.92 (1.75, 2.10)	<.0001	1.58 (1.44,1.75)	<.0001
Hispanic	1.10 (0.90, 1.35)	0.3359	.81 (.66,1.0)	.0487
Other	0.94 (0.81, 1.09)	0.3989	.73 (.63, .86)	<.0001
<b>Gender</b>				
Male	Referent	–	Referent	–
Female	1.28 (1.18, 1.39)	<.0001	1.22 (1.12,1.33)	<.0001
<b>Age (years)</b>				
> 12 years	Referent	–	Referent	–
4–12 years	.95 (.85, 1.05)	.2725	.95 (.85,1.05)	.3248
< 4 years	1.74 (1.58, 1.92)	<.0001	1.65 (1.48,1.83)	<.0001
<b>Health Insurance</b>				
Private	Referent	–	Referent	–
Self pay	1.70 (1.40,2.08)	<.0001	1.55 (1.27,1.89)	<.0001
Public	1.65 (1.52, 1.79)	<.0001	1.42 (1.30,1.55)	<.0001

Note: The analysis above does not include those patients identified with ‘unknown’ race nor those identified to have multiple types of insurance

income, health insurance, gender, crowding, and race/ethnicity (Table 4).

**Spatial densities of CA-MRSA and CA-MSSA infections**

The spatial distribution of CA-MRSA and CA-MSSA cover similar areas but the SIRs differ in areas affected: CA-MRSA encompasses larger areas with higher SIRs (and a significantly larger area where the ratio is > 2.5). CA-MRSA is much higher than CA-MSSA in the southeastern part of the eight counties that surround the Atlanta city limits. Similarly, this contrast in CA-MRSA and CA-MSSA are seen in southwestern parts of DeKalb county (one of the two counties where the population density is highest for the Atlanta MSA) and specifically, neighborhoods where the rate of ‘overcrowding’ is greater than 50%. In comparison, in areas where CA-MSSA SIRs are highest (northcentral DeKalb county), the rates of CA-MRSA SIR are still

relatively high. The highest SIRs are in areas not far from either hospital, where the patients sought care. However, the overall distribution and the shape of the two highest SIRs are quite distinct between those diagnosed with CA-MRSA compared to those with CA-MSSA (Fig. 3).

**Age, race, household crowding spatial relationship to CA-MRSA infections**

We found statistically significant measures of positive association between the spatial filter-level proportion of the *S. aureus* cases that were CA-MRSA and the age ( $r = 0.137, p < 0.001$ ), race ( $r = 0.517, p < 0.001$ ), and household crowding covariates ( $r = 0.320, p < 0.001$ ). These associations indicate that CA-MRSA is elevated in areas where a high proportion of children are younger than age 5 years, where a high proportion of the population are black, and where a high proportion of the households are crowded. These findings are corroborated in our spatial statistical models (Fig. 4). The geographic boundary correlation with race makeup of a geographic area among children with CA-MRSA infections has not been previously reported; previous spatial analyses of CA-MRSA have focused primarily with MRSA colonization of environment or household members, or the hospital’s physical boundaries [52, 53].

**Discussion**

Our goal for this spatial analysis was to use electronic health record data to determine which block group-level neighborhoods were most ‘at risk’ for CA-MRSA infections and therefore, would benefit most from targeted primary interventions. We found that the highest proportions of CA-MRSA infections occurred within areas that have the highest levels of household crowding, highest proportion of younger children, and highest proportion of blacks. Although there is some geographic overlap in the CA-MRSA and CA-MSSA densities, the geographic patterns of CA-MRSA were different from that of CA-MSSA even though both types presumably have transmission patterns following that of infectious spatial diffusion. This suggests that the factors driving risk for antibiotic resistant disease (specifically CA-MRSA) may be quite different than those which drive risk for

**Table 3** Neighborhood-Level Factors for CA-MRSA Compared to CA-MSSA Infection for Crude and Adjusted Odds Ratios

Variable	Crude OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
<b>Household Crowding</b>				
≤ 1 person/room	Referent	–	Referent	–
> 1 person/room	1.22 (1.10, 1.34)	<.0001	1.03 (.92,1.14)	.6543
<b>Black Population Concentration</b>				
< 40.46%	Referent	–	Referent	–
≥ 40.46%	2.16 (1.93, 2.41)	<.0001	2.01 (1.79, 2.27)	<.0001
% Below Poverty	1.014 (1.011,1.018)	<.0001	1.006 (1.002, 1.010)	.0033

**Table 4** Multi-factor Model for CA-MRSA Compared to CA-MSSA for Adjusted Odds Ratios

Variable	Adjusted OR (95% CI)	p-value
Race		
White	Referent	–
Black	1.68 (1.45, 1.94)	<.0001
Hispanic	.81 (.62, 1.05)	0.1169
Other	.67 (.55, .81)	<.0001
Gender		
Male	Referent	–
Female	1.27 (1.15, 1.41)	<.0001
Age (years)		
> 12 years	Referent	–
4–12 years	.93 (.81, 1.07)	.0260
< 4 years	2.06 (1.79, 2.36)	<.0001
Health Insurance		
Private	Referent	–
Self pay	1.72 (1.33, 2.23)	<.0001
Public	1.51 (1.35, 1.70)	<.0001
Black Concentration		
< 40.46%	Referent	–
≥ 40.46%	1.30 (1.13, 1.49)	.0003

non-antibiotic resistant CA-MSSA. This finding suggest that ‘overcrowding’ is a risk factor for CA-MRSA disease risk.

Fulton and DeKalb counties are very densely populated, and like many core urban areas, have areas with high rates of poverty. ‘Crowding’ and poverty may be factors which mediate antibiotic resistance risk by creating geographically closed communities with more substandard housing conditions and thus, increased opportunities for person-to-person transmission. Over 75% of our CA-MRSA cases came from two of the most densely populated counties in Georgia (Fulton and DeKalb counties). Black residents in these counties also have been described as racially segregated from other populations [54], and this segregation may preclude them from social and economic advantages available to other racial/ethnic groups which are not as segregated.

Our finding of CA-MRSA association with poverty is consistent with other published reports, where higher rates of CA-MRSA infection are identified among urban neighborhoods with higher poverty rates [4]. After adjusting for their individual sociodemographic characteristics, these results show no significant racial differences among the cases (CA-MRSA) and controls (CA-MSSA). This confirms that it is the socio-economic and environmental characteristics of the neighborhood that explain the geographic distribution of MRSA. From a prevention standpoint, providing

resources, which otherwise would not exist, to these neighborhoods might be a first step towards meaningful community level intervention.

One strength of this study is the use of a large multi-year dataset of geocoded EHR data to detect spatial patterns in the proportion of CA-MRSA. Another strength is our use of aggregated block group data to create spatially adaptive filters, which results in maps where: 1) all locations have approximately equally reliable estimates of disease risk, and 2) the CA-MRSA and CA-MSSA measures were computed within the same set of spatial filters. We also calculated the covariates using the same spatial filter areas used to estimate the risk of both types of *S. aureus*, which enabled us to measure the association between disease risk and the covariates.

#### Limitations

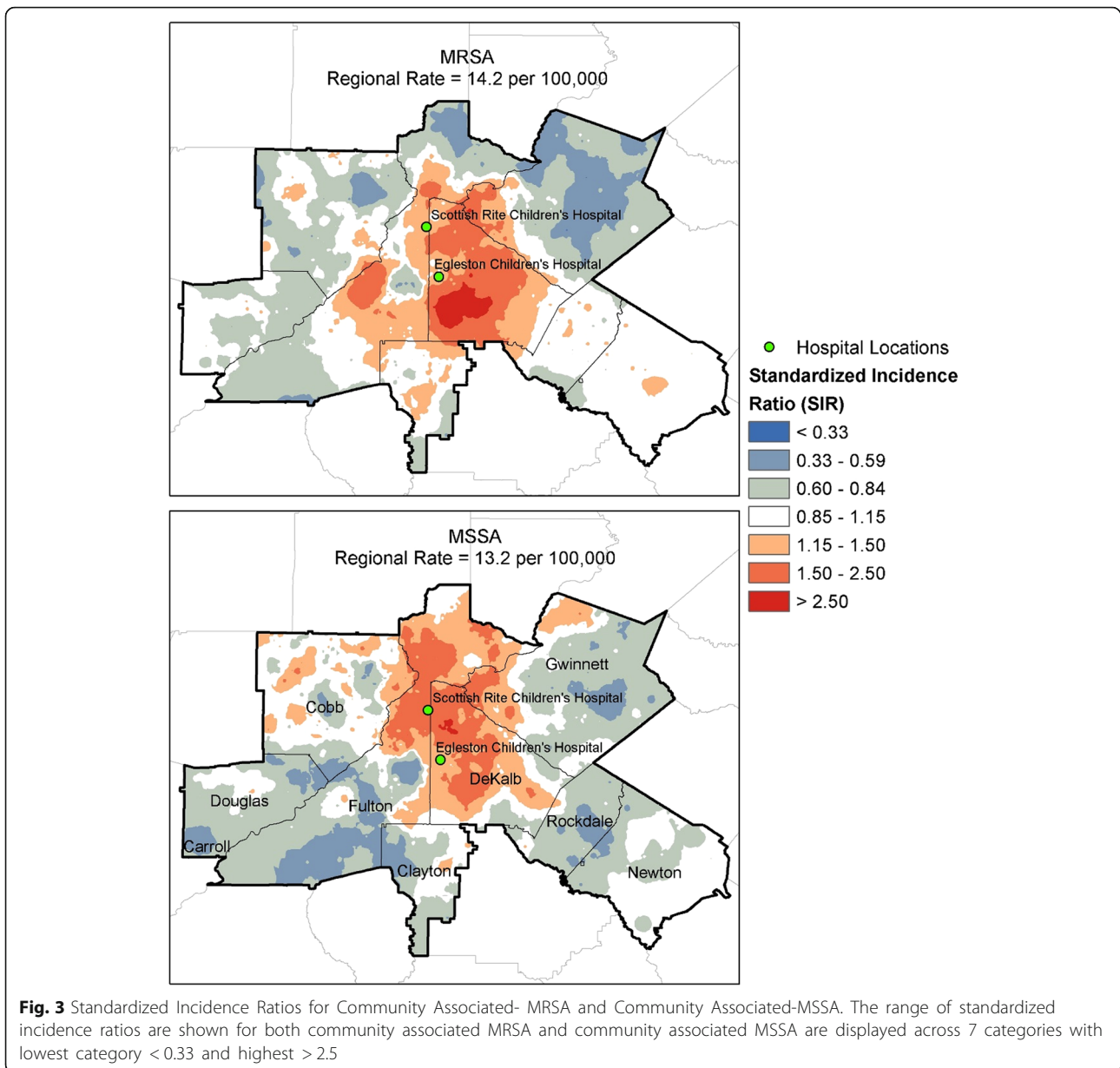
Similar to the limitations associated with using administrative health records [55, 56], the primary limitation of this study is that children who sought care and acquired CA-MRSA from the two hospitals may not be representative of the entire population of children in need of care. For this study, we selected children who accessed care from two pediatric hospitals which represents more than 80% of pediatric admissions in the Atlanta MSA. Although these two hospitals are major healthcare providers for the pediatric population residing within the MSA of Atlanta, this population may not reflect all children infected by *S. aureus* for the geographic region studied. The catchment area of these hospitals may not extend well beyond metropolitan Atlanta counties; and thus, cases closer to other large hospitals may be under reported. Another source of under-reported cases is related to transportation concerns, which may cause some parents to take their child to the nearest hospital rather than a pediatric hospital. Interventions developed would need to factor in the health behaviors characteristic of the populations living in a targeted community.

Another limitation is related to the geospatial correlation analysis of the proportion of *S. aureus* that were CA-MRSA and the covariates. Given spatial dependence inherent in geographic data, which is magnified when using the overlapping spatially adaptive filters, we have violated a central assumption of Pearson’s R correlation that observations be independent. Even though we detected statistically significant associations, the correlation that we measured is higher than what would be observed if the data was not overlapping. Future research using spatially adaptive filters to conduct correlation analysis needs to account for this inflation.

#### Future research

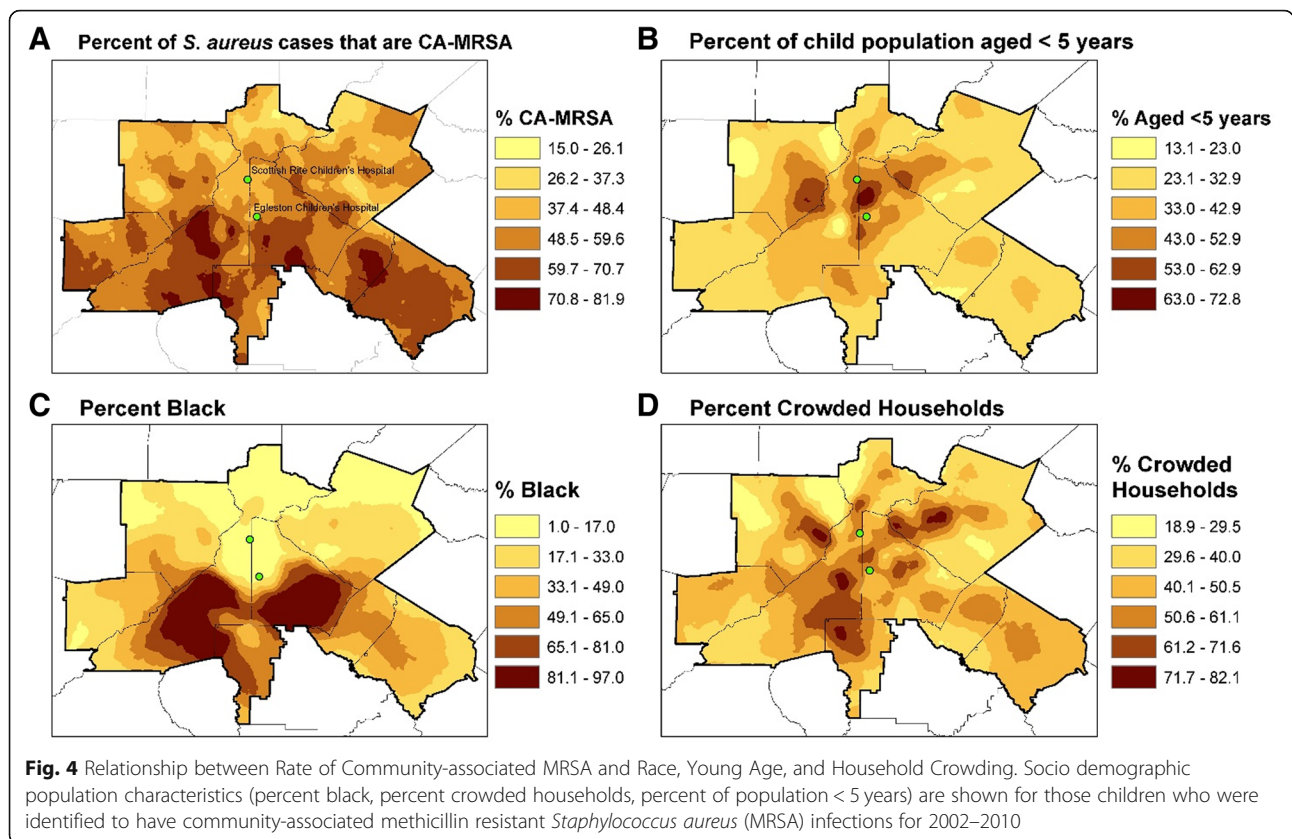
The development of targeted community level interventions could be supported by creating community specific





reports for the areas with the highest risk of CA-MRSA by characterizing the community according to its unique combination factors known to be associated with antibiotic resistant bacterial transmission. These include neighborhood-level measures of racial composition, household crowding, and age of children as well as information about its built environmental and other social factors. Such a report could be supported using spatial regression modeling to identify where CA-MRSA risk is most associated with these covariates [57]. For example, if types of health insurance and easier access to broad spectrum antibiotics are identified as significantly contributing to CA-MRSA (relative to CA-MSSA) and poor,

overcrowded neighborhoods are shown along with their locations of pharmacies and distribution of household income and poverty levels, then at the neighborhood level, we could come up with neighborhood-level (public health) strategies for prevention. Although this study did not address the relationship between CA-MRSA carriage and infection, future studies may include prospectively screening for persistence of carriage among household members and contaminated environmental surfaces associated with patients with recurring infection. Finding ways to decrease bacterial burden on a continuum in these communities may also contribute to an effective prevention strategy. Finally, refining how to spatially ‘weigh’ in socio-



and ecological factors, based on geographic place, could combat the spread of antibiotic resistant infections.

**Conclusions**

Hospital-based electronic health data systems in conjunction with spatial analyses can provide an effective surveillance system, which could be used to develop strategies for preventing CA-MRSA transmission. This study demonstrates how geocoded EHR data can be used to identify areas of excess risk for *S. aureus* infections which is important for developing interventions to prevent the spread of antibiotic resistant infectious conditions. These methods can be used to identify specific areas to target public health intervention strategies. Such methods have a tremendous potential to prevent the transmission of antibiotic resistant and virulent pathogens in our communities and neighborhoods. While the findings are particular to Atlanta, the methods can be applied globally, where electronically recorded data about *S. aureus* is available.

**Abbreviations**

aOR: adjusted odds ratio; CA-MRSA: Community-associated methicillin resistant *Staphylococcus aureus*; ED: Emergency department; GIS: Geographic information systems; MRSA: Methicillin resistant *Staphylococcus aureus*; MSA: Metropolitan Statistical Area; MSSA: Methicillin sensitive *Staphylococcus aureus*; OR: Odds ratio

**Acknowledgements**

We would like to extend gratitude to Dr. Monica Farley (Emory University School of Medicine), Dr. Lance Waller (Emory University Rollins School of Public Health), Dr. Shun Zhang (Morehouse School of Medicine, National Center for Primary Care), and Dr. Alexander Quarshie (Morehouse School of Medicine) for their guidance and review of manuscript. We appreciate and acknowledge the copy editing provided by Ms. Angela Wimes (Morehouse School of Medicine). We also appreciate the edits and manuscript feedback provided by Dr. Peter Baltrus (Morehouse School of Medicine, National Center for Primary Care), Dr. Anne Gaglioti (Morehouse School of Medicine, National Center for Primary Care) and Ms. Anaam Mohammed (Pediatric Emergency Medicine Associates, LLC). We appreciate the technical assistance provided by Mr. Anthony Giarrusso and Mr. Yun Sang from Georgia Institute of Technology's Center for GIS. We appreciate the help from Dr. Ling Liu (Georgia Institute of Technology's College of Computing) in data mining; Ms. Althea Lewis, and Mr. Alan Nadler from the Children's Healthcare of Atlanta's (CHOA) Clinical Microbiology Department, and Mr. Robert Massey (CHOA's Information & Technology Department) for electronic health record data acquisition and clean up.

**Funding**

Research reported in this publication was supported in part by funds received from Georgia Tech's Health Institute, Children's Healthcare of Atlanta Friends' Fund; PHS Grant UL1TR002378 from the Clinical and Translational Science Award Program, National Institute of Health, as part of the Georgia Clinical & Translational Science Alliance (Georgia CTSA) (formerly, UL1RR025008, Atlanta Clinical and Translational Science Institute (ACTSI)); Grant Number G12-RR03034, a component of the National Institute of Health; Grant Number HS024338-01, K-08 Mentored Clinical Scientist Award, Agency for Healthcare Research & Quality.

**Availability of data and materials**

The data that support the findings of this study are available from Children's Healthcare of Atlanta and Grady Health Center, but restrictions apply to the

availability of these data, which were used under license for the current study and contain personal identifiers, and so are not publicly available. Data, which is de-identified of personal information are however available from the authors upon reasonable request and with permission of Children's Healthcare of Atlanta and Grady Health Center.

#### Authors' contributions

LCI, RJ, GSR made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; LCI, GSR, TL, KM1 (Khusdeep Malhotra) and all authors (LCI, TL, KM1, TCP, FA, RJ, GSR) have given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content; and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

#### Authors' information

LCI is a general pediatrician and pediatric infectious disease specialist who has been caring for children with MRSA infections for over two decades. This work reflects data collected over the past 5 years and is part of a K-08 Mentored Clinical Scientist Development Award she received. (Dr. George S. Rust and Dr. Lance Waller are her mentors for this project).

#### Ethics approval and consent to participate

This research has been approved by the Institutional Review Boards of Morehouse School of Medicine, Children's Healthcare of Atlanta, and Grady Health Center.

#### Consent for publication

This research is based on electronic health records collected on patients as part of standard of care, and therefore, no individual patients were directly contacted for the data. Not applicable.

#### Competing interests

**CDC Disclaimer:** The findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the Centers for Disease Control and Prevention or the Centers for Medicare and Medicaid Services.

The authors have indicated they have no competing interests to disclose.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Department of Microbiology/Biochemistry/Immunology, Department of Pediatrics and Clinical Research Center, Morehouse School of Medicine, 720 Westview Drive, SW, Atlanta, GA 30310, USA. <sup>2</sup>Children's Healthcare of Atlanta, 1405 Clifton Road NE, Atlanta, GA 30322, USA. <sup>3</sup>Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA 30322, USA. <sup>4</sup>National Center for Primary Care, Morehouse School of Medicine, 720 Westview Drive, SW, Atlanta, GA 30310, USA. <sup>5</sup>Department of Pathology, Emory University, 1364 Clifton Road Northeast, Atlanta, GA 30322, USA. <sup>6</sup>Florida State University College of Medicine, 1115 W. Call St, Tallahassee, FL 32306, USA.

Received: 22 October 2018 Accepted: 4 January 2019

Published online: 18 February 2019

#### References

- Bazemore AW, Cottrell EK, Gold R, et al. "Community vital signs": incorporating geocoded social determinants into electronic records to promote patient and population health. *J Am Med Inform Assoc.* 2016;23:407–12.
- Gustafsson E, Ringberg H, Johansson P. MRSA in children from foreign countries adopted to Swedish families. *Acta Paediatr.* 2006;95:105–8.
- Tirabassi M, Wadie G, Moriarty K, et al. Geographic information system (GIS) localization of community acquired MRSA soft tissue abscesses. *J Pediatr Surg.* 2005;40:962–6.
- Hota B, Ellenbogen C, Hayden MK, Aroutcheva A, Rice T, Weinstein R. Community-associated methicillin resistant *Staphylococcus aureus* skin and soft tissue infections at a public hospital: do public housing and incarceration amplify transmission? *Arch Int Med.* 2007;167:1026–33.
- Sandora T. Prevention of healthcare associated infections in children: new strategies and success stories. *Curr Opin Infect Dis.* 2010;23:300–5.
- Klevens R, Morrison M, Nadle J, et al. Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA.* 2007;298:1763–71.
- Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. *Annual Rev Pub Health.* 2015;36:345–59.
- Miranda ML, Ferranti J, Strauss B, Neelon B, Califf RM. Geographic health information systems: a platform to support the 'triple aim'. *Health Aff.* 2013;32:1608–15.
- Blatt AJ. Geospatial Medicine. Health, Science, and Place. In: Doe RK, editor. *Geotechnologies and Environment Series*: Springer; 2015. p. 101–10.
- Bell CN, Bowie JV, Thorpe RJ Jr. Race and "hotspots" of preventable hospitalizations. *J Health Dispar Res Pract.* 2017;10:5.
- Westgard B, Warren D. 240 hotspots of triple aim opportunity: identifying geographic clusters of potentially preventable emergency department visits and costs. *Ann Emer Med Annals.* 2014;64:S85–S6.
- Horst MA, Coco AS. Observing the spread of common illnesses through a community: using geographic information systems (GIS) for surveillance. *J Amer Board Fam Med.* 2010;23:32–41.
- Laranjo L, Rodrigues D, Pereira AM, Ribeiro RT, Boavida JM. Use of electronic health records and geographic information systems in public health surveillance of type 2 diabetes: a feasibility study. *JMIR Pub Health Surv.* 2016;2:e12.
- Spratt SE, Batch BC, Davis LP, et al. Methods and initial findings from the Durham diabetes coalition: integrating geospatial health technology and community interventions to reduce death and disability. *J Clin Translat Endocrin.* 2015;2:26–36.
- Zhang K, McClure J, Elsayed S, Louie T, Conly J. Novel multiplex pcr assay for characterization and concomitant subtyping of staphylococcal cassette chromosome mec types I to V in methicillin-resistant *Staphylococcus aureus*. *J Clin Micro.* 2005;43:5026–33.
- Naimi T, LeDell K, Como-Sabetti K, et al. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. *JAMA.* 2003;290:2976–84.
- Fridkin S, Hageman J, Morrison M, et al. Methicillin-resistant *Staphylococcus aureus* disease in three communities. *New Engl J Med.* 2005;352:1436–44.
- Groom A, Wolsey D, Naimi T, et al. Community-acquired methicillin-resistant *Staphylococcus aureus* in a rural American Indian community. *JAMA.* 2001;286:1201–5.
- Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Micro Rev.* 2015;28:603–61.
- Limbago B, Fosheim G, Schoonover V, et al. Characterization of methicillin-resistant *Staphylococcus aureus* isolates collected in 2005 and 2006 from patients with invasive disease: a population-based analysis. *J Clin Micro.* 2005;47:1344–51.
- Fritz S, Garbutt J, Elward A, Shannon W, Storch G. Prevalence of and risk factors for community-acquired methicillin-resistant and methicillin-sensitive *Staphylococcus aureus* colonization in children seen in a practice-based research network. *Pediatr.* 2008;121:1090–8.
- Fritz SA, Epplin EK, Garbutt J, Storch GA. Skin infection in children colonized with community-associated methicillin-resistant *Staphylococcus aureus*. *J Inf Secur.* 2009;59:394–401.
- Hota B, Lyles R, Rim J, et al. Predictors of clinical virulence in community-onset methicillin-resistant *Staphylococcus aureus* infections: the importance of USA300 and pneumonia. *Clin Infect Dis.* 2011;53:757–65.
- Landrum M, Neumann C, Cook C, et al. Epidemiology of *Staphylococcus aureus* blood and skin and soft tissue infections in the US military health system, 2005–2010. *JAMA.* 2012;308:50–9.
- Georgia Hospital Association Report-Discharge Data -Market Share. <https://www.gha.org/GDDS>. Accessed 22 Dec 2014.
- Gorwitz R, Jernigan D, Powers J, Jernigan J. Strategies for clinical management of MRSA in the community: summary of an experts' meeting convened by the Centers for Disease Control and Prevention. Atlanta: Centers for Disease Control and Prevention; 2006. p. 1–23.
- Miller L, Eells S, Taylor A, et al. *Staphylococcus aureus* colonization among household contacts of patients with skin infections: risk factors, strain discordance, and complex ecology. *Clin Infect Dis.* 2012;54:1523–35.

28. Institute Clinical and Laboratory Standards. Performance standards for antimicrobial susceptibility testing. Wayne: Clinical Laboratory Standards Institute; 2008.
29. Jones-Webb R, Wall M. Neighborhood Racial/Ethnic Concentration, Social Disadvantage, and Homicide Risk: An Ecological Analysis of 10 U.S. Cities. *J Urb Health*. 2008;85:662–75.
30. Tsai PJ, Lin ML, Chu CM, Perng CH. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*. 2009;9:464.
31. Cohen SS, Sonderman JS, Mumma MT, Signorello LB, Blot WJ. Individual and neighborhood-level socioeconomic characteristics in relation to smoking prevalence among black and white adults in the southeastern United States: a cross-sectional study. *BMC Pub Health*. 2011;11:877.
32. US Census Bureau. Census 2010. Summary File 1. DP-1. Profile of General Demographic Characteristics: 2010 by block group, "Georgia". American FactFinder. Washington D.C.: U.S. Census Bureau; 2010.
33. Manson S, Schroeder J, Riper DV, Ruggles S. IPUMS National Historical Geographic Information System. In: Minnesota Univ, ed. version 12.0 ed. Minneapolis: University of Minnesota; 2017.
34. State of Arizona Consolidated Plan 2010–2014. 2010. (Accessed 22 Aug 2012, at [https://housing.az.gov/sites/default/files/documents/files/2010-2014\\_Con\\_Plan.pdf](https://housing.az.gov/sites/default/files/documents/files/2010-2014_Con_Plan.pdf).)
35. US Census Bureau. Census 2010. American Community Survey 1-Year Estimates, Table; using American FactFinder: <<http://factfinder2.census.gov>>. Accessed 22 Aug 2012.
36. US Census Bureau. Census 2000. Poverty thresholds for 2000 by size of family and number of related children under 18 years. Washington D.C.
37. Banthin J, Selden T. Income measurement in the medical expenditure panel survey. Rockville: Agency for Healthcare Research and Quality; 2006.
38. Brewer C. Basic mapping principles for visualizing Cancer data using geographic information systems (GIS). *Am J Prev Med*. 2006;30:S25–36.
39. Myers D, Baer W, Choi S. The changing problem of overcrowded housing. *J Am Plan Assoc*. 1996;1996:6266–84.
40. Krieger N, Chen J, Waterman P, Soobader M, Subramanian S, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter. *Am J Epid*. 2002;156:471–82.
41. Talbot TO, Kuldorff M, Forand SP, Haley VB. Evaluation of spatial filters to create smoothed maps of health data. *Stat Med*. 2000;19:2399–408.
42. Choynowski M. Maps based on probabilities. *J Amer Stat Assoc*. 1959; 54:385–8.
43. Matthews K. Spatiotemporal correlation analysis of colorectal cancer late-stage incidence, mortality, and survival: Iowa, 1999 to 2010: University of Iowa; 2018. p. 15–29.
44. Pickle LW, Mungiole M, Jones GK, White AA. Exploring spatial patterns of mortality: the new atlas of United States mortality. *Stat Med*. 1999;18: 3211–20.
45. Kafadar K. Smoothing geographical data, particularly rates of disease. *Stat Med*. 1996;15:2539–60.
46. Mungiole M, Pickle LW, Simonson KH. Application of a weighted head-banging algorithm to mortality data maps. *Stat Med*. 1999;18:3201–9.
47. Langford IH. Using empirical Bayes estimates in the geographical analysis of disease risk. *Area*. 1994;142–9.
48. Casper M, Kramer MR, Quick H, Schieb LJ, Vaughan AS, Greer S. Changes in the geographic patterns of heart disease mortality in the United States: 1973 to 2010. *Circulat*. 2016;133:1171–80.
49. Vaughan AS, Quick H, Pathak EB, Kramer MR, Casper M. Disparities in temporal and geographic patterns of declining heart disease mortality by race and sex in the United States, 1973–2010. *J Amer Heart Assoc*. 2015;4: e002567.
50. Cai Q, Rushton G, Bhaduri B. Validation tests of an improved kernel density estimation method for identifying disease clusters. *J Geograph Sys*. 2012;14: 243–64.
51. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance (ABCs) Report Emerging Infections Program Network Methicillin-Resistant *Staphylococcus aureus*, 2008. Atlanta: Centers for Disease Control and Prevention. p. 2008.
52. Rossini CJ, Moriarty KP, Tashjian DB, Garb JL, Wait RB. Geographic distribution of community-acquired methicillin-resistant *Staphylococcus aureus* soft tissue infections. *J Pediatr Surg*. 2011;46:1089–92.
53. Parrish KL, Hogan PG, Clemons AA 2nd, Fritz SA. Spatial relationships among public places frequented by families plagued by methicillin-resistant *Staphylococcus aureus*. *BMC Res Notes*. 2018;11:692.
54. Hayes MM. The building blocks of Atlanta: racial residential segregation and neighborhood inequity. Atlanta: Georgia State University; 2006.
55. Strauss BW, Valentiner EM, Bhattacharya S, et al. Improving population representation through geographic health information systems: mapping the MURDOCK study. *Amer J Translat Res*. 2014;6:402.
56. Dixon BE, Gibson PJ, Comer KF, Rosenman M. Measuring population health using electronic health records: exploring biases and representativeness in a community health information exchange. 15th World Congress on Health and Biomedical Informatics, MEDINFO 2015: IOS Press; 2015.
57. Charlton M, Fotheringham S, Brunsdon C. Geographically weighted regression. UK: Economic and Social Research Council.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

