Accepted Manuscript

Advances in spatial epidemiology and geographic information systems

Russell S. Kirby, PhD, MS, Eric Delmelle, PhD, Jan M. Eberth, PhD, MSPH

PII: S1047-2797(16)30495-1

DOI: 10.1016/j.annepidem.2016.12.001

Reference: AEP 8050

To appear in: *Annals of Epidemiology*

Received Date: 19 November 2016

Revised Date: 30 November 2016

Accepted Date: 4 December 2016

Please cite this article as: Kirby RS, Delmelle E, Eberth JM, Advances in spatial epidemiology and geographic information systems, *Annals of Epidemiology* (2017), doi: 10.1016/j.annepidem.2016.12.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Advances in spatial epidemiology and geographic information systems

Russell S. Kirby, PhD, MS[1]; Eric Delmelle, PhD[2]; Jan M. Eberth, PhD, MSPH[3]

1 Department of Community and Family Health, University of South Florida, Tampa, FL

2 Department of Geography and Earth Sciences, University of North Carolina-Charlotte, Charlotte, NC

3 Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC

Corresponding Author Information:
Russell S. Kirby, PhD, MS, FACE
Department of Community and Family Health
College of Public Health, University of South Florida
13201 Bruce B. Downs Blvd, MDC56
Tampa FL 33612
direct line: 813-396-2347
email: rkirby@health.usf.edu

Keywords: Spatial Epidemiology, Geographic Information Systems (GIS), Geocoding, Geographic Mapping, Health Services Accessibility, Spatial Regression, Spatial Analysis, Spatiotemporal Analysis

Conflict of Interest Statement: We confirm that this manuscript has not been previously published, either in whole or in part, nor have the findings been posted online, and is not under consideration for publication elsewhere. All authors have approved the manuscript and agree with submission to the Annals of Epidemiology.

Financial Disclosure: The authors have no financial or other material conflicts of interests with might be considered influential to the writing of this manuscript. Drs. Kirby and Eberth are Associate Editors for the Annals of Epidemiology, and paid an honorarium for their editorial services by Elsevier. Dr. Eberth is also supported in part by grants from the Federal Office of Rural Health Policy under a Cooperative Agreement with the South Carolina Rural Health Research Center (5U1CRH0311-12-00) and the American Cancer Society (MRSG-15-148-01-CPHPS).

ACCEPTED MANUSCRIPT

Abstract

The field of spatial epidemiology has evolved rapidly in the past two decades. This paper serves as a brief introduction to spatial epidemiology and the use of geographic information systems (GIS) in applied research in epidemiology. We highlight technical developments, and highlight opportunities to apply spatial analytic methods in epidemiologic research, focusing on methodologies involving geocoding, distance estimation, residential mobility, record linkage and data integration, spatial and spatio-temporal clustering, small area estimation, and Bayesian applications to disease mapping. The articles included in this issue incorporate many of these methods into their study designs and analytical frameworks. It is our hope that these papers will spur further development and utilization of spatial analysis and GIS in epidemiologic research.

Introduction

Defining spatial epidemiology

All students of epidemiology learn that descriptive epidemiology focuses on the triad of person, place and time [1]. While epidemiologic research focusing on place or location historically received considerably less attention, modern epidemiology increasingly incorporates the spatial perspective into research designs and models. Spatial factors have also become prominent features in etiologic research, especially concerning host-vector-agent interactions, but also in guiding social and environmental epidemiologic investigations. Spatial methods are also progressively incorporated into health services research focused on specific diseases, health conditions or risk factors.

The field of spatial epidemiology has evolved over the years. Elliot et al. [2] identified four types of spatial analyses in epidemiology: 1) disease mapping, 2) geographical correlation studies, 3) risk assessment in relation to point or line sources, and 4) cluster detection and disease clustering. Only a few years earlier, English [3] had defined geographical epidemiology as "the description of spatial patterns of disease incidence and mortality". More recently, in a widely used spatial statistics text, Lawson [4] states that spatial epidemiology "concerns the analysis of the spatial/geographical distribution of the incidence of disease". Here, we argue that spatial epidemiology encompasses research that incorporates the spatial perspective into the design and analysis of the distribution, determinants, and outcomes of all aspects of health and well-being across the continuum from prevention to treatment.

Spatial epidemiology is not synonymous with health/medical geography. Spatial epidemiology refers to inquiries that use epidemiologic study designs that involve spatial data or spatially-derived information about study subjects, health facilities or sources of exposure. Health or medical geography, a subdiscipline of human geography, encompasses research applying geographic analytical methods to health, disease or health care issues. Its distinguishing feature is the primary focus on spatial patterns and context, while spatial epidemiology is inherently focused on populations [5]. While many studies, especially those involving interprofessional research teams, combine the methods of both disciplines in creative and innovative ways, far more often spatial epidemiologic research does not fully incorporate the geographic perspective, and vice versa [6].

In this brief overview, we argue for an expanded role for spatial epidemiology within our discipline, and demonstrate the importance of a broader scope for spatial perspectives in the

study of epidemiology through the series of articles that follow in this symposium on "Geographic Information Systems and Spatial Methods in Epidemiology".

The concepts of place and neighborhood

Health outcomes are influenced by an interplay of different factors, such as individual attributes, the physical and social environments an individual interacts with, cultural norms and both the provision and utilization of health services [7]. It is widely recognized that the place where an individual lives or works should be considered as a potential disease determinant [8]. For example, women living in rural areas may have to travel longer distances to reach mammography facilities, potentially leading to a decrease of mammographic breast screening or a delay in their diagnosis [9-11]. Children living in a pedestrian friendly environment where parks and playgrounds are readily accessible are more likely to engage in physical activity, reducing odds of obesity [12, 13]. Along the same vein, higher risk factors for obesity are generally observed in food desert areas, which are characterized by poor quality food environments and a lack of supply of supermarkets with fresh food [14, 15].

Residents located in proximity of major traffic corridors are exposed to particulate matter and diesel exhausts, causing a variety of respiratory and cardiovascular diseases [16]. Individuals residing in high-crime neighborhoods may indirectly develop stress-related behaviors, such as anxiety and higher blood pressure [17, 18]. These instances illustrate the breadth of pathways through which both physical and social environments, as well as provision of health services, give rise to health disparities. Documenting the role of the geographic environment where individuals live and interact (often called "activity spaces") will improve our understanding of health outcomes. This has deep policy implications for local health interventions and resource allocation decisions, ultimately leading to a reduction of health disparities.

A neighborhood is typically defined as the geographic area relevant to the specific health outcome being studied [8]. However, neighborhoods can be delineated by the extent of the individual's spatial interaction, or administratively delimited at the scale at which policies are implemented (county, state). Secondary datasets -such as the American Community Survey published by the U.S. Census Bureau- help to overcome the absence of socioeconomic data in most U.S. medical records [19]. The scale at which the primary or secondary datasets are available often dictates the spatial granularity at which the analysis is conducted [20]. Spatial analyses with individual level data from public health databases are also often limited by privacy policies required by data managers.

Patient information may be aggregated at the ZIP Code level to comply with the Health Insurance Portability and Accountability Act (HIPAA) [21]. As such, uniformity is assumed within the unit of analysis, but sharp contrasts may occur among adjacent units. Further, the aggregation of neighborhoods does raise the issue of the *modifiable areal unit problem* (MAUP); using different boundaries may lead to significantly different analytical results [22]. As a rule of thumb, analyses should be conducted at different levels of granularity to test the robustness of the spatial relationships and the effect of different artificial boundaries.

What defines a neighborhood and the concept of scale (aggregated, disaggregated) will influence the choice of methodologies and ultimately impact the results. Spatially-based regression, contextual and multi-level modeling are some of the key methods developed to incorporate neighborhood effects [8, 23]. These approaches allow researchers to estimate the impact of neighborhood effects after controlling for individual characteristics.

Role of geographic information systems in spatial epidemiology

Defined by Cromley and McLafferty [24] as "computer-based systems for the integration and analysis of geographic data," geographic information systems (GIS) can describe, analyze, and predict patterns using feature (cartographic) and attribute data. GIS have been used in many epidemiologic applications, including disease mapping, rate smoothing, cluster or hot spot analysis, and spatial modelling. In its simplest form, GIS is often employed to create spatially-explicit variables such as availability and accessibility scores (e.g., food access), built environment measures (e.g., land use), environmental exposures (e.g., air pollutant concentrations), and demographic indicators (e.g., percent of persons in poverty). Measuring and describing the extent of spatial relationships is also a key function of GIS, which can be as simple as calculating the distance between two points or as complex as quantifying spatial dependencies in analytic models or identifying locally-varying predictors. As described by Thornton et al. [25] GIS offers opportunities to integrate data across multiple databases and spatial scales for display, management, and analysis. In Table 1, we identify several ways to apply GIS to human immunodeficiency virus (HIV).

Table 1. Functions of GIS and related epidemiologic applications in HIV research

GIS Functions	**Application to HIV Research**
Store and measure spatial relationships	Distance between homes of newly diagnosed HIV cases and the closest Ryan White Clinic
Display spatial relationships	Bivariate map showing the relationship between county HIV prevalence and poverty rate
Analyze attribute and feature data	Count of the number of HIV cases within 30 miles of each Ryan White Clinic

simultaneously	
Manage data from multiple sources	Create a geodatabase with county HIV incidence and prevalence rates, Ryan White Clinic locations, community-based organizations providing HIV services, and county demographic indicators
Identify spatial patterns	Conduct a hot spot analysis of newly diagnosed HIV cases in U.S. counties
Explain spatial patterns	Determine where the prevalence of men who have sex with men (MSM) is associated with the incidence rate for HIV in U.S. counties using GWR

Footnote: A variety of studies have used GIS approaches in the context of HIV patient and service provider data [26-32].

In descriptive epidemiology, thematic or color-shaded maps produced with GIS are useful for identifying areas at high risk for epidemics, highlighting population health disparities, examining resource needs, and ultimately, formulating hypotheses that lead to generation of explanatory models. Analytically, GIS tools can be used to explore spatial or spatiotemporal clustering, investigate locally-varying relationships, and explicitly model or adjust for spatial dependencies in one's data.

Just as research has shown that interventions are most effective when implemented at multiple levels [33], we can gain more insight into individual-level outcomes by considering the context in which people live and work [8]. In the context of healthcare, it is equally important to consider the location and characteristics of where people seek their care (i.e., provider- and system-level factors). Contextual information can be considered in both descriptive and analytic ways. For example, one could explore the distribution of greenspace or recreational facilities within a region, as well develop a greenspace index to predict the probability of meeting physical activity guidelines or individual weight status. When both individual and area-level data are available, hierarchical models can be used to examine how outcomes vary at both the individual

and/or group level (see example from Mobley et al., this issue [34]), as well as what factors outside an individual's immediate control impact their health and wellbeing. Generalized estimating equations, which also account for the correlation of persons within groups, treat this group-level effect as a nuisance that simply needs to be adjusted for in one's analysis. As epidemiologists, we need to be cautious about using methods to adjust away potentially meaningful information about the role of "place" – whether operationalized at the workplace, school, or residential level, on our health.

Study Design Perspectives

Cross-sectional studies typically examine the distribution of exposures and outcomes simultaneously, frequently through use of population-based administrative health data or representative sample surveys. Studies using cross-sectional designs rarely integrate the spatial perspective directly, because spatial referents for each observation are not collected or are at a spatial resolution that does not permit localized investigations. For example, the National Health Interview Survey includes a data element for broad U.S. Census regions (East, South, Midwest, West) [35], but nothing more spatially specific, while the public use version of the National Survey of Children's Health has a data element for state of residence and another for residence in an Metropolitan Statistical Area (MSA) [36], but these do not permit spatial analyses within a state. Similarly, surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) [37] will often mask county-level information due to small sample sizes. However, some states such as Florida routinely oversample to support ZIP Code-level analyses every third year. Ecological studies are a form of cross-sectional study design in which the unit of analysis is grouped, by political unit (e.g. nation, state, county, ZIP Code, census tract), health facility, school, or other organizational unit. Most ecological studies have the potential to incorporate the spatial

perspective into their study designs and analyses. However, while researchers frequently map data from ecological studies using administrative or political units, the statistical analyses performed are typically aspatial (i.e., do not account for similarity or clustering between neighboring units). With tools such as spatial smoothing techniques, spatial regression, and multi-level modeling readily available to epidemiologists, more sophisticated analyses that take into account the spatial dependence present in many datasets should be considered.

Case-control study designs compare exposure histories of persons with a disease or health condition (cases) with persons who do not have that disease or health condition but are representative of the population from which cases are identified (controls) [1]. In general, the case-control study does not lend itself well to spatial epidemiologic analysis, unless cases and controls are selected from a population-based sampling frame that incorporates location into its design. However, there are several areas where the use of GIS tools and spatial statistical methods can enhance case-control studies. These include estimation of measures of access (e.g. distance to a health facility, travel times), neighborhood amenities, and other local estimates derived from spatial surfaces (e.g., food swamps and deserts, residential segregation, inequalities in income and wealth or measures of neighborhood deprivation). If addresses are available for all cases and controls, these can be geocoded (see Section 3), and linked to Census level socio-demographic variables or environmental data (e.g. air pollutants, water quality, remotely sensed data on land use). For example, Lupo et al. found that neighborhood deprivation, measured at the census tract level, was associated with presence of cleft lip (with or without cleft palate) among Texas residents in a birth defects registry [38]. If it is feasible to incorporate collection of spatially-specific identifiers into case-control study designs, these studies may become more informative.

Cohort studies follow study subjects over time, comparing those with and without exposures of interest to determine their outcomes (disease, mortality) [1]. These study designs can take many forms, and utilize a variety of sophisticated biostatistical techniques. As with cross-sectional and case-control studies, epidemiologists typically do not incorporate spatial analysis directly into their study designs. Cohort study data can be integrated with Census, environmental, and other administrative-level data when geocoded to specific addresses or administrative units (e.g., Zip Code centroids). More often than not, these studies are spatially-enabled, in that the researcher uses spatial referents to acquire additional data to associate with each study subject, but do not incorporate spatial analysis into their study designs. A common use of GIS-derived data in cohort studies is the use of environmental data such as air pollution and other environmental contaminants on health outcomes such as mortality, cancer, or respiratory illnesses [39, 40]. Additionally, studies have linked changes in the neighborhood food environment with body mass composition in prospective studies [41-43].

Methodologies

Geocoding and uncertainty

Locational information, for instance in the form of addresses, can be transformed into geographic coordinates through a process known as geocoding [44]. The procedure requires that addresses from the epidemiological dataset are standardized and comprised of an address number, street name, city or town name, state and ZIP Code [45]. Using a reference dataset - typically a street database-, geographic coordinates can be estimated by comparing and interpolating the address to the range of addresses for each segment of the reference dataset. The

procedure is sensitive to the completeness of the addresses such as the presence of “P.O. Box” and on the quality of local and regional street road network files [46, 47].

The quality of the geocoding is evaluated based on the merits of three components [48]: the match rate (percentage of records being geocoded), the individual match score (how well the standardized address matches the street database) and the match type (e.g. geocoding at the street level or Zip Code). Low individual match score and low match type accuracy will increase positional errors, which are typically larger in rural areas. Such errors pose a serious challenge in spatial analysis, since it may result in (1) an underestimation of local risk, (2) the misplacement of high-risk areas of a disease, (3) a misevaluation of spatial association, and (4) biased evidence for decision makers.

Geocoding can be implemented using either commercial GIS software or online. Although the set-up costs to prepare the reference dataset and standardize the addresses can be expensive when using commercial GIS software, this approach allows the researcher to geocode large amount of records. Online geocoders are free and typically use more recent street network data, which is likely to result in lower positional errors if one’s data is also fairly recent (if working with older data, the opposite is true). However, online geocoders are often limited as to how many records can be processed and raise important issues of confidentiality, since addresses are uploaded online [48].

Due to privacy concerns such as HIPAA, health departments may release geocoded address information only in aggregate form, such as for five-digit ZIP Codes [24]. Alternatively, researchers can geomask the geocoded data to protect the privacy of individuals. Geomasking modifies the geographic coordinates of an individual event by displacing the original location to

some distance while maintaining spatial resolution for mapping purpose and cluster detection [49, 50].

Distance estimation

Geographic accessibility to a health service depends on its availability and the impedance to access that service. Greater travel impedance will affect both the provision and utilization of health services [51]. A low level of geographic accessibility may have dramatic consequences in terms of health outcomes. Due to recent advances in GIS, different metrics can be used to estimate travel impedance such as travel distance (Euclidean or network-based), travel time and travel cost [52, 53]. While estimating travel impedance by means of Euclidean distances is extremely straightforward, this approach can lead to serious underestimation of distances. When modeling travel, the ideal route between two points will be the one that minimizes the selected measure of impedance. Travel modeling should ideally incorporate speed limits, honor one-way restrictions, and reflect connectivity among roads - most of which can be specified in a GIS-based analysis. Travel time may be a more precise measure of impedance since it incorporates en-route conditions, such as congestion.

Online mapping providers such as GoogleMaps or MapQuest offer unprecedented support for estimating travel impedance. However, similar to geocoding services, the number of analytical queries that can be requested is usually limited, the accuracy of the travel estimates has not been extensively studied, and such approaches raise issues concerning data confidentiality. Through its Network Analyst extension, ArcGIS desktop software can produce distance and travel time estimates that account for road features with relative ease. Similarly, ArcGIS Pro can be used to calculate distance and travel time using ESRI's cloud-based road network data, although users are limited to no more than 1,000 origin-destination pairs.

While estimation of distance and/or travel time can now be done fairly simply, epidemiologists must consider which measures are most relevant to their population/region. The choice of origin and destination locations is also important, as variation in spatial scales, accuracy, and context can affect the results. For example, Khang et al. (this issue) [54] explores the probability of diagnostic resolution after an abnormal mammogram as a function of several operational definitions of distance: distance from the patient's residence to the screening facility, the diagnosing facility and the closest mammography facility. Similarly, in a comparative analysis of GIS-based measures of access to mammography, Lian et al. [55] found that travel time and facility density were poorly correlated with odds of late-stage breast cancer, while spatial accessibility scores that considered both supply and demand (i.e., two-step floating catchment area) were significantly associated with late-stage diagnosis.

Residential mobility

Epidemiologists often include neighborhood factors in models examining determinants of individual health outcomes. In a GIS, neighborhood characteristics are assigned to individuals based on their residential location, a process known as *spatial join*. Results from this process are particularly informative to establish whether individual disease conditions may be caused by exposure to hazardous environment [56].

While residential mobility is widely recognized in the research literature, typically only the attributes of the neighborhood where the individual resides at the time of diagnosis are incorporated into spatial epidemiologic analyses [57]. Failure to include longitudinal residential profiles has serious implications since the location where the case was diagnosed may not necessarily coincide with the place where exposure started or occurred. In a seminal study by Pershagen et al. [58], the authors assessed the residential history of all study subjects, basing

their exposure assessment on radon measurements at all dwellings in which the subject lived for 2 or more years over a period of 30+ years. The issue of residential histories is further exacerbated for (1) individuals with high mobility, (2) diseases with long periods of latency such as cancer and (3) places where the conditions of the environment may change rapidly [59].

In maternal and child health, most studies assume that residential mobility does not occur during pregnancy; as such the use of maternal residence at birth may result in exposure misclassification, biased estimates of association and imprecise risk estimation [60, 61]. In a study by Miller [62] in the metropolitan Atlanta region, approximately 22% of pregnant women changed their residential location from the date of conception to delivery [61], although most relocation distances were short and residential location remained generally within the same county. Exposures that are homogenous within a community may be well estimated with limited residential data, but this assumption may not hold for longer distances.

Although the U.S. Census Bureau publishes migration information between counties and the proportion of residents who lived in the same residence one year ago [63], the lack of data at a more disaggregated level may be one reason why most epidemiological do not incorporate residential histories. Hughes and Pruitt (this issue) [64] suggest that electronic medical records (EMR) derived address histories have the potential to alleviate this pitfall. More work is needed to develop spatiotemporal (cumulative) models of exposure [59].

Data linkages

Geocoding is a form of record linkage, in which address information is converted into latitude-longitude (X-Y) coordinates or administrative unit codes (e.g., Federal Information Processing Standard/FIPS codes in the U.S.). For example, Barnes et al. (this issue) converted the self-reported address of each survey respondent to an X-Y coordinate for the purpose of

measuring the presence, types of, and distance to food retailers around survey participants. More commonly however, researchers link data from various databases across spatial units using geocodes. In a recent example [65], the association between income inequality and infant mortality was examined, using U.S. states as units of analysis. While this study was aspatial, data from vital records were linked with socioeconomic data using state-level geocodes. Health services researchers and epidemiologists should be familiar with the Area Health Resources Files (AHRF) [66], providing hundreds of data fields with socio-economic, demographic and health services provider and utilization data by county. Similarly, epidemiologists frequently use geocodes for census units to access data from U.S. Census and associated American Community Survey databases. Increasingly, researchers also use geocodes to link individual records or census units surrounding their residences to environmental databases, using GIS technologies to integrate data on ambient air pollution, soil contamination, land use cover, temperature fluctuations, and other measures of phenomena in the physical environment with data on health behaviors and outcomes.

Overview of spatial and spatio-temporal clustering

According to McLafferty (2015) [67], a spatial cluster –also termed hotspot- can be defined as "…an unusual number of cases within a population, place and time period…". Practically, an epidemiologist may be interested to determine whether disease rates around a hazard site are elevated, or if the number of infected individuals is higher than would be expected. Accurate cluster information provides practical knowledge for public health interventions, such as screening, prevention and surveillance. A series of robust statistical approaches have been developed to detect spatial or spatiotemporal clusters of diseases.

According to Walter Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things [68]." This principle, otherwise known as *spatial autocorrelation* or dependence, is a key component of spatial epidemiology. Traditional generalized linear models that assume the independence of modeled observations will likely produce inaccurate estimates in the presence of spatial autocorrelation. It is generally desirable to estimate the amount of clustering using spatial search strategies, test its significance (for instance by means of Monte Carlo simulations) and finally visualize location of clusters along with their magnitude or extent. Global clustering tests evaluate the presence of clustering over the study area, while local clustering statistics tests for specific clusters at a finer scale. With the increasing technology and data available to incorporate contextual or locational data into our epidemiologic models, epidemiologists would be remiss to ignore the potential for spatial autocorrelation or non-stationarity in their data.

Spatial clustering

Evaluating the clustering of individual disease cases is typically implemented using a count statistic [69]. The *K*-function is a clustering statistic which estimates the magnitude of clustering at different scales [70]. It uses a circular or elliptical search window over each case to count how many other cases occur in the near proximity. The window is then moved to the next points and the process repeated. The entire procedure is conducted at different search windows. The scale at which clustering is the greatest is then determined and can serve as an input for the kernel density estimation technique, which is essentially a point density function helping to visualize the location of clusters on the map [69]. Note that variable-size, adaptive filters can be used instead of static search strategies [67]; this is particularly important when there is a

substantial amount of heterogeneity in the underlying population, such as the small numbers problem [71, 72].

For clustering of cases aggregated into spatial units, spatial autocorrelation-based methods such as *Moran's I* [73], tell us whether nearby units tend to exhibit similar rates. Spatial autocorrelation statistics measure the degree to which spatial features (e.g. Zip Codes, census tract) and their associated values tend to be clustered together in space or dispersed. *Moran's I* value range from -1 to +1, with a value of -1 denoting that units with low rates are located near other units with high rates, while a *Moran's I* value of +1 indicates a concentration of spatial units exhibiting similar rates. Although the *Moran's I* statistic informs on the presence of clustering, it does not indicate where such clusters occur. The local version of the *Moran's I* (local indicators of spatial association or *LISA* [74]) identifies both the locus and shape of these clusters. Another commonly used local clustering approach is the *Kulldorff* 's spatial scan statistic [75], which identifies the most likely disease clusters in a study area by maximizing the likelihood that disease cases are located within a set of concentric circles that are moved across the study area. Both number of observed and expected cases are counted and the most "unusual" excess of observed cases is reported. The approach is flexible to incorporate heterogeneous background population densities and adjust for multiple testing. Several packages (i.e., R, GeoDa, ArcGIS, SaTScan) can implement and visualize both global and local tests for clustering.

Space-time clustering

Several statistical approaches have been proposed to detect spatiotemporal clusters of diseases. For individual data, the *Knox* test for space-time interaction evaluates the presence of a

space-time cluster at given spatial and temporal distances [76]. Knox's method is limited due to its arbitrary definition of closeness and the critical distance does not account for population heterogeneity [77, 78]. The *Mantel* test [79] incorporates the notion of distance decay in that nearby pairs of events are more important than distant pairs. However, the *Mantel* statistic assumes a linear association between space and time, which is violated, especially for infectious diseases. The *Jacquez's k-Nearest Neighbor k-NN* statistic [78] addresses the weaknesses of the *Knox* and *Mantel* statistics, by counting the number of pairs of events that are nearest neighbors in both space and time. The space-time *Ripley's K* function evaluates the magnitude of space-time clustering at different spatial and temporal scales [69]. Results from the *Knox*, *Mantel*, *k-NN* or the *space-time K-function* can serve as an input to visualize space-time patterns, for instance by means of the space-time kernel density estimation [80, 81].

Local space-time clustering techniques include the space-time permutation statistic [82], which uses a cylinder with a circular (or elliptical) base, where the vertical axis represents time. *Kulldorff* 's space-time scan statistic [83] or the extension of the *LISA* statistic in time are generally used to estimate space-time clustering of aggregated data, such as space-time change in disease rates across a study region.

Although statistical tests of clustering are useful for retrospective analysis, prospective approaches such as the cumulative sum (cusum) method are designed for time series data which are updated on a regular basis, or when new events are observed [84]. These techniques are able to quickly detect deviations from the mean in a series of events to detect clusters as quickly as possible [85]. Software products such as SaTScan [86] and Geosurveillance [84] incorporate prospective tests.

Small area estimation

Local health behavior and outcome data are important for program planning and evaluation, resource allocation, and policy-making activities. Although interest in local-level data is growing, population-based surveys powered to make national or state-level inferences are rarely adequate to support direct estimation for small areas such as counties. Unit-level small area estimation (SAE) techniques have become the optimal method to derive rates for counties and other small geographic areas given these known limitations. Indirect estimates for small areas are usually derived using synthetic or multilevel modeling. Although both approaches are based on prediction models, indirect estimates based on multilevel modeling also account for random area-level effects not explained by the covariates alone [87]. Multilevel model-based approaches to small area estimation have been extensively applied to population-based health surveys such as BRFSS on topics ranging from COPD, periodontitis, smoking prevalence, obesity, and HPV vaccine uptake [88-92]. A multilevel, post-stratification approach developing for polling research has been used in more recent years to provide granular and precise disease estimates at a variety of spatial scales. The post-stratification approach is unique in that it accounts for the underlying distribution of population such as age, race, and gender at the small area level (*k*).[93] In essence, estimates are created for every possible subgroup combination and then averaged by their weighted distribution in small area *k*. In this issue, Lin et al. [94] applies this multilevel, post-stratification approach to estimation health-related quality of life in adults aged 65 and older using data from the Behavioral Risk Factor Surveillance System. One can also use the spatial structure of the data to improve the small area estimates utilizing the conditional autoregressive (CAR), intrinsic conditional autoregressive (ICAR) or simultaneous autoregressive (SAR) modelling frameworks.

Bayesian approaches to disease mapping

The primary focus of disease mapping is on identifying local variations in the incidence or prevalence of a health condition, disease, or health outcome. At its most elementary level, an analyst can plot the distribution of outcomes as a choropleth map, classifying each areal unit into one of a series of categories based on its relative frequency. Increasingly however, spatial epidemiologists are turning to Bayesian methods for disease mapping and small area estimation. These methods have become increasingly sophisticated as tools for simulation analyses now run very efficiently on desktop computers. Lawson provides a useful introduction to these approaches.[4] Two papers in this issue demonstrate how these methods are being extended to improve small area estimates of health outcomes (Neyens et al. [95]) and improve modeling of rare outcomes through spatiotemporal mixture modeling (Carroll et al. [96]).

Spatial regression models

Standard regression models such as the Ordinary Least Squares (OLS) capture the average strength and significance of the predictor variables, but makes the assumption that rates at location *i* are independent of rates at neighboring units *j* and that the residuals are normally distributed. As such, OLS does not consider small-scale spatial variation and spatial autocorrelation in the residuals are indicative that those assumptions are violated [97]. There exist several techniques to incorporate spatial effects, such as the spatial regression [98] or the Geographically Weighted Regression (GWR) technique [99]. In spatial regression, autocorrelation in the dependent variable -such as when rates in location *i* are influenced by rates at location *j* -, justifies the use of a spatial lag model, which incorporates an autoregressive term for the dependent variable. Autocorrelation in the residuals however, indicates that variables may have been omitted and the prescription is to use a spatially lagged error model. Unlike OLS or

spatial regression that provides global estimates, GWR is a local regression technique used to measure how the strength of the relationships among the dependent and explanatory variables differ from location to location. In essence, for each location, GWR pools data from adjacent neighborhoods to conduct a local regression, resulting in an estimate of local regression coefficients for each of the predictor variables.

Outlook

There is increasing interest by policy-makers, public health practitioners, community planners and researchers in understanding the social, environmental and structural features of neighborhoods, which have been linked to a wide array of health outcomes. Identifying neighborhood-level assets and deficits provide a baseline for targeting resources and interventions appropriately to the unique needs of the underlying population. The stated goals of Healthy People 2020 [100], however, demonstrate a move away from simply identifying health disparities, towards identifying and testing potential solutions that maximize each neighborhood's opportunity to achieve health equity.

At a foundational level, GIS software packages and online mapping applications offer a range of tools for stakeholders to visualize, analyze, and report on risk factors and outcomes at multiple geographic levels. These levels can have traditionally been defined based on administrative boundaries; however, empirical approaches based on the underlying data distribution rather than somewhat arbitrary geopolitical boundaries have also been proposed [101]. Using boundaries that mimic the actual travel patterns of the residents within areas or clustering of outcomes across areas seeks to minimize the modifiable area unit problem and understand how geography impacts areas differently. For example, when designing facility

service areas or hospital markets, one can use patient referral data to draw hospital market boundaries [102] or use local provider competition data (e.g., number of medical providers around a locality) to draw variably-sized market boundaries that account for the likelihood of travel outside the area [103].

Web-based mapping applications have also been on the rise, as technological advances such as application programming interfaces (APIs) including Google Maps API, ArcGIS Web APIs, and Bing Maps API, now readily enable programmers to build interactive web-based mapping tools. On the user end, these applications (such as HealthLandscape [104], HealthMap [105], Community Commons [106], and Dartmouth Atlas [102]) allow for the creation of customized maps and reports, though rarely allow for the direct export of the underlying data.. The vast majority of these on-line mapping applications have been restricted to the visualization of individual or aggregated data –partly due to protection of privacy and computational capabilities- but in recent years, several initiatives have been undertaken to incorporate spatial and spatio-temporal analytical capabilities over the web [107-109], such as kernel density estimation and space-time linkages.

Fortunately, the era of big data and enhanced data sharing is here; epidemiologists working in many sectors will be affected and can positively impact the trajectory of the field. As evidenced by Vice President Joe Biden's Moonshot Cancer Initiative, enhanced data sharing is a major need for clinical and epidemiological research in the next decade. With small area (even address) level data being shared more readily with researchers (Table 2), more analyses are enabled that utilize a geospatial lens. Researchers looking to investigate geographic trends in provider availability, treatment patterns, and health outcomes may find resources such as www.healthindicators.gov offer a one-stop shop for access to geospatial data and is searchable

by geographic level or risk factor/disease of interest. Many of the indicators listed on www.healthindicators.gov are cross-referenced to their respective Healthy People 2020[100] objective as well.

Table 2. Sample of data sources that permit geographic analysis of provider availability, treatment patterns, and health outcomes

Data Source	Geographic Levels	Description	Outside Information
Healthcare Cost and Utilization Project (HCUP)	State, County, Zip	Discharge-level information on inpatient, outpatient, and emergency department visits including charges, procedures and diagnosis codes, and demographics	http://www.hcup-us.ahrq.gov/
Medicare Fee-For-Service Provider Payment and Utilization Data	State, Zip, Address	Provider-level utilization and average payment information on services/procedures provided to Medicare beneficiaries	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html
County Health Rankings	County	County-level rates of risk behaviors and health outcomes such as premature death, adult smoking and teen births; online mapping tool available	http://www.countyhealthrankings.org/
Kids Count	County, Region, Congressional Districts	Various demographic, economic, public safety, health behavior and outcome data available	http://datacenter.kidscount.org/
Area Health Resources Files	County	County-level health status and provider availability indicators; online mapping tool available	http://ahrf.hrsa.gov/
RTI Spatial Impact Factors Dataset	County, ZCTA, PCSA, Tract, MSSA	Various health, transportation, environmental, and demographic indicators for multiple geographies	https://rtispatialdata.rti.org/
National Provider Identifier Registry	State, Zip, Address	Provider-level information including medical specialty	https://npiregistry.cms.hhs.gov/
Hospital Compare	State, Zip, Address	Hospital-level information including complications and	https://data.medicare.gov/data/hospital-compare

		readmissions, payment data, quality of care measures, and patient satisfaction scores	
Medicare Geographic Variation Public Use File	State, HRR, County	Demographic, spending, utilization and quality indicators for various geographic levels (some mapping tools available)	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html

Footnotes: ZCTA = Zip code tabulation area, PCSA = Primary care service area; MSSA = Medical service study area (California only), HRR = Hospital referral region

With the growth in public data sharing and technologies to measure the neighborhood environment [110] (e.g., social media, crowdsourcing, Google Street View), epidemiologists have more tools at hand to monitor population health and identify actionable targets for intervention. Where possible, researchers must get away from measuring contextual influences based only on the residential environment, in static space and time. Life-course approaches, that consider the cumulative exposures people face as they age, make occupational shifts, and relocate over time are starting to surface, as are activity-space approaches that account for where people choose to engage in daily activities (see example from Duncan et al., this issue [111]) [112, 113]. With its inherent ability to integrate data from multiple systems, users, geographies, and time points, GIS holds great potential to help health researchers explore these challenges.

References:

1. Szklo, M. and J. Nieto, *Epidemiology: Beyond the Basics*. Vol. 3rd Edition. 2012, Sudbury, MA: Jones & Bartlett Learning.
2. Elliot, P., et al., *Spatial epidemiology: methods and applications*. 2000: Oxford University Press.
3. English, D., *Geographical epidemiology and ecological studies.* Geographical and environmental epidemiology (P. Elliott, J. Cuzick, D. English, R. Stern, eds.), 1992: p. 1.
4. Lawson, A.B., *Statistical methods in spatial epidemiology*. 2013: John Wiley & Sons.

5. Dummer, T.J.B., *Health geography: supporting public health policy and planning.* CMAJ : Canadian Medical Association Journal, 2008. **178**(9): p. 1177-1180.
6. Kirby, R.S., *Toward congruence between theory and practice in small area analysis and local public health data.* Statistics in medicine, 1996. **15**(17): p. 1859-1866.
7. Kwan, M.-P., *Geographies of health, disease and well-being: recent advances in theory and method*. 2016: Routledge.
8. Diez Roux, A.V., *Investigating neighborhood and area effects on health.* American journal of public health, 2001. **91**(11): p. 1783-1789.
9. Nattinger, A.B., et al., *Relationship of distance from a radiotherapy facility and initial breast cancer treatment.* Journal of the National Cancer Institute, 2001. **93**(17): p. 1344-1346.
10. Celaya, M.O., et al., *Travel distance and season of diagnosis affect treatment choices for women with early-stage breast cancer in a predominantly rural population (United States).* Cancer Causes & Control, 2006. **17**(6): p. 851-856.
11. Bennett, K.J., J.C. Probst, and J.D. Bellinger, *Receipt of cancer screening services: surprising results for some rural minorities.* The Journal of Rural Health, 2012. **28**(1): p. 63-72.
12. Gordon-Larsen, P., et al., *Inequality in the built environment underlies key health disparities in physical activity and obesity.* Pediatrics, 2006. **117**(2): p. 417-424.
13. Singh, G.K., M. Siahpush, and M.D. Kogan, *Neighborhood socioeconomic conditions, built environments, and childhood obesity.* Health affairs, 2010. **29**(3): p. 503-512.
14. Cummins, S. and S. Macintyre, *Food environments and obesity—neighbourhood or nation?* International journal of epidemiology, 2006. **35**(1): p. 100-104.
15. Davis, B. and C. Carpenter, *Proximity of fast-food restaurants to schools and adolescent obesity.* American Journal of Public Health, 2009. **99**(3): p. 505-510.
16. McEntee, J.C. and Y. Ogneva-Himmelberger, *Diesel particulate matter, lung cancer, and asthma incidences along major traffic corridors in MA, USA: A GIS analysis.* Health & Place, 2008. **14**(4): p. 817-828.
17. Chandola, T., *The fear of crime and area differences in health.* Health & place, 2001. **7**(2): p. 105-116.
18. Lorenc, T., et al., *Crime, fear of crime, environment, and mental health and wellbeing: mapping review of theories and causal pathways.* Health & place, 2012. **18**(4): p. 757-765.
19. Krieger, N., *Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology.* American journal of public health, 1992. **82**(5): p. 703-710.
20. Krieger, N., et al., *Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures-the public health disparities geocoding project.* American journal of public health, 2003. **93**(10): p. 1655-1671.
21. Tellman, N., et al., *The effects of the Health Insurance Portability and Accountability Act privacy rule on influenza research using geographical information systems.* Geospatial health, 2010. **5**(1): p. 3-9.
22. Openshaw, S. and S. Openshaw. *The modifiable areal unit problem*. 1984. Geo Abstracts University of East Anglia.
23. Kanaroglou, P., E.M. Delmelle, and A. Paez, *Spatial Analysis in Health Geography*. 2015: Ashgate Publishing, Ltd.
24. Cromley, E.K. and S.L. McLafferty, *GIS and public health*. 2011: Guilford Press.
25. Thornton, L.E., J.R. Pearce, and A.M. Kavanagh, *Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary.* International Journal of Behavioral Nutrition and Physical Activity, 2011. **8**(1): p. 1.

26. Zulu, L.C., E. Kalipeni, and E. Johannes, *Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010.* BMC infectious diseases, 2014. **14**(1): p. 1.
27. Pierce, S.J., et al., *Identifying HIV prevention service needs of African American men who have sex with men: an application of spatial analysis techniques to service planning.* Journal of Public Health Management and Practice, 2007. **13**: p. S72-S79.
28. Fulcher, C. and C. Kaukinen, *Mapping and visualizing the location HIV service providers: an exploratory spatial analysis of Toronto neighborhoods.* AIDS care, 2005. **17**(3): p. 386-396.
29. Hanchette, C.L., et al., *A national, geographic database of CDC-funded HIV prevention services: development challenges and potential applications.* International journal of health geographics, 2005. **4**(1): p. 1.
30. Buehler, C.P., et al., *Assessing spatial patterns of HIV knowledge in rural Mozambique using geographic information systems.* Tropical Medicine & International Health, 2015. **20**(3): p. 353-364.
31. Goswami, N.D., et al., *Geographic information system-based screening for TB, HIV, and syphilis (GIS-THIS): a cross-sectional study.* PloS one, 2012. **7**(10): p. e46029.
32. Shepard, C.W., et al., *Using GIS-based density maps of HIV surveillance data to identify previously unrecognized geographic foci of HIV burden in an urban epidemic.* Public health reports, 2011: p. 741-749.
33. Gorin, S.S., et al., *Multilevel Interventions and Racial/Ethnic Health Disparities.* JNCI Monographs, 2012. **2012**(44): p. 100-111.
34. Mobley, L.R., et al., *Using Residential Segregation to Predict Colorectal Cancer Stage at Diagnosis: Two Different Approaches.* Annals of Epidemiology.
35. *National Health Interview Survey*. 2015 September 27, 2016 [cited 2016 October 10]; Available from: http://www.cdc.gov/nchs/nhis/index.htm.
36. *National Survey of Children's Health*. 2012 October 30, 2013 [cited 2016 October 10]; Available from: http://www.cdc.gov/nchs/slaits/nsch.htm.
37. Prevention, C.f.D.C.a. *Behavioral Risk Factor Surveillance System*. 2015 August 26, 2016 [cited 2016 October 10]; Available from: http://www.cdc.gov/brfss/.
38. Lupo, P.J., et al., *Neighborhood-Based Socioeconomic Position and Risk of Oral Clefts Among Offspring.* American Journal of Public Health, 2015. **105**(12): p. 2518-2525.
39. Escamilla-Nuñez, M.-C., et al., *Traffic-related air pollution and respiratory symptoms among asthmatic children, resident in Mexico City: the EVA cohort study.* Respiratory research, 2008. **9**(1): p. 1.
40. Hoek, G., et al., *Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study.* The lancet, 2002. **360**(9341): p. 1203-1209.
41. Chen, H.-J. and Y. Wang, *Changes in the Neighborhood Food Store Environment and Children's Body Mass Index at Peripuberty in the United States.* Journal of Adolescent Health, 2016. **58**(1): p. 111-118.
42. Gibson, D.M., *The Neighborhood Food Environment and Adult Weight Status: Estimates From Longitudinal Data.* American Journal of Public Health, 2011. **101**(1): p. 71-78.
43. Zenk, S.N., et al., *Longitudinal Associations Between Observed and Perceived Neighborhood Food Availability and Body Mass Index in a Multiethnic Urban Sample.* Health Education & Behavior, 2016.
44. Goldberg, D.W., J.P. Wilson, and C.A. Knoblock, *From text to geographic coordinates: The current state of geocoding.* URISA-WASHINGTON DC-, 2007. **19**(1): p. 33.
45. Jacquez, G.M., *A research agenda: Does geocoding positional error matter in health GIS studies?* Spatial and spatio-temporal epidemiology, 2012. **3**(1): p. 7-16.

46. Cayo, M.R. and T.O. Talbot, *Positional error in automated geocoding of residential addresses.* International journal of health geographics, 2003. **2**(1): p. 1.
47. Zandbergen, P.A., *A comparison of address point, parcel and street geocoding techniques.* Computers, Environment and Urban Systems, 2008. **32**(3): p. 214-232.
48. DeLuca, P. and P.S. Kanaroglou, *An Assessment of Online Geocoding Services for Health Research in a Mid-Sized Canadian City*, in *Spatial Analysis in Health Geography* P. Kanaroglou, E. Delmelle, and A. Paez, Editors. 2015. p. 31.
49. Wieland, S.C., et al., *Revealing the spatial distribution of a disease while preserving privacy.* Proceedings of the National Academy of Sciences, 2008. **105**(46): p. 17608-17613.
50. Hampton, K.H., et al., *Mapping health data: improved privacy protection with donut method geomasking.* American journal of epidemiology, 2010: p. kwq248.
51. Lovett, A., et al., *Car travel time and accessibility by bus to general practitioner services: a study using patient registers and GIS.* Social science & medicine, 2002. **55**(1): p. 97-111.
52. Frizzelle, B.G., et al., *The importance of accurate road data for spatial applications in public health: customizing a road network.* International journal of health geographics, 2009. **8**(1): p. 1.
53. Delmelle, E.M., et al., *Modeling travel impedance to medical care for children with birth defects using Geographic Information Systems.* Birth Defects Research Part A: Clinical and Molecular Teratology, 2013. **97**(10): p. 673-684.
54. Khang, L., et al., *Travel distance to screening facilities and completion of abnormal mammographic follow-up among disadvantaged women.* Annals of Epidemiology.
55. Lian, M., J. Struthers, and M. Schootman, *Comparing GIS-based measures in access to mammography and their validity in predicting neighborhood risk of late-stage breast cancer.* PLoS One, 2012. **7**(8): p. e43000.
56. Urayama, K.Y., et al., *Factors associated with residential mobility in children with leukemia: implications for assigning exposures.* Annals of epidemiology, 2009. **19**(11): p. 834-840.
57. Nordsborg, R.B., et al., *Investigation of spatio-temporal cancer clusters using residential histories in a case–control study of non-Hodgkin lymphoma in the United States.* Environmental Health, 2015. **14**(1): p. 1.
58. Pershagen , G., et al., *Residential Radon Exposure and Lung Cancer in Sweden.* New England Journal of Medicine, 1994. **330**(3): p. 159-164.
59. Sabel, C.E., et al., *Modelling individual space–time exposure opportunities: A novel approach to unravelling the genetic or environment disease causation debate.* Spatial and spatio-temporal epidemiology, 2009. **1**(1): p. 85-94.
60. Polissar, L., *The effect of migration on comparison of disease rates in geographic studies in the United States.* American journal of epidemiology, 1980. **111**(2): p. 175-182.
61. Bell, M.L. and K. Belanger, *Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures.* Journal of Exposure Science and Environmental Epidemiology, 2012. **22**(5): p. 429-438.
62. Miller, A., C. Siffel, and A. Correa, *Residential mobility during pregnancy: patterns and correlates.* Maternal and child health journal, 2010. **14**(4): p. 625-634.
63. *Migration/Geographic Mobility: County-to-County Migration Flows*. [cited 2016 October 13]; Available from: http://www.census.gov/hhes/migration/data/acs/county-to-county.html.
64. Hughes, A.E. and S.L. Pruitt, *The utility of EMR address histories for assessing neighborhood exposures.* Annals of Epidemiology.
65. Kershenbaum, A., et al., *The Pattern of Association between U.S. Economic Indicators and Infant Mortality Rates at the State Level.* Journal of Health Care for the Poor and Underserved, 2014. **25**(3): p. 1432-1448.

66. *Area Health Resources Files*. 2016 [cited 2016 October 10]; Available from: http://ahrf.hrsa.gov/.
67. Mclafferty, S., *Disease cluster detection methods: recent developments and public health implications.* Annals of GIS, 2015. **21**(2): p. 127-133.
68. Tobler, W.R., *A Computer Movie Simulating Urban Growth in the Detroit Region.* Economic Geography, 1970. **46**: p. 234-240.
69. Bailey, T. and Q. Gatrell, *Interactive Spatial Data Analysis*. 1995, Edinburgh Gate, England: Pearson Education Limited.
70. Ripley, B.D., *The second-order analysis of stationary point processes.* Journal of applied probability, 1976: p. 255-266.
71. Tiwari, C. and G. Rushton, *Using spatially adaptive filters to map late stage colorectal cancer incidence in Iowa*, in *Developments in spatial data handling*. 2005, Springer. p. 665-676.
72. Shi, X., *Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds.* International Journal of Geographical Information Science, 2010. **24**(5): p. 643-660.
73. Moran, P.A., *Notes on continuous stochastic phenomena.* Biometrika, 1950: p. 17-23.
74. Anselin, L., *Local indicators of spatial association—LISA.* Geographical analysis, 1995. **27**(2): p. 93-115.
75. Kulldorff, M., *A spatial scan statistic.* Communications in Statistics-Theory and methods, 1997. **26**(6): p. 1481-1496.
76. Kulldorff, M. and U. Hjalmars, *The Knox method and other tests for space-time interaction.* Biometrics, 1999: p. 544-552.
77. Robertson, C., et al., *Review of methods for space–time disease surveillance.* Spatial and Spatio-temporal Epidemiology, 2010. **1**(2): p. 105-116.
78. Jacquez, G.M., *A k nearest neighbour test for space–time interaction.* Statistics in medicine, 1996. **15**(18): p. 1935-1949.
79. Mantel, N., *The detection of disease clustering and a generalized regression approach.* Cancer research, 1967. **27**(2 Part 1): p. 209-220.
80. Nakaya, T. and K. Yano, *Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics.* Transactions in GIS, 2010. **14**(3): p. 223-239.
81. Delmelle, E., et al., *Visualizing the impact of space-time uncertainties on dengue fever patterns.* International Journal of Geographical Information Science, 2014. **28**(5): p. 1107-1127.
82. Kulldorff, M., et al., *A space-time permutation scan statistic for disease outbreak detection.* PLoS medicine, 2005. **2**(3): p. 216.
83. Kulldorff, M., et al., *Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico.* American journal of public health, 1998. **88**(9): p. 1377-1380.
84. Yamada, I., P.A. Rogerson, and G. Lee, *GeoSurveillance: a GIS-based system for the detection and monitoring of spatial clusters.* Journal of Geographical Systems, 2009. **11**(2): p. 155-173.
85. Rogerson, P.A., *Surveillance systems for monitoring the development of spatial patterns.* Statistics in Medicine, 1997. **16**(18): p. 2081-2093.
86. *SaTScan: software for the spatial, temporal, and space-time scan statistics*. [cited 2016 November 15]; Available from: http://www.satscan.org/.
87. Rao, J.N. and I. Molina, *Small area estimation*. 2015: John Wiley & Sons.
88. Eberth, J.M., et al., *Human papillomavirus vaccine coverage among females aged 11 to 17 in Texas counties: an application of multilevel, small area estimation.* Women's Health Issues, 2013. **23**(2): p. e131-e141.

89. Eke, P., et al., *Predicting Periodontitis at State and Local Levels in the United States.* Journal of dental research, 2016. **95**(5): p. 515-522.
90. Zhang, X., et al., *Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system.* American journal of epidemiology, 2014. **179**(8): p. 1025-1033.
91. Zhang, X., et al., *Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system.* American journal of epidemiology, 2015. **182**(2): p. 127-137.
92. Song, L., et al., *Peer Reviewed: Using Small-Area Estimation to Calculate the Prevalence of Smoking by Subcounty Geographic Areas in King County, Washington, Behavioral Risk Factor Surveillance System, 2009–2013.* Preventing chronic disease, 2016. **13**.
93. Zhang, X., et al., *Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system.* Am J Epidemiol, 2015. **182**(2): p. 127-37.
94. Lin, Y.-H., et al., *Health-related quality of life among adults 65 years and older in the United States, 2011–2012: a multilevel small area estimation approach.* Annals of Epidemiology.
95. Neyens, T., et al., *Disease mapping of zero-excessive mesothelioma data in Flanders.* Annals of Epidemiology.
96. Carroll, R., et al., *Space-time variation of respiratory cancers in South Carolina: a flexible multivariate mixture modeling approach to risk estimation.* Annals of Epidemiology.
97. Anselin, L., *Under the hood issues in the specification and interpretation of spatial regression models.* Agricultural economics, 2002. **27**(3): p. 247-267.
98. Anselin, L., *Spatial regression.* The SAGE handbook of spatial analysis, 2009. **1**: p. 255-276.
99. Fotheringham, A.S., C. Brunsdon, and M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*. 2003: John Wiley & Sons.
100. People, H., *2020.(2010).* Healthy people, 2011. **2020**.
101. Wang, F., D. Guo, and S. McLafferty, *Constructing geographic areas for cancer data analysis: a case study on late-stage breast cancer risk in Illinois.* Applied Geography, 2012. **35**(1): p. 1-11.
102. Dartmouth, A. 2016 [cited 2016 August 3]; Available from: http://www.dartmouthatlas.org/data/region/.
103. McGrail, M.R. and J.S. Humphreys, *The index of rural access: an innovative integrated approach for measuring primary care access.* BMC Health Services Research, 2009. **9**(1): p. 1.
104. *HealthLandscape*. 2016 [cited 2016 October 14]; Available from: http://www.healthlandscape.org/.
105. *Health Map*. 2016 [cited 2016 October 14]; Available from: http://www.healthmap.org/en/.
106. *Community Commons*. 2016 [cited 2016 October 14]; Available from: http://www.communitycommons.org/.
107. Delmelle, E.M., et al., *A web-based geospatial toolkit for the monitoring of dengue fever.* Applied Geography, 2014. **52**: p. 144-152.
108. Dominkovics, P., et al., *Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in Barcelona, Spain.* International journal of health geographics, 2011. **10**(1): p. 1.
109. Padmanabhan, A., et al., *FluMapper: A cyberGIS application for interactive analysis of massive location-based social media.* Concurrency and Computation: Practice and Experience, 2014. **26**(13): p. 2253-2265.

110. Schootman, M., et al., *Emerging technologies to measure neighborhood conditions in public health: implications for interventions and next steps.* International Journal of Health Geographics, 2016. **15**(1): p. 1.
111. Duncan, D.T., et al., *Quantifying Spatial Misclassification in Exposure to Noise Complaints Among Low-Income Housing Residents Across New York City Neighborhoods: A Global Positioning System (GPS) Study.* Annals of Epidemiology.
112. Hirsch, J.A., et al., *Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis.* International Journal of Health Geographics, 2014. **13**(1): p. 51.
113. Logan, J.J., A.M. Jolly, and J.I. Blanford, *The Sociospatial Network: Risk and the Role of Place in the Transmission of Infectious Diseases.* PLoS ONE, 2016. **11**(2): p. e0146915.

ACCEPTED MANUSCRIPT