HHS Public Access

Author manuscript

Cancer. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

Cancer. 2019 August 01; 125(15): 2544–2560. doi:10.1002/cncr.32052.

GIScience and Cancer: State of the Art and Trends for Cancer Surveillance and Epidemiology

Liora Sahar, PhD[1], **Stephanie L. Foster, MPH, MA**[2], **Recinda L. Sherman, MPH, PhD, CTR**[3], **Kevin A. Henry, PhD**[4,5], **Daniel W. Goldberg, PhD**[6,7], **David G. Stinchcomb, MS, MA**[8], and **Joseph E. Bauer, PhD**[9]

[1]Geospatial Research, Statistics and Evaluation Center, American Cancer Society, Atlanta, Georgia; [2]Geospatial Research Analysis and Services Program, Centers for Disease Control and Prevention, Atlanta, Georgia; [3]Data Use and Research, North American Association of Central Cancer Registries, Springfield, Illinois; [4]Department of Geography and Urban Studies, Temple University, Philadelphia, Pennsylvania; [5]Cancer Prevention and Control Program, Fox Chase Cancer Center, Philadelphia, Pennsylvania; [6]Department of Geography, College of Geosciences, Texas A&M University, College Station, Texas; [7]Department of Computer Science and Engineering, College of Engineering, Texas A&M University, College Station, Texas; [8]Health Studies, Westat, Rockville, Maryland; [9]Statistics and Evaluation Center, American Cancer Society, Atlanta, Georgia.

Abstract

Maps are well recognized as an effective means of presenting and communicating health data, such as cancer incidence and mortality rates. These data can be linked to geographic features like counties or census tracts and their associated attributes for mapping and analysis. Such visualization and analysis provide insights regarding the geographic distribution of cancer and can be important for advancing effective cancer prevention and control programs. Applying a spatial approach allows users to identify location-based patterns and trends related to risk factors, health outcomes, and population health. Geographic information science (GIScience) is the discipline that applies Geographic Information Systems (GIS) and other spatial concepts and methods in research. This review explores the current state and evolution of GIScience in cancer research by addressing fundamental topics and issues regarding spatial data and analysis that need to be considered. GIScience, along with its health-specific application in the spatial epidemiology of cancer, incorporates multiple geographic perspectives pertaining to the individual, the health care infrastructure, and the environment. Challenges addressing these perspectives and the synergies among them can be explored through GIScience methods and associated technologies as integral parts of epidemiologic research, analysis efforts, and solutions. The authors suggest GIScience is a powerful tool for cancer research, bringing additional context to cancer data analysis and

Corresponding author: Liora Sahar, PhD, Geospatial Research, Statistics and Evaluation Center, American Cancer Society, 250 Williams Street, Atlanta, GA 30303; liora.sahar@cancer.org.

CONFLICT OF INTEREST DISCLOSURES

Joseph E. Bauer is a scientific reviewer and serves on the *Cancer Journal* Editorial Advisory Board. Liora Sahar is the Scientific Director for Geospatial Research within the American Cancer Society. The remaining authors had no disclosures.

potentially informing decision-making and policy, ultimately aimed at reducing the burden of cancer.

Keywords

cancer surveillance; geographic information science (GIScience); Geographic Information Systems (GIS); mapping and visualization; spatial epidemiology; spatial statistics

INTRODUCTION

Geographic Information Systems (GIS) are hardware, software, technologies, and tools that enable the storage, retrieval, visualization, and analysis of geographic features and associated data. Oftentimes, GIS is superficially understood as merely mapping data. Historically, the early epidemiologic use of mapping provided the foundation for understanding the relationship between geography and health, and examples date back to the 1800s. Dr. John Snow's map of the 1854 London cholera outbreak, with its clustering of deaths around the Broad Street water pump, is likely the most well known historic map.[1–3] Although the earliest maps were dominated by infectious diseases like typhoid and cholera, there are several early maps illustrating the distribution of cancers. These include the 1870 map by Haviland of cancer mortality rates in Britain,[4,5] Power's map illustrating the precise location of cancer cases in a small British village from 1872 to 1888, and Green's 1908 map illustrating cancer mortality in relation to coal-burning and wood-burning areas in France.[4] The goal of those early cancer maps was to reveal disease patterns in relation to local environmental factors with the hope of shedding light on disease etiology.

Over time, GIS has become much more information-rich and scientifically rigorous, and GIS technologies have greatly simplified the compilation of health maps. However, the power of GIS is not just in the aesthetic cartographic display of health data but also in the ability to link attribute data (properties of the feature) with the geography. GIS provides the capability to visualize such attributes beyond traditional charts and graphs alongside features on the map. Linking these attributes by geography makes understanding and interpretation simpler, allowing consumers of mapping products the ability to identify social and demographic patterns and trends. Furthermore, with advanced spatial statistical methodologies, users may avoid the subjective visualization and interpretation of data and instead have objective, quantitative measures that support and reveal the underlying spatial relationships of both risk and potential confounders.

GIS has greatly evolved while incorporating new technologies, such as computers, computer-aided design, and databases along with the integration of methodologies from disciplines such as statistics, economics, computer science, and others,[6,7] to emerge as the field of geographic information science (GIScience).[8,9] GIScience is often defined using the bylaws of the University Consortium for Geographic Information Science as "the development and use of theories, methods, technologies, and data for understanding geographic process, relationships, and patterns. The transformation of geographic data into useful information is central to geographic information science."[9] This collaborative, interdisciplinary approach provides opportunities to further examine relationships and

interactions among health outcomes, the physical environment, and various socioeconomic and other risk factors to advance cancer-related and other health-related research.[10]

In this review, we present and discuss relevant geospatial concepts for consideration when planning and designing cancer research. The sections below provide an overview of popular topics in the spatial epidemiology of cancer with the viewpoint that an interdisciplinary approach is required to advance cancer research. We review important topics, including spatial data for cancer research, cancer mapping and visualization, and advanced spatial analysis. These topics help to address questions about what is special or unique about spatial data, which types of spatial data can be used for cancer research, and how spatial data can be used for cancer research to complement traditional analysis approaches in epidemiology. Each section discusses the current state of the art, issues, limitations and potential solutions, and available resources.

This introduction to geospatial concepts is meant to enable researchers and practitioners to recognize the potential and value of incorporating a spatial approach into their work. It also allows researchers to evaluate whether they have sufficient knowledge and tools or whether they need to seek the assistance of a GIS professional. We lead the reader through the process of identifying and assessing potential sources of spatial data and their associated limitations, through visualization of the data, spatial analysis, and the advantages of utilizing advanced spatial-statistical analysis. Spatial analysis enables researchers to address geographic discrepancies, which often are driven by racial or social disparities, and to augment traditional exploratory mapping and visual interpretation by testing geographically based hypotheses. We also describe methodologies, available tools, and best practices to visualize, understand, and communicate cancer risk factors and disease burden.

SPATIAL DATA FOR CANCER RESEARCH

Geocoding Cancer Data

Spatial data for cancer research span numerous sources with inherently different characteristics, including type, format, and geographic scale. Such data can also be the result of geocoding addresses, which usually is the first step when using cancer registry data. Geocoding is often defined as the process of converting address information into geographic coordinates, such as latitude and longitude.[11] The process of geocoding has been well studied,[11–27] along with the effects that geocoding errors and inconsistencies can have on the analysis and visualization of cancer data.[16,24,28–54] Geocoding tools and services are now widely available to the cancer community,[17] but care needs to be taken when choosing a geocoding service,[24,33,40] passing address data confidentially to that service,[55–59] and interpreting the results.[16,28,32,34,38,39,44,48,52,60–62]

Although this may appear to be a straightforward process, researchers should be cognizant of factors that influence the geocoding results and should be careful when selecting geocoded records for analysis. To represent the geographic distribution of cancer registry data accurately and/or to have confidence in the results of a spatial analysis, certainty and spatial accuracy depend fully on the geocoding results. Users need to be aware of the elements of the geocoding system and should consider the associated quality and accuracy,

which may have important implications when interpreting results.[63] Therefore, proper geocoding techniques and careful attention to locational accuracy are fundamental to mapping and geospatial analysis.

The geocoding system—A geocoding system implements the geocoding process to produce geographic output (or geocoded data) by matching the input data (usually addresses) to a reference layer (such as streets). Most geocoding systems contain the components presented in Figure 1 and operate in a similar fashion.[11,14,17] These components work together to produce output geocodes, and each component affects the accuracy of each geocode and the overall level of accuracy of the geocoding system. The first component, *input data*, is the text of the postal address records about patients, hospitals, or other features or entities. The second component, *reference data*, consists of geographic data files describing roads, parcels, building points, ZIP codes, and other geographic objects used to compute a geographic output for a given address. A geocoding system may use 1 or multiple (composite) reference layers, such as address points, parcels, street segments, ZIP codes, etc. These data can be purchased and/or obtained for free. The quality of the input data and the reference data set greatly influences the completeness and accuracy of the geocoding results. *Address matching* is the third component, in which the text for each address is processed using several well established steps, referred to as *address parsing* and *address normalization/standardization*, followed by a matching algorithm.[11,14,17] The result is a set of potential match candidates (point locations) for a given address. This can be an iterative process, which may require relaxing the matching criteria until a match is identified or asking the user to match a point manually on the map. A match can be along a street segment, the centroid of a ZIP code, the centroid of a county, etc. The final component, *output data*, comprises the geocoded results. At a minimum, geocoding results include a geographic representation (commonly, latitude/longitude), and some form of metadata about the quality of the geocode.

Geocoding quality—Most geocoders provide match rates and match type as metrics[18,39,52,64,65] to describe the quality of the results. Match rates characterize the percentage of geocodes that the system produced out of the total number of records that it was asked to process and are not useful for assessing fitness for use at the per-record level. Match types are important for each geocode, especially because a composite geocoder may produce more than 1 result. For example, a geocode marked as a *county-level* match type is not as accurate as a *building centroid-level* match. Clearly, including records matched to a building with records matched to the centroid of a county can affect the spatial accuracy and can have significant implications for the results from any visualization or spatial analysis activities. However, evaluating these metrics is an important methodological step that is often overlooked.

Types of geocoding systems: Standards?—Currently, there are no standards for geocoding systems. Each system/service processes data differently, has different reference files and algorithms, and returns different associated metadata. The various geocoding services available to cancer researchers differ mainly in the location where geocoding occurs (standalone, desktop/online, cloud) and the cost of geocoding (free/fee-based). To the best of

our knowledge, the North American Association of Central Cancer Registries (NAACCR) is the first and only organization in the United States (and possibly the world) that has attempted to undertake a geocoding standardization initiative to ensure that all cancer data in the United States is processed in the same manner, with the same reference data and algorithms, and that the results are reported in the same fashion to ensure comparison and consolidation.

Once created, geocoded data provide the basis for visualizations and analyses useful for different applications, including understanding the geographic variation of cancer burden, interactions between risk factors and disease development, and identifying gaps in health services. For example, in the Atlanta metropolitan area, researchers quantified travel times to geocoded mammography and cancer screening program clinics using public transportation routes.[66] This spatial analysis, called *network analysis*, is commonly used and provides important information for identifying disparities and improving access to screening and treatment services. NAACCR researchers developed a web-based application for processing travel time and distance using the US road network. Other systems exist that incorporate bicycle or public transportation routes and times.

Cancer Risk Factors

Data characteristics—In addition to spatial data resulting from geocoding, spatial data sets for characterizing cancer risk factors are available and are commonly used by researchers. Several key characteristics should be considered in describing and selecting these data. Of primary interest is the geographic level or scale of the data. Individual point location data are often available for some physical and social environmental factors. Cancer risk-factor data often are aggregated and are available at the state level, with some data available at the county, census tract, and ZIP code levels. When using aggregated data, researchers should be aware that original measurements may be lost, and the administrative boundaries may or may not align with project needs. Analysis at the ZIP code level (or the census-defined ZIP Code Tabulation Areas) can be problematic,[67] mainly because ZIP codes are defined for mail delivery and may pose issues when used for other applications. Users should seek data at the geographic level that best aligns with the scale of the analysis. Careful consideration should be taken when integrating data at different geographic levels to refrain as much as possible from unnecessary assumptions, such as uniform distribution across a region. For example, instead of assuming uniform distribution of cancer cases across a county, the analyst can use census information at a subcounty level to identify areas of higher proportions of certain sex or age groups, as appropriate for the specific cancer, and accordingly assign cases and rates. Another important consideration is the time-period or temporal coverage of the data. Ideally, risk-factor data should be temporally aligned with the associated cancer data. If there is a temporal lag in the hypothesized effect, then historic risk-factor data should be considered. Recent data systems, such as the National Historical GIS,[68] have begun to address issues of availability and harmonization of geospatial data sources over time. Data quality and reliability also are key attributes of cancer risk-factor data. Many data sources are based on sampled data, and the quality and reliability of such data can cause issues at smaller geographic levels.[69] Another characteristic to consider is the availability and conditions of use. Much of the cancer risk-factor data are available for free

as public-use data sets. Other data sources may require a fee, a data use agreement, or may be available only in a controlled environment of a research data center. Finally, many public-use data sets are modified to protect confidentiality by applying statistical methods, such as data swapping (switching values between records), cell suppression (excluding data), top coding (reporting values as "above" a certain threshold), and rounding.[70]

Analytic methods to develop cancer risk-factor data—Often, various analytic methods are used to develop geospatial cancer risk-factor data. When the original data are in the form of measurements at individual point locations, spatial techniques can be used to interpolate values for locations between measurement points.[71] Alternatively, a spatial model can be constructed that predicts values for arbitrary geographic locations using the measured data points to construct and validate the model.[72] When the original data are in the form of survey results or other area-based measures, spatial models and spatial smoothing methods (also described below; see Spatial Analysis), can be applied to fill gaps in the data.[73] Small-area estimation methods can be used to develop estimates for smaller geographic areas by combining information from multiple surveys.[74,75] Statistical dimension reduction methods, such as principle component analysis and factor analysis, often are used to develop a single index or a set of factors to capture complex social environmental risk factors that are multidimensional in nature.[76] Multilevel regression methods can be used to assess the impact of cancer risk factors that operate at different spatial scales.[77,78] Because many cancer risk factors operate over an extended period, spatial-temporal analysis methods are needed to assess exposure as individuals progress through daily travel and residential mobility.[79]

Types of spatial risk-factor data—There are several important types of spatial data available for characterizing behavioral risk factors. Data on the geographic differences in cancer screening behavior are an important explanatory factor in the analysis of late-stage diagnosis rates and cancer mortality rates.[80,81] Because smoking is a significant behavioral risk factor for many cancers, geospatial data on the rates of tobacco product use are important. Similarly, geographic data on tobacco policy regulations for smokefree workplaces, restaurants, and bars can provide information on possible levels of exposure to secondhand tobacco smoke.[82] Because dietary behavior can be a risk factor for cancer, geospatial data on access to healthy foods also are important.[83] Likewise, geographic differences in exercise rates, fitness levels, and obesity rates are important measures of key behavioral cancer risk factors. Finally, with the advent of human papillomavirus (HPV) vaccines and their potential for reducing cancer rates, geographic differences in HPV vaccination rates should be included as a key measure of behavioral risk factors.[84]

Spatial data for characterizing physical environmental cancer risk factors include data on various types of toxins and contaminants with either established or hypothesized carcinogenic properties. These types of data are generally categorized by their transport mechanism: air-borne, water-borne, and soil-based. Different methods are needed to develop estimates of exposure from each.[85] Two environmental cancer risk factors that do not fit neatly into these categories are exposure to ultraviolet radiation and its link with melanoma[86] and exposure to naturally occurring radon gas (Fig. 2) and its link with lung

and other cancers.[87] Finally, the effects of physical environmental cancer risk factors often are moderated by gene/environment interactions.[88]

A growing area of geospatial cancer research is studying cancer risk factors related to the social environment. These *social determinants* of health refer to characteristics of an individual's neighborhood and social context that influence health outcomes independent of individual characteristics. These social risk factors include specific neighborhood demographic and socioeconomic measures, such as poverty,[89] or composite index measures of a group of demographic and socioeconomic factors.[76] Key risk factors, such as access to cancer screening, diagnosis, and treatment services[90] as well as obesity, access to healthy food, and exercise rates,[91] are all important aspects and have been discussed before. Similarly, geospatial measures of neighborhood walkability[92] and of sprawl[93] often are used to assess social environmental risk factors. Other important neighborhood characteristics include urban/rural environment, levels of crime, neighborhood cohesion, and measures of social inequity, such as segregation, diversity, depravation, and discrimination.[94]

CANCER MAPPING AND VISUALIZATION

Cancer and cancer risk-factor data can be translated into points (eg, hospitals and patients), lines (eg, routes to treatments, roads), and polygons (eg, county cancer mortality rates) and represented on a map. Maps are a powerful means for visualizing data for cancer research and can illustrate spatial patterns and elucidate connections that may be incomprehensible in other formats. Visualizing the spatial distribution of populations in relation to screening and treatment centers or the patterns of cancer mortality and incidence in the context of place-based factors furthers our understanding of the cancer burden and stimulates research hypothesis generation. An important example of the power of visualizing the cancer burden in the United States dates back to the *Atlas of Cancer Mortality for US Counties*.[95,96] These maps allowed researchers, for the first time, to identify unusual geographic patterns in cancer mortality, subsequently stimulating studies to generate etiologic hypotheses and identify cancer sites that warranted special study. For example, high mortality rates were identified in counties with shipbuilding industries using US mortality atlases from the 1950s and 1960s. This led to the discovery of asbestos exposure as the cause of a specific type of lung cancer in World War II shipyard workers.[97] With increasing amounts of geographically enabled cancer data and more sophisticated visualization methods, mapping continues to be an invaluable method for understanding cancer burden.

Mapping Qualitative and Quantitative Data

Maps can display both qualitative and quantitative data. Qualitative data express differences in the kinds of information collected, whereas quantitative data reflect amounts. For instance, quantitative maps can display the distribution of cancer rates and provide opportunities to explore whether rates fall within the norm for a given population. Qualitative maps can allow visualization of the types of available services in an area to evaluate access to care. Furthermore, examining qualitative and quantitative data simultaneously can be a powerful technique to recognize gaps in service or for allocation of resources, as depicted in Figure 3.[98]

Given the facility of visualization, it is important to be considerate of cartographic standards. Cartographic guidelines[98–101] provide practical and fundamental concepts for sound mapmaking. They present an overview of concepts, such as map scale, projections, data classification, and visual hierarchy. For example, individuals who are not familiar with GIScience may not be aware of the distortion incurred when converting the 3-dimensional globe into a 2-dimensional map. Different projections that display different distortions are available, whereas some are more commonly used.[102] An improperly projected map may incorrectly portray the geographic distribution and density of data, leading to erroneous interpretation.

Mapping a Snapshot: Points to Polygons

A common first step in visualizing cancer data is generating dot (ie, point) maps, which can be generated when street address data are thoughtfully geocoded. Dot maps provide a first pass at visually assessing the distribution of data and should be designed carefully to protect confidential information. It is preferred that dot maps of exact addresses not be published and only used for inhouse, exploratory analysis. In addition, such maps can be misleading when not combined with population data, because a high concentration of points simply may be reflecting high population density. Although data-masking techniques exist to help protect confidentiality,[98,103–105] not publishing dot maps is preferable.

The most popular way of visualizing cancer data is aggregating point data to some geographic boundary, such as county boundaries. One challenge when mapping aggregated cancer data is the modifiable areal unit problem (MAUP), in which the choice of areal unit can change the observed geographic patterns. For example, maps created at the census tract aggregation unit may produce different geographic patterns than those aggregated to county or ZIP codes.[106]

Aggregated data typically are visualized using choropleth mapping methods, with specific colors assigned to specified rate ranges based on defined groups (eg, quartiles). Several useful tools are available for color selection. Color Brewer[107–109] (available at: http://colorbrewer2.org/, Accessed March 21, 2019) provides easily distinguishable, predefined color ramps. Complimentary tools, such as Color Oracle (available at: http://www.colororacle.org/) or other tools listed in https://www.color-blindness.com/2008/12/23/15-tools-color-blindness/, simulate color-impaired–specific outputs (Fig. 4). Ultimately, the onus is on the researcher to make sound judgements regarding the color choices and breakpoints used to distinguish classes of rates based on the data's distribution and levels of significance.

Rates based on small numbers of cases should not be displayed to protect confidentiality or to avoid unreliable estimates.[110] Options are to observe the standard deviation of the range of rates (eg, confidence intervals) or to present only statistically significant, meaningful results. Other options include methods to aggregate or merge neighboring geographic units together until a userdefined population and/or number of cases is reached, minimizing the standard error.[111–113] Several such algorithms are discussed below (see Spatial Analysis). Often, researchers need to display multiple layers of data and results from complicated epidemiologic and/or geospatial statistical analyses. Bivariate mapping techniques and

bivariate choropleth maps, depicting 2 variables, allow researchers to overlay additional data pertinent to understanding the context for cancer risk.[107,114–116] Figure 5 provides an example of bivariate mapping of smoking rates and estimates of radon-attributable lung cancer mortality. GIS, online tools, and tutorials are available for creating bivariate choropleth maps.[117,118]

Mapping Trends Over Time

In addition to visualizing a snapshot in time, exploring trends over time is often important. Researchers can create a series of maps if data of the same quality and spatial scale are available across time. Micromaps[119] offer additional means for linking statistical information to features and visualizing and evaluating spatial data patterns and temporal trends. An interactive, online tool developed by the National Cancer Institute (available at: https://gis.cancer.gov/tools/micromaps/, Accessed August, 2018) facilitates comparisons of multiple variables and associated risk factors across regions and time. Micromaps and graphs created using such tools are easily linked to identify trends in changes over space and time and can provide comparisons between rates (Fig. 6).

Interactive Mapping

The visualization of data does not have to be limited to static maps and there are numerous, easy to use interactive mapping software available. Besides free online tools, commercial products such as InstantAtlas (available at: http://www.instantatlas.com/, Accessed March, 2018) and ArcGIS Online (available at: https://www.arcgis.com/home/index.html, Accessed March, 2018) facilitate the production of interactive maps and allow users to share and publish cancer data with interactive features. Users can link map data to graphs, tables, and charts and may generate animations to explore patterns over time. In addition, there is a burgeoning field of innovative data visualization techniques to explore relationships between multiple views, simultaneously exploring geographic patterns of cancer rates along with potential cancer risk factors, such as environmental exposures. For example, the New York State Department of Health offers an interactive dashboard enabling users to explore environmental facilities and cancer (available at: https://apps.health.ny.gov/statistics/cancer/environmental_facilities/mapping/map/, Accessed March 25, 2019). Numerous online interactive cancer mapping applications are available, such as State Cancer Profiles, the American Cancer Society Cancer Atlas (available at: http://canceratlas.cancer.org/data/#?view=map, Accessed March 20, 2019), NAACCR Cancer Maps (available at: http://www.cancer-rates.info/naaccr/, Accessed March 20, 2019), and the US Cancer Statistics Data Visualizations (available at: https://gis.cdc.gov/Cancer/USCS/DataViz.html, Accessed March 25, 2019).

Beware of Mapping Limitations

Of course, there are limitations to visualizing cancer data. One of the greatest challenges is the spatial scale of available data. Often, cancer research is limited to aggregated rates with limited locational specificity. When analyzing rates at a county level, for example, there is the underlying assumption that rates across the entire geographic area are homogenous. Another limitation associated with most geographic enumeration units, such as county-level data, is inconsistent size (or area) throughout a state or the country. In addition, researchers

must use careful consideration when faced with incomplete case reporting or analyzing rare cancer data. Finally, depending on the spatial scale of the data, we may not be able to answer important questions crucial to understanding the distribution of the disease.

With the expansion of spatial statistical techniques and methods, such as spatial cluster analysis, geographically weighted regression,[120,121] and mixed modeling methods, researchers can now explore complex relationships between multiple risk factors and changing patterns of disease over space and time. Through GIS, the results of these highly sophisticated methods can more easily and effectively be communicated. In the section below, we further explore spatial analysis methods that allow researchers to transition from the traditional, visual and subjective interpretation of data to more quantitative spatial statistical methods.

SPATIAL ANALYSIS

Cancer incidence, mortality, treatment, and survival vary by geography. These deviations have important implications for the development and implementation of prevention strategies as well as for further understanding the etiology of cancer.[122] Spatial analysis is a statistical approach that can be applied to further understand the complex pathway of cancer development by integrating physical, social, and cultural environmental factors into the analysis.[123] Researchers can apply a spatial approach to epidemiology to identify geographic patterns and test geographic hypotheses, postulate about a community's health, focus public health action, and choose suitable prevention interventions.

Key Concept: Spatial Autocorrelation

A key precept in geography is Tobler's law or the "first law of geography," which states that "everything is related to everything else, but near things are more related than distant things".[124] In statistics, this is known as autocorrelation; and, in spatial statistics, it is known as spatial autocorrelation. Spatial autocorrelation is incorporated into different spatial analysis methods.[125] General global tests, such as the Moran's I and the Geary's C are designed to assess spatial autocorrelation. These tests generally are used when the focus of inquiry is not on place itself but on determining whether the analysis needs to be adjusted for location. Epidemiologists might use this approach to assess the impact of poverty on neighborhood health.

Traditionally, epidemiologists mapped disease rates to identify high-risk populations, but rates in sparsely populated areas can be outliers or may be statistically insignificant, leading to unwarranted alarm or inappropriate disregard.[126] Also, areas with small numbers of cases or small populations may not meet the threshold for statistical stability, but the differences still may have public health significance. Researchers may choose to adjust estimates toward neighboring values or toward a local mean using smoothing algorithms that incorporate data from neighboring or adjacent areas.[127] As an extension of Tobler's law, spatial smoothing assumes that rates are more similar and will not vary much between areas close to each other; therefore, differences among neighbors are likely because of random variations. Numerous smoothing methods exist[128–130] that reduce random variation to more clearly demonstrate and evaluate spatial patterns, such as the true underlying distribution of cancer

rates. In addition, several methods have been developed for creating spatially adaptive cancer incidence, mortality, and survival map data.[110,131–133] Unfortunately, using spatial smoothing to manage the variability of small numbers can sometimes mask true cancer patterns.

Identify and Assess Disease Patterns and Clusters

Other spatial analysis techniques are well suited for identifying and assessing geographically based disease patterns, such as addressing concerns about potential disease clusters. Spatial statistical methods like spatial regression algorithms and Bayesian space-time models also can quantify patterns and trends over space and time (spatiotemporal) and are available in different statistical applications (SAS [SAS Institute Inc, Cary, NC], R [R Foundation for Statistical Computing, Vienna, Austria], etc). Researchers at population-based cancer registries frequently respond to public cancer cluster concerns. SaTScan cluster detection free-ware (developed by Martin Kulldorff and Information Management Services Inc, Calverton, MD) uses spatial scan statistics[134] to evaluate geographically based disease risk. This method generates circles (or ellipses) of various sizes and evaluates observed versus expected rate ratios (risk within vs outside the circles) to identify statistically significant "clusters" of disease rates, including clustering over time.[135] Models to evaluate clusters are available for different data types, including the Poisson model for cancer rates, the Bernoulli and ordinal models for proportions like early versus late, and the exponential model for survival data. Recent work aims to extend the algorithm to detect linear,[136] empty-center circular, and ring-shaped hotspots.[137] Figure 7 illustrates how different models can help inform cancer control, depicting areas with higher than expected (Fig. 7, dark blue) versus lower than expected (Fig. 7, light blue) rates of colorectal cancer in Miami-Dade County, Florida, using the Poisson method and areas with higher than expected rates of late-stage versus early stage colorectal cancer (Fig. 7, purple hashing) using the Bernoulli method. Areas of lower or average expected incidence but high rates of late-stage versus early stage disease indicate areas that would benefit from increased population-based screening are circled in yellow. Areas with high rates of disease and late-stage disease are circled in orange and also may be good target populations for increased screening and important populations to evaluate, with the objective of gaining a better understanding of the risks of colorectal cancer. Colorectal cancer screening rates are well below public health targets, and such a combined approach can refine the focus of interventions and research to reduce cancer burden most efficiently.

Another common cluster analysis method is the Getis-Ord Gi* statistic,[138,139] which is available within the ArcGIS software package (Esri, Redlands, CA). It identifies coldspots/hotspots based on the "neighborhood" of each feature as derived from modeling spatial relationships among all features (like surrounding counties). Often, when using cancer mortality and incidence rates, we need to account for variations in feature (like county) size and/or the exclusion of some features because of suppression or missing data.[140] One option is to quantify spatial relationships based on both userdefined distance and the minimum number of required neighbors.[141] The result of the analysis includes the associated Z-score and *P* value, indicating the statistical significance of the cluster.[139] If possible, researchers may consider multiple analyses using different methodologies to assess the consistency and

reliability of the results. The results from these analyses can be used further to identify focus areas for interventions. Figure 8 depicts potential areas for target screening interventions by overlaying Federally Qualified Health Centers (FQHC) over identified pockets of elevated mortality rates.[140]

In contrast to global tests, focused tests evaluate clustering around specific geographic locations. A focused test evaluates the pattern of disease frequency based on proximity to a specific geographic coordinate.[142] The Lawson & Waller focused test, which is available in the proprietary software ClusterSEER (BioMedware, Ann Arbor, MA), detects clustering around a suspected point source of exposure.[143] For example, Figure 9 illustrates the results from an analysis of bladder cancers in Michigan that can accommodate residential mobility, [143] in which red squares indicate industrial sites with statistically significant, higher rates of bladder cancer in close communities. Although it may be useful in some applications, this approach is relevant only if distance is a valid proxy for exposure.

Spatial analysis also can be applied to answer neighborhood-level research questions. For example, is a risk merely a reflection of the aggregate risk of individuals (composition), or do different areas have different risks (context)? Traditional statistical approaches in epidemiology can assess individual-level behaviors and outcomes and can be used in tandem with spatial analysis to assess compositional effects. For instance, to identify a high-risk community for targeted intervention, a standard logit model can be applied. However, if area-based and case-level variables are used to describe the community, then hierarchical modeling is most appropriate, because it includes a random effect to account for both the direct (composition) and contextual effects. Such models are available in most statistical packages. For instance, Bayesian spatial regression analysis is available in spatial packages in R (eg, R INLA, CARBayes, R2BayesX; R Foundation for Statistical Computing),[144] and PROC GLIMMIX in SAS (SAS Institute, Inc), and several R packages can be used for hierarchical logistic regression models to model both census tract and county as both random and mixed effects (eg, nlme, lme4), with tracts nested within counties.

Spatial Analysis: Limitations

Spatial analysis is not without limitations. Today, the availability of spatial statistics software allows researchers to conduct prompt spatial analyses. However, care must be taken to understanding the underlying assumptions about the data to avoid erroneous results, and users should carefully and methodologically interpret the results. For instance, as mentioned above, the MAUP comprises 2 interrelated, geographically based problems.[125] First, the size and shape of the study area affect the results: this is known as the *zoning effect*.[28] The zoning effect is problematic because, like mapped results, results of spatial analysis can change, depending upon scale of the analysis. The spatial scan in SaTScan produces different results for different maximum scan window sizes. Because there is no clear optimal setting for scaling parameters,[145] multiple scans should be run at various maximum circle sizes to identify the most persistent core for each cluster.[145] Second, different results can be obtained at different units of analysis, such as block group versus census tract. This is known as the *aggregation effect* and can result in the loss of statistical power to detect clusters.[146] A focused test can be used to test for spurious clusters caused by aggregation errors, such as

lumping together based on ZIP code centroid versus actual street address. Missing data also can affect spatial analysis, resulting in geographic confounding. It is becoming common for researchers to use multiple imputation methods to impute missing values like stage at diagnosis or insurance status and to use geographic imputation methods to impute missing locational data.[147–149]

DISCUSSION

This review discusses the evolution, current state, and trends of geospatial science for cancer research and serves as a high-level overview of important topics. A geographic approach is a natural companion to epidemiologic research, and the use of spatial epidemiology has increased as health data are now commonly geocoded and healthfocused spatial computer applications are available. This trend is expected to continue and grow as such applications become more user-friendly and as more professionals are exposed to spatial thinking earlier in their careers and/or as students. However, relevant conclusions hinge on understanding the limitations of the data and the methods as well as the suitability of a spatial approach to epidemiologic research.

Results and conclusions from a spatial approach can inform evidence-based decision making and public policy and can support the implementation of communitylevel interventions and efficient resource allocation. As discussed in the paper, a researcher can arrive at their results using a multitude of methods that also vary in sophistication, emphasizing the usefulness of the progression from visual interpretation of static and interactive maps to spatial analysis and spatial statistical methods. Maps are very useful tools of communication and can provide easy to share visualizations for identifying focus areas and gaps in service as a snapshot in time as well as for examining spatial and temporal trends. Spatial analysis can enhance cancer control activities by identifying geographic areas with high-risk populations to target public health interventions in communities. Incorporating spatial statistical methods, such as cluster detection, into existing disease surveillance activities allows programs to use results and base decisions on the distribution of disease to respond to the public’s concern about potential cancer clusters in a scientifically rigorous manner.

Spatial epidemiology affords taking an interdisciplinary approach to cancer research, and a “thoughtful research” approach should be used, recognizing the strengths but also the limitations and constraints of the data, methods, and technology. Spatial analysis and spatial statistical software packages evolve, making it easier than ever to execute complex and specialized spatial analyses. Although it is an encouraging trend, researchers should consider collaborating with a geospatial scientist and/or spatial statistician to ensure that results can be interpreted correctly and to avoid misinformation and unintended policy decisions and intervention outcomes. Regardless, the advantages of applying GIScience to cancer research and “spatially enabling” cancer researchers, can have a profound impact on understanding patterns and trends in incidence and mortality, providing screening and treatment services, implementing effective prevention programs, and addressing geographic disparities.

FUNDING SUPPORT

No specific funding was disclosed.

REFERENCES

1. Price M Dr John Snow and an early investigation of groundwater contamination In: Mather JD, ed. 200 Years of British Hydrogeology. Special Publication 225. London, UK: Geological Society; 2004:31–49.
2. Cameron D, Jones IG. John Snow, the Broad Street pump and modern epidemiology. Int J Epidemiol. 1983;12:393–396. [PubMed: 6360920]
3. Koch T Cartographies of Disease: Maps, Mapping, and Medicine. Redlands, CA: ESRI Press; 2005.
4. Koch T Disease Maps: Epidemics on the Ground. Chicago, IL: University of Chicago Press; 2011.
5. Haviland A The Geographical Distribution of Disease in Great Britain. 2nd ed London, UK: Swan Sonnenschein; 1892.
6. Goodchild MF. Twenty years of progress: GIScience in 2010. J Spat Inform Sci. 2010;1:3–20.
7. Shekhar S, Chawla S. Spatial Databases: A Tour. Upper Saddle River, NJ: Prentice Hall; 2003.
8. Mark DM. Geographic information science: defining the field In: Duckham M, Goodchild MF, Worboys MF, eds. Foundations of Geographic Information Science. New York: Taylor & Francis; 2003:3–18.
9. University Consortium for Geographic Information Science (UCGIS). UCGIS bylaws. 2016 version. Washington, DC: UCGIS; 2016 Available at: http://www.ucgis.org/assets/docs/ucgis_bylaws_march2016.pdf. Accessed February 5, 2018.
10. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect. 2004:998–1006. [PubMed: 15198920]
11. Rushton G, Armstrong MP, Gittler J, et al. Geocoding in cancer research: a review. Am J Prev Med. 2006;30(2 suppl):S16–S24. [PubMed: 16458786]
12. Abe T, Stinchcomb DG. Geocoding practices in cancer registries In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data—The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/Taylor & Francis Group; 2008:195–223.
13. Bakshi R, Knoblock CA, Thakkar S. Exploiting online sources to accurately geocode addresses In: Cruz IF, Pfoser D, eds. Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems; November 12–13, 2004; Washington, DC. New York: ACM Press; 2004:194–203. https://dl.acm.org/citation.cfm?id=1032251. Accessed March 20, 2019.
14. Boscoe FP. The science and art of geocoding In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data— The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/Taylor & Francis Group; 2008:95–109.
15. Christen P, Churches T. A probabilistic deduplication, record linkage and geocoding system. In: Proceedings of the Australian Research Council Health Data Mining Workshop. Canberra, Australia: The Australian National University; 2005 http://users.cecs.anu.edu.au/~Peter.Christen/publications/arc-health-dm-2005-paper.pdf. Accessed March 20, 2019.
16. Davis CA Jr, Fonseca FT. Assessing the certainty of locations produced by an address geocoding system. Geoinformatica. 2007;11:103–129.
17. Goldberg D A Geocoding Best Practices Guide. Springfield, IL: North American Association of Central Cancer Registries (NAACCR); 2008.
18. Goldberg D Improving geocoding match rates with spatially-varying block metrics. Trans Geogr Inform Syst. 2011;15:829–850.
19. Goldberg D, Ballard M, Boyd JH, et al. An evaluation framework for comparing geocoding systems. Int J Health Geogr. 2013;12: 50. [PubMed: 24207169]
20. Goldberg DW, Cockburn MG. Toward quantitative geocode accuracy metrics In: Tate NJ, Fisher PF, eds. Accuracy 2010 Leicester, UK: Accuracy; 2010:329–332.
21. Goldberg D, Wilson J, Knoblock C. From text to geographic coordinates: the current state of geocoding. Urisa J. 2007;19:33–47.
22. Hutchinson M, Veenendall B. An agent-based framework for intelligent geocoding. Applied Geomatics. 2013;5:33–44. https://link.springer.com/article/10.1007/s12518-011-0063-z. Accessed March 20, 2019.

23. O'Reagan RT, Saalfeld A. Geocoding Theory and Practice at the Bureau of the Census Statistical Research Report Census/SRD/RR-87–29. Washington, DC: US Census Bureau; 1987.
24. Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. Comput Environ Urban Syst. 2008;32: 214–232.
25. Zandbergen PA. Influence of street reference data on geocoding quality. Geocarto Int. 2011;26:35–47.
26. Armstrong MP, Greene BR, Rushton G. Using geocodes to estimate distances and geographic accessibility for cancer prevention and control In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data—The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/Taylor & Francis Group; 2008:11–36.
27. Beyer KMM, Schultz AF, Rushton G. Using ZIP codes as geocodes in cancer research In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data—The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/Taylor & Francis Group; 2008:37–68.
28. Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. Epidemiology. 2003;14:408–411. [PubMed: 12843763]
29. Cayo MR, Talbot TO. Positional error in automated geocoding of residential addresses [serial online]. Int J Health Geogr. 2003;2:10. [PubMed: 14687425]
30. Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL. Evaluation of the positional difference between 2 common geocoding methods. Geospat Health. 2011;5:265–273. [PubMed: 21590677]
31. Fulcomer MC, Bastardi MM, Raza H, Duffy M, Dufficy E, Sass MM. Assessing the accuracy of geocoding using address data from birth certificates: New Jersey, 1989 to 1996 In: Williams RC, Howie MM, Lee CV, Henriques WD, eds. Geographic Information Systems in Public Health: Proceedings of the Third National Conference (1998, San Diego). Atlanta, GA: US Agency for Toxic Substances and Disease Registry; 2000:547–560.
32. Geronimus AT, Bound J, Neidert LJ. On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics National Bureau of Economic Research (NBER) Technical Working Papers 0189. Cambridge, MA: NBER; 1995.
33. Gilboa SM, Mendola P, Olshan AF, et al. Comparison of residential geocoding methods in population-based study of air quality and birth defects. Environ Res. 2006;101:256–262. [PubMed: 16483563]
34. Goldberg D, Cockburn D. The effect of administrative boundaries and geocoding error on cancer Rates in California. Spat Spatiotemporal Epidemiol. 2012;3:39–54. [PubMed: 22469490]
35. Jacquez GM, Rommel R. Local indicators of geocoding accuracy (LIGA): theory and application [serial online]. Int J Health Geogr. 2009;8:60. [PubMed: 19863795]
36. Karimi HA, Durcik M, Rasdorf W. Evaluation of uncertainties associated with geocoding techniques. J Comput Aided Civil Infrastruct Eng. 2004;19:170–185.
37. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, 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? The Public Health Disparities Geocoding Project. Am J Epidemiol. 2002;156:471–482. [PubMed: 12196317]
38. Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. Am J Public Health. 2001;91:1114–1116. [PubMed: 11441740]
39. Krieger N, Waterman P, Chen JT, et al. ZIP code caveat: bias due to spatiotemporal mismatches between ZIP codes and US censusdefined areas—the Public Health Disparities Geocoding Project. Am J Public Health. 2002;92:1100–1102. [PubMed: 12084688]
40. Lovasi GS, Weiss JC, Hoskins R, et al. Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree [serial online]? Int J Health Geogr. 2007; 6:12. [PubMed: 17367520]
41. Mazumdar S, Rushton G, Smith BJ, Zimmerman DL, Donham KJ. Geocoding accuracy and the recovery of relationships between environmental exposures and health. Int J Health Geogr. 2008;7:13. [PubMed: 18387189]

42. Oliver MN, Matthews KA, Siadaty M, Hauck FR, Pickle LW. Geographic bias related to geocoding in epidemiologic studies [serial online]. Int J Health Geogr. 2005;4:29. [PubMed: 16281976]
43. Ratcliffe JH. Geocoding crime and a first estimate of a minimum acceptable hit rate. Int J Geogr Inform Sci. 2004;18:61–72.
44. Skelly C, Black W, Hearnden M, Eyles R, Weinsgtein P. Disease surveillance in rural communities is compromised by address geocoding uncertainty: a case study of campylobacteriosis. Aust J Rural Health. 2002;10:87–93. [PubMed: 12047502]
45. Strickland MJ, Siffel C, Gardner BR, Berzen AK, Correa A. Quantifying geocode location error using GIS methods [serial online]. Environ Health. 2007;6:10. [PubMed: 17408484]
46. Vieira V, Fraser A, Webster T, Howard GJ, Bartell S. Accuracy of automated and E911 geocoding methods for rural addresses [abstract]. Epidemiology. 2008;19:S352.
47. Ward MH, Nuckols JR, Giglierano J, et al. Positional accuracy of 2 methods of geocoding. Epidemiology. 2005;16:542–547. [PubMed: 15951673]
48. Wey CL, Griesse J, Kightlinger L, Wimberly MC. Geographic variability in geocoding success for West Nile virus cases in South Dakota. Health Place. 2009;15:1108–1114. [PubMed: 19577505]
49. Zandbergen PA. Improving environmental exposure analysis using cumulative distribution functions and individual geocoding [serial online]. Int J Health Geogr. 2006;5:23. [PubMed: 16725049]
50. Zandbergen PA. Geocoding quality and implications for spatial analysis. Geogr Compass. 2009;3:647–680.
51. Zandbergen PA. Geocoding accuracy considerations in determining residency restrictions for sex offenders. Criminal Justice Policy Rev. 2009;20:62–90.
52. Zhan FB, Brender JD, De Lima I, Suarez L, Langlois PH. Match rate and positional accuracy of 2 geocoding methods for epidemiologic research. Ann Epidemiol. 2006;16:842–849. [PubMed: 17027286]
53. Zimmerman DL. Statistical methods for incompletely and incorrectly geocoded cancer data In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data—The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/ Taylor & Francis Group; 2008:165–180.
54. Zimmerman DL, Fang X, Mazumdar S, Rushton G. Modeling the probability distribution of positional errors incurred by residential address geocoding [serial online]. Int J Health Geogr. 2007;6:1. [PubMed: 17214903]
55. Armstrong MP, Ruggles AJ. Geographic information technologies and personal privacy. Cartographica. 2005;40:63–73.
56. Beresford AR. Privacy issues in geographic information technologies In: Rana S, Sharma J, eds. Frontiers of Geographic Information Technology. Berlin, Germany: Springer; 2006:257–277.
57. Cho G Geographic information science, personal privacy, and the law In: Wilson JP, Fotheringham AS, eds. The Handbook of Geographic Information Science. Malden, MA: Blackwell; 2008: 519–539.
58. Gittler J Cancer registry data and geocoding—privacy, confidentiality, and security issues In: Rushton G, Armstrong MP, Gittler J, et al., eds. Geocoding Health Data—The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: CRC Press/ Taylor & Francis Group; 2008:210–211.
59. Onsrud HJ, Johnson JP, Lopez X. Protecting personal privacy in using geographic information systems. Photogrammetric Eng Remote Sensing. 1994;60:1083–1095.
60. Kriege N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health. 2005;95:312–323. [PubMed: 15671470]
61. Schootman M, Sterling DA, Struthers J, et al. Positional accuracy and geographic bias of 4 methods of geocoding in epidemiologic research. Ann Epidemiol. 2007;17:379–387.
62. Zandbergen PA. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads [serial online]. BMC Public Health. 2007;7:37. [PubMed: 17367533]
63. Krieger N Place, space, and health: GIS and epidemiology. Epidemiology. 2003;14:384–385. [PubMed: 12843759]

64. Bichler G, Balchak S. Address matching bias: ignorance is not bliss. Policing Int J Police Strategies Manag. 2007;30:32–60.

65. Drummond WJ. Address matching: GIS technology for mapping human activity patterns. J Am Plan Assoc. 1995;61:240–251.

66. Peipins LA, Graham S, Young R, et al. Time and distance barriers to mammography facilities in the Atlanta metropolitan area. J Community Health. 2011;36:675–683. [PubMed: 21267639]

67. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data [serial online]. Int J Health Geogr. 2006;5:58. [PubMed: 17166283]

68. Minnesota Population Center. National Historical Geographic Information System. Version 2.0. Minneapolis, MN: University of Minnesota; 2011.

69. Spielman SE, Folch D, Nagle N. Patterns and causes of uncertainty in the American Community Survey. Appl Geogr. 2014;46:147–157. [PubMed: 25404783]

70. Hundepool A, Domingo-Ferrer J, Franconi L, et al. A Network Excellence in the European Statistical System in the Field of Statistical Disclosure Control (ESSNet SDC) Handbook on Statistical Disclosure Control. Version 1.2 Brussels, Belgium: European Commission; 2010.

71. Tatalovich Z, Wilson JP, Cockburn M. A comparison of Thiessen polygon, Kriging, and Spline models of potential UV exposure. Cartography Geogr Inform Sci. 2006;33:217–231.

72. Goovaerts P Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging [serial online]. Int J Health Geogr. 2006;5:52. [PubMed: 17137504]

73. Pickle L, Su Y. Within-state geographic patterns of health insurance coverage and health risk factors in the United States. Am J Prev Med. 2002;22:75–83. [PubMed: 11818175]

74. Raghunathan TE, Xie D, Schenker N, et al. Combining information from 2 surveys to estimate county-level prevalence rates of cancer risk factors and screening. J Am Stat Assoc. 2007;102:474–486.

75. Zhang X, Holt JB, Yun S, Lu H, Greenlund KG, Croft JB. 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;182:127–137. [PubMed: 25957312]

76. Yu M, Tatalovich Z, Gibson JT, Cronin KA. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. Cancer Causes Control. 2014;25:81–92. [PubMed: 24178398]

77. Johnson CM, Wei C, Ensor JE, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control. 2013;24:1207–1222. [PubMed: 23563998]

78. Pruitt SL, Leonard T, Zhang S, Schootman M, Halm EA, Gupta S. Physicians, clinics, and neighborhoods: multiple levels of influence on colorectal cancer screening. Cancer Epidemiol Biomarkers Prev. 2014;23:1346–1355. [PubMed: 24732630]

79. Sloan CD, Jacquez GM, Gallagher CM, et al. Performance of cancer cluster Q-statistics for case-control residential histories. Spat Spatiotemporal Epidemiol. 2012;3:297–310. [PubMed: 23149326]

80. Anderson AE, Henry KA, Samadder NJ, Merrill RM, Kinney AY. Rural vs urban residence affects risk-appropriate colorectal cancer screening. Clin Gastroenterol Hepatol. 2013;11:526–533. [PubMed: 23220166]

81. Henry KA, McDonald K, Sherman R, Kinney AY, Stroup AM. Association between individual and geographic factors and nonadherence to mammography screening guidelines. J Women Health (Larchmt). 2014;23:664–674.

82. Tatalovich Z, Stinchcomb DG, Lyman JA, Hunt Y, Cucinelli JE. A geo-view into historical patterns of smoke-free policy coverage in the USA [serial online]. Tobacco Prev Cessation. 2017;3:134.

83. Walker RE, Keane CR, Burke JG. Disparities and access to healthy food in the United States: a review of food deserts literature. Health Place. 2010;16:876–884. [PubMed: 20462784]

84. Henry KA, Stroup AM, Warner EL, Kepka D. Geographic factors and human papillomavirus (HPV) vaccination initiation among adolescent girls in the United States. Cancer Epidemiol Biomarkers Prev. 2016;25:309–317. [PubMed: 26768989]

85. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect. 2004;112:1007–1015. [PubMed: 15198921]

86. Tatalovich Z, Wilson JP, Mack T, Yan Y, Cockburn M. The objective assessment of lifetime cumulative ultraviolet exposure for determining melanoma risk. J Photochem Photobiol B. 2006;85:198–204. [PubMed: 16963272]

87. Teras LR, Diver WR, Turner MC, et al. Residential radon exposure and risk of incident hematologic malignancies in the Cancer Prevention Study-II nutrition cohort. Environ Res. 2016;148:46–54. [PubMed: 27015563]

88. Institute of Medicine, Board on Health Sciences Policy; Roundtable on Environmental Health Sciences, Research, and Medicine. Chapter 3: The links between environmental factors, genetics, and the development of cancer In: Wilson S, Jones L, Couseens C, Hanna K, eds. Cancer and the Environment: Gene-Environment Interaction. Washington, DC: The National Academies Press; 2002:25–35.

89. Boscoe FP, Johnson CJ, Sherman RL, Stinchcomb DG, Lin G, Henry KA. The relationship between area poverty rate and site-specific cancer incidence in the United States. Cancer. 2014;120:2191–2198. [PubMed: 24866103]

90. Henry KA, Sherman R, Farber S, Cockburn M, Goldberg DW, Stroup AM. The joint effects of census tract poverty and geographic access on late-stage breast cancer diagnosis in 10 US States. Health Place. 2013;21:110–121. [PubMed: 23454732]

91. Berrigan D, Hipp AJ, Hurvitz PH, et al. Geospatial and contextual approaches to energy balance and health. Ann GIS. 2015;21:157–168. [PubMed: 27076868]

92. Hoehner CM, Handy SL, Yan Y, Blair SN, Berrigan D. Association between neighborhood walkability, cardiorespiratory fitness and body-mass index. Soc Sci Med. 2011;73:1707–1716. [PubMed: 22030212]

93. Berrigan D, Tatalovich Z, Pickle LW, Ewing R, Ballard-Barbash R. Urban sprawl, obesity, and cancer mortality in the United States: cross-sectional analysis and methodological challenges [serial online]. Int J Health Geogr. 2014;13:3. [PubMed: 24393615]

94. Krieger N, Singh N, Waterman PD. Metrics for monitoring cancer inequities: residential segregation, the Index of Concentration at the Extremes (ICE), and breast cancer estrogen receptor status (USA, 1992–2012). Cancer Causes Control. 2016;27:1139–1151. [PubMed: 27503397]

95. Arnold K From crayons to computers: mapping cancer moves on. J Natl Cancer Inst. 2000;92:524–526. [PubMed: 10749903]

96. Mason TJ; National Cancer Institute (US), Epidemiology Branch. Atlas of Cancer Mortality for US Counties, 1950–1969. Department of Health, Education, and Welfare publication no. DHEW 75–780. Bethesda, MD: US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health; 1975.

97. Blot WJ, Fraumeni JF Jr. Lung cancer mortality in the United States: shipyard correlations. Ann N Y Acad Sci. 1979;330:313–315. [PubMed: 294182]

98. Centers for Disease Control and Prevention (CDC). Cartographic Guidelines for Public Health. Atlanta, GA: CDC; 2012.

99. Robinson A, Morrison JL, Muehrcke PC, Kimerling AJ, Guptill SC. Elements of Cartography. 6th ed New York: John Wiley & Sons, Inc; 1995.

100. Krygier J, Wood D. Making Maps: A Visual Guide to Map Design for GIS. New York: Guilford Press; 2016.

101. Dent BD. Cartography: Thematic Map Design. 5th ed. Little Rock, AR: William C. Brown Publishing; 1999.

102. Snyder JP. Map Projections Used by the US Geological Survey Bulletin 1532. Washington, DC: Department of the Interior, US Geological Survey; 1982.

103. Armstrong MP, Rushton G, Zimmerman DL. Geographically masking health data to preserve confidentiality. Stat Med. 1999; 18:497–525. [PubMed: 10209808]

104. Leitner M, Curtis A. Cartographic guidelines for geographically masking the locations of confidential point data. Cartographic Perspect. 2004;49:22–39.

105. Zandbergen PA. Ensuring confidentiality of geocoded health data: assessing geographic masking strategies for individual-level data [serial online]. Adv Med. 2014;2014:567049. [PubMed: 26556417]
106. Cressie NA. Change of support and the modifiable areal unit problem. Geogr Syst. 1996;3:159–180.
107. Brewer CA. Basic mapping principles for visualizing cancer data using geographic information systems (GIS). Am J Prev Med. 2006;30(2 suppl):S25–S36. [PubMed: 16458787]
108. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series. Ann Assoc Am Geogr. 2002;92:662–681.
109. Harrower M, Brewer CA. ColorBrewer.org: an online tool for selecting colour schemes for maps. Cartographic J. 2003;40:27–37.
110. Beyer KMM, Tiwari C, Rushton G. Five essential properties of disease maps. Ann Assoc Am Geogr. 2012;102:1067–1075.
111. Wang F, Guo D, McLafferty S. Constructing geographic areas for cancer data analysis: a case study on late-stage breast cancer risk in Illinois. Appl Geogr. 2012;35:1–11. [PubMed: 22736875]
112. Black R, Sharp L, Urquhart J. Analysing the spatial distribution of disease using a method of constructing geographical areas of approximately equal population size. IARC Sci Publ. 1996;135: 28–39; discussion 155–162.
113. Mu L, Wang F. A scale-space clustering method: mitigating the effect of scale in the analysis of zone-based data. Ann Assoc Am Geogr. 2008;98:85–101.
114. Fisher H Mapping Information: The Graphic Display of Quantitative Information. Cambridge, MA: Abt Books; 1982.
115. Leonowicz A Research on 2-variable choropleth maps as a method for portraying graphical relationships In: Proceedings of the 21st International Cartographic Conference (ICC): Cartographic Renaissance; August 10–16, 2003. Durban, South Africa Available at: https://icaci.org/files/documents/ICC_proceedings/ICC2003/Papers/400.pdf.
116. Tyner JA. Principles of Map Design. New York: Guilford Press; 2010.
117. Hallisey EJ, Henry J. Bivariate Choropleth Maps: Overview and “How To” for ArcGIS Paper presented at: GeoSWG Forum; 4 1–4, 2010; Atlanta, Georgia.
118. Buckley A ArcGIS Bivariate Mapping Tools Paper presented at: North American Cartographic Information Society (NACIS) Conference; 10 10–11, 2017; Greenville, South Carolina.
119. Carr DB, Pickle LW. Visualizing Data Patterns With Micromaps. Boca Raton, FL: Chapman & Hall/CRC Press; 2010.
120. Fotheringham AS, Brunsdon C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. New York: John Wiley & Sons, Inc; 2003.
121. Goovaerts P, Xiao H, Adunlin G, et al. Geographically-weighted regression analysis of percentage of late-stage prostate cancer diagnosis in Florida. Appl Geogr. 2015;62:191–200. [PubMed: 26257450]
122. Seidman CS. An introduction to prostate cancer and geographic information systems. Am J Prev Med. 2006;30(2 suppl):S1–S2. [PubMed: 16458783]
123. Graves BA. Integrative literature review: a review of literature related to geographical information systems, healthcare access, and health outcomes [serial online]. Perspect Health Inform Manag. 2008;5:11.
124. Tobler WR. A computer movie simulating urban growth in the Detroit region. Econ Geogr. 1970;46(supp1):234–240.
125. Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Vol 368 Hoboken, NJ: John Wiley & Sons, Inc; 2004.
126. Richards TB, et al. Choropleth map design for cancer incidence, part 1 [serial online]. Prev Chronic Dis. 2010;7:A24. [PubMed: 20040239]
127. Waller L, Carlin BP. Chapter 14. Disease mapping In: Gelfand AE, Diggle PJ, Fuentes M, Guttorp P, eds. Handbook of Spatial Statistics. Boca Raton, FL: Chapman & Hall/CRC Press; 2010:217–243.

128. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res. 2005;14:35–59. [PubMed: 15690999]
129. Mungiole M, Pickle LW, Simonson KH. Application of a weighted head-banging algorithm to mortality data maps. Stat Med. 1999;18:3201–3209. [PubMed: 10602145]
130. Osnes K, Aalen OO. Spatial smoothing of cancer survival: a Bayesian approach. Stat Med. 1999;18:2087–2099. [PubMed: 10441765]
131. Brunsdon C Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. Comput Geosci. 1995;21:877–894.
132. Tiwari C, Rushton G. Using spatially adaptive filters to map late stage colorectal cancer incidence in Iowa In: Fisher P, ed. Developments in Spatial Data Handling. New York: Springer-Verlag US; 2005:665–676.
133. Talbot TO, Kulldorff M, Forand SP, Haley VB. Evaluation of spatial filters to create smoothed maps of health data. Stat Med. 2000;19:2399–2408. [PubMed: 10960861]
134. Kulldorff M A spatial scan statistic. Commun Stat Theory Methods. 1997;26:1481–1496.
135. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model [serial online]. Int J Health Geogr. 2009;8:58. [PubMed: 19843331]
136. Tang X, Eftelioglu E, Oliver D, Shekhar S. Significant linear hotspot discovery. IEEE Trans Big Data. 2017;3:140–153.
137. Eftelioglu E, Shekhar S, Kang JM, Farah CC. Ring-shaped hotspot detection. IEEE Trans Knowl Data Eng. 2016;28:3367–3381.
138. Getis A, Ord JK. Local spatial statistics: an overview In: Longley P, Batty M, eds. Spatial Analysis: Modelling in a GIS Environment. New York: John Wiley & Sons, Inc; 1996:261–277.
139. Mitchell A The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics. Vol 2 Redlands, CA: ESRI Press; 2005.
140. Siegel RL, Sahar L, Robbins A, Jemal A. Where can colorectal cancer screening interventions have the most impact? Cancer Epidemiol Biomarkers Prev. 2015;24:1151–1156. [PubMed: 26156973]
141. ESRI. An overview of the Spatial Statistics toolbox. Redlands, CA: ESRI; 2018 Available at: www.resources.arcgis.com. Accessed March 20, 2019.
142. Lawson AB, Williams FL. An Introductory Guide to Disease Mapping. New York: John Wiley & Sons, Inc; 2001.
143. Jacquez GM, Shi C, Meliker JR. Local bladder cancer clusters in southeastern Michigan accounting for risk factors, covariates and residential mobility [serial online]. PLoS One. 2015;10:e0124516. [PubMed: 25856581]
144. Rue H, Rikebler A, Sorbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: a review. Annu Rev Stat Appl. 2017;4:395–421.
145. Chen J, Roth RE, Naito AT, Lengerich DJ, Maceachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality [serial online]. Int J Health Geogr. 2008;7:57. [PubMed: 18992163]
146. Ozonoff A, Jeffeery C, Manjourides J, White LF, Pagano M. Effect of spatial resolution on cluster detection: a simulation study [serial online]. Int J Health Geogr. 2007;6:52. [PubMed: 18042281]
147. Walter SR, Rose N. Random property allocation: a novel geographic imputation procedure based on a complete geocoded address file. Spat Spatiotemporal Epidemiol. 2013;6:7–16. [PubMed: 23973177]
148. Howlader N, Noone AM, Yu M, Cronin KA. Use of imputed population-based cancer registry data as a method of accounting for missing information: application to estrogen receptor status for breast cancer. Am J Epidemiol. 2012;176:347–356. [PubMed: 22842721]
149. Henry KA, Boscoe FP. Estimating the accuracy of geographical imputation [serial online]. Int J Health Geogr. 2008;7:3. [PubMed: 18215308]

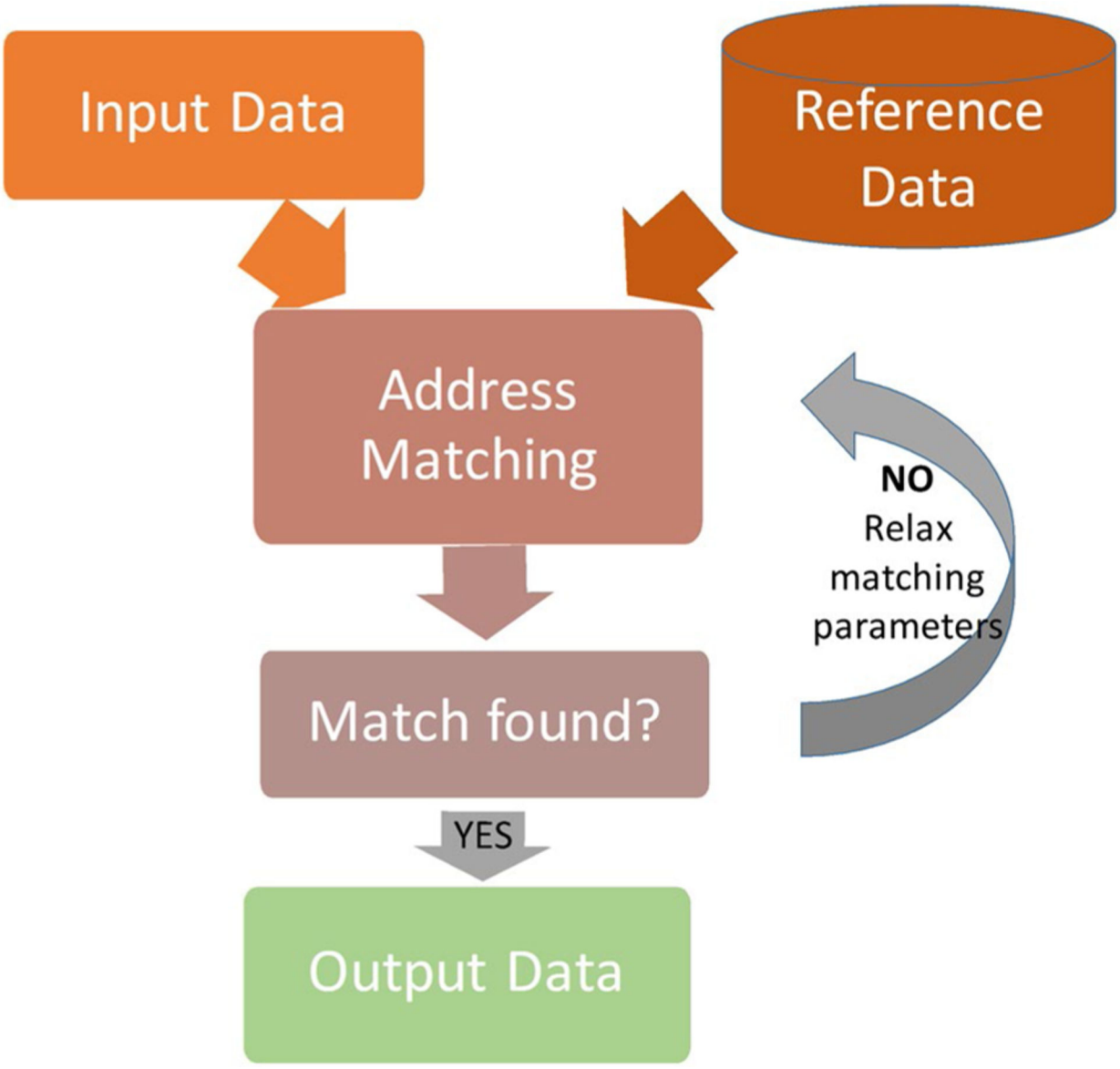

Figure 1.
This is a generalized schematic of the geocoding process.

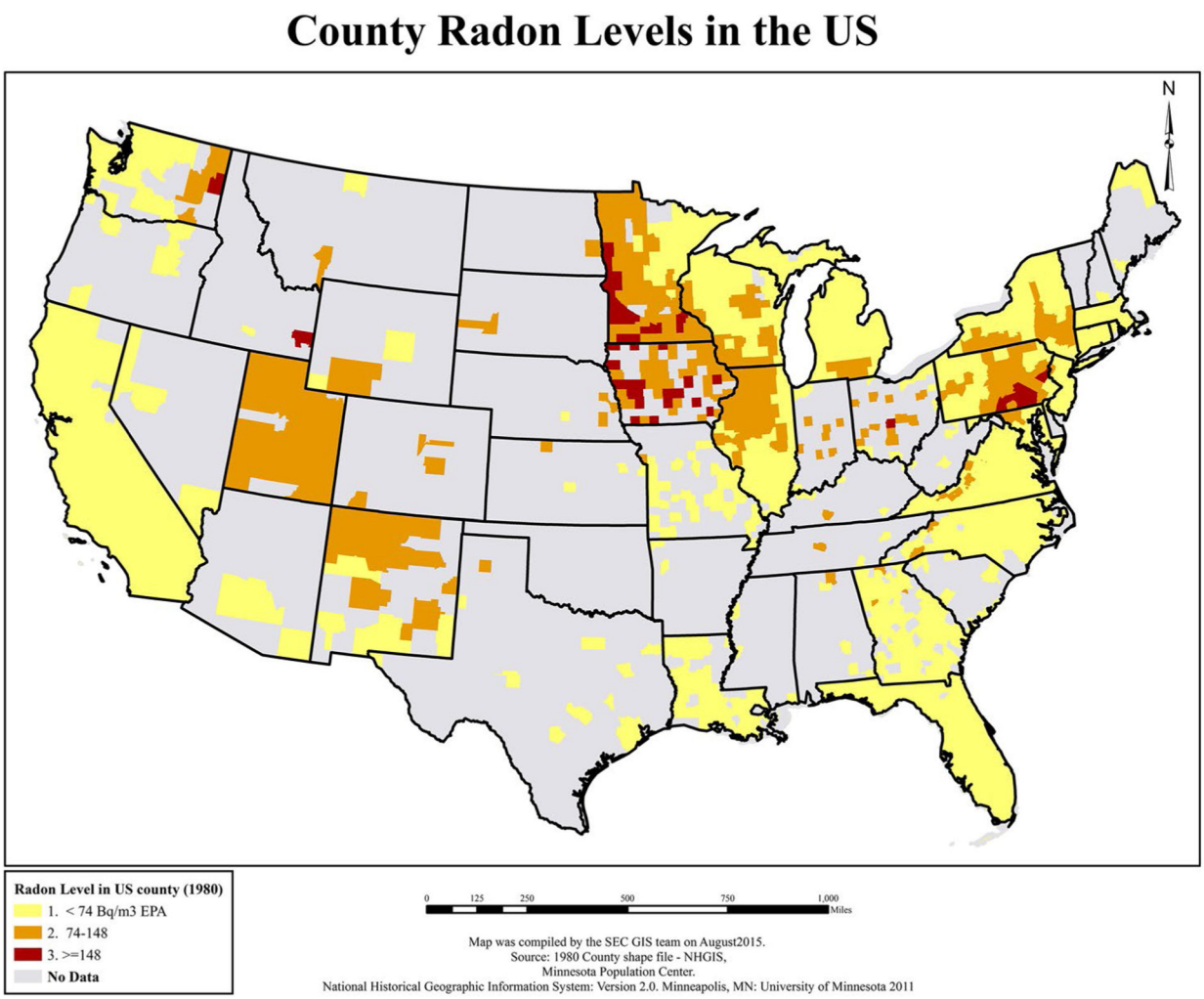

Figure 2.

Radon levels are illustrated according to 1982 US county of residence from the Cancer Prevention Study-II. EPA indicates Environmental Protection Agency; GIS, geographic information systems; NHGIS, National Historical Geographic Information Systems; SEC, Statistics and Evaluation Center. Reprinted from: Teras LR, Diver WR, Turner MC, et al. Residential radon exposure and risk of incident hematologic malignancies in the Cancer Prevention Study-II nutrition cohort. *Environ Res*. 2016;148:46–54, with permission from Elsevier.[87]

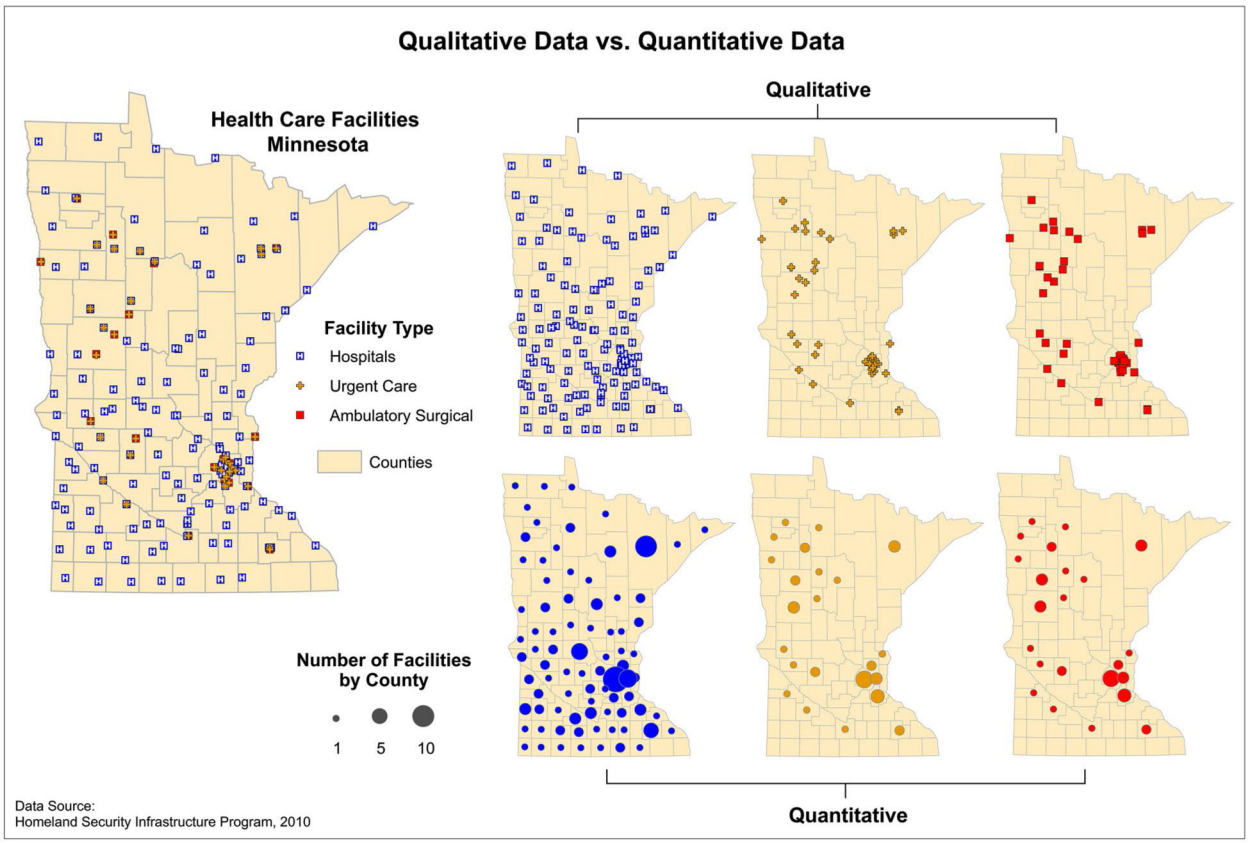

Figure 3.

Qualitative versus quantitative data are displayed. On the right, the *top* 3 maps (qualitative) depict facilities by type, and the *bottom* 3 maps (quantitative) depict the number of facilities by county. Source: Centers for Disease Control and Prevention Cartographic Guidelines for Public Health.[98]

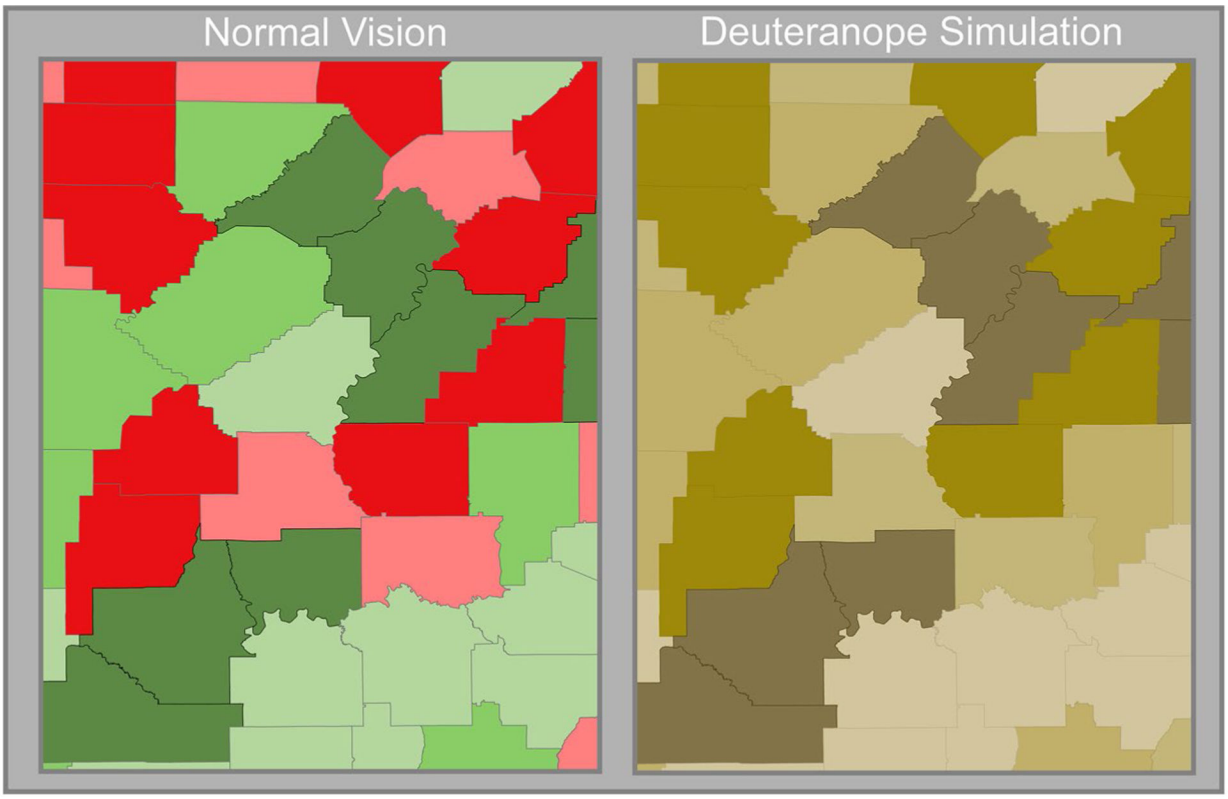

Figure 4.

Color schemes are depicted on a choropleth map of county lung cancer rates. Note that readers with deuteranopia cannot easily differentiate the red-green scheme.

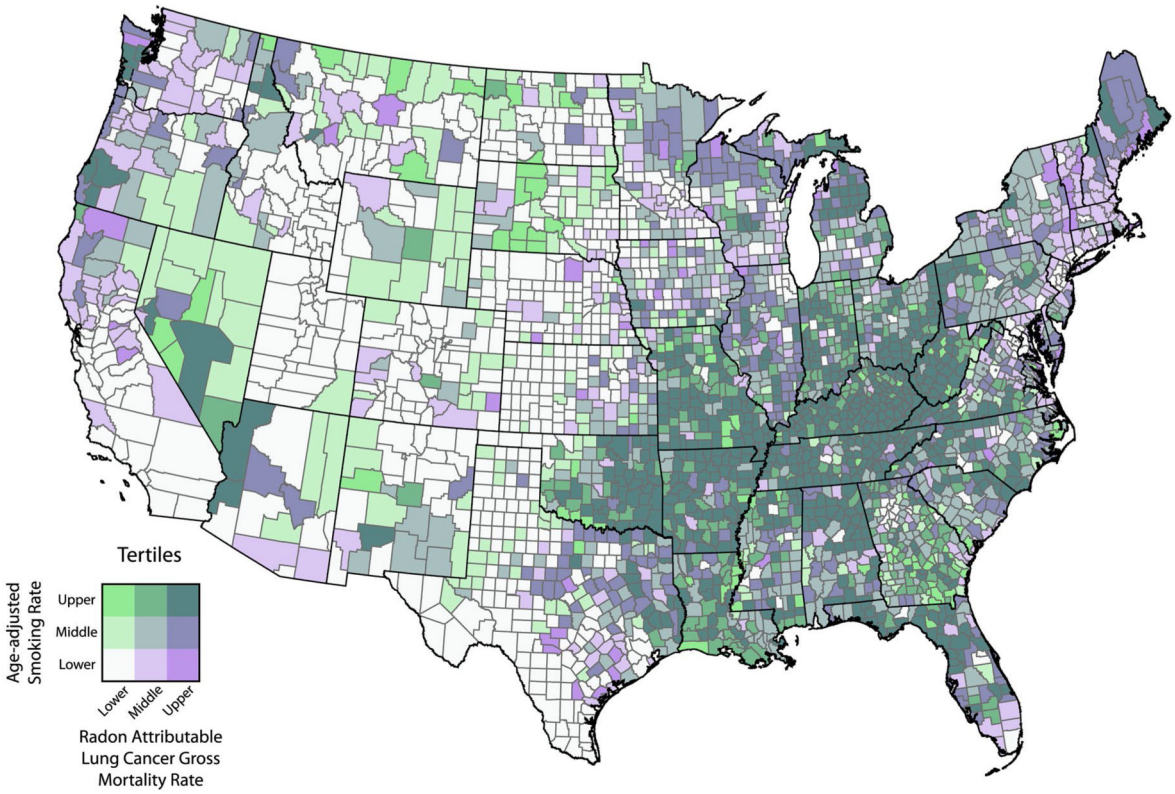

Figure 5.

Age-adjusted smoking rates and estimates of radon-attributable lung cancer mortality are illustrated. The darker purple counties represent the areas with the lowest smoking rates yet the highest estimated radon-attributable lung cancer deaths. The dark green counties are those areas with the highest smoking rates and the highest radon-attributable lung cancer mortality. The map was created by Andrew S. Berens at the Geospatial Research Analysis and Services Program (GRASP), Division of Toxicology and Human Health Sciences, Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention, Atlanta, GA.

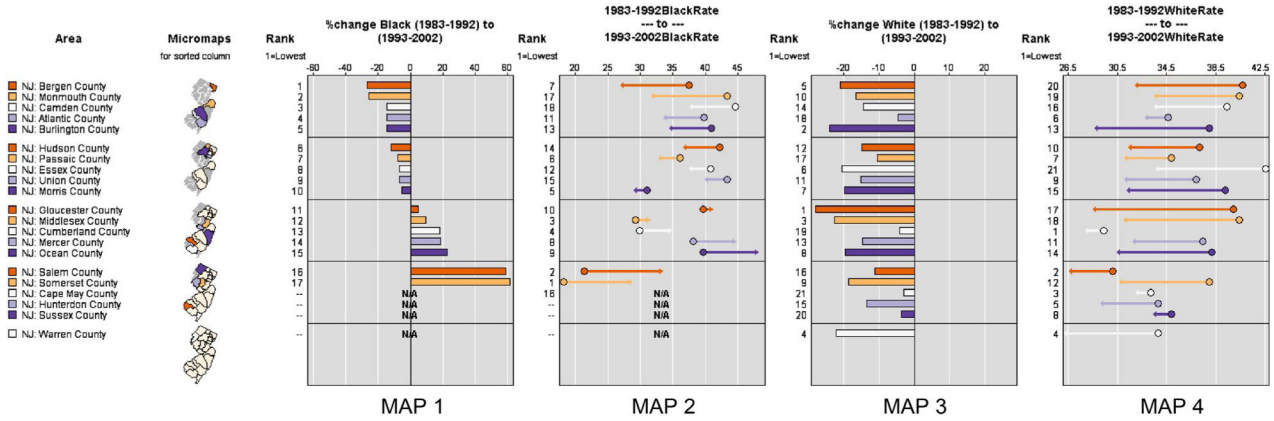

Figure 6.

This series of micromaps illustrates changes in breast cancer mortality rates in New Jersey (10-year aggregated rates) over 20 years. Maps 1 and 3 depict the percent change (%change) in blacks and whites over the 2 time periods, and maps 2 and 4 depict the actual change in rates. N/A indicates not applicable. Data source: National Center for Health Statistics, CDC.

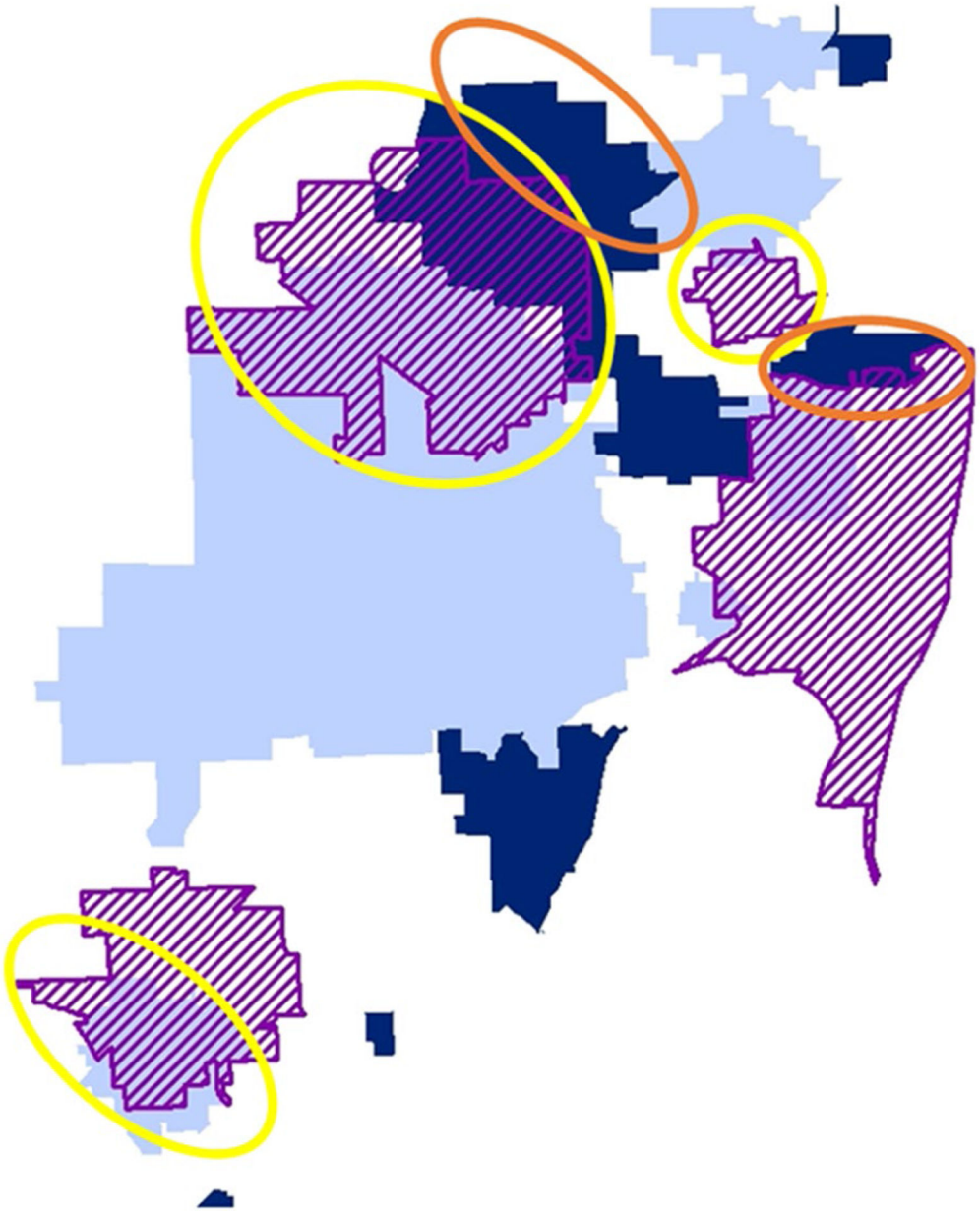

Figure 7.

Combined results from Poisson and Bernoulli cluster detection are depicted. Circled areas were identified for colorectal cancer screening prevention based on the overall risk of colorectal cancer and the risk of late-stage disease.

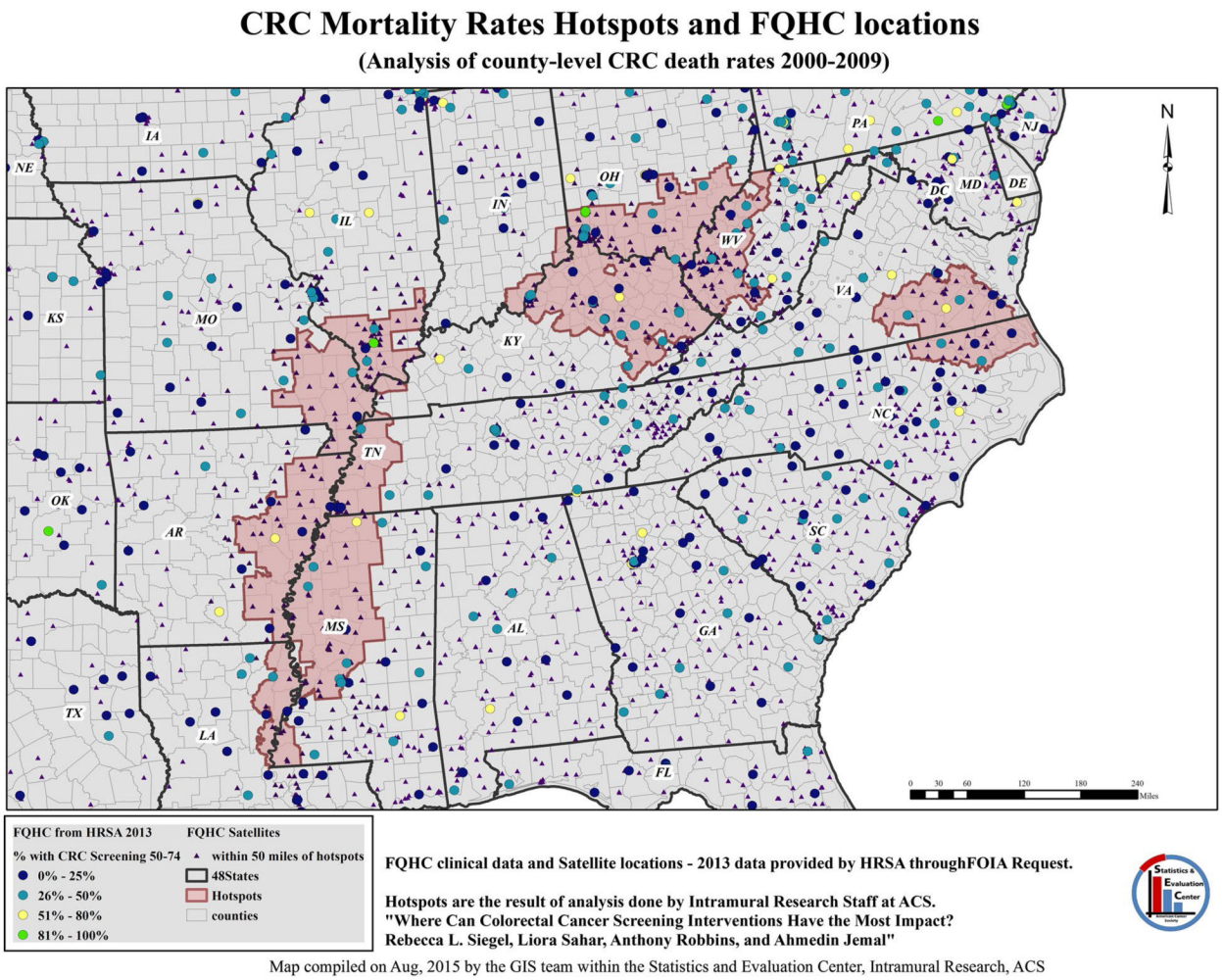

Figure 8.

Colorectal cancer (CRC) mortality rate hotspots are illustrated with a Federally Qualified Health Center (FQHC) location overlay. ACS indicates American Cancer Society; FOIA, Freedom of Information Act; GIS, Geographic Information System; HRSA, Health Resources and Services Administration.

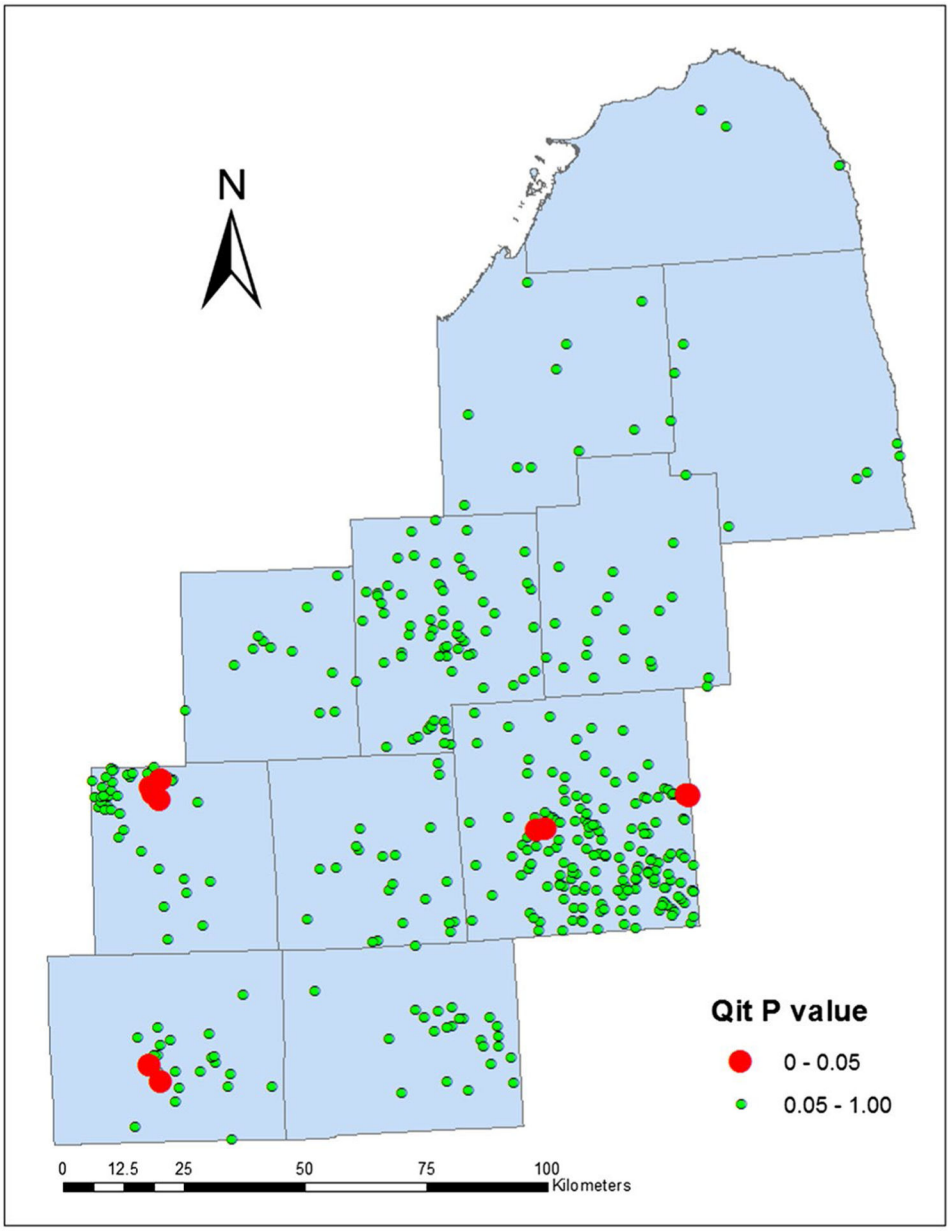

Figure 9.

Results of focused cluster detection of bladder cancers around industrial sites are illustrated. Q_{it} P value refers to the significance of the local space and time Q-statistics. Adapted from: Jacquez GM, Shi C, Meliker JR. Local bladder cancer clusters in southeastern Michigan accounting for risk factors, covariates and residential mobility. *PLoS One*. 2015;10:e0124516.[143]