Case-Control Geographic Clustering for Mobile Populations: Methods for Infectious and Chronic Diseases

Geoffrey M. Jacquez[1] and Katie E. Jacquez[2]

[1]BioMedware Inc, 516 North State Street
Ann Arbor, MI 48104-1236
Tel: +1 734-913-1098
FAX: +1 734-913-2201
Email: jacquez@biomedware.com

[2]Michigan State University, 5620 Cherry Hill Road
Ypsilanti, MI 48197 USA
Email: jacquezk@msu.edu

Abstract

Humans are mobile and constructs of Geographic Information Science have been used to model daily and weekly activity patterns, as well as residential and work spaces. But geographic epidemiology often ignores human mobility and employs methods that assume humans are sessile rather than mobile. This paper first quantifies how relaxing the assumption of sessile individuals might impact case-control cluster tests, and finds the results are highly sensitive to when the system is observed. Recently developed tests for case-control clustering that account for human mobility are then presented, along with extensions to the analysis of infectious disease data. We conclude by revisiting an analysis of bladder cancer in south eastern Michigan, and demonstrate the ability of the new techniques to detect global and local clustering in case-control data for residential histories. Statistical techniques that account for human mobility are needed for chronic and infectious diseases where causative exposures occur at locations different from ones location at time of diagnosis.

1. Introduction

Population surveys in the United States estimate that adults spend the majority of their day inside (87%), 69% of this at home, and 6% in a vehicle (Reuscher et al 2002). People are highly mobile, and this mobility evinces daily, weekly, and seasonal patterns. In addition, substantial portions of the US population frequently move their place of residence, with an average of once every 5-7 years. But most published disease cluster investigations ignore the dynamics of human mobility and instead assume static geographies in which individuals are immobile. Examples include the use of geocoded place of residence at time of diagnosis, death, and at time of birth, as well as the address of the admitting hospital to record locations of health events. A substantial body of literature in spatial epidemiology thus ignores human mobility, even though most researchers acknowledge that residential mobility should be accounted for, especially for diseases with long latencies such as cancer.

Present-day GIS software is not well suited to the representation of mobile individuals, nor to the

handling of information from temporally dynamic spatial systems (AvRuskin et al 2004). Goodchild (2000) called this the "static world view", and one tangible consequence is a lack of statistical methods for disease clustering that are suited to mobile individuals (Jacquez 2000). Space Time Intelligence Systems or STIS (Jacquez et al 2005a; Greiling et al 2005) address this weakness by implementing space-time data structures and constructs from time geography for representing human mobility. The STIS software is now being used to reconstruct exposure and provides a powerful visualization and analysis platform for undertaking space-time analyses in epidemiology (AvRuskin et al 2004; Meliker et al 2005). The technology also supports temporally dynamic geostatistical analysis (Goovaerts and Jacquez 2005) and the analysis of health disparities (Goovaerts 2005). Jacquez et al (2005b) recently proposed global, local and focused tests for case-control data with residential histories. This paper extends these methods to the analysis of infectious diseases, and revisits his analysis of bladder cancer in south eastern Michigan using more current data.

1.1 Setting the Problem

When considering temporal change in the geographic distribution of cases relative to controls, one might use place of residence of individuals from T years ago, and then allow T to vary in a range of several decades. The addresses of place of residence then will change through time, and one could simply apply a purely spatial cluster method to each change point. How might results vary depending on when one looks at the system (e.g. on selection of T)? Jacquez et al (2005b) demonstrated that this naïve approach can be misleading since it ignores the duration of each geographic "slice", and does not take the temporal dynamic into account when assessing cluster probabilities. They analyzed data from a population-based bladder cancer case-control study currently underway in south eastern Michigan. Cases are recruited from the Michigan State Cancer Registry and diagnosed in the years 2000-2004. Controls are frequency matched to cases by age (±5 years), race, and gender, and recruited using a random digit dialing procedure from an age-weighted list. Using Cuzick and Edwards (1990) Tk statistic with k=5 nearest neighbours they then analyzed these data at every point in time when the topology of place of residence of the cases and controls changed. The graph of Tk through time (Figure 1) is ascending, reflecting the larger number of cases and controls residing in the study area in later time periods. Clearly, results of cluster analyses that rely on single locations may be highly sensitive to the choice of the time for which the analysis is conducted.

2. Methods

We are interested in two types of methods, the first for infectious disease processes and the second for chronic diseases such as cancer. The first approach uses activity spaces and infection traces to model the individual contacts that spread infectious diseases such as SARS. We use SARS as a concrete example but note this approach could be extended to vector borne diseases without undue difficulty. The second set of methods was presented by Jacquez et al (2005b) and will be briefly summarized.

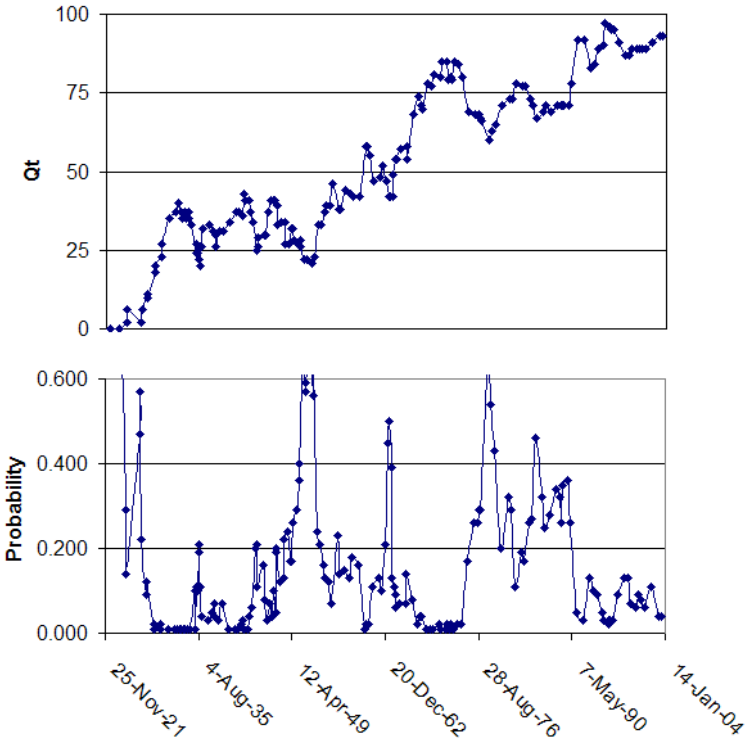

Figure 1. Graph of Cuzick and Edward's Tk statistic (top) and its Probability (bottom) through time. From Jacquez et al (2005b).

2.1 Methods for Infectious Diseases

We are interested in developing analysis approaches that make explicit the role super spreaders and infection foci play in epidemic spread. SARS transmission requires close contact between infected and susceptible individuals. Miller (2005) and others have defined activity spaces that represent the space-time locations of mobile individuals as they move throughout their day. We build on this construct to model *infection traces* defined as those portions of the activity space in which cases were infectious. Infection transmission events are possible only when the infection traces for cases intersect with the activity spaces of susceptible individuals in a fashion (e.g. of sufficient duration) that supports infection transmission. For example, by documenting and modelling the space-time geometry of activity spaces for SARS cases and controls, one could identify those characteristics of human mobility that are associated with infection transmission events in the Beijing SARS outbreak. We now present the modelling approach, and then define statistics for the clustering of infection traces

2.1.1 Methods for modelling activity spaces

Hagerstrand (1970) conceptualized the space time path as an individual's continuous physical movement through space and time, and visually represented this as a 3-dimensional graph. Hornsby and Egenhofer (2002) recognized that space-time paths mediate individual-level exposure to pathogens and environmental toxins, and that practical application would require a mechanism for representing location uncertainty. A space time prism refers to the possible locations an individual could feasibly pass through in a specific time interval, given knowledge of their actual locations in the

times bracketing that interval. The potential path area (Miller, 2005) shows the locations the individual could occupy given these constraints, and represents places where exposure events might occur. These constructs enabled new research approaches in diverse fields such as student life (Huisman and Forer, 1998), sports analysis (Moore et al, 2003), social systems (Kwan, 2000), transportation (Miller 1991), the analysis of disparities in gender accessibility in households (Kwan 2003), and the modelling of human activity spaces for both chronic and infectious diseases (Sinha and Mark 2005, Jacquez 2005).

2.1.2 Notation

Define the coordinate $\mathbf{u}_{i,t} = \{x_{i,t}, y_{i,t}\}$ to indicate the geographic location of the i^{th} case or control at time *t*. Activity spaces can then be represented as the set of space-time locations as:

$$\mathbf{L}_i = (\mathbf{u}_{i0}, \mathbf{u}_{i1}, ..., \mathbf{u}_{iT}) \qquad \text{(Equation 1)}$$

This defines individual *i* at location $\mathbf{u}_{i0}$ at the beginning of the study (time 0), and moving to location $\mathbf{u}_{i1}$ at time *t*=1. At the end of the study individual *i* may be found at $\mathbf{u}_{iT}$. *T* is defined to be the number of unique location observations on all individuals in the study. Activity spaces can be associated with time-dependent attributes such as infection status, case control status, and so on. We now define a case-control identifier, c_i, to be

$$c_i = \begin{cases} 1 \text{ if and only if } i \text{ is a case} \\ 0 \text{ otherwise} \end{cases} \qquad \text{(Equation 2)}$$

Define n_a to be the number of cases and n_b be the number of controls. The total number of individuals in the study is then $N=n_a+n_b$.

2.1.3 Nearest neighbour relationships for activity spaces

Let *k* indicate the number of nearest neighbours to consider when evaluating nearest neighbour relationships (see for example Jacquez 1996), and define a nearest neighbour indicator to be:

$$h_{i,j,k,t} = \begin{cases} 1 \text{ if and only if } j \text{ is a } k \text{ nearest neighbor of } i \text{ at time } t \\ 0 \text{ otherwise} \end{cases} \qquad \text{(Equation 3)}$$

We then can define a binary matrix of k^{th} nearest neighbour relationships at a given time *t* as:

$$\mathbf{?}_{k,t} = \begin{bmatrix} 0 & h_{1,2,k,t} & \cdot & \cdot & h_{1,N,k,t} \\ h_{2,1,k,t} & 0 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & h_{N-1,N,k,t} \\ h_{N,1,k,t} & \cdot & \cdot & h_{N,N-1,k,t} & 0 \end{bmatrix} \qquad \text{(Equation 4)}$$

This matrix enumerates the *k* nearest neighbours (indicated by a 1) for each of the *N* individuals. The entries of this matrix are 1 (indicating that *j* is a *k* nearest neighbour of *i* at time *t*) or 0 (indicating *j* is not a *k* nearest neighbour of *i* at time *t*). It may be asymmetric about the 0 diagonal

since nearest neighbour relationships are not necessarily reflexive. Since two individuals cannot occupy the same location, we assume at any time *t* that any individual has *k* unique *k*-nearest neighbours. The row sums thus are equal to *k* ($h_{i,\bullet,k,t} = k$) although the column sums vary depending on the spatial distribution of case control locations at time *t*. The sum of all the elements in the matrix is *Nk*.

There exists a 1 x *T*+1 vector of times denoting those instants in time when the system is observed and the locations of the individuals are recorded. We can then consider the sequence of *T* nearest neighbour matrices defined by

$$?_k^T = \{h_{k,t} \ \forall \ t = 0..T\} \qquad \text{(Equation 5)}$$

This defines the sequence of *k* nearest neighbour matrices for each unique temporal observation recorded in the data set, and thus quantifies how spatial proximity among the *N* individuals change through time.

2.1.4 Definition of Infection Traces

Infection traces are those portions of an infected individual's activity space that were traversed while that individual was infectious. An infected individual is defined as infectious during an infectious period (Δ_E). Depending on the disease, this may or may not be preceded by a latent disease period (Δ_L) in which the infected individual is not yet infectious. Given the activity space for case *i*, $\mathbf{L}_i$, denote the space-time coordinate at time of diagnosis as $\mathbf{u}_{i,t_D}$, noting that $\mathbf{u}_{i,t_D} \in \mathbf{L}_i$. We can then define that subset of the activity space $\mathbf{L}_i$ over which the infectious period occurred as:

$$\mathbf{L}_i^E = \{\mathbf{u}_{i,t} \ \forall \ (t_{i,D} - \Delta_E) > t > (t_{i,D} - \Delta_L - \Delta_E) \qquad \text{(Equation 6)}$$

Here $t_{i,D}$ is the time of diagnosis for individual *i*. The term $(t_{i,D} - \Delta_E)$ is the time when the infectious period began and $(t_{i,D} - \Delta_L - \Delta_E)$ indicates the time prior to diagnosis when the latent period began. Hence equation 6 denotes that portion of case *i*'s activity space in which s/he could have infected susceptible individuals. Call this the *infection trace*. Notice this infection trace assumes a natural history of infection in which the infection event occurs, is followed by a latency period, and then by a period in which the individual is infectious. Upon diagnosis we assume the patient is treated and his/her activities are curtailed to prevent infection of others. This is easily modified to fit other models of the natural history of infection, including those in which infectivity continues after diagnosis.

2.1.5 Definition of Sampling Distributions for Infection Traces

Once we know the sampling distributions for infection traces we can define statistics to identify super-spreaders in which infection traces cluster about certain individuals, as well as geographic areas traversed by many infection traces (locations of high infection transmission). To do this denote the distribution of infectious periods for the cases as Ψ_E. Notice this is a distribution of durations. This may be defined empirically as:

$$\hat{\Psi}_E = \{\Delta_{i,E} \ \forall \ i = 1,..,n_a\} \qquad \text{(Equation 7)}$$

Further, define the distribution of times of diagnosis as Ψ_D. This may be defined empirically as:

$$\hat{\Psi}_D = \{t_{i,D} \ \forall \ i = 1,..,n_a\} \qquad \text{(Equation 8)}$$

This is the distribution of points in time defined by the times of diagnosis of the cases. Finally, define the distribution of latency periods as Ψ_L. This may be defined empirically as:

$$\hat{\Psi}_L = \{\Delta_{i,L} \ \forall \ i = 1,..,n_a\}. \qquad \text{(Equation 9)}$$

In order to evaluate whether infection traces of the cases cluster we first must construct a procedure for generating representative times of diagnosis, latent periods, and infectious periods for the controls. Once this is accomplished we will be able to determine whether the infection traces for the cases cluster relative to those so constructed for the controls. Given the activity space of a control, steps involved to accomplish this are:

(1) Set the "time of diagnosis" for each control to the time of diagnosis for the matched case.

(2) Define the exposure window and latency period for each control based on the covariates for each control as was accomplished for that control's matched case. Completion of steps (1) and (2) will result in infection traces for both cases and controls

(3) Randomly assign case-control identifiers across the residential histories with equiprobability conditioned on the total number of cases and the total number of controls.

(4) Calculate the desired test statistic for clustering of infection traces.

(5) Repeat steps 3 and 4 a desired number of times to construct the reference distribution of the statistic under randomization.

2.1.6 Statistics for Modelling Super Spreaders

Super spreaders are those individuals who infect many other individuals. The infection trace for case i ($\mathbf{L}_i^E$) records those places where that individual was while s/he was infectious. Now define an indicator, $e_{i,t}$, as:

$$e_{i,t} = \begin{cases} 1 \text{ if and only if time } t \text{ is within the infection trace for individual } i \\ 0 \text{ otherwise} \end{cases} \qquad \text{(Equation 10)}$$

When $e_{i,t}$ is 1, let us say the infection trace is "active". A local case-control test for spatial clustering of infection traces at time t is then:

$$Q_{i,k,t}^E = c_i \ e_{i,t} \sum_{j=1}^{N} h_{i,j,k,t} \ c_j \ e_{j,t} \qquad \text{(Equation 11)}$$

This is the count, at time t, of the number of k nearest neighbours of case i's infection trace that are

cases (and not controls) and whose infection traces also are active. Hence the statistic will be large when infection traces of a group of cases are active at about the same time and cluster about case *i*.

We can explore whether infection traces of cases tend to cluster spatially about certain individuals (e.g. super spreaders) through time. A statistic sensitive to this pattern is:

$$Q_{i,k}^{E} = \sum_{t=0}^{T} Q_{i,k,t}^{E} \qquad \text{(Equation 12)}$$

$Q_{i,k}^{E}$ will tend to be large when active exposure traces for cases tend to cluster around the active exposure trace of the i^{th} case.

2.1.7 Statistics for Modelling Sites of High Infection Transmission

We can also ask whether the infection traces of cases cluster about specific locations (e.g. mixing sites) that we refer to as a focus:

$$Q_{F,k,t}^{E} = \sum_{j=1}^{N} h_{F,j,k,t}\, c_j\, e_{j,t} \qquad \text{(Equation 13)}$$

Here $h_{F,j,k,t}$ is 1 if individual j is a k nearest neighbour of the focus at time t, and 0 otherwise. The statistic $Q_{F,k,t}^{E}$ is the count of the number of cases whose infection traces are k nearest neighbours of the focus at time t. Significance of this statistic may be evaluated by constructing infection traces for the controls as described earlier, and by then repeatedly allocating case-control identifiers across the N activity spaces that are k nearest neighbours of the focus in order to construct the reference distribution for $Q_{F,k,t}^{E}$. Notice this statistic can also be implemented through time in a manner analogous to equation 12.

2.2 Methods for Chronic Diseases

Jacquez et al (2005b) developed global, local and focused tests for case-control clustering of residential histories for use with chronic diseases such as cancer. These statistics are similar in concept to the ones presented above, using nearest neighbour relationships, case-control identifiers and activity spaces as defined earlier in Equations 1-5. But rather than having the activity space denote destinations or "stays" in a person's day, the locations recorded in Equation 1 are places of residence over the last 20 or more years during which causative exposures might have occurred. Jacquez et al (2005b) presented dozens of cluster statistics for assessing different aspects of space-time patterns. We will employ the duration-weighted versions of their statistics for global, local and focused clustering.

To determine whether there is statistically significant case clustering of residential histories throughout the study area and when the entire study time period is considered (a spatially and temporally global test) we will use statistic Q_k^w as defined in Equation A5 of their appendix. This will tell us whether there is overall global clustering of residential histories when the residential histories over the entire study period are considered simultaneously.

Should significant global clustering be found, we next use statistic $Q_{i,k}^{w}$, as defined in their Equation A6, to identify local clusters of residential histories. This statistic will be evaluated for each of the

cases to identify those cases with low p-values. Notice these local statistics are a decomposition of the global statistic into local contributions, and the sum of the local statistics is equal to the global statistic.

We will use statistic $Q^w_{F,k}$ from Equation A8 to determine whether bladder cancer cases cluster near the business addresses of industries known to emit bladder cancer carcinogens. This will allow us to evaluate whether there was statistically significant clustering about a given industry F (e.g. a specific metal-plating business) over the life of its operation. We will use statistic Q_{F,k,w_o} from Equation A7 to identify those time intervals when there was case clustering about industry F.

3. Bladder Cancer Data

A population-based bladder cancer case-control study is underway in south eastern Michigan. Cases diagnosed in the years 2000-2004 are recruited from the Michigan State Cancer Registry. Controls are frequency matched to cases by age (±5 years), race, and gender, and are being recruited using a random digit dialing procedure from an age-weighted list. To be eligible for inclusion in the study, participants must have lived in the eleven county study area for at least the past 5 years and had no prior history of cancer (with the exception of non-melanoma skin cancer). Participants are offered a modest financial incentive and research is approved by the University of Michigan IRB-Health Committee. The data analyzed here are from 219 cases and 437 controls. As part of the study, participants complete a written questionnaire describing their residential mobility. The duration of residence and exact street address were obtained, otherwise the closest cross streets were provided. See Jacquez et al (2005b) for geocoding procedures and accuracy.

Address histories were collected for those industries believed to emit contaminants associated with bladder cancer. These were identified using the Toxics Release Inventory (TRI 2000) and the Directory of Michigan Manufacturers (Manufacturer Publishing Co., 1946, 1953, 1960, 1969, 1977, 1982). Standard Industrial Classification (SIC) codes were adopted, but prior to SIC coding, industrial classification titles were selected. Characteristics of 268 industries, including, but not limited to, fabric finishing, wood preserving, pulp mills, industrial organic chemical manufacturing, and paint, rubber, and leather manufacturing, were compiled into a database. Each industry was assigned a start year and end year, based on best available data. Industries were geocoded following the same matching procedure as for residences: 89% matched to the address, 5% were placed on the road using best informed guess, and as a last resort, and 6% were matched to town centroid.

4. Results

Following Jacquez et al (2005b) we used k=5 nearest neighbours when evaluating the global, local and focused tests. The test for global clustering of residential histories under the duration weighted statistic was highly significant (Q^w_k =4853780063.0 case seconds, p=0.001 using 999 Monte Carlo simulations). This means there is significant clustering of residential histories of cases relative to controls over the entire study area and duration of the study. We thus are justified in decomposing this statistic into local contributions in order to identify those cases whose residential histories tend to be near other cases.

Using the local statistic we identified 6 cases that are the nexus of local case clusters ($Q^{w}_{i,k}$ ranged from 6027091200.0 to 4370889600.0 case seconds; p ranged from 0.004 to 0.009). We considered a local cluster significant only when its p-value was less than 0.01. Of these 6, five resided in Oakland county, and one in Lansing. The map of cases and controls in January, 1993 is shown in Figure 2. At that time four of the cases were on municipal water supplies that are monitored for elevated arsenic levels, and 2 were on private wells. Because they are subject to monitoring and government regulation, municipal sources do not have the high levels of arsenic that are observed in certain ground waters in the study area. It thus is unlikely these cases can be attributed solely to high drinking water arsenic concentrations, since their sources of supply were primarily municipal rather than private wells.

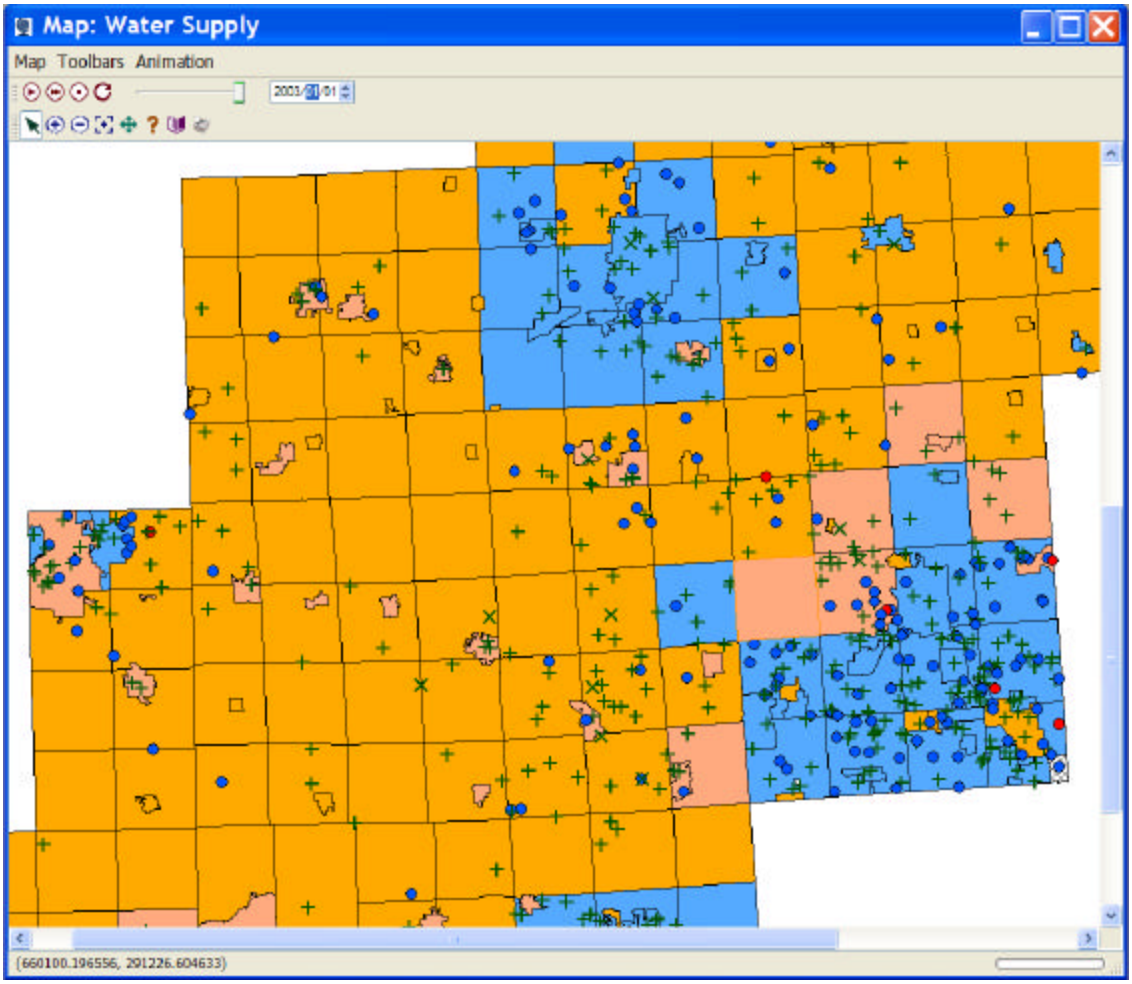

Figure 2. Map of significant local clusters of residential histories at t=01/01/2003. Circles are cases, crosses are controls. Red indicates cases whose residential histories have p-values less than 0.01. Underlying geography is the water supply source, with gold indicating private wells, and blue and tan municipal supplies.

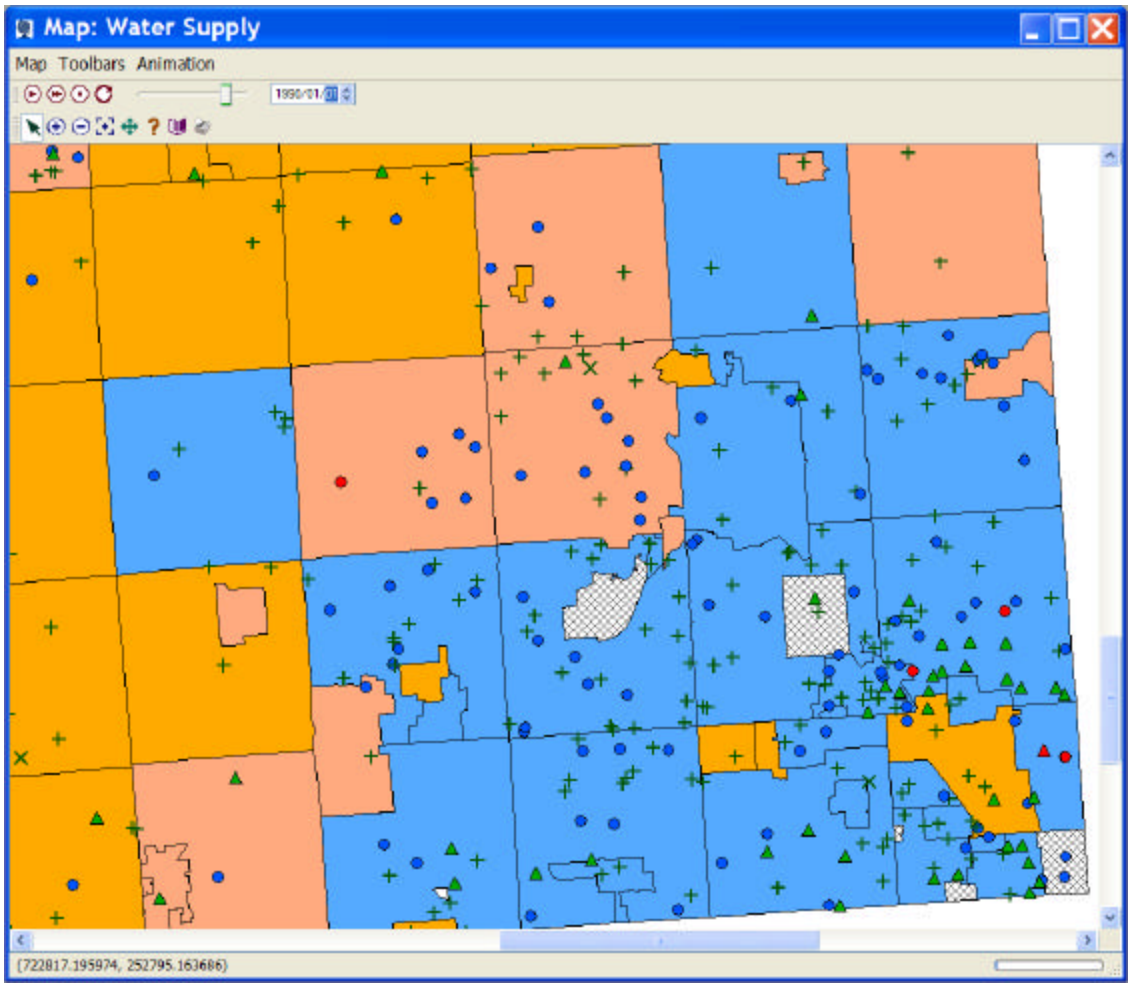

Figure 3. Map of significant focused clusters of residential histories at t=01/01/1990. Circles are cases, crosses are controls. Red indicates cases and industries whose address histories have p-values less than 0.01. The industry that is the focus of a significant cluster of residential histories of cases is shown as a red triangle.

Could the spatial and temporal patterns of these cases be attributed to industrial sources? To address this question we used the duration-weighted focus tests. We used the business address histories of the 268 industries described earlier and found one industry with a focused p-value less than 0.01 (Q_{F,k,w_o} =2398550400.0 case seconds; p=0.004; 999 Monte Carlo runs). This industry is located in Oakland County and has one of the significant bladder cancer cases as a first nearest neighbour (Figure 3).

5. Discussion

The results presented above are tentative as the study is in progress and cases and controls are still being enrolled. It is entirely possible that the significant clustering found from these incomplete data is an artefact of the order in which cases are being collected and will disappear when the complete and fully validated data set is available. Our purpose in analyzing these data has been for demonstration only, specifically to illustrate how these novel global, local and focused tests for

clustering of residential histories are applied.

This paper derived new statistics for the identification of super spreaders and foci of infection for infectious diseases, and revisited an analysis of a chronic disease (bladder cancer) to illustrate the identification of global, local and focused clustering of residential histories. The major difference between the methods for infectious vs. the methods for chronic diseases is in the definition of the activity space from which the nearest neighbour relationships are calculated. For infectious diseases the time span of the infection process is a few days, for cancer the time lag between causative exposure and diagnosis can be decades. Thus for infectious diseases we concern ourselves with the quantification of daily and weekly activity patterns; for chronic diseases we consider residential histories. The research described in this paper uses super sets of nearest neighbour matrices to define the space-time geometry of cases and controls. Constructs based on the notion of geospatial lifelines are used to create "threads" corresponding to daily activity patterns (for infectious diseases) and residential histories (for the bladder cancer example). Logical extensions to this approach include the mathematical definition of geometries representing weekly and seasonal activity patterns. The movement model used in this paper is a step function whereby an individual is considered to move instantaneously from one location to another. While this is a reasonable abstraction for residential histories, where the time spent moving is very brief compared to the time spend residing in a home, it may not be reasonable for modelling daily and weekly activity patterns. Further research is needed to develop modelling approaches appropriate for assessing daily exposures in the workplace and while commuting.

6. Acknowledgements

This research was funded by grant R43CA117171 from the National Cancer Institute. Data were provided by Dr. Jerome Nriagu under grant R01CA96002from the National Cancer Institute. The perspectives stated in this article are those of the author and do not necessarily represent the official views of the funding agency. Gillian AvRuskin geocoded the residential history data.

7. References

AvRuskin, GA, GM Jacquez, JR Meliker, MJ Slotnick, A Kaufmann and JO Nriagu. 2004. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *International Journal of Health Geographics* **3**:26.

Cuzick, J., and R. Edwards, 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society Series B* **52**: 73-104.

Greiling DA, Jacquez GM, Kaufmann AM, Rommel RG, 2005. Space time visualization and analysis in the Cancer Atlas Viewer. *Journal of Geographical Systems* 7:67-84

Goodchild, M. 2000. GIS and Transportation: Status and Challenges. *GeoInformatica* **4**: 127-139.

Goovaerts, P. 2005. Analysis and detection of health disparities using geostatistics and a space-time information system: The case of prostate cancer mortality in the United States, 1970-1994. In: *Proceedings of GIS Planet 2005, May 30-June 2, Estoril, Portugal.*

Goovaerts, P and G. M. Jacquez. 2005. Detection of temporal changes in the spatial distribution of cancer rates using LISA statistics and geostatistically simulated spatial neutral models. *Journal of Geographical Systems* **7**: 137-159.

Greiling, D.A., G.M. Jacquez, A.M. Kaufmann, and R.G. Rommel. 2005. Space time visualization and analysis in the Cancer Atlas Viewer. *Journal of Geographical Systems* **7**:67-84.

Hagerstrand,T. 1970. What about people in regional science? *Papers in Regional Science*, 24: 7-21.

Hornsby K, Egenhofer M. 2002. Modelling moving objects over multiple granularities, special issue on Spatial and Temporal Granularity. *Annals of Mathematics and Artificial Intelligence*, **36**:177-194

Huisman O, Forer P. 1998. Computational agents and urban life spaces: A preliminary realization of the time-geography of student lifestyles. In *Proceedings of the 3rd International Conference on GeoComputation*. University of Bristol.

Moore AB, Whigham P, Holt A, Aldridge C, Hodge K. 2003. A time geography approach to the visualisation of sport. *Proceedings of the 7th International Conference on GeoComputation* University of Southampton, United Kingdom 8 - 10 September, 2003. CD-ROM

Kwan MP. 2000. Extensibility and individual hybrid accessibility in space time: A multi scale representation using GIS. In *Information, Place and Cyberspace: Issues in Accessibility*. Edited by Janelle DG, Hodge DC. New York: Springer.

Kwan MP, Janelle DG and Goodchild MF. 2003. Accessibility in space and time: A theme in spatially integrated social science. *Journal of Geographical Systems*, **5**:1-3.

Sinha G, Mark DM. 2005. Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems* **7**: 115-136.

Jacquez, G. M., 2000, Spatial analysis in epidemiology: Nascent science or a failure of GIS?, *Journal of Geographical Systems*, **2**:91-97.

Jacquez, G. M., D. Greiling and A. Kaufmann. 2005a. Design and Implementation of a Space-Time Intelligence System for Disease Surveillance. *Journal of Geographical Systems* **7**:7-23

Jacquez GM, Kaufmann, J Meliker, P Goovaerts, G AvRuskin and J Nriagu. 2005b. Global, local and focused geographic clustering for residential histories. *Environmental Health: A Global Access Science Source* **4**:4.

Meliker J, Slotnick M, AvRuskin GA, Kaufmann A, Jacquez GM, Nriagu JO, 2005. Improving exposure assessment in environmental epidemiology: application of spatio-temporal visualization tools. *Journal of Geographical Systems* **7**:49-66.

Miller, H. 2005. A measurement theory for time geography. *Geographical Analysis* **37:** 17-45.

Reuscher TR, Schmoyer RL, Hu PS. 2002. Transferability of Nationwide Personal Transportation Survey data to regional and local scales. *Transportation Research Record* **1817**: 25-32

Toxics Release Inventory (TRI) Data Files, 2000. http://www.epa.gov/tri/tridata/tri00/data/index.htm