Chapter 10

SURVEILLANCE AND EPIDEMIOLOGY OF INFECTIOUS DISEASES USING SPATIAL AND TEMPORAL CLUSTERING METHODS

TA-CHIEN CHAN and CHWAN-CHUEN KING*

CHAPTER OVERVIEW

In the control of infectious diseases, epidemiologic information and useful clustering algorithms can be integrated to garner key indicators from huge amounts of daily surveillance information for the need of early intervention. This chapter first introduces the temporal, spatial and spatio-temporal clustering algorithms commonly used in surveillance systems – the key concepts behind the algorithms and the criteria for appropriate use. This description is followed by an introduction to different statistical methods that can be used to analyze the clustering patterns which occur in different epidemics and epidemic stages. Research methods such as flexible analysis of irregular spatial and temporal clusters, adjustment of personal risk factors, and Bayesian approaches to disease mapping and better prediction all will be needed to understand the epidemiologic characteristics of infectious diseases in the future.

Keywords: Infectious disease epidemiology; Geographical information system; Spatial epidemiology; Tempo-spatial cluster methods; Dengue; Influenza; Emerging infectious disease; Taiwan

---

* *Institute of Epidemiology, College of Public Health, National Taiwan University, 17 Xu-Zhou Road, Taipei (100), Taiwan, chwanchuen@gmail.com*

1.   INTRODUCTION

Spatial epidemiology has been commonly utilized to describe and to analyze the geographical distributions of diseases in recent decades. The distribution patterns of diseases are further investigated by several risk factors including demographic variables, levels of social economic status, environmental factors, genetic variations, exposure-related behaviors, contact patterns, specific niche of the etiologic agent, and modes of transmission [1, 2]. In general, descriptive epidemiologic studies present the mortality or incidence rate of an interesting disease by using thematic maps. The best example is John Snow's cholera map used in 1854 (Figure 10-1). Snow plotted all fatal cholera cases on the map to find that the contaminated pump was located on Broad Street in London, United Kingdom [http://www.ph.ucla.edu/epi/snow/snowmap1_1854_lge.htm]. In recent decades, the geographic information system (GIS) has been applied to understand the epidemiology of infectious diseases, particularly the relationship among agent, host and environment [3, 4]. And it even helped to eliminate cholera outbreaks in Bangladesh [5].
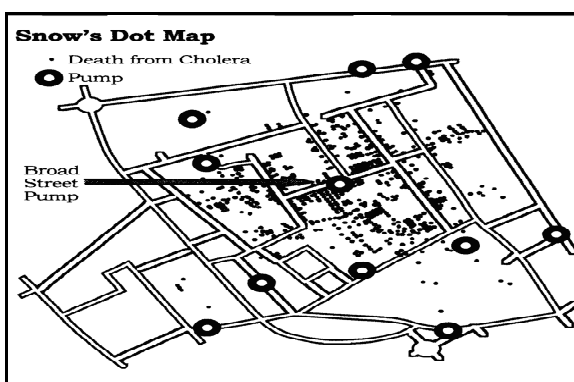

Figure 10-1. John Snow's dot map of cholera cases in 1854 (Source: http://www.ph.ucla.edu/epi/snow/snowmap1_1854_lge.htm).

Surveillance, a public health endeavor to monitor health data regularly by searching for evidence of a change, is the most cost-effective way to provide early warning signals and then to prevent outbreaks of infectious diseases [6]. The traditional analysis of geographical distribution of disease cases is generally to mark darker colors in a choropleth map[1] with the location of cluster cases that can be identified visually. This approach is easily misled

---

[1] **Choropleth Map:** A thematic map in which areas are distinctly colored or shaded to represent classed values of a particular phenomenon.

by the misclassification of symbology[2] or by neglecting temporal factors. On the other hand, much progress has been made in spatial techniques, which are frequently used to indicate the extent of "clustering" across a map. The follow-up spatial analysis can determine whether the increase in each epidemiologic measure is localized or general and even where high risk areas are located with statistically significant increases [7]. Furthermore, development of spatial and temporal clustering methods may provide a more integrated picture of the dynamic diffusion of disease cases that could block further transmission more effectively. In other words, the combination of surveillance, spatial techniques, and statistical methods – particularly the methods developed for characterizing the spatial and temporal clustering, can not only improve the surveillance system but can also enhance the effectiveness of the surveillance system to reach public health goals.

2.   CURRENT COMMONLY USED METHODS IN SPATIAL, TEMPORAL, AND TEMPO-SPATIAL CLUSTERING

Investigating disease clusters is an urgent task for public health authorities and professionals. If the disease happened non-randomly in temporal and spatial units, the clustering cases in time and place would be observed. Since outbreaks of emerging infectious diseases (EID) have been increasing rapidly in the past 2–3 decades, infectious disease surveillance becomes the most important task in public health. With the advances of information technology, electronic disease reporting systems have been established in many parts of the world. The real-time collection of disease information through the Internet is becoming more feasible [8]. However, numerous data need to be summarized. Therefore, the development of more convenient algorithms to detect temporal and spatial clustering is necessary to help public health staff with routine monitoring. In general, temporal clustering algorithms focus on setting up the baseline data for determining the threshold cut-off values. When the observation value exceeds the expected value, the alert signal will be triggered [8–10]. Spatial clustering algorithms test the null hypothesis, which assumes the disease is randomly distributed. If the null hypothesis is rejected by the predefined confidence level, the so-called "spatial clusters" would occur. Since time and place are the two most important epidemiological

---

[2] **Symbology:** The set of conventions, rules, or encoding systems that define how geographic features are represented with symbols on a map. A characteristic of a map feature may influence the size, color, and shape of the symbol used.

characteristics in infectious disease outbreaks, recently efforts have tried to consider both simultaneously.

2.1   Temporal Clustering Methods

Public health professionals have used three main methods – historical limit, cumulative sum (CUSUM), and time series to detect cases with temporal clustering.

2.1.1   Historical Limit, the Concept of Moving Average, and Scan Statistics

Historical Limit

Historical limit is a method that was frequently used to monitor infectious disease surveillance data in the United States before 2001 by requiring historical information – generally at least 5 years of background data – to serve as the upper baseline data for statistical aberration detection. If the observed value is higher than the 95% confidence limit of this upper baseline data, it is assumed that an outbreak would occur [10]. Therefore, the levels of baseline data in this method are easily influenced by the large-scale epidemic(s) of the past.

The Early Aberration Reporting System (EARS), developed by the Centers for Disease Control and Prevention in the United States of America (US-CDC), consists of a class of quality control (QC) charts, including the Shewhart chart (P-chart), moving average (MA), exponentially weighted moving average (EWMA), and variations of cumulative sum (CUSUM) [10]. In temporal analysis of syndromic surveillance data, a common approach is the use of a sample estimate for obtaining the baseline mean and standard deviation (SD) to circumvent the possible difficulties associated with the baseline trend that may be complicated by the seasonality and daily fluctuation of the syndromic data [9].

The Application of Moving Average

In 1989, Stroup et al. [11] used three simple moving average measures, moving average in mean, moving average in medium, and scan statistics, to implement historical limit methods on notifiable infectious diseases. The concept of analysis adopted the general form shown in Equation 10-1. The numerator X_0 was the observation value at the current time point (the temporal unit that can be daily or weekly or monthly, defined by the users). The denominator, serving as a baseline, was calculated as a mean or a median

value of the same time period plus the pre- and post-periods within the past 5 years (e.g., total 15 time points) (Figure 10-2).

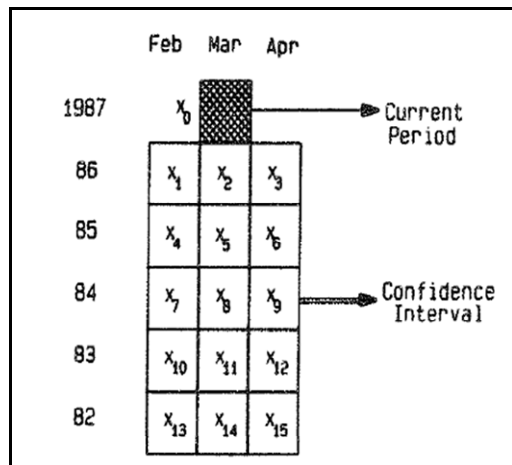$$\frac{X_0}{Baseline} \tag{10-1}$$

Figure 10-2. Baseline for comparison cases reported for March 1987 [11].

Since 1989, the historical limit method has been employed in the summarized surveillance results of the U.S.A. published in the Morbidity and Mortality Weekly Report (MMWR). The case numbers of a reported specific disease for a given health outcome in the three most recent time periods (pre-, current, post-) are compared with historical incidence data on the same health outcome from the same three time periods of the preceding 5 years. The results are shown by comparing the ratio of the current case numbers with the historical mean and SD. The historical mean and SD involve the 15 totals of the three time intervals, including the same pre-viously mentioned three periods (the current period plus the preceding one period, and the subsequent one period over the preceding 5 years) as the historical data. For example, if we want to know whether the influenza-like illness (ILI) cases in September of 2008 are unusual or not, the ILI case numbers of August, September and October in each year from 2003 to 2007 need to be added up to obtain a mean or a median for comparison. For an infectious disease with a strong seasonality trend, the seasonally adjusted CUSUM can be applied. That is, the positive one-sided CUSUM where the count of interest is compared with the 5-year mean $\pm$ SD for that specific period. Similarly, Taiwan's emergency department syndromic surveillance system can track diseases with strong holiday, post-holiday, or weekend effect because closures of most local clinics occur on most holidays/weekends.

To verify the accuracy and sensitivity of the outbreak detection, the epidemiologic investigation has to be followed. In 1993, Stroup et al. [12] compared the historical limit method with three other methods (bootstrap, jackknife, delta) for estimating standard error to detect abnormal time clusters. The results showed that the values estimated by using the historical limit method and delta method were close to the true value. The variance values estimated by the two methods were under-estimated, which might result in over-alert. Therefore, using bootstrap in the historical limit method to obtain an estimated confidence interval is a good statistical approach.

2.1.2   Cumulative Sum

Cumulative sum (CUSUM), a method initially used in quality control, has recently been applied to surveillance [13]. The original idea of CUSUM was to set up a control limit under a steady period. The strength of CUSUM, similar to the exponentially weighted moving average control chart, is to detect small shifts in the process mean even without historical data for 3–5 years. Two important parameters are used in CUSUM. First, an appropriate value for the control limit, h, is based on the average run length (ARL) of the CUSUM, while the failure rate is acceptable within a time interval for quality control that can be regarded as the upper limit of failure rate in quality control or the case number of a studied disease in surveillance. The other parameter is k, which represents the minimum standardized difference from the running mean. The traditional CUSUM chart generally uses the sum of differences both above and below the mean to detect anomalies in either direction. For biosurveillance, an upper sum S_H is used to look only for excessive counts in which small differences are ignored and only differences of at least 2k standard deviations above the mean μt (mean of day/week/other time unit t) are counted. A common practice is to set k at 0.5 to detect a shift of one SD.

Since the anthrax attack [10] **that** occurred shortly after the September 11, 2001 World Trade Center Attack, more interest has arisen in using public health approaches that could rapidly detect "unusual events" without requiring several years of background data. Therefore, newly developed non-historical aberration detection methods can analyze data as short as 1 week. To consider daily variation, the revised CUSUM method, using the measure-ments from C1, C2 to C3 to increase the sensitivity of the detection based on a positive one-sided CUSUM calculation from a week's information, was then developed [9]. For C1 and C2, the CUSUM threshold reduces to the mean plus 3 standard deviations (SD). The mean and SD for the C1 are based on the raw data from the past 7 days. The mean and SD for the C2 and C3 are based on the data from 7 days, ignoring the two most recent days to

minimize the bias. For C1 and C2, the test statistic on day t was calculated as $S_t = \max [0, (X_t - (\mu t + k * \sigma t))/\sigma t]$ where X_t is the count data (number of cases) on day t, k is the shift from the mean to be detected, and μt and σt are the mean and standard deviation of the counts during the baseline t time period. For C1, the baseline period is (t-7 to t-1); for C2 the baseline is (t-9 to t-3). The test statistic for C3 is the sum of $S_t + S_{t-1} + S_{t-2}$ from the C2 algorithm. Using these C1, C2 and C3, outbreaks of any infectious disease with a strong seasonal or regular fluctuation trend can be easily detected. This is particularly useful for an agent such as influenza virus, in which different types or subtypes of the virus are dominant each year in addition to continuous antigenic drifts.

2.1.3 Time Series

Based on the epidemiologic characteristics of each infectious disease, certain diseases have trends in the periodicity of epidemics. Therefore, researchers could use time series models such as the autoregressive integrated moving average model (ARIMA) or the Serfling model to predict the possible time and wave of the outbreak. The ARIMA models fit better with time series data that can be applied to better understand the characteristics of epidemiologic data or to predict future time points in a series. The fine-tuning characteristics of ARIMA involve adding lags of the different series and/or considering time lags of the forecast errors to the prediction equation to better predict the temporal trend. The Serfling model uses regression by adding sine and cosine functions to adjust the periodicity. It is frequently used in the excess mortality data analysis of influenza or pneumonia and influenza. Using these methods, the training period of the dataset to be selected is very critical. The cyclical pattern of time intervals such as seasons or months or other time units should be represented in the training data. Then, the dynamic pattern would be updated and predicted by the latest data. This time series model has recently been used in predicting the impact of several infectious diseases related to climate changes.

2.2 Spatial Clustering Methods

To analyze spatial data, the characteristic of the data – pointed data or regional data – needs to be examined first. In general, northern, southern, central and eastern Taiwan regional data are frequently used in routine surveillance for monitoring possible changes of several important infectious diseases in different geographical areas. Once the outbreaks occur, point data will be gathered by collecting the geo-coding information of the cases' addresses or by using a Global Positioning System (GPS) to locate any

important sites related to possible exposures of the cases for further detailed investigation.

The next step is to select appropriate methods for analysis of spatial clustering. Three methods of spatial clustering, including global cluster, local cluster and focused cluster, are frequently used for analyzing epidemiologic data [14]. Spatial autocorrelation, involving global indices and local indices as the degree of association between a factor of interest and its specific location, is a convenient approach to explore the degree of spatial clustering among cases and the possible associated spatial risk [14].

2.2.1   Global Clustering Test

Global cluster detection methods can help to determine whether or not spatial clustering exists in any place of the study period statistically [15]. Positive spatial autocorrelation reflects a "clustering" of points related to a particular variable of interest to be assessed. Negative spatial autocorrelation (e.g., spatial outliers) displays inverse correlation between the tested neighboring areas based on the attribute of interest. A zero spatial autocorrelation indicates a random distribution rather than a cluster or a dispersed distribution. This method is particularly useful if the source of infection is unknown or not easily identified. The limitation of this method is that it cannot identify the specific location(s) of spatial cluster(s).

Clustering tests involve four types – (1) area-based tests for global clustering, (2) point-based tests for global clustering, (3) area-based tests for local clustering, and (4) point-based tests for local clustering. Different statistical tests are used for each of these four types, depending on the type of data. Area data emphasize analysis on the relationship between the tested area and its neighboring area. Pointed data stress the distance between the two points. However, the central point of an area can be regarded as a point and then be tested in point data. Besides, both LISA and Moran's I spatial autocorrelation tests in Table 10-1 can be applied to point or polygon data, depending on the definition of the spatial relationship. If public health authorities have pointed data, more hypotheses can be tested and better diffusion dynamics of cases can be described. To protect patients' privacy, more area-type data are available than pointed-type data, particularly for those infectious diseases with higher social stigma.

Table 10-1 summarizes the clustering methods. The first two methods (Whittemore's test and K nearest neighbors) are global tests for pointed data and the next three methods are local tests for pointed data. Whittemore's test is to measure the mean distance of all cases divided by the mean distance of all individuals in that area. IF this ratio is less than 1, it reflects there is a cluster. In addition, the K nearest neighbor method assumes that the spatial

Table 10-1. Summary of the most commonly used spatial clustering algorithms.

Data format (point/polygon)	Type of method (global/local)	Statistical methods	Authors
Point	Global	Whittemore's test	Whittemore et al. (1987)
Point	Global	K nearest neighbors	Cuzick and Edwards (1990)
Point	Local	Geographical analysis machine (GAM)	Openshaw et al. (1987)
Point	Local	Besag and Newell test	Besag and Newell (1991)
Point	Local	Satscan	Kulldorff (1995)
Area	Global	Moran's I	Moran (1950)
Area	Global	Geary's C	Geary (1954)
Area	Local	Gi	Getis and Ord (1992)
Area	Local	Local indicator of spatial association (LISA)	Anselin (1995)

distribution of cases is random. If the observed value is higher than the expected value, it means a spatial cluster is present there. However, this test does not point out where the cluster is.

On the other hand, two global tests for area-type data are Moran's I and Geary's C tests. Moran's I statistic works by comparing the value at any one location with the value at all other locations. Moran's I is the most frequently used as the screening tool for clusters in global testing. It is generally used to reveal whether there is evidence of clustering or indication of the evidence of hot spots, shown by geographic boundary aggregated data. The results of Moran's I vary between −1.0 and +1.0. The Moran's $I > 0$, $=0$, and <0 indicate the positive spatial autocorrelation, random distribution, and negative correlation, respectively. If areas are close together with similar values, the Moran's I result is high. Geary's C statistic is used to describe differences at the local level by measuring the deviations in intensity values of each point with one another. The values of the C statistic vary between 0 and 2, where values equal to 1 represent spatial independence for each point, values less than 1 indicate evidence of positive spatial autocorrelation, and values between 1 and 2 indicate evidence of negative spatial autocorrelation.

2.2.2   Local Clustering Test

This method can provide definitive information on the specific location of clusters derived from local autocorrelation indices to evaluate clustering trends of an interested variable or factor, particularly under the condition with unidentified source of the infection, by determining whether the data

are spatially similar or different at that specific area/site [16]. Practically, public health personnel can use this method to define the risk areas of a disease for further prevention and control efforts. For example, the epidemic of dengue in 2002 in Taiwan was fast spreading. First, we investigated whether clustering occurred using the "global cluster" test. Then, the boundary between Kaohsiung City and Kaohsiung County was identified by "local cluster" test, and prevention and control efforts were immediately implemented. In infectious disease epidemiology, the local clustering test is very useful in investigating not only the source of the infection but also potentially unidentified risk areas that might facilitate subsequent diffusion and further spread of cases.

Satscan is a point-based test for local clustering whereas Local Indicator of Spatial Autocorrelation (LISA) is an area-based test for local clustering. Both methods are very frequently used. LISA divides the significant areas into four categories: (1) high-high for central area is high and neighboring area is also high, (2) high-low, (3) low-high and (4) low-low. The other area-type data local test is Gi and its calculation is quite simple. High Gi value represents the presence of clusters in high Gi value areas whereas low Gi value indicates the existence of cluster in low Gi value areas, similar to high-high and low-low areas in LISA, but it does not involve the other two categories in the LISA method.

Scan Statistic

Spatial scan method, initially used to detect clusters in cancer epi-demiology [17], has been applied to infectious diseases since 2000, such as bovine tuberculosis in Argentina, Toxoplasmosis in Southeast Asia, West Nile encephalitis in the United States, and human granulocytic ehrlichiosis near Lyme disease in Connecticut [18, 19]. The spatial scan can handle both point and area types of data, and it takes the central point of each polygon of the area-type data to be calculated. Nowadays, Satscan, which uses a circular window (circle centroid) to scan the entire study area to calculate the likelihood ratio, has become the most popular tool to detect diseases clusters. For any given location of the centroid, the radius of the screened window is continuously changed to take any value between zero to a certain upper limit.

The size of the circular window changes until the predefined population is screened. The maximum size of this circular window in the tested area has to be less than 1/2 of the target population to get a meaningful likelihood ratio in comparison with those in other areas and to avoid the overlap areas as well. After scanning the whole area, the area of the maximum likelihood ratio with statistical significance, so-called "clustering area," will be obtained.

However, the circular window in the spatial scan method is not the natural shape of most clusters. Therefore, an ellipse shape [20] and irregular shape [21] have been developed recently.

Local Indicator of Spatial Autocorrelation

In many studies, the LISA method has frequently served as the spatial risk index to identify both significant spatial clusters and outliers [22]. *Spatial outliers* present particular areas that have values of the tested variable opposite to its neighboring areas. The definition of a LISA index is:

$$I(i) = \frac{X_i - \bar{X}}{\delta} \times \sum_{j=1}^{n} W_{ij} \times \frac{X_j - \bar{X}}{\delta}$$

where I(i) = the LISA index for region i, W_{ij} = the proximity of region i to region j, where a value of 1 means the region i is next to the region j, and a value of 0 means the region i is far away from the region j, X_i = the value for the tested variable in region i, X_j = the value for the tested variable in region j, $\bar{X}$ = the mean value of the tested variable, δ = the standard deviation of X_i, and n = the total number of regions to be tested.

A positive *l(i)* value of the LISA index designates that a region and its neighboring areas tend toward local spatial dependency. In other words, area-specific cases of an interested infectious disease in the tested region and its neighboring areas approach homogeneity. In contrast, a negative *l(i)* value, which tends toward the opposite values between X_i and X_j (i.e., X_i = high, X_j = low or vice versa), specifies that the spatial dependency is negative, thereby suggesting that the region is a spatial outlier in relation to neighboring regions. In general, a Monte Carlo statistical test is used to evaluate the significance of spatial clusters and outliers [23]. Using LISA index values, risk areas of any infectious disease, such as dengue in southern Taiwan, have been classified into several different risk levels for implementing various control strategies to counteract outbreaks [24].

GAM and Besag and Newell Tests

The other two local clustering methods for pointed data are geographical analysis machine (GAM) and Besag and Newell tests. GAM is to test whether there is a statistically significant high disease rate by comparing each circle of the studied area with various radius values. The Besag and Newell test assumes that k is the minimum case number of the clustering area and then uses each case as a center to look for k-1 cases regarded as a cluster. In this way, the lacking neighboring cases force the investigator to

look for further areas so that the case number divided by larger searching area becomes a smaller value implying "cases without cluster." Both of these two methods result in the overlaps of sub-clusters (circles with different radius values or different k-1 cases), in which case the GAM method offers higher repetition without independence that may provide more false positives.

2.2.3   Focused Clustering Test

The "focused clustering" test is to assess the clustering of the observed cases around a fixed point – the smallest scope that is different from "general clustering" or "local clustering" without having any prior information on the centre of clustering. Therefore, this test has been used to investigate raised incidence of disease, particularly the rare disease or the beginning period of an outbreak of infectious disease, in the vicinity of pre-specified putative sources of increased risk. In addition, the focused clustering method is applied to detect whether there is an excess risk or a cluster of cases of a disease around a putative source of the infection [25]. Stone's test [26] is a very popular method used in testing "focused clustering" since it is based on traditional epidemiological estimates after adjusting the important confounders – standardized mortality ratio (SMR) or standardized incidence ratio (SIR).

The following summary Table 10-1 helps readers to firstly assess which type of spatial data – pointed format or area format – are collected. Then, global clustering tests can be employed to examine the presence of clusters or not. If the answer is "yes," subsequent local clustering tests will be followed to indicate the exact location of the case clustering areas. All these methods can be found in GIS software or free statistic test R packages.[3] Different statistic tests using the same datasets can also be simultaneously compared and evaluated to find out which one offers the best power. In general, spatial scan statistic has good power in detecting hot spot clusters.

2.3   Spatial and Temporal Clustering Methods

In addition to a spatial clustering method, temporal factors must also be taken into consideration. Analysis of spatial clustering data is quite similar to the data analysis in cross-sectional study design in epidemiology. When the distribution of the cumulative cases is displayed, it only explains the results of the overall pattern without definite conclusions on causal inference. Once

---

[3] **The R Package for Multidimensional and Spatial Analysis:** This is a group of programs (Macintosh and VAX/VMS) that allows public health data analyzer to perform with ease various complex multidimensional and spatial analysis procedures (http://www.r-project.org).

the temporal factors are included into the analysis, the results can clearly show the different waves of the epidemics, the transmission patterns, and possible risk factors that are involved in different time periods. Then, the three most important epidemiologic characteristics – person, place and time – can be simultaneously integrated to obtain more insights than each characteristic alone. Here we briefly introduce two methods, namely Knox method and the space-time scan statistic, which integrate spatial and temporal factors.

2.3.1 Knox Method

The Knox method is to test for space-time interactions, particularly when there are different impacts of time factor on the studied population in various regions [27, 28]. The time and geographical location of each case is obtained. For each possible pair of cases, the distances between them are also calculated in time and space. If many of the cases that are "close" in time are also "close" in space or vice versa, then there is a space-time interaction. Users can predefine how close the time period and the geographical distance are to one another of those studied cases in temporal and spatial units, based on their research questions. Then for each space-time combination, expected values will be calculated by a 2×2 contingency table [29]. Cases are assumed to be rare, independent events, distributed as a Poisson variable. The significance of the departure of the observed number of close pairs (O) from the expected number (E) is tested using d, where:

$$d = \frac{(O - E)}{\sqrt{Var(O)}}$$

A d value greater than 1.96 indicates that there is a statistically significant cluster at *p*-value 0.05.

The Knox test is attractive in epidemiologic data analysis because it is simple and straightforward to calculate the test statistic without requiring the knowledge of controls. However, the Knox test can be biased if the population growth is not constant for different geographical areas (e.g., distribution does not meet Poisson distribution). For detecting an "early" outbreak of infectious disease, such bias is not a major problem to be considered.

2.3.2 Space-Time Scan Statistic

Space-Time Scan Statistic [16], an improved version of the purely spatial scan method, is defined by a cylindrical window with a circular geographic base and with height corresponding to time. The base will vary the radius continuously. The height reflects any possible time interval of less than or equal to half the total study period. The likelihood in each cylinder will be

calculated. Using the cylinder with the maximum likelihood, and then selecting the tempo-spatial one with more than its expected number of cases, is denotes the most likely cluster.

Comparing the Knox and space-time distance methods, the Knox method categorizes the individual case's space-time distance into several groups and then uses a test similar to the Chi-square test. The temporal and spatial distance between the cases is determined by the user-based research questions. The space-time scan statistic, a purely spatial scan, uses the cylinder as the scanning window and the height is time. It scans over the study area by the different radii of the base to calculate the observed values in different areas. The expected value can be calculated by Monte Carlo simulation. Finally, the question on tempo-spatial clusters can be tested to determine whether the observed value exceeds the expected value. For example, if point data of individual cases from outbreaks of infectious diseases such as dengue or enterovirus-related cases are available, the Knox method is very suitable to apply. Alternatively, when an overall incidence or prevalence rate from different geographical regions rather than individual case data is available, the space-time scan method is more appropriate to use.

3.   CASE STUDIES USING SPATIAL CLUSTERING METHODS IN INFECTIOUS DISEASE EPIDEMIOLOGY

The following sections introduce the application of the above spatial and spatio-temporal methods to infectious diseases with public health significance, including respiratory spread, gastrointestinal-related (GI) transmission, vector-borne transmission, zoonotic and emerging infectious diseases.

3.1   Respiratory Spread

Epidemics of tuberculosis (TB) have reemerged in recent years due to the increasing cases of acquired immunodeficiency syndrome (AIDS) and multi-drug resistant tubercle bacilli. The incidence rate of TB in the Fukuoka Prefecture urban area of Japan (Figure 10-3a) in 2001 was higher than that of the nationwide data. Using local cluster tests for pointed data by spatial scan statistics and spatial-temporal scan statistics, the spatial analysis alone identified TB clusters in different geographical areas of Japan that occurred in different years (Figure 10-3b), including: (1) Chikuho coal mining area in 1999, 2002, 2003 and 2004, (2) Kita-Kyushu industrial area in 2000, and (3) Fukuoka urban area in 2001 [30]. However, using the space-time analysis, the most likely clusters were found in the Kita-Kyushu industrial area in

2000. In other words, clusters of cases had already appeared in the Kita-Kyushu industrial area in 2000 before the occurrence of other spatial clusters from 2002 to 2004. Further analysis found that the occurrence of TB in the clusters located in northern Fukuoka Prefecture in 2000 were also significantly higher than those clusters identified in other years. In conclusion, spatial method alone can be used to evaluate the cluster cases in each year whereas spatial-temporal methods can be applied to find out where cluster cases are within a specific time period and their dynamic changes over different time periods and places as well.
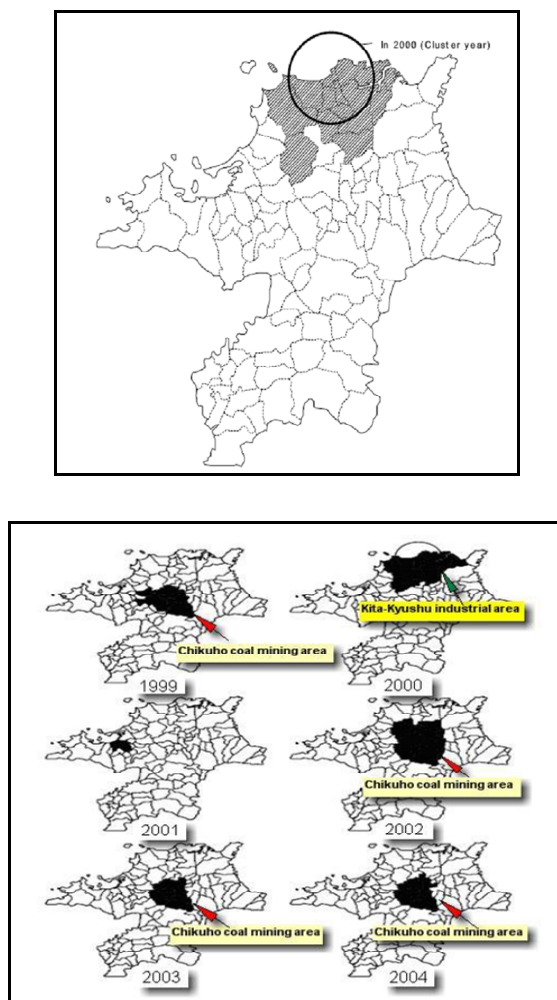
Figure 10-3. (**a**)The Space-time Analysis detected clusters of TB cases in Kita-Kyushu industrial area located in the northern Fukuoka Prefecture during 1999~2004, based on the historical data from 1999 to 2004 [30]. (**b**) Locations of the clusters of TB cases detected in Fukuoka Prefecture from 1999 to 2004, based on a purely spatial analysis [30].

3.2   GI-Related Transmission

Giardia lamblia is the most frequently identified human intestinal proto-
zoa in Canada with an estimated prevalence of 4–10%. The spatial scan
statistic was used to identify local spatial clusters of those cases with pointed
data, to measure the possible "rural" effect from the distribution of the giardiasis
and to explore the associations between the area-specific giardiasis rates and
the manure application on agricultural land and livestock density [31]. Finally,
giardiasis clusters in southern Ontario were identified (Figure 10-4a). How-
ever, neither livestock density (Figure 10-4b) nor manure application on agri-
cultural land plays an important role in the epidemiology of giardiasis there.
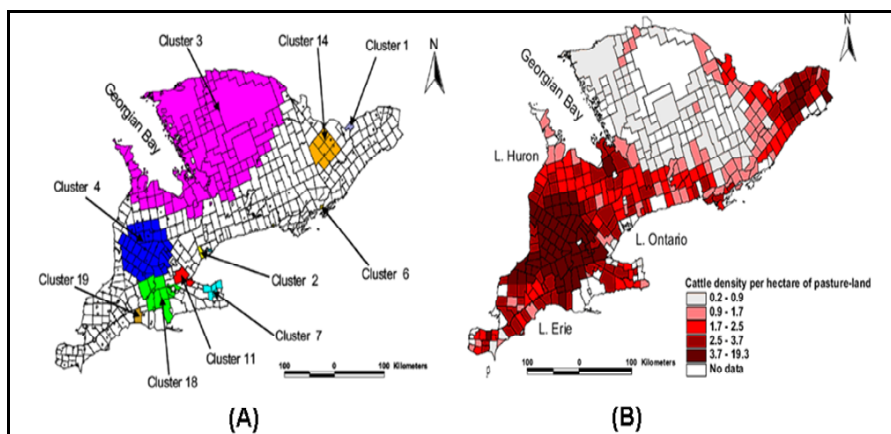

Figure 10-4. (**a**) Spatial distribution of giardiasis with significant high rate of giardiasis
clusters located in southern Ontario during 1990–1998, (**b**) Spatial distribution of cattle
density in southern Ontario[31].

3.3   Vector-Borne Transmission: Dengue as an Example

To retrospectively detect spatio-temporal dengue clusters at patients'
homes (point-type data) in Iracoubo, French Guiana and the disease onset
dates during 2001 [32], GIS integrated with the Knox method was employed.
Heterogeneity in the variations of relative risk (RR) in space and time was
found to be associated with mosquito factors, including mosquito feeding
cycle, host-seeking behavior, and life span of mosquitoes. Particularly,
higher RR values were more likely to be identified in the time periods and
areas with shorter temporal and spatial distances (Figure 10-5a) and more
clear suspected/confirmed dengue clusters were detected in shorter time
distances (Figure 10-5b). In addition, confirmed dengue cases showed
more clear higher risk (in red color) than suspected dengue cases, illustrating
the importance of laboratory diagnosis. The cluster analysis also proved that

the probability of observing a dengue case outside of 100 m around the dengue foci, a distance measured to correspond to a statistical threshold, was low. However, this threshold could vary if the case numbers increased with the improved surveillance system. By contrast, Taiwan's GIS analysis of confirmed dengue cases showed that the relocation diffusion occurred more frequently as the duration of the epidemic wave in that epidemic site became longer [33]. In other words, spatial limit of transmission, expanding distribution of mosquito vectors even after control efforts, and dynamic changes in populations at risk (e.g., susceptible) can be obtained more precisely once integrated temporal and spatial data are simultaneously analyzed. The situation might be even more complicated for malaria, which involves
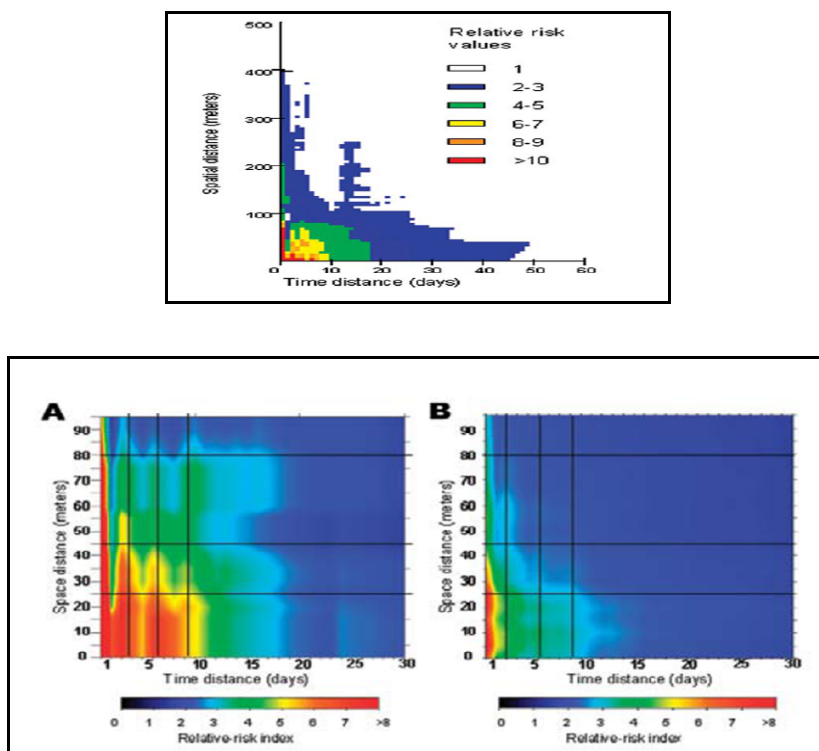
Figure 10-5. (**a**) The relative risk (RR) calculated from the confirmed dengue cases in Iracoubo, French Guiana during 2001, when space-distance and time-distance from the first index suspected dengue case varied from 0 to 500 m and from 0 to 60 days, respectively. Color areas indicated their RR values significantly greater than 1 (*p* < 0.001) [32]. (**b**) Main risk area for dengue fever (within 100 m and 30 days boundaries), derived from both the laboratory-positive dengue cases (A) and all suspected cases (B) in Iracoubo, French Guiana. *Vertical dark lines* indicate an apparent temporal periodicity, and *horizontal dark lines* correspond to apparent spatial breaks [32].

different species of mosquitoes and their variations in ecology [34], and for yellow fever, in which the immunity following vaccination means public health officials need to consider the "population at risk" as well as naturally acquired infection [35].

3.4   Zoonosis: Rabies as an Example

Zoonotic diseases involve the ecology of infectious diseases while animals in nature are sick. Therefore, the surveillance of zoonosis must start from the targeted animal population, its ecological niche and possible associated
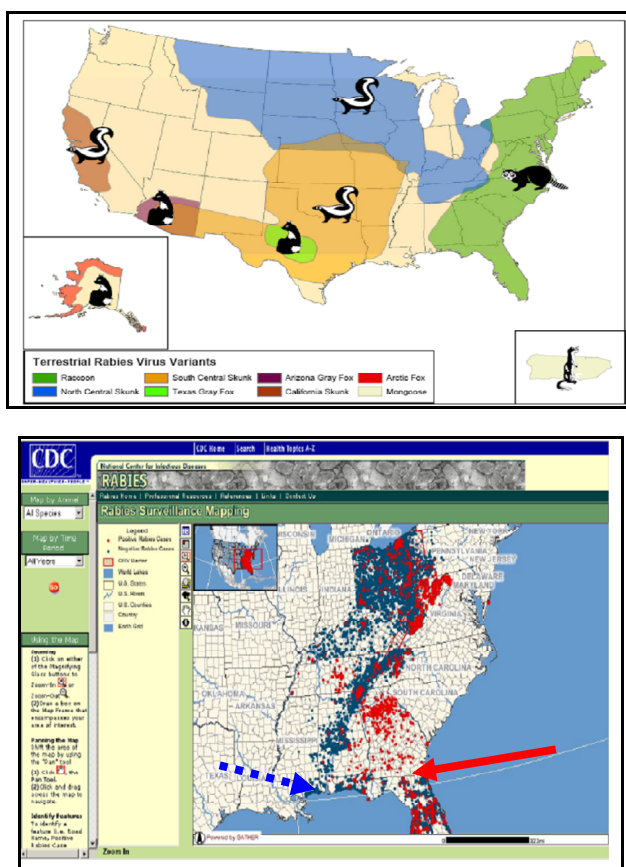

Figure 10-6. (**a**) Reservoirs for Rabies Virus in the United States. Geographic distribution of terrestrial rabies virus variants – defined by monoclonal antibody typing [36]. (**b**) The distribution of rabies in the USA, using complete web-based easily updated rabies RabID data of both animals and humans. Solid arrow (Right) indicates the geographical distribution of "positive" rabies identified by laboratory diagnosis. Dashed arrow (Left) indicates reported cases with the samples tested as "negative" for rabies [36].

environmental factors [4]. Particularly, a disease such as rabies would lead to higher case fatality rates if proper treatment and control were not implemented in a timely manner. Therefore, a GIS web-based platform for the surveillance of rabies named "RabID" was set up to rapidly map the animal rabies cases, to track the rabies reservoir and then to disseminate this information for public education at the US-CDC [36]. Several geographically discrete terrestrial wildlife reservoirs were identified (Figure 10-6a). In addition, real-time and web information on the type of animals infected and associated genotypes and strains of rabies virus identified by monoclonal antibodies have been shared among local animal and public health personnel across various geographical areas. This information will certainly facilitate the timely management of rabies control (Figure 10-6a and b). In other words, GIS information with integrated spatial epidemiologic characteristics is very useful for prevention and control on zoonotic or vector-borne infectious diseases, from public health planning to implementation and evaluation of the effectiveness of control.

3.5   EID: Avian Influenza as an Example

Avian influenza has been an increasing public health threat since the cross-country spread during 2003–2004. Between October 2005 and June 2006, 161 outbreaks of highly pathogenic avian influenza (HPAI) H5N1 occurred in poultry villages of Romania [37]. Using two combined temporal and geostatistical methods, Anselin's local indicator of spatial autocorrelation statistics (LISA) for area-type data and space-time permutation scan statistic for point-type data, the clusters of H5N1 were identified. The former method focuses only on spatial clusters and the latter method simultaneously considers temporal and spatial clusters. The space-time permutation scan statistic method is particularly useful in infectious diseases with shorter incubation period but closely associated with large-scope ecology and also in those situations where the numbers of populations at risk is unknown in syndromic surveillance [18]. The results found that the locations of the clusters were different by using the two different cluster algorithms (Figure 10-7a and b). The origin, evolution and increasing spread of the epidemic can be grasped more clearly. The outbreak first appeared in the region of the Danube River Delta by the introduction of the virus, implying the importance of landscape epidemiology. Then, the movement of poultry might facilitate its further spread to central Romania the next year. Using the spatio-temporal methods, the progression of the outbreak from a confined, local epidemic extended to a large, nationwide epidemic can be fully understood. Such efforts are very helpful to minimize the spread of the next H5N1 epidemic in other countries and the future global spread of HPAI H5N1 viruses.
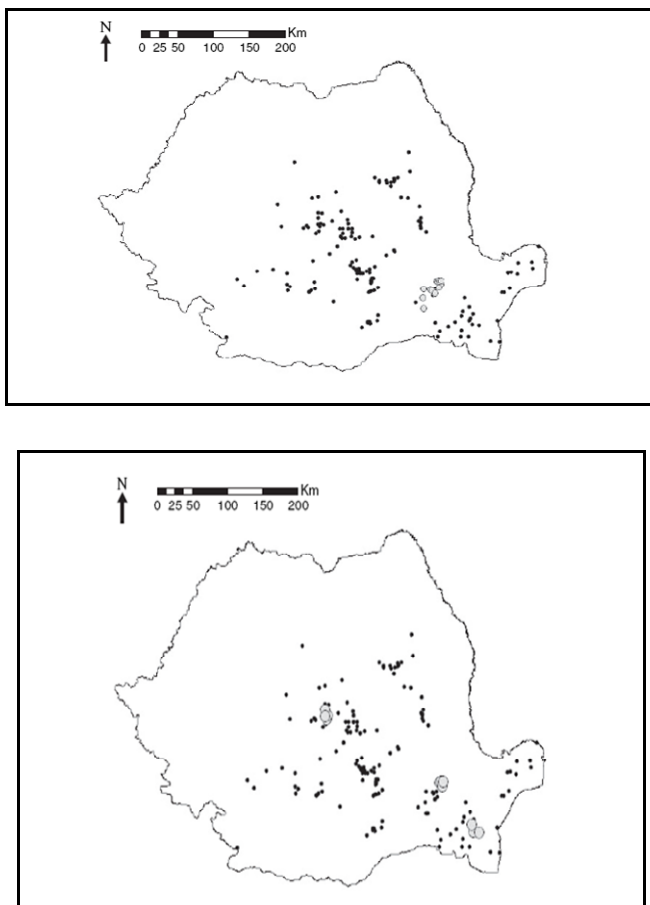
Figure 10-7. (**a**) One cluster (*open circle*) of HPAI subtype H5N1 and the villages with outbreaks (*filled circle*) in Romania, October 2005–June 2006, were identified by the local indicator of spatial autocorrelation statistic [37]. (**b**) Three clusters (*open circle*) of HPAI subtype H5N1 and the villages with outbreaks (*filled circle*) in Romania, October 2005–June 2006, were identified by the space-time permutation scan statistic [37].

   To summarize the analysis of temporal and spatial clusters for infectious diseases with different modes of transmission, both the transmissibility and pathogenicity of the microbial agents are the key factors to determine the best method to be selected. For the diseases with high transmissibility and high pathogenicity, the rapidly cross-geographical spread shown by a cross-sectional map with the appearance of cases might indicate an emerging infectious disease which needs to use integrated spatio-temporal clustering methods for further data analysis. For the diseases with low transmissibility

and high pathogenicity, the spatial clustering method can capture the distribution of the cases. For the diseases with high transmissibility and low pathogenicity, temporal clustering methods would need to be used to obtain warning signals as early as possible.

Until now, we have not had the perfect method to detect all kinds of clusters. Due to the unknown clustering pattern, it is better to use more than two methods to cross-validate the clusters before drawing a final conclusion.

4.   CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

4.1   What We Have Learned in the Past

Spatial and temporal clustering methods have been applied to prevention and control measures of infectious diseases, from improving surveillance systems, real-time integrating of clinical, microbiologic, environmental and epidemiologic data, to understanding the epidemiologic characteristics of infectious diseases and evaluating the effectiveness of control measures.

In routine surveillance systems, the algorithms such as CUSUM, ARIMA and Satscan have been widely used in different surveillance systems to detect early abnormal signals. The sensitivity and specificity of the algorithms for aberration detection need to be evaluated by each algorithm using different datasets. In general, the incorporation and integration of several different algorithms to complement each other can help to verify the occurrence of an outbreak.

For those infectious diseases with high communicability such as measles, smallpox or a disease with a high case fatality rate such as rabies, Ebola hemorrhagic fever and highly pathogenic avian influenza H5N1, or a disease with fast transmissibility such as the 2009 swine-origin H1N1 in human populations, the real-time integration of clinical, microbiologic, environmental and epidemiologic data is crucially important to increase the efficiency and accuracy of surveillance. Spatio-temporal analysis of the updated confirmed cases is frequently compared with that of the reported suspected cases for investigating how the epidemic expands rapidly and where analysis can be further improved. These analytic results can then provide positive feedback to improve the surveillance system and can also point out those high risk areas in need of more attention.

In the analysis of epidemiologic data of an infectious disease, spatial and temporal clustering algorithms can be applied after collecting the spatial information using GPSs of geo-coded addresses of the studied cases and their

exposure sites. Through fully understanding the epidemiologic characteristics of the outbreak disease, the specific prevention and control strategies can be formulated based on scientific data. This is most important for emerging infectious diseases when the etiologic agent is not known, such as the cross-country outbreak of severe acute respiratory syndrome (SARS) in 2003. For example, the modes of transmission, the time period that is most communicable, and the spatial patterns of the cases with different social contacts are unclear at the initial stage of disease outbreaks [38]. The subsequent cases after the introduction of prevention and/or control measures can also be carefully evaluated to verify the most effective strategy, using time-based integrated surveillance data. The visualized dynamic distributions of cases in various time periods and places at different levels of the public health system, from local, state/provincial to national and international, can be presented to generate hypotheses and to verify the success of containing the outbreak for decision-makers. Most importantly, evidence of spatial clustering along with other epidemiological findings and laboratory tests may indicate a possible infectious etiology for emerging infectious disease, similar to Epstein Barr virus for Hodgkin's disease [39].

4.2   Limitations of GIS Studies and Unsolved Problems

Several limitations of GIS studies need to be improved including data collection, quality of data to statistical methods and interpretation of the data.

4.2.1   Data Collection and Quality of GIS Data

In data collection, timely data and "modifiable area unit problem" (MAUP) – similar to ecological fallacies in epidemiology – are the two major barriers. Available timely data are important to fast-spreading infectious diseases such as most respiratory infections. In addition, the high quality of GIS data is another limitation in many developing countries. By contrast, those pointed address data of cases related to privacy are generally inaccessible in developed countries. Most importantly, for infectious diseases involving higher social stigma or patients' private life such as tuberculosis, sexual transmitted disease (STD) or AIDS, the pointed data for spatial cluster analysis will be very difficult to obtain. Then, the problem of spatial precision or polygon data will make it very hard to investigate the evolution of the outbreak by time and place simultaneously or to search for interesting hypotheses. Since most public health systems are governed by local departments, it is very likely those surveillance data are frequently aggregated into administrative units. Unfortunately, different densities and distribution patterns of disease, such as cholera in Figure 10-8, exhibited from different aggregated administration
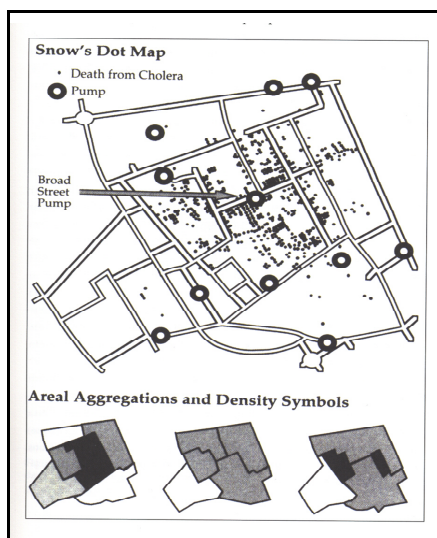
Figure 10-8. Different density maps and distribution patterns of cholera are shown by using different aggregative levels [40].

boundaries [40]. Therefore, researchers need to think about the most appropriate spatial units related to possible hypotheses at the initial stage of data collection.

4.2.2   Limitations in Statistical Methods and Interpretation of Data

Several unsolved statistical methods include too small value of relative risk to be detected, multiple covariate adjustment in spatial analysis, and better prediction during the occurrence of fast dynamic changes of cases in time and place. According to a simulated study [41], when the relative risk among study areas was lower than 1.5, the sensitivity of the detecting clusters dropped dramatically. However, the specificity can still keep a high performance level (above 95%). For an infectious disease with low pathogenicity, the relative risk is almost close to l and the sensitivity of the clustering algorithm should be low. Then, the false negative might be high. Under this circumstance, it is better to take specimens for laboratory diagnosis to increase the specificity. In addition, the demographic variables such as age structure and gender ratio are the frequently encountered confounding variables and other basic covariates should be adjusted for the risk. In dealing with fast-spreading infectious diseases, higher precision of the temporal and spatial units, real-time data gathering, and better statistical power all must be considered.

4.3   **Future Directions**

Many challenges of infectious diseases are common in different countries, including the impact of global warming on infectious diseases, emergency responses to EID, timely collection, and interchanges of high quality data to develop better control strategies. All these related issues need international collaboration. From our experiences, future global needs will involve flexible cluster methods to analyze irregular clusters, adjustment for personal risk factors, and application of Bayesian approaches to disease mapping and better prediction.

4.3.1   **Flexibility of the Cluster Method in Detecting Irregular Clusters**

Due to the natural barriers and the movement of humans, hosts and the vectors, the realistic shapes of clusters in most situations are irregular. If the algorithm intends to enhance the performance of detecting true clusters, flexibility of the shape will be needed. Risk-adjusted Nearest Neighbor Hierarchical clustering (RNNH), Support Vector Machine (SVM) [42], and ellipse shape Satscan [20] are all used to solve the problem of detecting clusters with irregular shape.

4.3.2   **Adjustment for Personal Risk Factors**

All ecologic data may involve the possible risk of "ecologic fallacy," and particularly the aggregated data might involve too many risk factors together. Detailed information is always difficult to collect through routine surveillance. In general, the demographic information such as age and gender, as total of ten covariates, would need to be adjusted using Satscan 8.0. In clustering algorithms, incorporating the data of important risk factors such as age, gender, occupation into analyses will help figure out the epidemiological conditions to form clusters.

4.3.3   **Bayesian Method for Better Prediction [43]**

Bayesian hierarchical spatial models have become widespread in disease mapping and ecologic studies of health-environment associations. In order to use posterior distribution of space–time interactions for predictions, information over space and time must be applied to estimate typical patterns for each area. Based on the extension of the Bayesian hierarchical models, the problems in detecting small numbers of events, particularly a small incidence of cases in the early wave of an outbreak, may soon be overcome in the future.

ACKNOWLEDGEMENTS

QUESTIONS FOR DISCUSSION

1. How can different clustering methods be applied to infectious diseases with various modes of transmission?
2. Are there any differences in using spatio-temporal analysis methods to analyze the data of an acute infectious disease versus a chronic disease?
3. Do you agree that the irregular clustering shapes and Bayesian model may enhance the capability to detect the true clusters?
4. Real-time syndromic surveillance is important for the early detection of abnormal events. Which clustering methods would you use to detect an early outbreak in a real-time manner?

REFERENCES

1. Elliott P, Wartenberg D: Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives* 2004, **112**(9):998–1006.
2. Spatial epidemiology [http://en.wikipedia.org/wiki/Spatial_epidemiology].
3. Gesler W: The uses of spatial analysis in medical geography: a review. *Social science & medicine (1982)* 1986, **23**(10):963–973.
4. Peterson AT: Ecologic niche modeling and spatial patterns of disease transmission. *Emerging infectious diseases* 2006, **12**(12):1822–1826.
5. Ali M, Emch M, Donnay JP, Yunus M, Sack RB: Identifying environmental risk factors for endemic cholera: a raster GIS approach. *Health & place* 2002, **8**(3):201–210.
6. Teutsch SM, Churchill RE: Principles and Practice of Public Health Surveillance, 2nd Edn. New York, NY: Oxford University Press; 2000.

7. Pascutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A, Elliott P: Statistical issues in the analysis of disease mapping data. *Statistics in medicine* 2000, **19**(17–18):2493–2519.

8. Wu TS, Shih FY, Yen MY, Wu JS, Lu SW, Chang KC, Hsiung C, Chou JH, Chu YT, Chang H *et al*: Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. *BMC public health [electronic resource]* 2008, **8**:18.

9. Jackson ML, Baer A, Painter I, Duchin J: A simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC medical informatics and decision making [electronic resource]* 2007, **7**:6.

10. Hutwagner L, Browne T, Seeman GM, Fleischauer AT: Comparing aberration detection methods with simulated data. *Emerging infectious diseases* 2005, **11**(2):314–316.

11. Stroup DF, Williamson GD, Herndon JL, Karon JM: Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in medicine* 1989, **8**(3):323–329; discussion 331–322.

12. Stroup DF, Wharton M, Kafadar K, Dean AG: Evaluation of a method for detecting aberrations in public health surveillance data. *American journal of epidemiology* 1993, **137**(3):373–380.

13. Williams SM, Parry BR, Schlup MM: Quality control: an application of the cusum. *BMJ (Clinical research ed.)* 1992, **304**(6838):1359–1361.

14. Longley PA, Goodchild MF, Maguire DJ, Rhind DW: Geographic Information System and Science. England: John Wiley & Sons, Ltd; 2001.

15. Kulldorff M: Statistical methods for spatial epidemiology: Tests for randomness. *GIS and Health* 1998.

16. Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR: Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American journal of public health* 1998, **88**(9):1377–1380.

17. Kulldorff M, Nagarwalla N: Spatial disease clusters: detection and inference. *Statistics in medicine* 1995, **14**(8):799–810.

18. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F: A space-time permutation scan statistic for disease outbreak detection. *PLoS medicine* 2005, **2**(3):e59.

19. Kleinman KP, Abrams AM, Kulldorff M, Platt R: A model-adjusted space-time scan statistic with an application to syndromic surveillance. *Epidemiology and infection* 2005, **133**(3):409–419.

20. Kulldorff M, Huang L, Pickle L, Duczmal L: An elliptic spatial scan statistic. *Statistics in medicine* 2006, **25**(22):3929–3943.

21. Tango T, Takahashi K: A flexibly shaped spatial scan statistic for detecting clusters. *International journal of health geographics [electronic resource]* 2005, **4**:11.

22. Anselin L: Local indicators of spatial association – LISA. *Geographical Analysis* 1995, **27**:93–116.

23. Kelsall JE, Diggle PJ: Non-parametric estimation of spatial variation in relative risk. *Statistics in medicine* 1995, **14**(21–22):2335–2342.

24. Wen TH, Lin NH, Lin CH, King CC, Su MD: Spatial mapping of temporal risk characteristics to improve environmental health risk identification: a case study of a dengue epidemic in Taiwan. *The Science of the total environment* 2006, **367**(2–3):631–640.

25. Tango T: Score tests for detecting excess risks around putative sources. *Statistics in medicine* 2002, **21**(4):497–514.

26. Stone R: Investigation of excess environmental risks around putative sources: statistical problems and a proposed test. *Statistics in medicine* 1988, **7**:649–660.

27. Knox G: The detection of space-time interactions. *Applied Statistics* 1964, **13**:25–29.

28. Pike MC, Smith PG: Disease clustering: a generalization of Knox's approach to the detection of space-time interactions. *Biometrics* 1968, **24**(3):541–556.

29. Gilman EA, McNally RJ, Cartwright RA: Space-time clustering of acute lymphoblastic leukaemia in parts of the U.K. (1984–1993). *European Journal of Cancer* 1999, **35**(1):91–96.

30. Onozuka D, Hagihara A: Geographic prediction of tuberculosis clusters in Fukuoka, Japan, using the space-time scan statistic. *BMC infectious diseases [electronic resource]* 2007, **7**:26.

31. Odoi A, Martin SW, Michel P, Middleton D, Holt J, Wilson J: Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International journal of health geographics [electronic resource]* 2004, **3**(1):11.

32. Tran A, Deparis X, Dussart P, Morvan J, Rabarison P, Remy F, Polidori L, Gardon J: Dengue spatial and temporal patterns, French Guiana, 2001. *Emerging infectious diseases* 2004, **10**(4):615–621.

33. Kan CC, Lee PF, Wen TH, Chao DY, Wu MN, Lin NH, Huang SY, Shang CS, Fan IC, Shu PY, Huang JH, Pai L, King CC: Two clustering diffusion patterns identified from the 2001–2003 dengue epidemics, Kaohsiung, Taiwan. *The American journal of tropical medicine and hygiene* 2008, **79**(3):344–352.

34. Guerra CA, Snow RW, Hay SI: Defining the Global Spatial Limits of Malaria Transmission in 2005. In: *Global Mapping of Infectious Diseases – Methods, Examples, and Emerging Applications.* Edited by Hay SI, Graham A, Rogers DJ. Oxford, United Kingdom: Academic Press; 2007.

35. Rogers DJ, Wilson AJ, Hay SI, Graham AJ: The Global Distribution of Yellow Fever and Dengue. In: *Global Mapping of Infectious Diseases – Methods, Examples, and Emerging Applications.* Edited by Hay SI, Graham A, Rogers DJ. Oxford, United Kingdom: Academic Press; 2007.

36. Blanton JD, Manangan A, Manangan J, Hanlon CA, Slate D, Rupprecht CE: Development of a GIS-based, real-time internet mapping tool for rabies surveillance. *International journal of health geographics* 2006, **5**:47.

37. Ward MP, Maftei D, Apostu C, Suru A: Geostatistical visualisation and spatial statistics for evaluation of the dispersion of epidemic highly pathogenic avian influenza subtype H5N1. *Veterinary research* 2008, **39**(3):22.

38. Chen Y-D, Tseng C, King CC, Wu TSJ, Chen H: Incorporating Geographical Contacts into Social Network Analysis for Contact Tracing in Epidemiology: A Study on Taiwan SARS Data. In: *Intelligence and Security Informatics: Biosurveillance.* Edited by Zeng D, Gotham I, Komatsu K, Lynch C, Thurmond M, Madigan D, Lober B, Kvach J, Chen H. Heiderberg, Germany: Springer-Verlag; 2007.

39. Alexander FE, Williams J, McKinney PA, Ricketts TJ, Cartwright RA: A specialist leukaemia/lymphoma registry in the UK. Part 2: Clustering of Hodgkin's disease. *British journal of cancer* 1989, **60**(6):948–952.

40. Koch T: Cartographies of Disease: Maps, Mapping, and Medicine. Redlands, CA: ESRI Press; 2005.

41. Aamodt G, Samuelsen SO, Skrondal A: A simulation study of three methods for detecting disease clusters. *International journal of health geographics [electronic resource]* 2006, **5**:15.

42. Zeng D, Chen H, Lynch C, Eidson M, Gotham I: Infectious Disease Informatics and Outbreak Detection. In: *Medical Informatics: Knowledge Management and Data Mining in Biomedicine.* Edited by Chen H, Fuller SS. New York: Springer; 2005.
43. Abellan JJ, Richardson S, Best N: Use of space-time models to investigate the stability of patterns of disease. *Environmental health perspectives* 2008, **116**(8):1111–1119.

SUGGESTED READING

1. International Journal of Health Geographics, http://www.ij-healthgeographics.com
2. Waller LA, Gotway CA: Applied Spatial Statistics for Public Health Data. Hoboken, NJ: John Wiley & Sons, 2004.
3. Lawson AB: Statistical Methods in Spatial Epidemiology, 2nd Edn. Hoboken, NJ: Wiley, 2006.
4. Lawson AB: Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Boca Raton: CRC Press, 2009.

ONLINE RESOURCES

Free Software:

1. Epi Info, http://www.cdc.gov/epiinfo/downloads.htm
2. Quantum GIS 0.9, http://download.qgis.org/downloads.rhtml
3. R, http://www.r-project.org/
4. Satscan, http://www.satscan.org/download.html
5. Geosurveillance, http://www.acsu.buffalo.edu/~rogerson/geosurv.htm
6. Online Periodic Regression Models, http://www.u707.jussieu.fr/periodic_regression/.
7. Geoda, http://geodacenter.asu.edu/

Free Maps:

1. World Shapefile, http://www.cdc.gov/epiinfo/shape.htm
2. Geography Network Explorer: http://www.geographynetwork.com/