Chapter 27
Spatial Analytic Approaches to Explaining the Trends and Patterns of Drug Overdose Deaths

Charlie DiMaggio, Angela Bucciarelli, Kenneth J. Tardiff, David Vlahov and Sandro Galea

Abstract To effectively utilize and interpret spatial analyses, substance use researchers, public health practitioners and policy makers should be familiar with some of the available data analytic techniques, each of which comes with advantages and drawbacks. In this chapter we first discuss three cluster detection tools and their associated software applications. We then present a Bayesian hierarchical approach, briefly reviewing its theoretical underpinnings, commonly used models, and how inferences may be drawn from a sample-based posterior distribution. We demonstrate the use of each approach on a set of substance abuse mortality data, comparing the results across the four tools. Our empiric illustration considers the role of neighborhood-level socioeconomic status (SES) in explaining opiate-related overdose deaths in New York City. We end with a discussion of the implications of the choice of technique and software on interpreting spatial analyses of substance abuse and conclude that the choice of a method will be driven by the question to be answered, data and software availability and the intended audience or context in which the research is being conducted.

Introduction

Mapping techniques and spatial analysis have been used in a number of studies seeking to describe and analyze substance abuse. Spatial analytic studies have demonstrated the correlation of drug use to deprivation indices (Squires, Bleeching, Schlecht, and Ruben 1995); the role social networks play in urban adolescent substance abuse (Mason, Cheung, and Walker 2004); the effect of ecologic level

S. Galea
Department of Epidemiology, University of Michigan, 1214 S. University, Room 243, Ann Arbor, MI 48104, USA
e-mail: sgalea@umich.edu

variables, such as legal prohibitions against alcohol sales (Schulte, Aultman-Hall, McCourt, and Stamatiadis 2003); and whether frequency and type of drug use are geographically located independent of neighborhood characteristics (Latkin, Glass, and Duncan 1998).

There are a number of spatial analytic tools available to epidemiologists, each having advantages as well as drawbacks. To effectively utilize and interpret spatial analyses of substance abuse, researchers, public health practitioners, and policy makers should be familiar with some of the available data analytic techniques.

In this chapter, we first discuss three cluster detection tools and their associated software applications: nearest neighbor index (NNI; ESRI 2005), Ripley's K-function (Levine 2004), and a space–time and time permutation scan statistic (Kuldorf 2005). We briefly describe these techniques and then demonstrate their use on a set of substance abuse mortality data, comparing the results across the three tools. We then introduce hierarchical spatial modeling (Imperial College and Medical Research Council 2004). We will discuss the advantages and disadvantages of a Bayesian approach, some commonly used models, and how to draw inferences from the sampled posterior distribution. We will demonstrate this approach on our data set and compare the results to those we obtained with cluster detection tools. As an empiric illustration, we consider the role of neighborhood-level socioeconomic status (SES) in explaining opiate-related overdose deaths in New York City (NYC). We conclude with a discussion of the implications of the choice of software and techniques on interpreting spatial analyses of substance abuse.

Data and Variable Definitions

We manually reviewed medical files at the Office of the Chief Medical Examiner (OCME) of NYC and identified all cases of fatal accidental drug overdose occurring in the city between 2000 and 2004. The OCME is responsible for assessing all deaths of persons believed to have occurred in an unnatural manner in NYC. Therefore, all overdose deaths in NYC would have been reviewed by the OCME and included in this chart abstraction.

The OCME medico-legal investigators use the decedent's medical history, the circumstances and environment of the fatality, autopsy findings, and laboratory data in attributing the cause of death and other criteria to each case being reviewed. The variables we abstracted for our analysis included the decedents' gender, age, address of residence, and location of injury. We geocoded residential and injury locations using ArcGIS, version 9.1 (ESRI 2005). For analysis purposes, place of injury (location of death) was used. In the analysis, only decedents who were successfully assigned an address of injury were included. Overdose deaths may include more than one drug being present; to increase the reliability of our measures across the different analytic techniques, we restricted our analyses to cases in which opiates were the only (in the case of scan statistics) or primary cause of death.

Table 27.1 Demographic characteristics of successfully geocoded opiate-related drug overdose deaths, New York city 2000–2004

	No.	%
Total	2426	
Sex		
Male	1883	77.6
Female	543	22.4
Age		
15–24 yrs	146	6.0
25–34 yrs	431	17.8
35–44 yrs	898	37.0
45–54 yrs	762	31.4
55–64 yrs	162	6.7
65–74 yrs	25	1.0
Over 74 yrs	1	0.0
Race/ethnicity		
White	1069	44.1
Black/not Hispanic	560	23.1
Hispanic	776	32.0
Asian	11	0.5
Other	10	0.4
Year of death		
2000	543	22.4
2001	455	18.8
2002	484	20.0
2003	506	20.9
2004	438	18.1

Descriptive Epidemiology

From 2000 through 2004, the OCME reported 3982 fatal overdose deaths within NYC. Of these, 3777 occurred among NYC residents, in which 2516 were determined to have opiate toxicity as the primary cause of death. Together, 96.4% (2426 out of 2516) cases were successfully geocoded. These cases constituted the study base for our subsequent analyses. Their demographic characteristics are presented in Table 27.1.

Cluster Detection Techniques

We conducted cluster analyses for all opiate-related deaths. We first described cross-sectional spatial distribution of all fatal opiate-related deaths in NYC that occurred between 2000 and 2004 using an average NNI statistic. We then utilized an L function transformation (derived from Ripley's K-function) to produce graphs to assess at what distance the clustering (if present) was observed to be the greatest for each year. We then used a space–time permutation model to assess both the spatial and temporal clustering opiate-related overdose deaths. This approach uses

count (case) data only, and assesses not only spatial clustering characteristics but also the role of time as a variable over the 5-year period. In this way, we determined if any clusters were statistically significant when adjusting for the year in which the deaths occurred. Finally, we applied a space–time scan statistic that differs from the permutation model in that it also adjusts for the underlying population at the census tract level.

First-order Clustering Technique: Average Nearest Neighbor Index

Description

The global presence or absence of clustered overdose incidences can be assessed using the average NNI. This index is a measure of how similar the mean distance of all cases is to the expected mean distance for a hypothetical random distribution (Mitchell 2005). The equation for calculating the average NNI is expressed as (Mitchell 2005):

$$d = \left(\frac{\sum_i C_i}{n} \right) - \left(\frac{0.05}{\sqrt{n/A}} \right)$$

where the average NNI (d) is equal to the summed distance to each feature nearest neighbor ($\sum_i C_i$) divided by the number of features (n) or the 'observed distribution' of mean features minus the product of 0.05 divided by the square root of the number of features (n) divided by the study area (A) or the 'expected mean distance for a random distribution' (Clark and Evans 1954). Clustering is suggested when the observed average distance is greater than the mean random distance ($d < 1$). An index value close to 1 indicates randomness, while a value greater than 1 indicates dispersion of cases. Within ArcGIS, version 9.1, tests of significance (a z-score and p-value) are included with the NNI output. If the z-score is negative, this suggests the cases are clustered. Conversely, if the z-score is positive, this suggests the cases are dispersed, while a value close to zero indicates the random nature of cases.

Application

Table 27.2 displays the average NNI results of all opiate-related drug overdose deaths. The results suggest that the greatest clustering of any overdose occurring in NYC was in 2000 (NNI = 0.7636; z-score $= -10.5474$). Given the statistically significant negative z-score values of all 5 years, individually, opiate-related drug overdose demonstrates significant clustering for all years between 2000 and 2005.

Table 27.2 Average nearest neighbor analysis of opiate-related drug overdose deaths, New York city, 2000–2004*

Year of Death	Average Nearest Neighbor Ratio	z-score (Standard Deviations)	p-value
2000–2004	0.7023	−27.8189	<0.0001
2000	0.7636	−10.5474	<0.0001
2001	0.7216	−11.2869	<0.0001
2002	0.7659	−9.9835	<0.0001
2003	0.8235	−7.1310	<0.0001

*Based on weighted counts of injury location and direct distance measurement.

Moreover, clustering became less dense toward the later years of the study period as displayed by the gradually increasing z-scores.

Second-order Clustering Technique: Ripley K-Function (L-Transformation) Statistic

Description

While the average NNI considers only the distance between one case and its nearest other case, the Ripley's K-function statistic is a second-order statistic that considers the complete distribution of all distances in the point pattern of cases (Levine 2004). It tests the cumulative distribution function of the entire set of inter-point distances among the point data. When K statistics are transformed into a square root function, the result is called a L function transformation (L(d)). The square-root transformation results in a linear function. This statistic can be very useful when exploring the nature, in terms of distance, of the case clustering within the entire study area. L function equation is expressed as (Levine 2004):

$$L(d) = \sqrt{\frac{A \sum_{i \neq j} \sum I_{ij}\, d_{ij}}{\pi n(n-1)}}$$

The numerator is the Ripley's K-function, where the distance (d) is measured between case (i) and every other case (j); then each distance is multiplied by the weight for the case paring (I_{ij}), and all the values are summed ($i \neq j$ indicates the distance between cases) are not included in the sum (Levine 2004). Finally, the result is multiplied by the study area (A) and divided by the number of cases (n) squared. The denominator is π multiplied by the number of possible case pairings (represented as $n-1$). The square root of the product is then taken. At any given distance (represented by the x-axis of the result graph), if the line of observed L

values is above that of the expected values [L = 0 or complete spatial randomness (CSR)], then the cases are more clustered than expected for a random distribution (with the peak of the graph representing the greatest clustering detected at a specified distance) (Levine 2004). Once the curve falls below the CSR line, cases at that point become dispersed at a given distance.

To test the null hypothesis of global spatial randomness of opiate-related overdose for the period of interest, we computed a 95% CI (referred to as the envelope) of the L-function (L(d)) using a Monte Carlo method of specifying 100 simulated random patterns (Levine 2004). At a given distance (represented on the x-axis), a value of L(d) (represented on the y-axis) outside the confidence interval (CI) envelope is interpreted as a significant departure from CSR toward clustering or dispersion. When the function peaks at the largest, most positive value and remains outside the CI envelope, this is considered to be the distance at which cases tend to be the most clustered.

Application

Figure 27.1 presents a graph of the L-function statistic for the entire 2000–2004 time period, and suggests that most clusters of opiate-related overdose are fairly compact and that the greatest clustering occurs at distance of approximately 12,007.34 feet or 2.27 miles [distance at which L(d) peaks in the output]. This suggests that most clusters occur with a radius of approximately 2.25 miles. Thereafter, the cases become more dispersed. The L(d) curve also remains outside the 95% confidence envelope and, therefore, remains statistically significant. Additional curves for individual
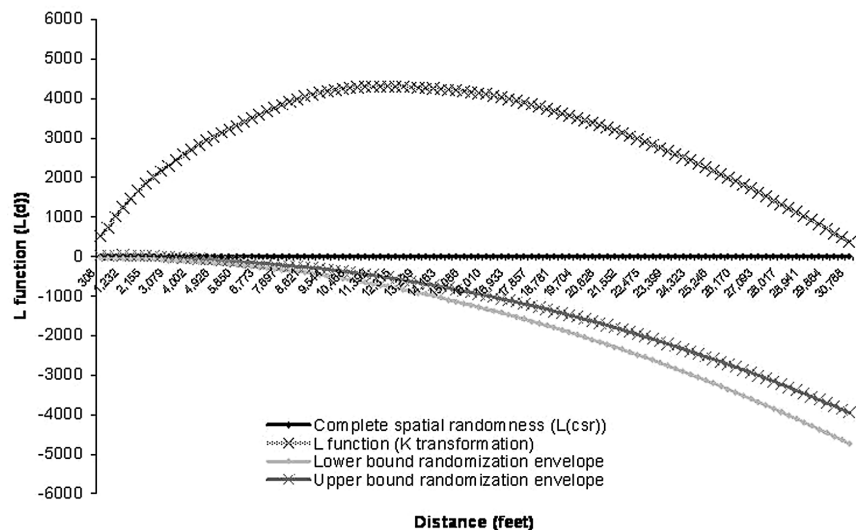
Fig. 27.1 Graph of L-function statistic. New York City opiate-related overdose deaths, 2000–2004

years (not presented) are also clustered at approximately 2–3 miles, with the tightest clusters observed to be in 2000 (10,775.82 feet or 2.27 miles).

Space–Time Modeling Techniques: Space–Time Permutation Statistic and Space–Time Scan Statistic (SaTScan)

Description

Pure spatial analyses, e.g., NNI and Ripley's K, are useful when exploring cross-sectional health outcomes. When the variable of 'time' (in units of hours, days, months, years, etc.) is of interest, we will need a model that assesses the trend of the outcome over both space and time. We are interested in whether the same areas experience clustering year after year, asking: Are the cases clustered and, if so, do they continue to cluster over time given the nature of the study area?

The space–time permutation scan statistic model uses only case data. There is no requirement for specifying the underlying population data. It makes minimal assumptions about the time, geographic location, or size of the potential case clustering. The model adjusts for what is termed as 'purely' spatial and temporal variation in the case data for a given area (Kulldorff, Heffernan, Hartman, Assuncao, and Mostashari 2005). Using a Poisson-based probability model, a series of overlapping scanning windows (cylindrical in form) move across the spatial plane (the base of the cylinder) while also scanning the point data for temporal clusters (the height of the cylinder). The circular base represents the geographical area or the study area while the height of the cylinder scans for time (in days, months, or years) clustering. For each location, the scanning window calculates the number of observed and expected cases. The statistical significance of an observed 'cluster' is then evaluated taking into account the multiple testing methods (0, 9, or 999 Monte Carlo replications). For each center and radius of the cylinder base, the method iterates over all possible temporal cylinder lengths. Cylinders can be geographically large and temporally short (forming a flat disc), or can be geographically small and temporally long (forming a pole), or any combination in between. The number of observed cases is divided by the calculated expected number of cases for each cylinder to the power of the observed inside the cylinder, and then multiplied by the observed, divided by the expected to the power of the observed outside the cylinder. The approximation, a Poisson generalized likelihood ratio, is expressed as (Kulldorff et al. 2005):

$$\left(\frac{C_A}{\mu_A}\right)^{C_A} \left(\frac{C - C_A}{C - \mu_A}\right)^{(C - C_A)}$$

where C is the total number of observed cases, μ_A represents the mean number of expected number of cases within the cylinder, and C_A represents the observed number of cases within the cylinder.

The space–time scan statistic is also based on Poisson modeling, just as the space–time permutation model, but allows for scanning of purely spatial, purely temporal, and special temporal clusters.

These models are most readily applied using the SaTScan software package, which is available for download (Kuldorf 2005) after registration and can be translated in ArcGIS for viewing of the cluster statistics.

Application

The space–time permutation scan statistic of opiate-related overdose was mapped to give a visual display of the model output (see Fig. 27.2a). In terms of spatiotemporal clustering for opiate-related overdose citywide, seven clusters were detected. The primary cluster was detected in the northwestern portion of Manhattan and the southern region of Bronx in 2001, while the other six secondary clusters were located in various parts of the city. It is notable that none of the clusters reached a level of statistical significance.

The majority of opiate-related drug clusters seem to have occurred in 2000, although the primary cluster in northern Manhattan/South Bronx was detected in 2001. It is suggested that opiate-related fatal overdose is not only changing incidence pattern, but also such cases are decreasing.

In our comparison space–time scan statistic model, we attempted to see how clusters may change when adjusting for the underlying population counts. In this analysis, we used population counts at the census tract level for the entire NYC area for a finer and more exact population adjustment. Using a small neighborhood unit of population count adjustments allows for a finer resolution of cluster detection.

The space–time scan statistic detected fewer clusters, all of which were statistically significant (Fig. 27.2b). What remains as the primary cluster (just as with the space–time permutation model cluster map) is the one located in the northern region of Manhattan and the southern region of Bronx. Compared to the space–time permutation cluster map, the central radius point of the cluster is shifted slightly west (encompassing more of Manhattan than Bronx), the radius of the cluster is approximately one mile larger (radius = 3.596 miles as compared to 2.664 miles), and the year in which cases were significantly detected was 2002–2003 as compared to 2001 for the space–time permutation model.

There are additional differences between the two approaches. A cluster in lower Manhattan detected by both methods is somewhat larger in the space–time scan, which additionally detected higher than expected cases in both 2000 and 2001 as compared to only 2000 in the space–time permutation scan.

Bayesian Hierarchical Models

Bayesian hierarchical modeling is frequently used in spatial epidemiological analyses, but may be unfamiliar to some substance abuse researchers. Multilevel spatial modeling, though, has been used to capture context in community studies
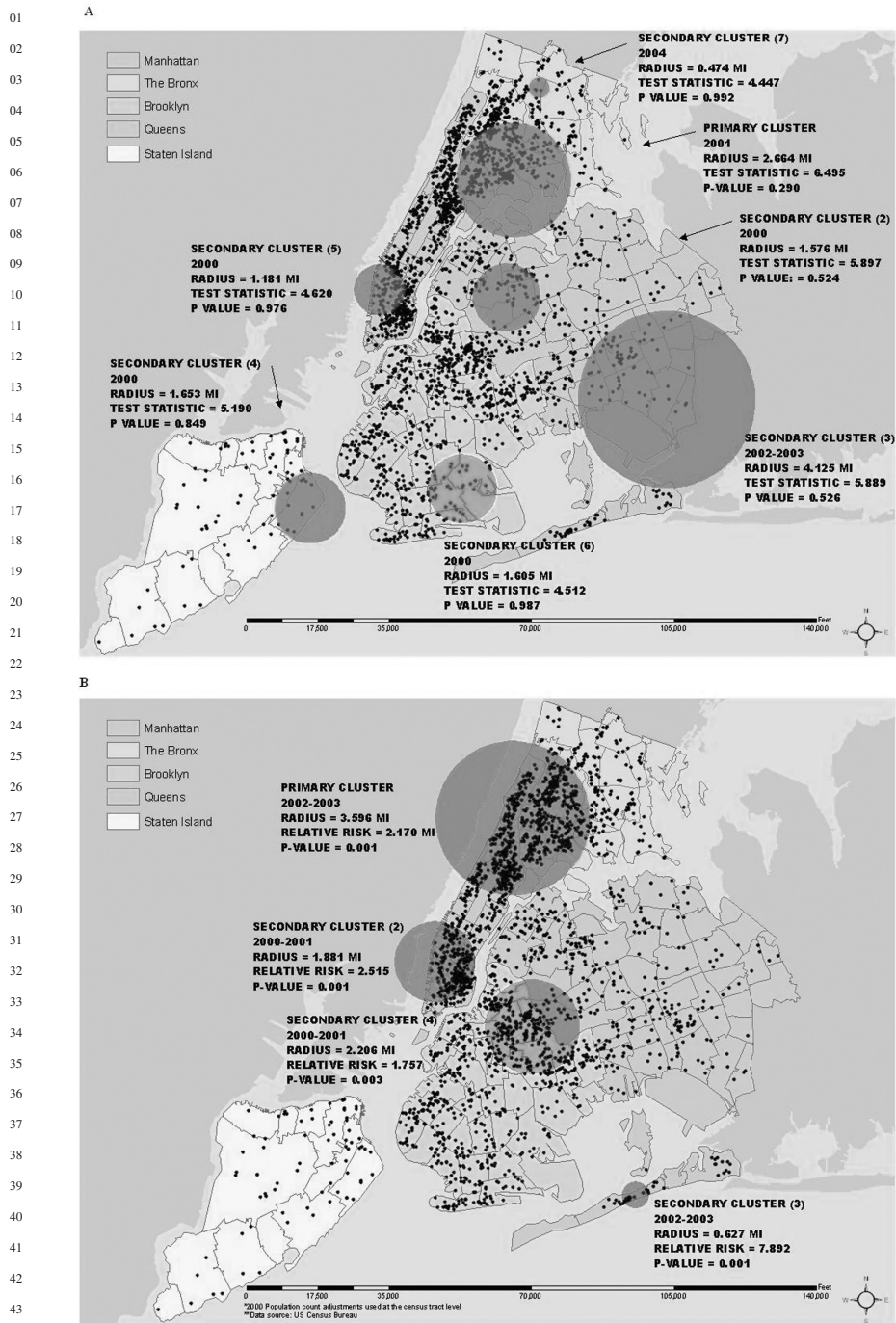
Fig. 27.2 Opiate-related deaths, New York City, 200–2004. (a) Space-time permutation scan indicating 7 clusters (b) Controlling for underlying population count (See also Plate 48 in the Colour Plate Section)

of substance abuse (Luke 2005) and in studies of drug-related crime (Law and Bayesian 2004). In this section, we review Bayesian methods, consider how they may address certain difficulties encountered in other spatial analytic techniques, and present the results of their use in our sample data set.

Description

Mapping issues and the Bayesian approach

In the classical maximum likelihood approach to risk measures, such as standardized mortality ratios, the risk estimate for each area j is given by the observed j/expected $j * 100$ with the standard error under an assumption of a Poisson distribution for each area given by the square root of the observed number divided by the expected number.

There are problems with this approach for spatial analyses. The map may be dominated by extreme values based on a few cases in small populations (Devine, Louis, and Halloran 1994). These rare events contribute to more heterogeneity than is assumed by a Poisson model (where μ is expected to be close to 1 and equal to S). A simple maximum likelihood approach also does not account for spatial correlation. Influential covariates of an outcome, which may be unmeasured, are likely to be similar in adjacent areas resulting in risk estimates that are also spatially correlated and similar. In situations when there are a small number of correlated cases relative to those at risk and Poisson 'noise' obscures the 'signal' of the spatial pattern in the data, hierarchical Bayesian modeling can be useful (Richardson, Abellan, and Best 2006).

In a Bayesian approach, our two main sources of information about the risk estimate for an area (θ) are our prior beliefs about θ, called the *prior distribution* or $p(\theta)$, and the *likelihood* of observing our data given θ or $L(y|\theta)$. We thus specify a probability distribution of risk estimates (θ) that vary across the areas of the map in some defined fashion, e.g., they may be normally distributed or Poisson distributed. This prior distribution may be based on previous studies, literature reviews, or expert opinions, and informs about θ through our beliefs or assumptions. The likelihood informs about θ via the data itself. When we have lots of data, the likelihood predominates our analysis, and our results will essentially be the maximum likelihood estimate. When we have less data, the prior has greater influence (Greenland 2006; Lawson, Browne, and Vidal 2003). The result of combining the prior distribution and the likelihood is called our posterior distribution.

Choice of our prior distribution is critical as it essentially indicates how we believe the parameter would behave if we had no data from which to make our decision. What, for example, might we expect is the probability that someone living within 3 miles of a certain location would die from an opiate-related overdose? Our best guess might be, for example, 1 in 20 or about 5%, and that this probability varies around this point estimate in a normal fashion with a variance of, say, 0.01 or 1%. This estimate may be based on previous studies, law enforcement

data, clinical experience, or a combination of sources. What if we conduct a study that indicates the risk of an opiate overdose within 3 miles of the location is 45%? How likely is our observed data given our postulated prior probability? Our posterior distribution combines our expectation with what we actually observe. In a very common sense way, it tells us, for example, that if the results of our study differ markedly from our best existing information, we should perhaps be somewhat skeptical.

The approach is hierarchical or mixed because we specify a distribution of *hyperparameters* (λ) for our risk parameter, θ allowing it to vary across each area. One could, for example, say that y_i is the empirical (observed) rate of substance abuse-related deaths in zip code i, θ is the true underlying rate, and λ how that true rate varies (Banerjee, Carlin, and Gelfand 2004).

As noted, the posterior distribution ($\Pr[\theta|y]$) is based on our prior assumptions and our observed data. It follows Bayes' theorem and is proportional to the likelihood times the prior: (Greenland 2006)

$$\Pr[\theta|y] \alpha \Pr[y|\theta] \,^* \Pr[\theta],$$

As described by Richardson, Abellon, and Best (2006), hierarchical Bayesian spatial models describe observed cases in an area as Poisson distributed with a mean equal to the expected number of cases times the risk for that area:

$$O_i \sim \text{Poisson}(\rho_i E_i).$$

At the second level of the model, the risk for each area (ρ) is transformed to a log scale (making relationships additive rather than multiplicative) and is described as an intercept term (a) and two random effects, one spatial (θ), the other non-spatial (Ψ):

$$\log \rho_i = a_i + \theta_i + \psi_i$$

The spatially structured component is described as a conditional autoregressive (CAR) Gaussian process [$\theta \sim$ CAR normal (W, τ_θ)] where the conditional distribution of each θ_i, given all the other θ_i's, is normal with $\mu =$ the average θ of its neighbors and a precision (τ_θ) proportional to the number of neighbors. W represents the matrix of neighbors that defines the neighborhood structure. The non-spatial component of the model (ψ_i) is defined as normally distributed with $\mu = 0$ and precision (τ_ψ). The model is completed by assigning hyper-priors to the precision terms τ_θ and τ_ψ.

The Poisson–gamma model

This hierarchical Bayesian approach most frequently described in the mapping literature is the Poisson–gamma model. In this formulation, the risk (θ) is described as

a set of parameters that may include any number of explanatory variables (Lawson et al. 2003). The prior distribution of the observed outcome y is described as $y|\theta \sim Po(\theta E)$, and the hyper-prior distribution of risk is $\theta|\acute{\alpha}, \beta \sim$ gamma $(\acute{\alpha}, \beta)$, with $\mu = \acute{\alpha}/\beta$ and $\sigma^2 = \acute{\alpha}/\beta^2$ (Banerjee et al. 2004; Lawson et al. 2003). We could further specify $\acute{\alpha}$ and β, but we assume that beyond a certain point, further model specification will have little practical effect on our results.

We commonly choose a non-informative (proper) or arbitrarily vague-prior that is uniform or 'flat' to allow the data to predominate and lead us to a posterior distribution that is dominated by the likelihood. A gamma (0.5, 0.0005) has been suggested as reasonable (Law and Bayesian 2004).

For simple models for which there is a closed form (i.e., they behave as true distributions and integrate to 1), we can estimate the posterior distribution directly via the maximum likelihood estimate, and a Bayesian approach is unnecessary. But for most reasonably realistic models, we will not be able to find a closed form and will need sample-based approaches.

Empirical Bayes methods approximate the posterior distribution (Devine et al. 1994). Full Bayes methods base inferences on a sample of the full posterior distribution. The results from such a sample are not as informative as the closed form itself, but are usually sufficient for inference. We increase the precision of our estimates by increasing the sample size (Banerjee et al. 2004).

One way to construct a sample from the posterior distribution is through Markov Chain Monte Carlo (MCMC) methods. Like a 'random walk' seen in time series analysis, the resulting series has no 'memory'. Subsequent values depend only on the current value, and the series converges to a stationary distribution assumed to be the posterior. Unlike traditional Monte Carlo methods, MCMC methods produce correlated samples, because they base subsequent values on current values. Methods such as thinning every other value may help decrease this correlation (Imperial College and Medical Research Council 2004).

Transition probabilities for selection into the series are typically determined through the use of the so-called Gibb's sampler. A special case of the Metropolis–Hastings algorithm, the Gibb's sampler generates conditional probability distributions of a parameter given all other parameters, and transition probabilities are generated that result in a proposal value that accepts or rejects the value with a probability of 1 or 0 (Lawson et al. 2003). The algorithm is useful in the context of Markov random fields where the joint posterior distribution is complicated but the full posterior prior distributions have simple forms.

As noted, the spatial Poisson process consists of two components: uncorrelated global heterogeneity (Ψ), usually due to unmeasured confounders or effects throughout the data, and correlated or specific heterogeneity, due to spatial correlation or local effects (θ) (Lawson et al. 2003; Richardson et al. 2006. To capture both spatial variation and non-spatial random effects in an additive fashion, we model the natural log of risk as the sum of these two components (da Silva and Melo 2004).

Setting the correlational structure

Since we want to model the spatial components so that geographically close areas present similar risks, we use information from other areas in the region to reduce random variation unrelated to the risk represented by our risk estimate. This should take spatial correlation into account and result in smoother informative maps. Here, we see the advantage of a Bayesian approach. Modeling θ as a random variable rather than a fixed variable allows us to set a spatial correlational structure.

We can describe this structure via Markov random fields where each θ, given all the other θ's, depends only on its neighborhood. A Markov random field is a locally specified joint distribution that can be determined by its full conditionals. Given a joint distribution, $\Pr[y_i \ldots y_n]$, the set of full conditional distributions, $\Pr[y_i|y_j]$, that we can create from it are uniquely determined. Brook's lemma tells us that we can go in the opposite direction. If we have a set of full conditional distributions, we can get the unique joint distribution from which they arose (Banerjee et al. 2004).

A locally determined, weighted structure can be represented by a CAR Gaussian model where the conditional distribution of each θ is given by:

$$\theta i | \theta j \sim N_1(\Sigma w_{ij}\theta_j/\Sigma w_{ij}, 1/\tau\Sigma w_{ij})$$

where j is not equal to i and is an element of d, the set of the neighbors of i (da Silva and Melo 2004).

The simplest and most commonly used definition of a neighborhood is the existence of a common border between areas. In this case, the weights are specified as $w_{ij} = 1$ if j is in d, and $w_{ij} = 0$ if j is not in d. In this case, the Σw_{ij} is simply the number of neighbors of area i. So, the conditional prior mean of θ is given by the arithmetic average of the spatial effects of its neighbors, and the conditional prior variance is proportional to the number of neighbors.

This structure has been used in a number of disease mapping studies (da Silva and Melo 2004), but other approaches are possible. We could set up a proximity matrix of weights, w_{ij}'s, based on the distance between or other relationships between the spatial units, e.g., a set of first order w_{ij}'s ($w_{ij}[1]$), if an areal unit is, say, less than some predefined distance (Banerjee et al. 2004). We could also adopt more complicated relationships between spatial units to represent, for example, the movement of commuters from one area to another. The more the neighbors we include, the more the smoothing we achieve. The limit would, of course, be averaging over all the spatial units, which would be the overall average and would not be very informative at the local level.

Interpreting results

An important consideration in MCMC methods is diagnosing convergence to the stationary Markov Chain. A commonly accepted approach is to run and dynamically monitor a specific number, e.g. 3, parallel chains and examine the trace plots for when they start to overlap as an indication of convergence. We then discard the

burn-in period samples and base inference on the stationary Markov Chain. The
Gellman Rubin statistic is useful in diagnosing convergence. It compares variation
within chains to those between chains for evidence of scale reduction. When the
scale reduction factor reaches 1, there is evidence of convergence. Other conver-
gence statistics are based on examining individual chains (Lawson et al. 2003).

Once the posterior distribution has been sampled, a Bayesian CI has a straightfor-
ward interpretation. In a 95% CI, we are 95% certain that the true value lies within
it. It is most easily obtained by chopping off the $\acute{\alpha}/2$ tails of the posterior probability
distribution.

Being a sample-based approach, in MCMC methods, no two analysts will end
up with exactly the same results. This makes variance assessment crucial. Recall
that MCMC produces correlated samples. This will lead to underestimates of the
variance. One may 'thin' the samples to decrease correlation, but this would result
in throwing out information. Variance estimates based on an effective sample size
(ESS), though, are available (Banerjee et al. 2004).

Application

We calculated standardized mortality ratios (SMR) for opiate-related deaths in NYC
for the years 2000–2004 using the expected number of overdose deaths in a zip
code tabulation area based on the mean number of such deaths in NYC throughout
the 5-year study period. We were interested, in this example, in drawing inference
about the potential role of SES as an explanatory variable for opiate-related overdose
deaths in NYC. We used zip code level median household income (MHI) as a proxy
for neighborhood-level SES throughout this example.

In the model, the likelihood of the observed values in the standardized morbidity
ratio is modeled as a Poisson distribution. The log of the observed value is a function
of the log of the expected value, an intercept, and the coefficient for a normally
transformed median household income measure. Random effects are represented
by a Gaussian intrinsic CAR model with the weights for adjacent neighbors set
at 1. Non-informative prior distributions are placed on the intercept, the coeffi-
cients, and on tau, the precision term for the CAR prior distribution for random
effects.

Our interest is in mapping the zip code level risk estimates while accounting
for the potential instability and autocorrelation of those rates and controlling for
MHI.

For this empiric illustration, we used WinBUGS (Imperial College and Medical
Research Council 2004) software to run three parallel MCMSs with over-dispersed
initial values for 120,000 iterations. The first 60,000 iterations were discarded as a
burn-in, and our inferences were based on the second 60,000 iterations. We assessed
convergence by examining trace histories for parallel chains, and we used R (20)
software to conduct the Brooks, Gelman and Rubin, and the Geweke convergence
diagnostics as well as the Heidleberger and Welch stationarity tests. We present our
results as median values for the coefficients with their associated 95% equal-tailed

Bayesian CIs, histograms (kernel density graphs) of the sampled distributions, and maps comparing raw and smooth risk estimates.

As indicated in Fig. 27.3a, the histogram for β1 (MHI coefficient) is smooth and normally distributed. The Gellman Rubin statistic was calculated to be 1 for most of the runs (Fig. 27.3b), and the trace history for β1 appears to reasonably overlap. Taken together, this information gives us confidence that the model appropriately converged to the posterior distribution and that our inferences based on this posterior distribution are valid.

The median value for the MHI coefficient was −0.3782 (95% CI 0.5681, −0.1855). The interpretation is not straightforward because the dependent variable in the model is the natural log of the SMR, and the MHI variable itself has been normalized to achieve appropriate convergence in WinBUGS. It does, though, indicate that SES, as measured by MHI, is strongly and significantly inversely related to the number of heroin overdose deaths in a zip code area. Essentially, the number of opiate-related deaths increases in a linear fashion as MHI declines. That this is as expected indicates, to a certain extent, the validity of the model. Also, and perhaps more importantly, the subsequent fitted SMR values now control for this important potentially confounding variable.
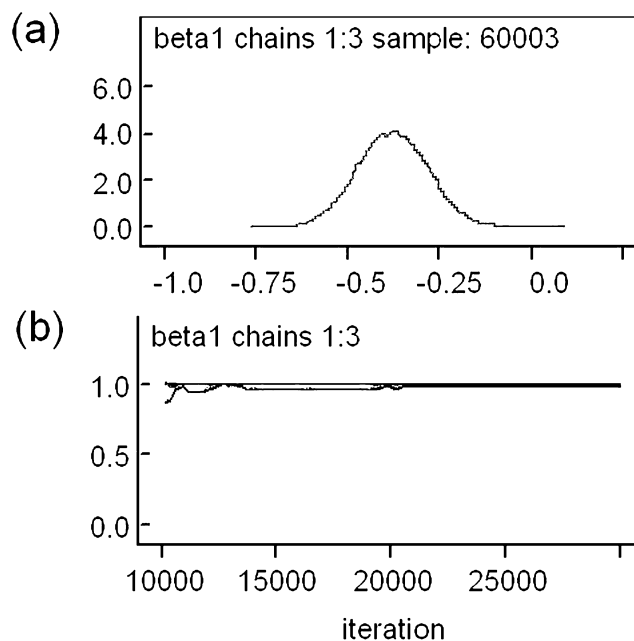
Fig. 27.3 Results from Markov Chain Monte Carlo run, association of median household income with opiate overdose standardized mortality ratios, New York City zip code tabulation areas, 2000–2004. (a) Histogram for Median Household Income beta coefficient. (b) Trace history for Median Household Income beta coefficient
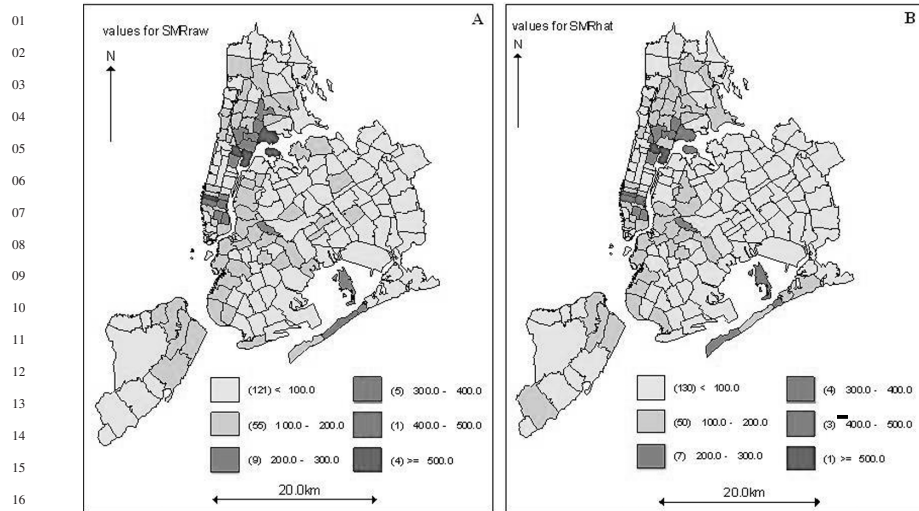
Fig. 27.4 Opiate-related standardized mortality ratios, New York City zip code tabulation areas, 2000–2004. (a) Unadjusted (b) Smoothed estimated adjusted for median household income and autocorrelation with Bayesian Hierarchical Modeling

Figure 27.4 presents the raw and fitted zip code level SMRs. It appears that if we had simply mapped the raw SMRs, we would infer greater than expected rates in such areas as South Bronx and Northern Queens. Looking at the fitted values, these potential clusters seem to become less severe when we take the underlying distribution of the population and its SES characteristics into account. The most evident area of continuing concern is Harlem in northern Manhattan which, despite controlling for MHI, continues to display much greater than expected numbers of heroin overdose deaths.

Discussion and Conclusions

The methods presented in this chapter have much to offer the substance use researcher. They can be viewed as offering incrementally more information and detail as one progresses from first-order cluster detection methods, such as the NNI through scan statistics, to more explanatory analytic techniques, such as hierarchical modeling.

While we did not spend much time on it, an essential first step in any spatial analysis is to describe the data in terms of summary statistics and simple plots. Not only does this provide key descriptive information, it also allows the researcher to assess whether the data meet the assumptions underlying subsequent tests, e.g., Poisson distributions of the outcome of interest.

First-order clustering methods offer a relatively straightforward and easily interpretable global assessment of whether clustering exists and how tight the clustering

appears to have been. These are important considerations, but most analysts will want to know exactly where and when the clustering occurred and whether any apparent clustering was simply due to chance. Scan statistics are an important public health tool in this regard. They are fairly straightforward in their application and interpretation, and allow the incorporation of a time variable.

When a population denominator is not available or appropriate, the space–time permutation model is very useful. Although none of the clusters detected through the use of space–time permutation model (Fig. 27.2a) were statistically significant, the results suggest that certain areas had higher counts than expected and that there were temporal changes in opiate-related drug overdose. Having a measure to simply detect higher than expected case counts is particularly worthwhile in public health research. In a setting where timely results based on possibly incomplete count data is a prime consideration, such as in syndromic surveillance (Heffernan et al. 2004), the scan statistic may be one's first choice.

When population estimates are available, adjusting for areas that are more highly populated is appropriate, and the space-time scan statistic is a better, more precise measurement of cluster points. In our example (Fig. 27.2b), we were able to more accurately describe the location and statistical importance of clusters detected by the space–time permutation model. Again, when public health concerns are uppermost, this method may be particularly useful.

Bayesian methods may be most appropriate when potential explanatory variables are available and one's interest is in assessing the *determinants* of health outcomes on a spatial level. It is important to appreciate that this is a smoothing method. When cluster detection is of uppermost concern, caution must be exercised that potential clusters are not smoothed away. While its appropriate utilization requires knowledge of MCMC and sample-based methods, full Bayesian analysis, as presented in our example, is not always necessary. Good empirical approximations are available and obviate the need to learn and use new complex statistical software (Devine et al. 1994; Greenland 2006). But, when data are sparse and highly correlated and there is concern over noise obscuring spatial signals, a full Bayesian approach can help describe both the determinants and the patterns of the outcome of interest at a finer level.

It should be noted that we do not, in this discussion, dwell extensively on the implications of the results of the specific example we have used here to illustrate the material being described in this chapter. However, we considered the role of neighborhood-level SES in explaining rates of heroin-related overdose in the largest US urban area. Conceptual frameworks that consider the complex etiology of substance use and its consequences (Galea, Rudenstine, and Vlahov 2005) have long suggested that a full consideration of the determination of substance use requires that we consider a range of individual- and group-level factors to understand population patterns of substance use. The hierarchical approach introduced here, suitably expanded, can be applied to test specific hypotheses and to pursue spatial etiologic questions, incorporating determinants at group and individual levels as necessary.

Ultimately, we note that the limitations of available data, including, for example, the use of zip codes as a neighborhood proxy, have been well discussed elsewhere in

the literature about the multilevel determination of population health (Osypuk and Galea 2007), and pertain just as much to spatial analyses as they do to all other types of epidemiologic analyses. Future work that makes use of the methods introduced here to address specific substance use-related etiologic hypotheses may benefit from application of these methods at different group levels of inference.

In conclusion, an appreciation for where and when health outcomes occur adds much to substance abuse epidemiology. A number of spatial and temporal methods are available, each with its own advantages and disadvantages in terms of complexity, data requirements, and underlying assumptions. The choice of a method will be driven by the question to be answered, data and software availability, and the intended audience or context in which the research is being conducted.