A Procedure for Disease Mapping: Bayesian Models and GIS

Thomas Jagger, PhD, Laurel Harduar-Morano, MPH, James Elsner, PhD, Chris Duclos, MS,
Greg Kearney , RS, MPH, David Johnson, MD, MS, DABT, FACOEM, Prakash Patel , MBBS, MPH, Emily Fogarty, MS

Background:

A traditional approach to mapping disease is to use the standardized incidence ratio (SIR), which is an estimate of incident number of disease cases in a population relative to what might be expected if the study population had the same disease incidence rate as some larger (or standard) comparison population (Fig.1). There are several limitations to this approach and other methods can make use of the correlation inherent in the data. For example, Bayesian models are flexible enough to incorporate spatial correlation and to adjust the overall mean ratio when relatively few cases exist.

Objective:

A problem with widespread use of Bayesian models in disease mapping is the lack of such tools within standard Geographic Information Systems (GIS). Here we demonstrate the tools needed. We use GIS for database management, estimating adjacency, and mapping, but use Bayesian software to develop and run a model of adult Leukemia rates. Our goal is to showcase the advantage of working with freely available and open source software to create a workflow for modern disease mapping.

Methods:

We build a model to explore the relationship between a known source of benzene exposure and Leukemia rates. This choice is predicated on results in which a significant relationship was found between acute childhood Leukemia and benzene exposure (Steffen et al. 2004). Here we choose an ecological approach that identifies the density of petrol stations (as a surrogate for benzene exposure) and total adult (20 years or older) Leukemia counts by county, race, age and sex (Fig. 2). We interpolate the 1990 and 2000 U.S census data to the years 1995-2002 and total the Leukemia counts by county and year over the same 8 year period. The initial study considers only petrol stations in existence from 1989 or earlier.

The process we use (Fig. 3) is shown in the flow from ArcMap to R to WinBUGS (Windows version of the Bayesian inference Using Gibbs Sampling) back to R then back to ArcMap. The WinBUGS and R software packages are described below. The process begins by exporting shapefiles from ArcMap. These files are read into R and converted to WinBUGS text files that are read by WinBUGS. Data is loaded and the model is compiled in WinBUGS. Initial values are specified and the model is run to generate samples that collectively form a posterior distribution for the risk ratio. The sample values are written by WinBUGS and read by R. R then generates and writes database files that are read into and displayed by ArcMap. Currently the process requires us to manually move between the 3 software platforms.

Results:

Figure 4 shows the posterior densities of the model parameters indicating the importance of the variable on Leukemia rates. The centers of distributional mass for white male, white female and other male are close to zero indicating little difference in these group's Leukemia risk relative to the entire population. With the other female group the center of mass is shifted to the right of zero indicating that this group might have a higher Leukemia rate than the population as whole though it is not statistically significant.

Figure 5 show the posterior mean Leukemia rate per 1000 people per year of exposure and the magnitude of the structured random effect. These rates can be used to identify counties at risk for Leukemia. The model does not indicate whether the differences are significant. However, further analysis using the generated posterior county rate samples can be used to determine the posterior probability that one county has a larger or smaller rate than another. By fixing a particular critical level for benzene the same set of data can be used to determine the probability that the posterior rate exceeds this level. The structured random effect is normalized so that $\Sigma(\log(\mathit{effect})) = 0$. Thus the values show the expected relative risk of living in a given region of Florida such as the Panhandle as compared to the risk of living elsewhere. It can also be used to determine if there is any regional trend in Leukemia rates after accounting for the petrol station density. From the figure we conclude that a person living in the Panhandle region has a lower relative risk of Leukemia after accounting for petrol station density, while those in the south central Peninsula region have a greater risk. In this case we estimate the ratio between the highest region and lowest region to be about 1.46.

Conclusions:

A statistical model was developed using a Bayesian approach to examine adult Leukemia rates in each of the 67 counties in Florida based on petrol station density as a surrogate for benzene concentrations in private drinking water wells. The results indicate little difference in white males, white females and other males risk relative to the entire population. The "other" female group indicated a slightly higher Leukemia rate than the population as a whole although not statistically significant. Based upon petrol facility density, people living in the Panhandle of Florida may have a lower relative risk compared to the southeast.

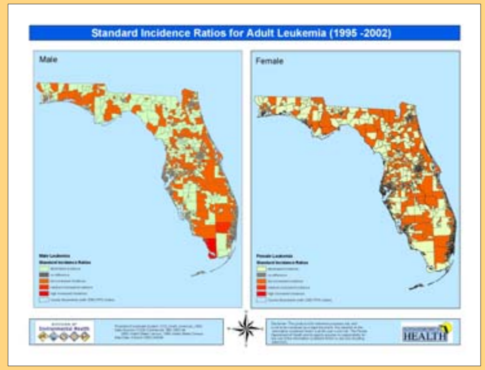

Fig. 1 Traditional SIR Maps by Census Tract

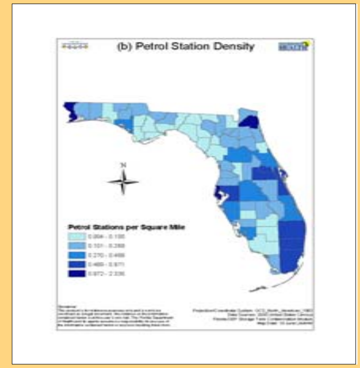

Fig. 2 Petrol Station Density by County

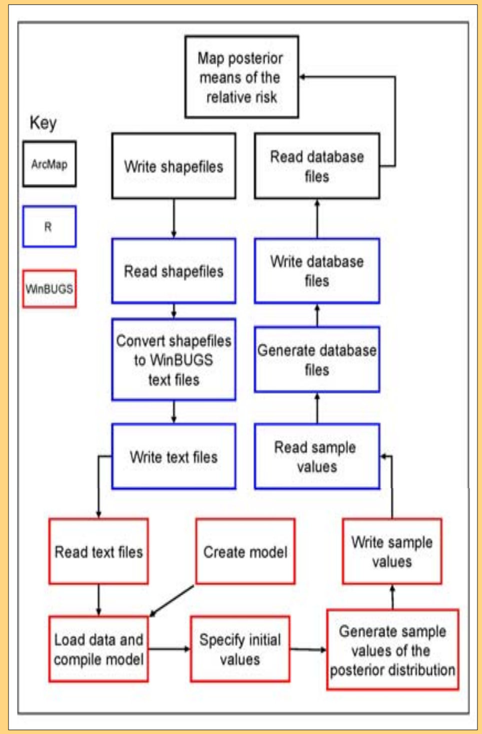

Fig. 3 Leukemia-Benzene Model: process of disease mapping using ArcMap, WinBUGS, and R. The process is layered with ArcMap on top, WinBUGS at the bottom and R in between. The flow is from ArcMap to R to WinBUGS back to R then back to ArcMap.

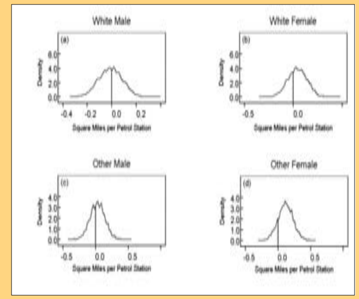

Fig. 4 Posterior densities of the model parameters indicating their importance on Leukemia rates by race and sex. (a) white male, (b) white female, (c) other male, and (d) other female. The distributions are generally centered about a parameter value of 0 indicating the factor is not important in statistically explaining Leukemia rates.

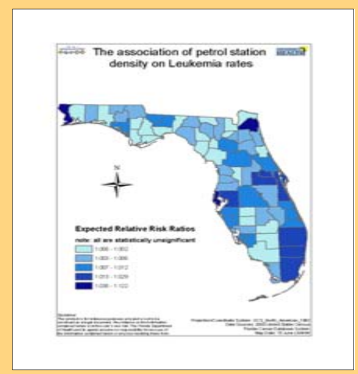

Fig. 5 The association between petrol station density and Leukemia rates based on Bayesian model. Map shows posterior mean of the relative risk ratio of developing adult Leukemia by county. The relative risk ratio is the population weighted risk of Leukemia relative to the risk if there are no petrol stations.

CDC | FLORIDA DEPARTMENT OF HEALTH | DIVISION OF Environmental Health | FLORIDA STATE UNIVERSITY Department of Geography

This poster was supported by Grant/Cooperative Agreement number US0/CCU 423288-01 from Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.