

BAYESIAN MODEL OF HIV/AIDS IN INDIA: A SPATIAL ANALYSIS

P. Venkatesan and R. Srinivasan

Tuberculosis Research Centre, ICMR, Chetpet, Chennai.

ABSTRACT

Background: Bayesian models are very flexible enough to incorporate spatial correlation and to adjust overall mean ratio when relatively few cases exist whereas our conventional Disease mapping has several limitation and it requires the Standardization Incidence Ratio which is derived from observed cases to what might be expected from larger population. This study explains the importance and advantage of Bayesian methods in Disease mapping

Aims

To study, the spatial distribution, spatio-temporal changes of HIV in India for the year 1996 to 2005. **Materials & Methods:** The data of HIV/AIDS prevalence from NACO and socio demographic data from census 2001 were used in this study. Bayesian disease mapping, Spatial filtering and spatial scan statistics is a useful technique for identifying areas that have higher or lower values than generally occur. In addition, Monte Carlo simulations are used to test the observed rates for statistical significance. The software was used for this Bayesian analysis through WinBUGS software and Disease map through Map Info software.

Introduction

Bayesian method have played major role in recent development in statistical models for spatial data. This visualizes the spatial pattern of disease risks to provide insight into geographic variation. It will give the most accurate rates in spatially defined small area.

And retain some level of geographic resolution (small areas) while maintaining statistical precision. It is keeping account for small numbers problem (variance instability) and analyses health disparities for population subgroups. In many studies, the spatial component is not looked in detail. In this work, we attempt to incorporate the spatial component along with other statistical methods to study the trend of HIV.

Data in close geographical proximity is more likely to be influenced by similar factors and thus affected in a similar way as spatial data are correlated in space. In the case of HIV/AIDS, spatial

correlation is present at both, short and large scales, reflecting the transmission of HIV/AIDS infection and the effects of environmental factors.

The analysis of spatial dispersion of the risk of occurrence of an event i.e., HIV/AIDS disease usually done via maps of incidence rate, where a set of areas is shaded according to the values of variable of interest. In spatial epidemiology this is called disease mapping. The main areas of Spatial epidemiology comprises into the following; 1). Disease mapping, 2). Disease clustering, and Ecological analysis. In the first case, usually, is used for highlighting the areas of high or low disease incidence. The second class, that of disease clustering, has particular importance in public health surveillance, where it may be important to be able to assess whether a disease map is clustered and where the clusters are located. The third class that of ecological analysis is of great relevance with epidemiological research, as its focus is the analysis of the geographical distribution of disease in relation to explanatory covariates, usually at an aggregated spatial level.

For rare diseases, small population sizes result in particularly unstable rate estimates. The statistical literature contains various methods of combining information or 'borrowing strength' between regions to achieve local rate stabilization without losing geographic resolution. The most common approaches involve hierarchical models with random effects for each region. The use of random effects presumes that regional rates are drawn from some common superpopulation of rates, and allows an analyst to combine information from several regions. The resulting estimates typically involve a weighted average of the specific region's crude rate and the rates from other regions.

In our study, for a map divided into 'i' regions, let y_i ($i=1, 2, \dots, n$) is the count of disease in the region and e_i ($i=1, 2, \dots, n$) is the expected count in the i th region. If the model follows the Poisson model θ_i ($i=1, 2, \dots, n$) is the relative risk in the i th region. Relative risk is a measure of how much of a risk factor influences the risk of a specified disease.

Maximum Likelihood Estimate for Relative Risk

Relative risk estimation is very important for disease mapping. In the classical approach, maximum likelihood estimate of the relative risks is given by $\theta_i = y_i / e_i$ and is usually called standardized mortality rate (SMR) which gives the highest likelihood for the data under a simple Poisson model. This method has several problems already identified by researchers. First, more extreme values of the estimate of relative risk may be based on few cases only in areas with small population. The second problem is that rare events in small areas can lead to extra Poisson model. i.e., there is more heterogeneity in population than is assumed by the Poisson variation model. Also zero counts aren't differentiated.

Hierarchical Bayesian Models for Relative Risks

The problems with MLE can be taken into account by allowing the relative risk to vary within each area and in this case the Bayesian approach is appropriate specifying prior distribution to the parameters. Modern approaches to relative risk estimation rely on smoothing methods; one way to produce smoother relative risk estimators is to assume that the risk has a distribution. In Bayesian terms this is called a prior distribution. In the Bayesian approach, further to information in the data, i.e., the number of disease occurrence in each area, we need to specify a prior distribution for the relative risks which provides information concerning their variability along the map. In the Poisson count, the commonest prior distribution is to assume that θ_i has a Gamma distribution.

The model is;

- $y_i \sim \text{Poisson}(e_i \theta_i)$
- $\theta_i \sim \text{Gamma}(\alpha, \beta)$
- α and β is unknown and we can assume that, i.e., $\alpha \sim \exp(\nu), \beta \sim \exp(\rho)$

Bayesian inference is then based on the combination of these two sources of information via the posterior distribution of the relative risks $P(\theta | y)$ obtained via Bayes theorem. The importance of Bayesian unlike the usual ML estimates of risk, a Bayesian model is described by a distribution and so a range of values of risk will arise (some more likely than others). Also Posterior distributions are sampled to give a range of these values (posterior sample) which contains a large amount of information about the parameter of interest.

Materials and Methods

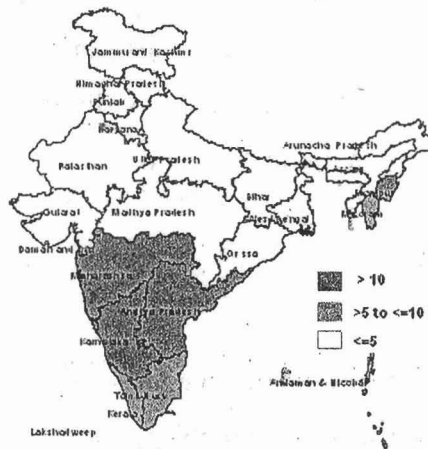
Data for this study was obtained from National AIDS Control Society (NACO), Census 2001 and National reports. NACO have been collected data through an annual HIV Sentinel surveillance survey over the years, in order to monitor trends of HIV infection in specific high-risk groups as well as low risk groups.

The main objectives of the study are to study, the spatial distribution of HIV in all the States and to forecast the trend of different States. It is also proposed to identify the significance variation in HIV/AIDS cases (2006) accounted by the other explanatory variables like HDI index for each district, Population in each district, Awareness of HIV, Literacy Rate(2001) etc.

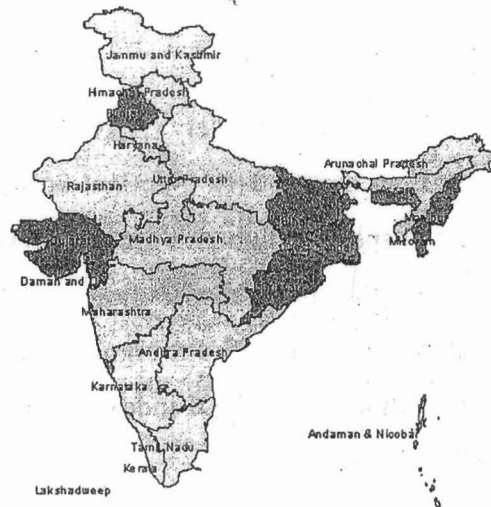
Bayesian disease mapping were used to reveal the spatial distribution of disease. Spatial regression analysis was used to forecast the trend in different States and it's also used to assess relationship between HIV/AIDS and other socio economic parameters. The software was used for this Bayesian analysis through Win BUGS software and Disease map through Map Info software.

Results

Bayesian Disease Mapping : The spread of HIV in India has been diverse, with much of it having a low rate of infection and the epidemic being most extreme in the southern half of the country in the far north-east. The highest HIV prevalence rates are found in Maharashtra in the west; Andhra Pradesh and Karnataka in the south; and Manipur and Nagaland in the north-east.



The spatio temporal changes between disease is very crucial, but there is no significant ($p > 0$) spatio-temporal changes between 1998 and 2005, however there is an increase in HIV prevalence between years.



Spatial Regression Analysis :

Spatial Regression analysis is used to find the variation in HIV/AIDS cases is accounted for by explanatory variable of HDI index, population of the states, Literacy Rate. From this calculation, found that R^2 value is 0.587, which is significant because the F statistics is significant. It may be that 58.70% of the total variation in Y is accounted for by the explanatory variable. Considering significance of individual regression coefficient using the t – statistics, it seen that only “Awarene

HIV" has significant regression coefficients. The negative slope of the equation revealed that the prevalence of HIV is in decreasing trend that is mentioned in yellow colour, and positive value of the slope describes that the prevalence of HIV is in increasing trend which is mentioned in red colour shade.

Conclusion:

The spread of HIV in India has been diverse, with much of India having a low rate of infection and the epidemic being most extreme in the southern half of the country and in the far north-east. The highest HIV prevalence rates are found in Maharashtra in the west; Andhra Pradesh and Karnataka in the south; and Manipur and Nagaland in the north-east.

The variation in HIV/AIDS cases is accounted for by the explanatory variable of HDI index, population of the states, Literacy Rate. It has been seen that only "Awareness of HIV" have significant regression coefficients. So HIV Awareness among people influences the control of HIV in India. Government may initiate again to create a high awareness about HIV through different media to control HIV in different States.

Conventional Disease mapping has several limitations and it requires the Standardization Incidence Ratio which is derived from observed cases to what might be expected from larger population, whereas Bayesian models are very flexible enough to incorporate spatial correlation and to adjust overall mean ratio when relatively few cases exist.

A Bayesian model consists of likelihood and prior distributions. The product of the likelihood and the prior distributions gives the most important distribution. In Bayesian modeling all the inference about parameters is made from the posterior distribution. The posterior distribution gives information about the distribution of parameters: not just about the most likely value. It is now relatively simple to obtain samples of parameters from posterior distributions.

References:

Cressie, N. (1993), *Statistics for Spatial Data* (2nd Ed.), New York: Wiley.

Nicky Best, Sylvia Richardson and Andrew Thompson, A Comparison of Bayesian Spatial models for disease mapping, *Statistical methods in Medical Research*, 2005;14;35.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2003): *Bayesian Data Analysis* (2nd ed.), London: Chapman & Hall.

GeoBUGS Manual. Version 1.2

Kelsall J, Wakefield J: Modelling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association* 2002;97:692-701.

Bayesian models of... P.Venkatesan and R.Srinivasa

Shin, D.R.K, Tsutakawa, H. Kin, and Z, He. Spatial-temporal interaction with disease mapping. *Statistics in Medicine* 200; 19:2015-2035.

Waller, L.A., B.P. Carlin, H. Xia, and A.E. Gelfand. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 1997; 438:607-617.

WinBUGS user manual. Version 1.4.

Xia, H. and B.P. Carlin. Spatial-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in Medicine* 1998; 2025-2043