

# Probability Mapping to Determine the Spatial Risk Pattern of Acute Gastroenteritis in Coimbatore District, India, Using Geographic Information Systems (GIS)

Pawlin Vasanthi Joseph, Brindha Balan<sup>1</sup>, Vidhyalakshmi Rajendran<sup>1</sup>, Devi Marimuthu Prashanthi<sup>2</sup>, Balasubramanian Somnathan<sup>3</sup>

Associate Professor in Zoology, Nirmala College for Women (Autonomous), Redfields, Coimbatore, <sup>1</sup>Research Scholar, <sup>2</sup>Assistant Professor, Department of Environmental Management, Bharathidasan University, Tiruchirapalli, Tamil Nadu, <sup>3</sup>Director of Research, Jagadguru Sri Shivarathreeswara Medical University, Mysore, Karnataka, India

## ABSTRACT

**Background:** Maps show well the spatial configuration of information. Considerable effort is devoted to the development of geographical information systems (GIS) that increase understanding of public health problems and in particular to collaborate efforts among clinicians, epidemiologists, ecologists, and geographers to map and forecast disease risk. **Objectives:** Small populations tend to give rise to the most extreme disease rates, even if the actual rates are similar across the areas. Such situations will follow the decision-maker's attention on these areas when they scrutinize the map for decision making or resource allocation. As an alternative, maps can be prepared using *P*-values (probabilistic values). **Materials and Methods:** The statistical significance of rates rather than the rates themselves are used to map the results. The incidence rates calculated for each village from 2000 to 2009 is used to estimate  $\lambda$ , the expected number of cases in the study area. The obtained results are mapped using Arc GIS 10.0. **Results:** The likelihood of infections from low to high is depicted in the map and it is observed that five villages namely, Odanthurai, Coimbatore Corporation, Ikkaraiبولuvampatti, Puliakulam, and Pollachi Corporation are more likely to have significantly high incidences. **Conclusion:** In the probability map, some of the areas with exceptionally high or low rates disappear. These are typically small unpopulated areas, whose rates are unstable due to the small numbers problem. The probability map shows more specific regions of relative risks and expected outcomes.

**Keywords:** Disease estimates, geographic information systems, probabilistic values, poison distribution, relative risk

## Introduction

Maps have an important role to play alongside the epidemiological use of geographical data, in the presentation and emphasis of conclusions drawn from

the analysis. They have a great potential as a method for communicating results. Maps show well the spatial configuration of information.<sup>(1)</sup> Considerable effort is devoted to the development of geographical information systems that increase understanding of public health problems and in particular to collaborate efforts among clinicians, epidemiologists, ecologists, and geographers to map and forecast disease risk.<sup>(2)</sup>

Growing public awareness of environmental hazards has led to an increased demand for public health authorities to investigate geographical clustering of diseases. Although such cluster analysis is nearly

Access this article online	
Quick Response Code:	Website: www.ijcm.org.in
	DOI: 10.4103/0970-0218.158865

### Address for correspondence:

Dr. Pawlin Vasanthi Joseph, Assistant Professor in Zoology, Nirmala College for Women (Autonomous), Redfields, Coimbatore - 641 018, Tamil Nadu, India. E-mail: pawl\_06@rediffmail.com

**Received:** 26-06-14, **Accepted:** 04-11-14

always ineffective in identifying causes of diseases, it often has to be used to address public concern about environmental hazards. The most common way of analyzing clustering in area health data is to prepare choropleth maps of disease incidence or prevalence rates. When the areas differ in population size, however, as is typically the case, the calculated rates of disease for those areas have different degrees of reliability. Rates for small areas — areas with small populations — vary more and are less reliable than those for large areas. For small areas, a difference of one or two cases can make a huge difference in incidence or prevalence rates. This is known as the small numbers problem.

In disease surveillance, the problem of making multiple comparisons can be overcome by testing for clustering and autocorrelation. When rates of disease are illustrated in disease maps undue focus on areas where random fluctuation is greatest can be minimized by smoothing techniques.<sup>(3)</sup> Mapping of disease is an activity closely related to disease surveillance and cluster detection. It is widely used for descriptive purposes to identify patterns of geographical variation in diseases and to develop new ideas about the cause of disease. Probability mapping based on a Poisson distribution model was adopted to identify areal units within regions having significantly high or low disease rates.<sup>(4)</sup>

Probability mapping is the most suitable method for detecting and monitoring spatial clusters of cancer.<sup>(5)</sup> Such clusters represent the areas where the occurrences of the diseases are statistically significant and meaningful. Probability mapping was used for analysing and visualizing the statistics of death caused by lung cancer, stomach cancer, leukemia, and skin cancer in 18 provinces of Iran.

Spatial autocorrelation contained in a disease map has its negative components fade and its positive components strengthen, through time. The hypothesis was tested by evaluating the regression coefficients of spatial filter eigenvectors as annual West Nile virus data become available.<sup>(6)</sup> An analyst can choose from a variety of analytical spatial statistical tools to study a disease map. Recent quantitative geography methodological developments have supplemented this approach with the spatial filter model specification.<sup>(7)</sup> A Markov Chain Monte Carlo methodology enables the use of the binomial, poisson, or negative binomial probability models with a spatial autoregressive specification.<sup>(8,9)</sup>

Understanding the spatial distribution of a disease is accomplished through applying statistical methods to data collected during disease surveillance and generating a map that describes spatial variation in risk. Disease

mapping has a long history for human diseases.<sup>(10-13)</sup> The goal of mapping disease is to isolate and display the role of location as a risk factor. Observation of disease rates across a region can vary due to differences in age or sex structure of the sample in addition to any location-specific variation. Hierarchical Bayesian methods were applied for disease mapping to understand the spatial and temporal distribution of chronic wasting disease in Wisconsin.<sup>(14)</sup> Practical application of Bayesian methods in large scale tropical disease control programs have been reported.<sup>(15,16)</sup> Bayesian probability maps were produced for each sex and age-group for a study on the mapping of the probability of schistosomiasis and associated uncertainty in West Africa.<sup>(17)</sup> Methods were suggested for the analysis of aggregate count data in the context of disease mapping and spatial regression using male lip cancer incidence data from Scotland.<sup>(18)</sup>

Disease occurrences are usually expressed as standardized rates for different areas. If the population of an area is small, the rate of disease estimates is more. Small populations will therefore tend to give rise to the most extreme disease rates, even if the actual rates are similar across the areas. Such situations will follow the viewers or decision-maker's attention on these areas when they scrutinize the map for decision-making or resource allocation.

As an alternative, maps can be prepared using *P* values (probabilistic values) from the tests of whether the rate in each region differs significantly from the overall rate. This will overcome the problem when the areas with largest population will dominate the map results.

The objective of the study was to spatially describe small area clustering of diseases based on rates using Probability Mapping methods in Coimbatore district of Tamil Nadu, India for the period 2000 to 2009.

### Probability mapping of diseases

Probability mapping is a well-established statistical method for addressing the small numbers problem based on poisson distribution.<sup>(19-21)</sup> In these maps, the probability of occurrence of an event *x* depends only on the space considered (area, volume, time, or inhabitants):

$$p(x) = u^x e^{-u} / x!$$

where *u* is assumed as a constant density and equals:  
# events *x*/space.

To build the maps the number of expected events *E* in a region *i* are defined as:

$$E_i = n_i \hat{p}$$

where  $\hat{p} = \sum H_i / \sum n_i$ ;  $H_i$  is the number of observed Acute Gastroenteritis cases in a region  $i$  and  $n_i$  is the population of region  $i$ . Under the assumption that  $H_i$  are independent Poisson random variables with expected values  $E_i$ , and that  $h_i/n_i = \dots = h_{125}/n_{125} = p$ , an index of deviation from equal  $h_i/n_i$  can be defined:

$$r_{oi} = \left. \begin{array}{l} \sum_{x \geq H_i} \frac{E_i^x e^{-E_i}}{x!} \text{ for } H_i \geq E_i \\ \text{Or} \\ \sum_{x < H_i} \frac{E_i^x e^{-E_i}}{x!} \text{ for } H_i < E_i \end{array} \right\}$$

A choropleth map based on  $r_{oi}$  is called a "probability map." Values of  $r_{oi} < 0.01$  indicate that region  $i$ 's Acute Gastroenteritis rate departs from expected Poisson values, being unusually high (for  $H_i \geq E_i$ ) or low (for  $H_i < E_i$ ). In addition to using the rate as "cases per inhabitant" ( $n$ ) as defined earlier, other denominators were employed:  $n$  defined as population ( $np$ ),  $n$  defined as area ( $na$ ), and as the product of area and population ( $nap$ ).

## Materials and Methods

In the present study, we used the statistical significance of rates rather than the rates themselves to map the results. Statistical significance is measured by probability values that show the likelihood of a rate occurring given the normal rate of disease in the corresponding national or regional population. We refer to this rate as the population rate,  $P$ . The probability value for an area indicates the likelihood that the rate observed in that area would occur by chance if the underlying risk of disease was equal to  $P$ . Probability values close to 0 or 1 indicates rates that are significantly different from the population rate.

There are many statistical methods for computing probability values. One of the most common is the Poisson test, used for modelling the probability of rare binary (present/absent) events in large populations. Many health problems (e.g., cancers, birth defects) fit this definition because they are rare, occurring in only a small fraction of the population, and binary, either present or absent in an individual. Consider a small area containing a population,  $n$  and  $k$  cases of disease. We want to find out whether the presence of those  $k$  cases in a population of size  $n$  is unusual. In other words, is the actual number of cases significantly higher than expected based on the national or regional prevalence rate. For the present study, incidence rates calculated for each village from 2000 to 2009 in Coimbatore district is used to estimate  $\lambda$ .

If the regional incidence rate is  $P$ , the expected number of cases in the study area,  $\lambda$ , is the

study area population,  $n$ , multiplied by the national or regional rate:

$$\lambda = np$$

For example, if the study area contains 40000 people and the national prevalence rate is 1 per 10000, we would expect 4 cases in the study area because  $\lambda = 0.0001 (40,000) = 4$ .

If we know the number of cases,  $k$ , occurring in a study region population, we can use the Poisson distribution to determine the probability,  $P(k)$ , that the observed number of cases would occur in a population of the study region's size. The Poisson distribution states that in a population of size  $n$ , the probability of  $x$  cases occurring is  $P(x) = e^{-\lambda} (\lambda^x / x!)$ . From this calculation, we can determine the probability of  $k$  or more cases occurring by chance,  $P(x \geq k)$ , if the true rate of disease in the population were  $p$ . That probability is calculated as:

$$P(x \geq k) = 1 - \sum_{x=0}^{k-1} P(x)$$

For mapping the probable Acute Gastroenteritis incidence cases in the study area, the average incidence rate (1.0068) already estimated for each village was taken as  $\lambda$ , as the rates varied for each village. Using the above formula, the probability of infected cases prevalent in each village was calculated. From the results, it was observed that the probability of one case would be 0.0367 [Table 1]. For example, if there are 6 cases of disease in the study region where only 1 case were expected based on incidence rates, the corresponding probability value would be 1-0.0005 or 0.9995. This means that here is a 99% chance of 6 or more cases occurring by chance if the underlying prevalence is 1 per the existing population, which means high prevalence rates. The closer this value is to 0, the smaller the likelihood that it would arise by chance alone. If in any case, the probability is not particularly small such as 21% or 0.021, we infer that the rate of disease is not

**Table 1: Poisson probabilities,  $\lambda = 1.0068$**

No. of cases (x)	Probability $P(x)$	Cumulative probability $P(\leq x)$
0	0.3654	0.037
1	0.3679	0.404
2	0.1852	0.590
3	0.0621	1.037
4	0.0156	1.052
5	0.0031	1.055
6	0.0005	2.037
7	0.0001	2.037
8	0.0000	2.037
9	0.0000	3.037
10	0.0000	3.037

unusually high. In general, probabilities less than 0.05 or 0.01 are considered to indicate significantly high prevalence rates.

## Results and Discussion

Using the above method, similar probabilities were estimated for all the villages and the probable chances of high disease prevalence were identified. The obtained results were mapped using Arc GIS 10.0. The likelihood of infections from low to high was depicted in the map and it was observed that five villages namely, Odanthurai, Coimbatore Corporation, Ikkaraboluvampatti, Puliakulam, and Pollachi Corporation were more likely to have significantly high incidences [Figure 1].

This map shows dominance of low standardized ratio in the largely dispersed rural villages. High Dominance is observed at Coimbatore Corporation and Pollachi Corporation which are urban areas and some rural villages at the periphery. These may be areas of apparent clustering inspite of low standardization. As these villages have low population, an excess risk may be generated by just one or two cases. Hence when choropleth map is prepared, the random variability is removed and the map becomes much simpler and all the extreme values disappear. This clearly shows an evidence of variability in incidence rates ( $P \leq 0.01$ ); however, there is little evidence of spatial clustering.

In the probability map, some of the areas with exceptionally high or low rates disappear. These are typically small unpopulated areas, whose rates are unstable due to the small numbers problem. However, regions with significantly high rates are observed to be scattered with a clustering observed in moderate rates. This approach has standardized the incidences by re-expressing them as a ratio of the estimated number of

positives and the number that would have been expected in a standard population. It was pointed out that when a population is large (therefore the expected number of cases is large), the Poisson assumption is no longer significant and a small probability is more due to lack of fit in the model than the relative risk in the local area.<sup>(20)</sup> The probability map shows more specific regions of relative risks and expected outcomes.

## Conclusion

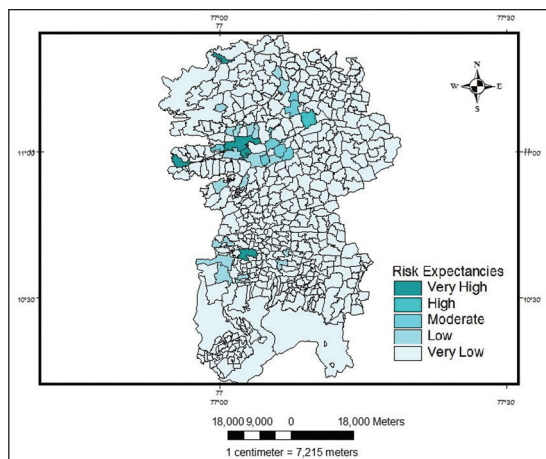
Probability mapping is a useful way of addressing the small numbers problem when mapping area health data. For an area with a large population, a rate that is slightly higher than the expected rate, will often be statistically significant because the size of the population increases the statistical power. This means it is easier to reject the null hypothesis that there is no difference in rates. Analysts need to look beyond statistical significance by examining the disease rates, the locations of high-rate areas, and any additional information that might assist in interpreting high rate areas. Through GIS a more comprehensive view of statistical significance can be observed.

## Acknowledgement

The author would like to thank the District Directorate of Health and Preventive Medicine, Coimbatore for providing the disease data.

## References

1. Brown P, Hirschfield A, Marsden J. Analysing spatial patterns of disease: Some issues in the mapping of incidence data for relatively rare conditions. In: De Lepper MD, et al. editors. The Added Value of Geographic Information Systems in Public and Environmental Health. New York City: Kluwer Academic Publishers; 1995. p. 145-63.
2. Croner CM, Sperling J, Broome FR. Geographic Information Systems (GIS): New perspectives in understanding human health and environmental relationships. *Stat Med* 1996;15:1961-77.
3. Olsen FS, Martuzzi M, Elliott P. Cluster analysis and disease mapping — why, when and how. A step by step guide. *BMJ* 1996;313:863-6.
4. Gesler W. The use of spatial analysis in Medical geography: A review. *Soc Sci Med* 1986;23:963-73.
5. Mesgari MS, Masoomi Z. GIS applications in Public Health as a decision making support system and its limitation in Iran. *World Appl Sci J* 2008;3:73-7.
6. Griffith DA. A comparison of six analytical disease mapping techniques as applied to West Nile virus in the coterminous United States. *Int J Health Geogr* 2005;4:18.
7. Griffith D. A linear regression solution to the spatial autocorrelation problem. *J Geogr Syst* 2000;2:141-56.
8. Gilks R, Richardson S, Spiegelhalter J. Markov chain Monte Carlo in practice. London: Chapman and Hall; 1996.
9. Huffer F, Wu H. Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 1998;54:509-24.



**Figure 1:** Expected disease risk in Coimbatore, India, through probability mapping



10. Bernardinelli LD, Clayton C, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space time variation in disease risk. *Stat Med* 1995;14:2433-43.
11. Waller LA. Hierarchical spatio temporal mapping of disease rates. *J Am Stat Assoc* 1997;92:607-17.
12. Elliott P, Wakefield JC, Best N, Briggs DJ. *Spatial epidemiology: Methods and applications*. Oxford: Oxford University Press; 2000.
13. Lawson AB, Browne WJ, Vidal Rodeiro CL. *Disease mapping with Win Bugs and MLwiN*. Chichester: John Wiley and Sons; 2003.
14. Osnas EE, Heisey DM, Rolley RE, Samuel MD. Spatial and temporal patterns of chronic wasting disease: Fine-scale mapping of a wildlife epidemic in Wisconsin. *Ecol Appl* 2009;19:1311-22
15. Clements AC, Moyeed R, Brooker S. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* 2006;133:711-9.
16. Clements AC, Lwambo NJ, Blair L, Nyandini U, Kaatano G, Kinunghi S, *et al.* Bayesian spatial analysis and disease mapping: Tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop Med Int Health* 2006;11:490-503.
17. Clements AC, Garba A, Sacko M, Toure S, Dembele R, Landoure A, *et al.* Mapping the probability of schistosomiasis and associated uncertainty, West Africa. *Emerg Infect Dis* 2008;14:1629-32.
18. Wakefield J. Disease mapping and spatial regression with count data. *Biostatistics* 2007;8:158-83.
19. Choynowski M. Maps based on probabilities. *J Am Stat Assoc* 1959;54:385-8.
20. Cressie NA. *Statistics for Spatial Data*, revised edition. New York: John Wiley and Sons Inc; 1993.
21. Bailey T, Gatrell A. *Interactive Spatial Data Analysis*. Addison Wesley Longman Limited, Edinburgh. Barreto ML (1993). The dot map as an epidemiological tool: A case study of *Schistosoma mansoni* infection in an urban setting. *Int J Epidemiol* 1995;22:731-41.

**How to cite this article:** Joseph PV, Balan B, Rajendran V, Prashanthi DM, Somnathan B. Probability mapping to determine the spatial risk pattern of acute gastroenteritis in Coimbatore District, India, using Geographic Information Systems (GIS). *Indian J Community Med* 2015;40:188-92.

**Source of Support:** Nil, **Conflict of Interest:** None declared.