Methodology

Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh

Mohammad Ali*[1], Pierre Goovaerts[2], Nushrat Nazia[3], M Zahirul Haq[4], Mohammad Yunus[4] and Michael Emch[5]

Address: [1]International Vaccine Institute, SNU Research Park, San 4-8 Bongcheon-7 dong, Kwanak-gu, Seoul, Korea, [2]BioMedware, Inc., Ann Arbor, MI, USA, [3]University of Texas at Dallas, USA, [4]ICDDR,B: Centre for Health and Population Research, Dhaka, Bangladesh and [5]University of North Carolina at Chapel Hilll, USA

Email: Mohammad Ali* - mali@ivi.int; Pierre Goovaerts - goovaerts@biomedware.com; Nushrat Nazia - adinann@yahoo.com; M Zahirul Haq - mzhaq@icddrb.org; Mohammad Yunus - myunus@icddrb.org; Michael Emch - emch@email.unc.edu

* Corresponding author

Abstract

Background: Disease maps can serve to display incidence rates geographically, to inform on public health provision about the success or failure of interventions, and to make hypothesis or to provide evidences concerning disease etiology. Poisson kriging was recently introduced to filter the noise attached to rates recorded over sparsely populated administrative units. Its benefit over simple population-weighted averages and empirical Bayesian smoothers was demonstrated by simulation studies using county-level cancer mortality rates. This paper presents the first application of Poisson kriging to the spatial interpolation of local disease rates, resulting in continuous maps of disease rate estimates and the associated prediction variance. The methodology is illustrated using cholera and dysentery data collected in a cholera endemic area (Matlab) of Bangladesh.

Results: The spatial analysis was confined to patrilineally-related clusters of households, known as *baris*, located within 9 kilometers from the Matlab hospital to avoid underestimating the risk of disease incidence, since patients far away from the medical facilities are less likely to travel. Semivariogram models reveal a range of autocorrelation of 1.1 km for dysentery and 0.37 km for cholera. This result translates into a cholera risk map that is patchier than the dysentery map that shows a large zone of high incidence in the south-central part of the study area, which is quasi-urban. On both maps, lower risk values are found in the Northern part of the study area, which is also the most distant from the Matlab hospital. The weaker spatial continuity of cholera versus dysentery incidence rates resulted in larger kriging variance across the study area.

Conclusion: The approach presented in this paper enables researchers to incorporate the pattern of spatial dependence of incidence rates into the mapping of risk values and the quantification of the associated uncertainty. Differences in spatial patterns, in particular the range of spatial autocorrelation, reflect differences in the mode of transmission of cholera and dysentery. Our risk maps for cholera and dysentery incidences should help identifying putative factors of increased disease incidence, leading to more effective prevention and remedial actions in endemic areas.

Background

Disease maps can serve to display incidence rates geographically, to inform on public health provision about the success or failure of interventions, and to make hypothesis or to provide evidences concerning disease etiology [1]. Understanding social/human ecological influence on health may help trigger an effective mechanism to reduce the risk of the disease. However, making a perfect disease risk map is challenging. Recent reviews of disease mapping have been provided by Kulldroff [2], Bithel [3], Diggle [4], Lawson [5], and Lawson and Clark [6]. Most of the proposed methods to create disease maps were tested on data aggregated within large geographical units and applied on non-infectious (cancer-related) diseases. Since ecological process in a large geographical unit would eventually be seen as flat, associating it with the spatial pattern of diseases may yield spurious outcomes [7].

Cholera and other diarrheal diseases remain a problem in many developing countries [8-11]. Endemic cholera has displayed a particular geographical pattern in many nations of the world [11-16]. Although several geographic studies were conducted for studying disease risk, only a few studies were concerned with risk in endemic areas that creates a constant public health problem. Identifying zones of elevated risk in endemic areas and studying socioecological conditions of these zones may provide useful clues for controlling the disease burden and improving the endemic situation. However, assessing risk for diarrheal diseases requires knowledge not only of environment but also of individual's lifestyle and health behavior that may influence the disease risk. Emch [17] illustrated that diarrheal disease risk is a dynamic interaction of biological, socioeconomic, behavioral, cultural and environmental factors over time and space. Heterogeneity in these factors among individuals may influence normal process of the disease phenomena.

Since cholera and other diarrheal diseases are primarily driven by environmental factors [18], and since environmental processes are spatially continuous in nature [19], these disease rates should display some spatial pattern. In other words, rates measured in households that are close geographically should be more similar than the rates recorded over larger separation distances. Geostatistics provides tools to analyze spatially distributed data, capitalizing on the correlation between observations to interpolate the attribute of interest and to delineate areas with high disease risk [1,19,20]. Traditional geostatistical methods are, however, not suited to the analysis of disease rates since they ignore the fact that rate data consist of a denominator and a numerator, which means that observations have varying degree of reliability that should be incorporated into the analysis. Goovaerts [21,22] recently introduced a methodology that is based on Poisson kriging and semivariogram estimators developed by Monestiez et al. [23] for mapping the relative abundance of species in the presence of spatially heterogeneous observation efforts and sparse animal sightings. Simulation studies indicated that this approach outperforms simple population-weighted averages and empirical Bayesian smoothers to estimate underlying risk in cancer mortality maps. The analysis was however confined to rate data aggregated to the level of large geographical entities (i.e. counties) that covered the entire study area, thereby eliminating the need for predicting disease risks at unmonitored locations. The creation of continuous risk maps was addressed in Monestiez et al. [23] with an application to the mapping of the expected number of whale sightings per hour.

This paper presents the first application of Poisson kriging to the spatial interpolation of local disease rates. Cholera and dysentery data were collected at the household level in an endemic area of Bangladesh. Both the disease risk and associated measure of reliability are mapped using geostatistics. These results allow then the mapping of the probability that locally the incidence rate differs from the average rate recorded over the region.

Methods

The study area

The study was carried out in Matlab, a field research station of the ICDDR,B (International Centre for Diarrhoeal Disease Research, Bangladesh): Centre for Health and Population Research, Bangladesh. Matlab is situated near the lower Meghna River, one of the main rivers of Bangladesh, and is approximately 50 km southeast of Dhaka (the capital of Bangladesh). The field research station has been involved with cholera studies since 1963. Approximately 200,000 people live in the 184 square kilometers study area. Currently, there are 142 villages; 128 are predominantly Muslim, the rest are predominantly Hindu. The study area is almost entirely rural and most people's occupations are in agriculture or fisheries. The monsoon climate of the study area is characterized by high temperatures, heavy rainfall and marked seasonal variation [24,25].

Population and disease data

In 1994, a GIS database for the study area was created in order to facilitate spatial analysis for health and population studies. The database includes *baris*, flood-control embankments, rivers, village boundaries, and health facilities. *Baris* are patrilineally-related clusters of households. An average of five households lived in a *bari*. The details on how the geographic database was created can be found elsewhere [26]. This database facilitates the linking of population and disease surveillance data to *bari* locations. The ICDDR,B's Health and Demographic Surveillance

Systems (HDSS) databases are maintained at the individual level, thus, the data are aggregated to the *baris*-level for spatial analysis. There were 7,181 points representing *baris* and 207,576 individuals in the study area in 1994. The number of individuals living in a *bari* ranges from 1 to 469, with a median number of 19.

The population data were obtained from the HDSS database. Initiated in 1966, the HDSS maintains records of all vital demographic events [27-29]. The residents in the study area are identified by a registration number given in the surveillance system. To identify the household locations of the individuals, the HDSS maintains a *bari* identification, which is unique and a permanent number in the system.

The disease data were obtained from the Matlab hospital surveillance system, which has in- and out-patient services as well as a laboratory for the identification of pathogens. The Matlab hospital treats approximately 7,000 to 8,000 diarrhea cases per year at no cost to patients. The hospital maintains motorized boats, which function as a free ambulance service for diarrhea patients; access to the hospital is excellent throughout the study area. Stool samples are collected from all patients who are admitted to the hospital who live in the ICDDR,B area. All stool samples are screened for enteric pathogens in the laboratory.

We obtained cholera and dysentery data for 1994, the year when the GIS database was created for conducting the study. Note that these two diseases are endemic in that research setting. All cholera cases were laboratory confirmed for *V. cholerea* isolated from subject's stool. A dysentery case was defined as the presence of blood in stool and sought for medical treatment at the Matlab hospital. The spatial distribution of cholera and dysentery cases is shown in Figure 1.

Disease mapping

Disease mapping requires the interpolation of bari-level rate data to the nodes of a grid covering the study area, in this case a grid with 100 meters spacing. Although spatial interpolation methods, such as kriging, are now fairly common, they cannot be applied directly to disease data since they ignore the population size underlying the computation of rate data. In other words, rates computed from small population sizes (e.g. half of the baris include no more than 19 individuals) tend to be less reliable than rates computed in densely populated areas, and this effect should be incorporated into the estimation algorithm. Poisson kriging allows one to account for population size in the analysis of disease data. The risk of contracting a disease at a given location $\mathbf{u}_\alpha$, denoted $r(\mathbf{u}_\alpha)$, is estimated as a linear combination of K neighboring data:

$$\hat{r}_{PK}(\mathbf{u}_\alpha) = \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) z(\mathbf{u}_i) \qquad (\text{Equation 1})$$

where $z(\mathbf{u}_i)$ is the rate observed at location $\mathbf{u}_i$. The kriging weights are found by solving the following system of (K+1) linear equations:

$$\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha)\left[C_R(\mathbf{u}_i - \mathbf{u}_j) + \delta_{ij} \frac{m^*}{n(\mathbf{u}_i)} \right] + \mu(\mathbf{u}_\alpha) = C_R(\mathbf{u}_i - \mathbf{u}_\alpha) \quad i = 1,...,K$$

$$\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha) = 1$$

where δ_{ij} = 1 if $\mathbf{u}_i = \mathbf{u}_j$ and 0 otherwise, $n(\mathbf{u}_i)$ is the population size at $\mathbf{u}_i$, and m^* is the population-weighted mean of the set of N rates computed as:

$$m^* = \frac{\sum_{\alpha=1}^{N} n(\mathbf{u}_\alpha) z(\mathbf{u}_\alpha)}{\sum_{\alpha=1}^{N} n(\mathbf{u}_\alpha)} \qquad (\text{Equation 3})$$

The quantity $m^*/n(\mathbf{u}_i)$ is an error variance term that represents the variability arising from population size and is derived directly under the Poisson model for the counts [21]. The incorporation of this term for a zero distance (i.e. $\mathbf{u}_i = \mathbf{u}_j$) leads one to assign smaller kriging weights to rates that are computed from smaller populations and deemed less reliable. The term $\mu(\mathbf{u}_\alpha)$ is a Lagrange parameter that results from the minimization of the estimation variance subject to the unbiasedness constraint on the estimator.

To solve the kriging system (Equation 2), one needs to have a model of the spatial covariance of the unknown risk, $C_R(\mathbf{h})$, or equivalently its semivariogram $\gamma_R(\mathbf{h})$ = $C_R(0)$-$C_R(\mathbf{h})$. The experimental semivariogram of the risk is computed using the following estimator developed by Monestiez et al. [23]:

$$\hat{\gamma}_R(\mathbf{h}) = \frac{1}{2\sum_{\alpha=1}^{N(\mathbf{h})} \frac{n(\mathbf{u}_\alpha)n(\mathbf{u}_\alpha+\mathbf{h})}{n(\mathbf{u}_\alpha)+n(\mathbf{u}_\alpha+\mathbf{h})}} \sum_{\alpha=1}^{N(\mathbf{h})} \left\{ \frac{n(\mathbf{u}_\alpha)n(\mathbf{u}_\alpha+\mathbf{h})}{n(\mathbf{u}_\alpha)+n(\mathbf{u}_\alpha+\mathbf{h})} \left[z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha+\mathbf{h}) \right]^2 - m^* \right\} \qquad (\text{Equation 4})$$

where $N(\mathbf{h})$ is the number of pairs of baris separated by a vector $\mathbf{h}$. The different spatial increments $[z(\mathbf{u}_\alpha)-z(\mathbf{u}_\alpha+\mathbf{h})]^2$ are weighted by a function of their respective population sizes, $n(\mathbf{u}_\alpha)n(\mathbf{u}_\alpha+\mathbf{h})/(n(\mathbf{u}_\alpha)+n(\mathbf{u}_\alpha+\mathbf{h}))$, a term which is inversely proportional to their standard deviation. More importance is thus given to the more reliable data pairs (i.e. smaller standard deviations). A permissible model, $\gamma_R(\mathbf{h})$, is then fitted to the experimental semivariogram in order to derive the semivariogram, or covariance value, for any possible distance h. In this paper, the modeling was conducted using the weighted least-square
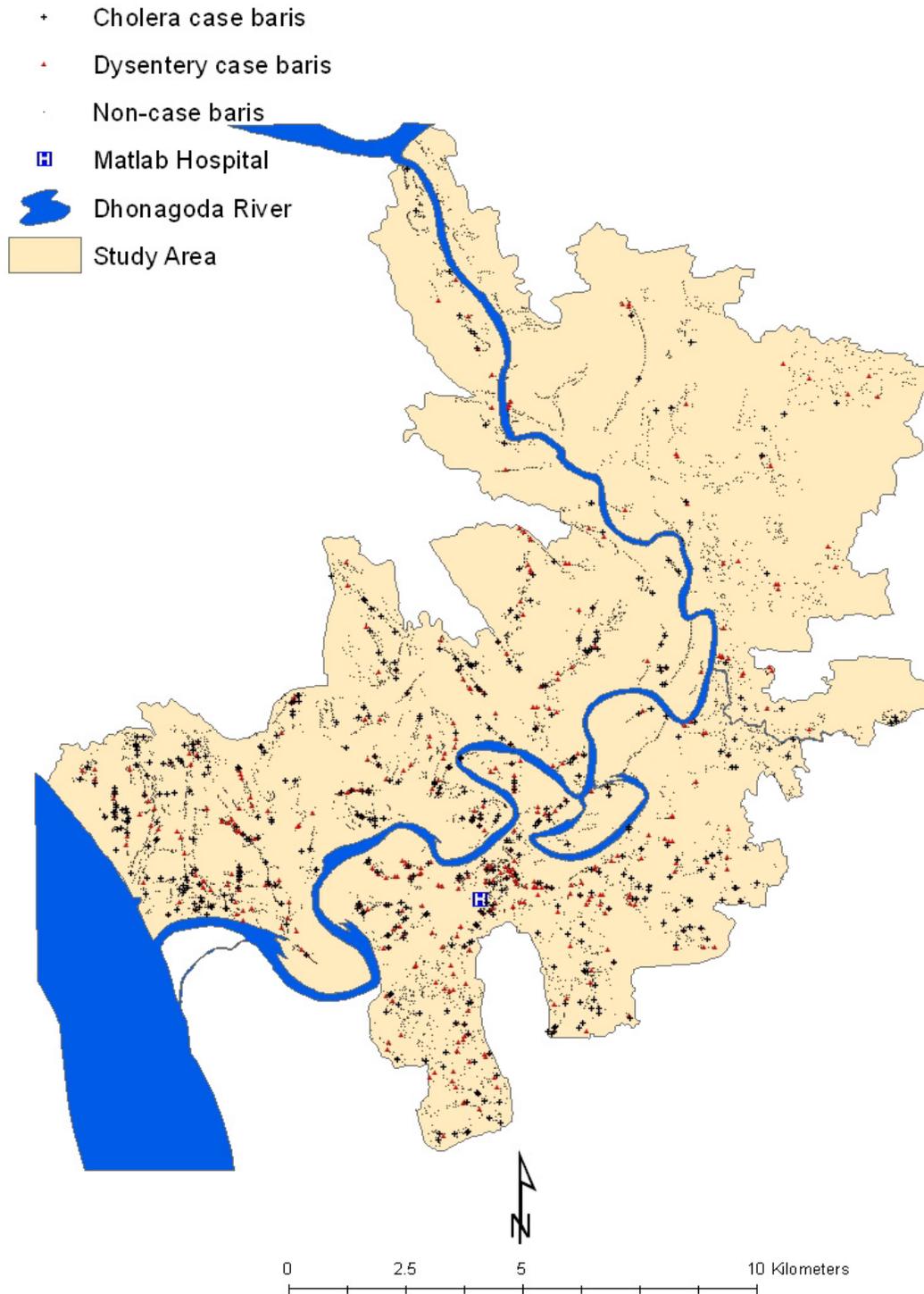
Figure 1
Spatial distribution of cholera and dysentery in Matlab, 1994. There were 7181 baris in the study area. In 1994, a total of 706 cholera cases were observed in 455 baris, and the 510 dysentery cases were observed in 421 baris.

regression procedure implemented in the public-domain software **poisson_kriging.exe** [21].

A probabilistic model of the uncertainty about the risk of incidence of the disease at a given location $\mathbf{u}_\alpha$ denoted $r(\mathbf{u}_\alpha)$, is provided by the random variable $R(\mathbf{u}_\alpha)$. This variable is fully characterized by its conditional cumulative distribution function (ccdf) defined as:

$$F(\mathbf{u}_\alpha; r \mid (K)) = \text{Prob}\{R(\mathbf{u}_\alpha) \le r \mid (K)\} = G\left(\frac{r - r_{PK}(\mathbf{u}_\alpha)}{\sqrt{\sigma_{PK}^2(\mathbf{u}_\alpha)}}\right) \qquad (\text{Equation } 5)$$

where $G(.)$ is the cumulative distribution function of the standard normal random variable. The notation "$\mid(K)$" expresses conditioning to the local information, say, K neighboring observed rates. The ccdf is modeled as a Gaussian distribution [22] with the mean and variance corresponding to the Poisson kriging estimate (Equation 1) and the kriging variance that is computed as:

$$\sigma_{PK}^2(\mathbf{u}_\alpha) = C_R(0) - \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) C_R(\mathbf{u}_i - \mathbf{u}_\alpha) - \mu(\mathbf{u}_\alpha) \qquad (\text{Equation } 6)$$

The ccdf (Equation (5)) allows the computation and mapping of the probability that the local incidence rate exceeds any particular threshold. In absence of regulatory threshold in the present case study, this function is here used to assess whether the local incidence rate at $\mathbf{u}_\alpha$ is significantly different from the regional rate m* (Equation 3). The *p*-value for the test is calculated as:

$$p(\mathbf{u}_\alpha) = 2\left[1 - G\left(\left|\frac{m* - r_{PK}(\mathbf{u}_\alpha)}{\sqrt{\sigma_{PK}^2(\mathbf{u}_\alpha)}}\right|\right)\right] \qquad (\text{Equation } 7)$$

Because the test is repeated over a large number of grid node $\mathbf{u}_\alpha$, some locations are expected to be tested significant even if there is no significant differences present (i.e. multiple testing). Rather than using the very conservative Bonferroni correction, *p*-values were here simply used to rank locations according to their likelihood of having much smaller or larger incidence rates without any formal classification as significant locations or not.

Results and discussion

Figure 2 shows admission rates for both cholera and dysentery patients, as well as death due to acute watery diarrhea, as a function of the distance to the Matlab hospital. Mortality data were collected through routine demographic surveillance programs that ensure completeness of the data for the entire study area. Unlike mortality rates due to acute watery diarrhea, the two incidence rates decline to a near zero level when the bari is more than 9 km away from the hospital. This strong contrast between mortality and disease incidence rates highlights the impact of proximity to medical facilities on hospital admittance; hence the risk of bias in mapping disease rates in remote areas. The following analysis has thus been restricted to data collected in bari within a radius of 9 kilometers from the hospital.

Experimental risk semivariograms were computed from the 6,017 bari-level incidence rates using the population-weighted estimator (Equation 4). Since no systematic difference was observed between directional semivariograms, the spatial variability was considered isotropic and only the omnidirectional semivariograms are displayed in Figure 3. For both diseases, the semivariogram was estimated using 50 lags of 50 meters. The modeling was performed using weighted least-square regression where the weights account for both the number of data pairs and the semivariogram value (i.e. weighting option #2, $w(\mathbf{h}_l) = \sqrt{N(\mathbf{h}_l)} / \gamma(\mathbf{h}_l)$, in the program **poisson_kriging.exe**).

The cholera semivariogram model is exponential with a practical range of 370 meters, while a spherical model with a range of 1,100 meters was fitted for dysentery. The larger range of autocorrelation for the second model indicates a better spatial structure of dysentery incidence rates, which can be explained by the differences in the mode of transmission for the two diseases. Cholera is usually transmitted to a person through oral-fecal-oral process, while dysentery can be spread through hand-to-hand contacts. In the past there have been a number of interventions carried out in the study area to reduce the risk of cholera, while dysentery received much less attention. Therefore, local differences in the socioeconomic condition and health behavior among neighboring households due to health interventions may explain the shorter range of spatial autocorrelation displayed by cholera incidence rates relatively to dysentery.

The risk of incidence of both diseases was mapped over a grid with a 100-meter spacing. All 13,689 grid nodes are within a 9 km distance from the Matlab hospital. At each node, up to a maximum of 64 closest observations within a radius of 5 km were used for Poisson kriging. Figure 4 shows the maps of cholera and dysentery risk. In both cases, the boundaries of the color classes correspond to the deciles of the histogram of risk estimates. Table 1 indicates that, on average over all grid nodes, the kriged risks are slightly smaller than the observed rates. Yet, the extreme high, and likely unstable, incidence rates (i.e. 500 per 1,000 habitants for cholera and 250 per 1,000 habitants for dysentery) are not reproduced by the risk maps that display a smaller range of variation. The cholera risk map is patchier than the dysentery map that shows a large zone of high incidence in the south-central part of the
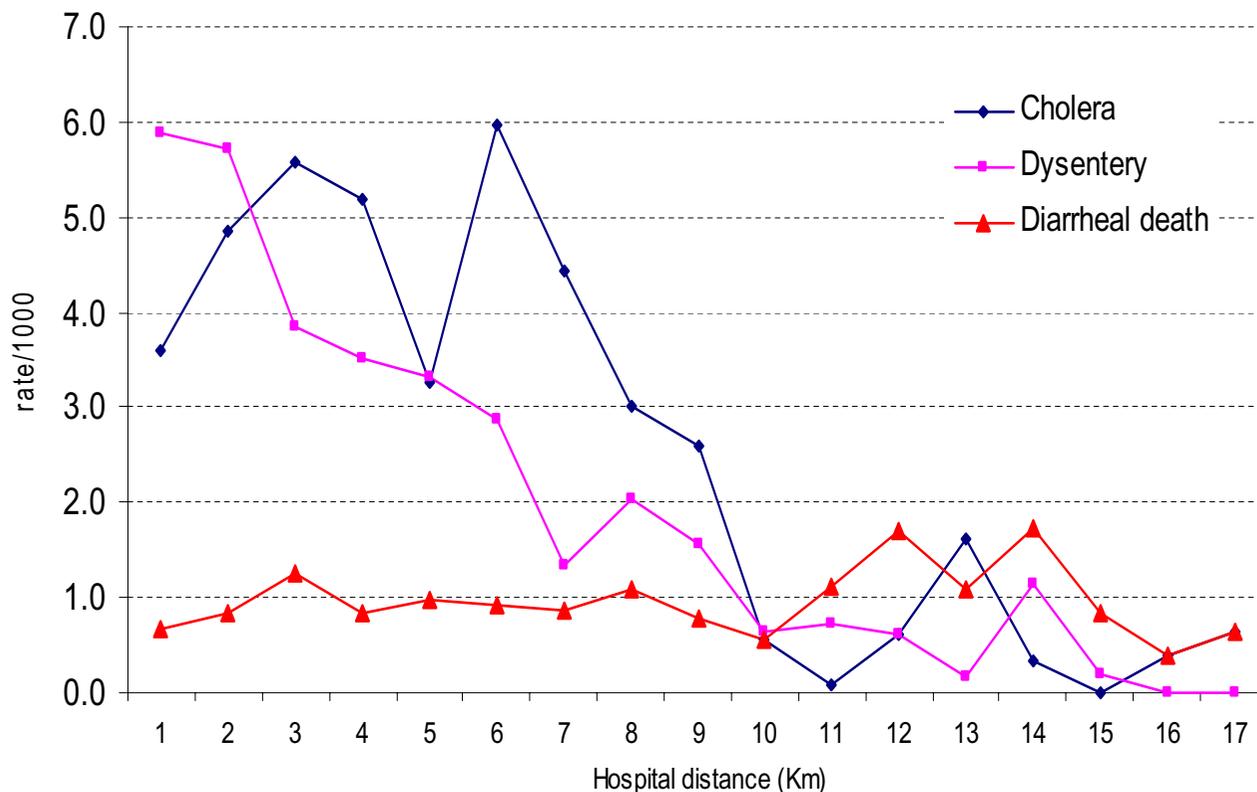
Figure 2
Hospital admittance rate for cholera and dysentery, and death due to diarrheal death (watery) in Matlab as a function of the distance to the Hospital, 1994. The morbidity data were obtained from hospital-based disease surveillance, while the mortality data were obtained from community based demographic surveillance. All the rates are expressed as number of cases per 1,000 habitants.

study area, which is a quasi-urban zone. This could be due to household level differences in the socioeconomic conditions and health behavior described above. On both maps, lower risks are found in the Northern part of the study area, which is also the most distant from the Matlab hospital. This can partially be explained by the inverse relationship between hospital admittance and distance to the medical facility which is illustrated in Figure 2.

The kriging variance maps (Figure 5) capture both the spatial distribution of the population across the study area and the spatial pattern of the incidence rates. The greater magnitude of the kriging variance computed for cholera is caused by the shorter range of autocorrelation of its semivariogram model: 0.37 km versus 1.1 km for dysentery.

The larger kriging variance alongside the border of the study area highlights lower density of the population.

The kriging variance maps were combined with the risk maps to assess whether the local disease rate is significantly different from the regional rate. For both diseases, only a few grid nodes have a p-value (Equation 6) smaller than 0.05; see Figure 6. Because of multiple testing, these positive results are likely caused by chance and we can safely assume that no local incidence rate departs significantly from the regional rates. Yet, the maps in Figure 6 can be used to rank areas according to the likelihood that the local rate is much smaller or larger than the regional rate. This information should prove useful to guide future control activities.
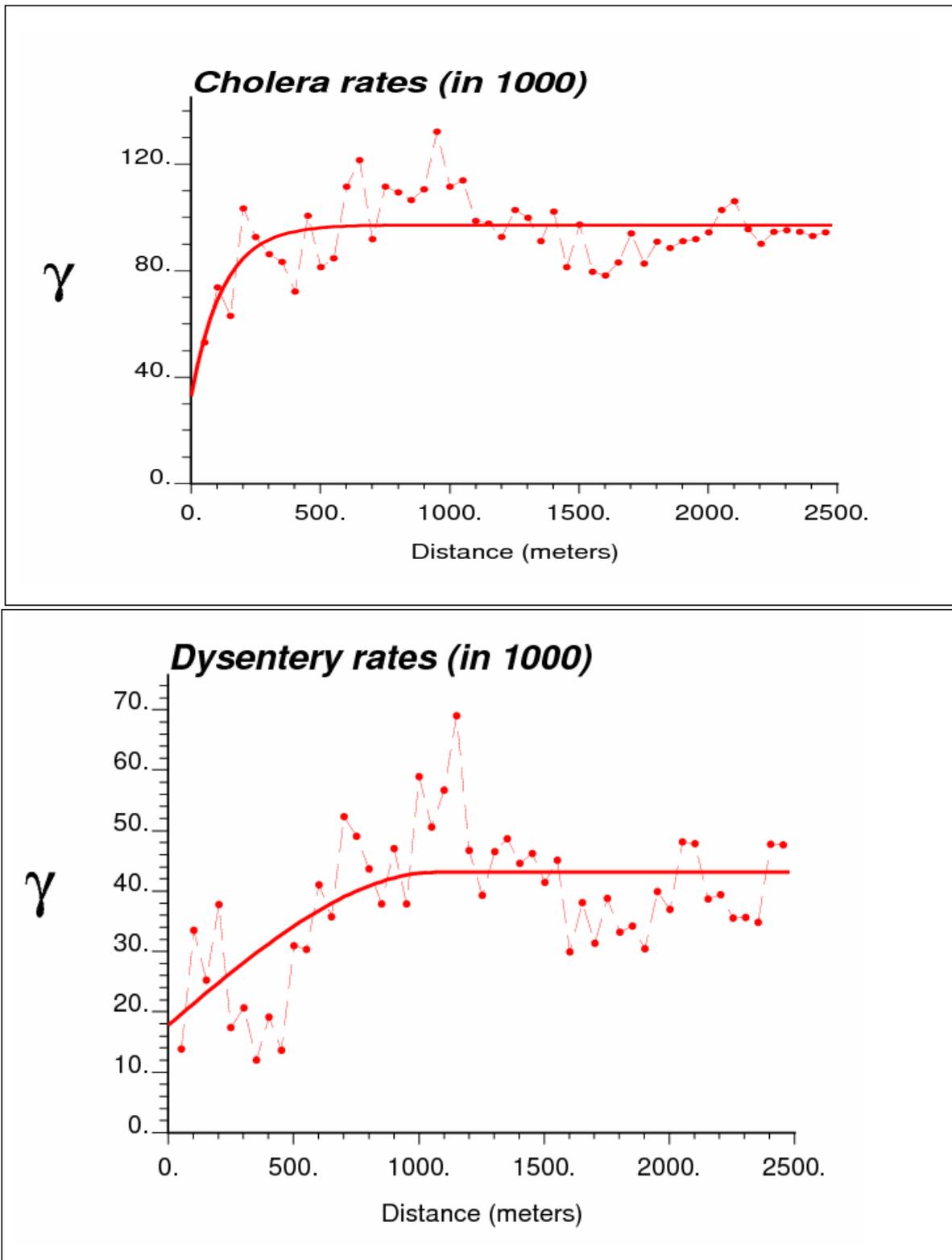
Figure 3
Omnidirectional semivariograms of cholera and dysentery incidence rates, with the model fitted.

Table 1: Descriptive statistics of the distribution of incidence rates and kriged risks for cholera and dysentery in Matlab, 1994.

Statistics	Cholera		Dysentery	
	Kriged risk	**Observed rate**	**Kriged risk**	**Observed rate**
N	13689	6017	13689	6017
Mean	3.7238	3.8512	2.6841	2.7732
Std. Deviation	3.5083	20.134	2.4847	14.776
Minimum	0.00	0.00	0.00	0.00
Maximum	40.93	500.00	15.92	250

Units are number of cases per 1,000 habitants.

Conclusion

In Matlab, disease maps have never been used to isolate areas of high cholera or dysentery incidence and to guide for control activities. The health authority of the area could however greatly benefit from the availability of such maps for their disease control programs. Spatial interpolation of incidence rates can be performed using a variety of simple techniques, including inverse square distance methods, spline or moving average. The main limitation of these deterministic methods is that they fail to provide
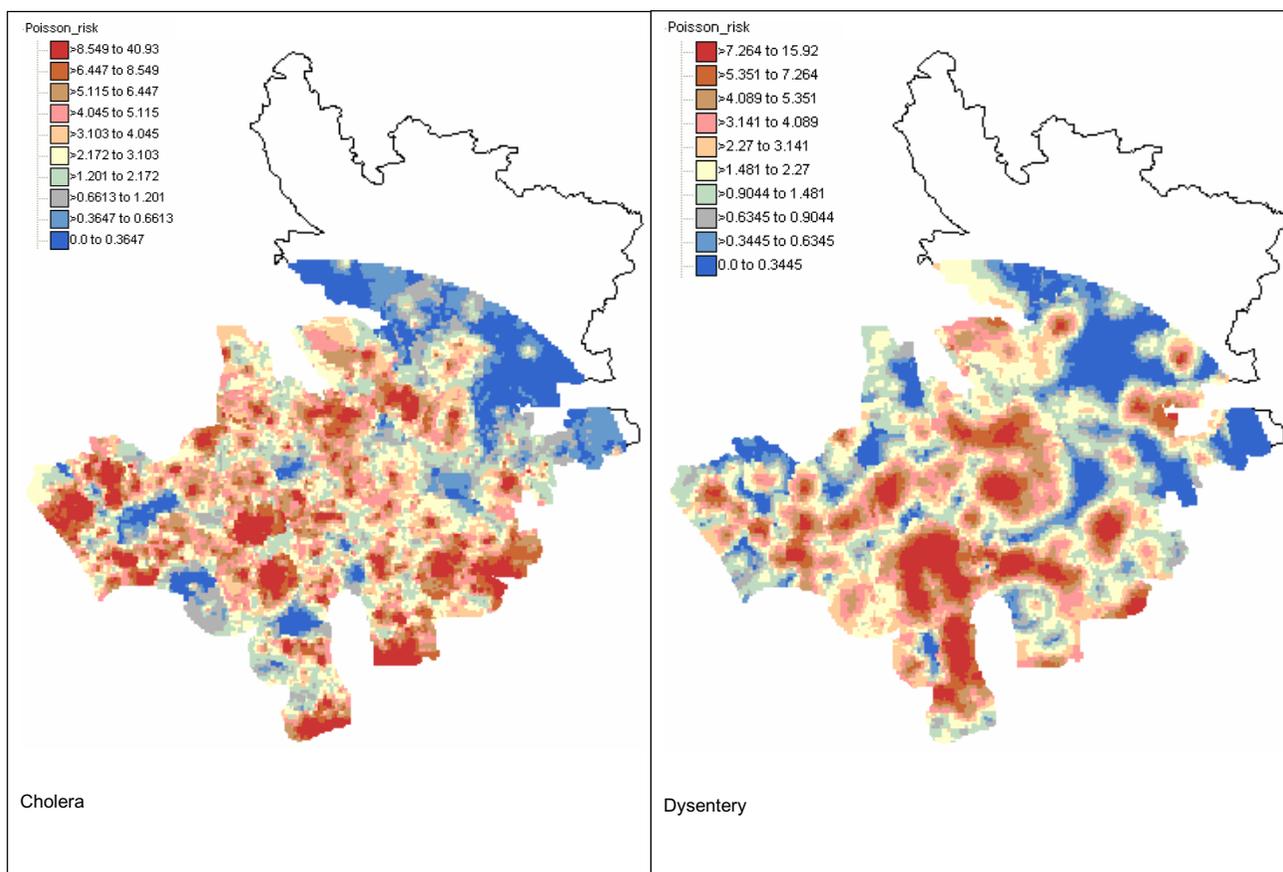
Figure 4
Maps of cholera and dysentery incidence risk in Matlab 1994 estimated by Poisson kriging. The class boundaries of the color legend correspond to the deciles of the histogram of risk estimates.
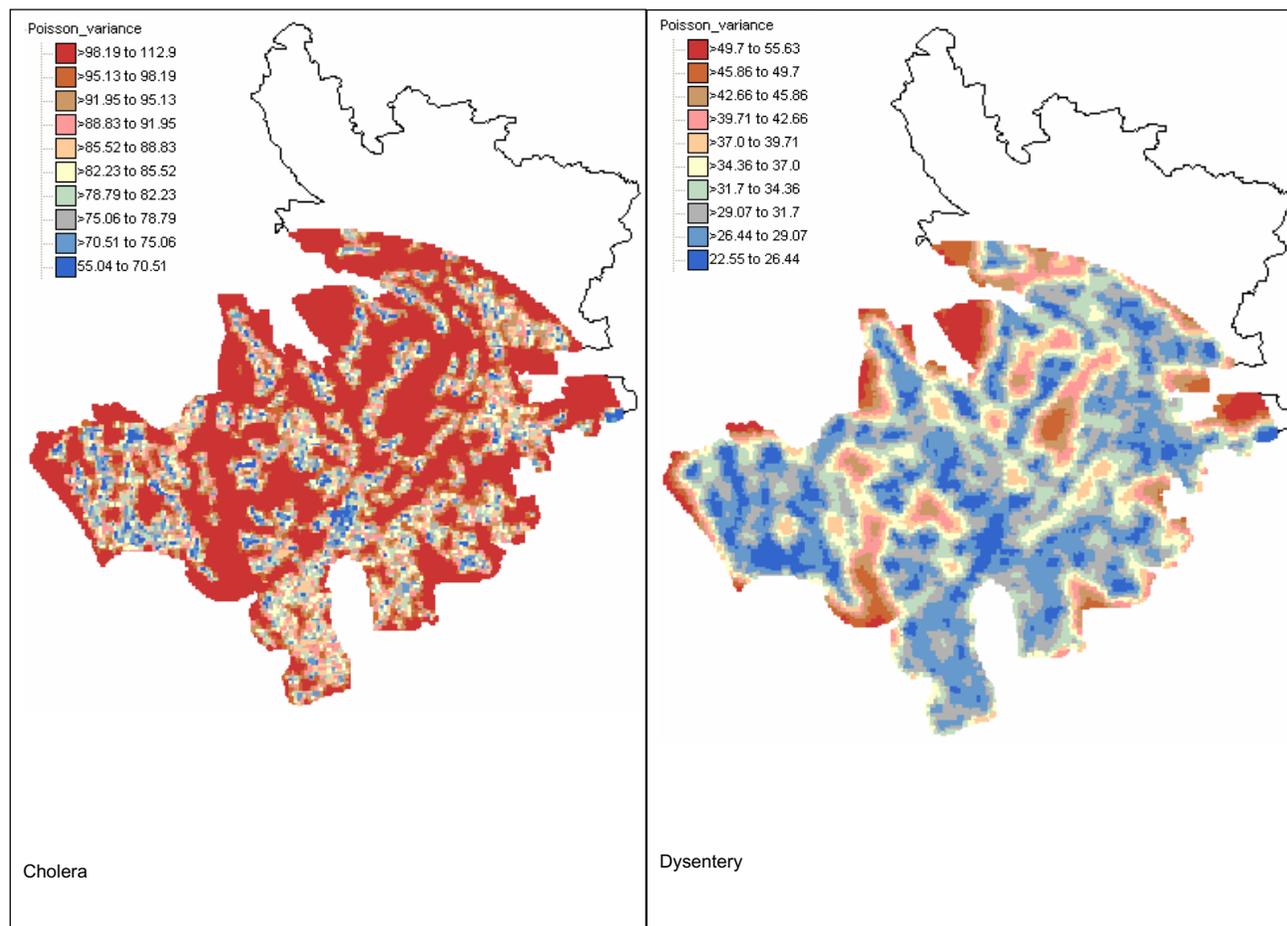
Figure 5
Maps of Poisson kriging variance for cholera and dysentery incidence risk in Matlab 1994. The class boundaries of the color legend correspond to the deciles of the histogram of kriging variances.

a measure of the reliability of the prediction. Kriging is increasingly preferred because it accounts for spatial patterns and data configuration (i.e. clustering of observations) in the prediction, while providing both an estimate and a measure of the error variance. The kriging variance can be combined with the kriging estimate to compute confidence intervals or the probability of exceeding specific thresholds, under the assumption of Gaussianity.

Interpolation of bari-level data cannot be performed using conventional geostatistical tools that ignore the population size underlying the computation of incidence rates. Because the number of individuals living in bari greatly fluctuates across the study area, local incidence rates have various degrees of reliability that need to be

accounted for in the variography and spatial interpolation. Poisson kriging, which was introduced in the health science literature for filtering the noise in cancer mortality maps, is here used to create continuous maps of incidence risks from punctual bari-level data.

Unlike traditional semivariogram estimators, the version used in this paper directly accounts for population size in the computation of the semivariograms, attenuating the influence of less reliable rates recorded in sparsely populated areas. Cholera and dysentery display semivariograms with very different ranges of autocorrelation. The much larger range (1.1 km versus 0.370 km) observed for dysentery indicates that its incidence rates vary more continuously in space than cholera rates across the study area.
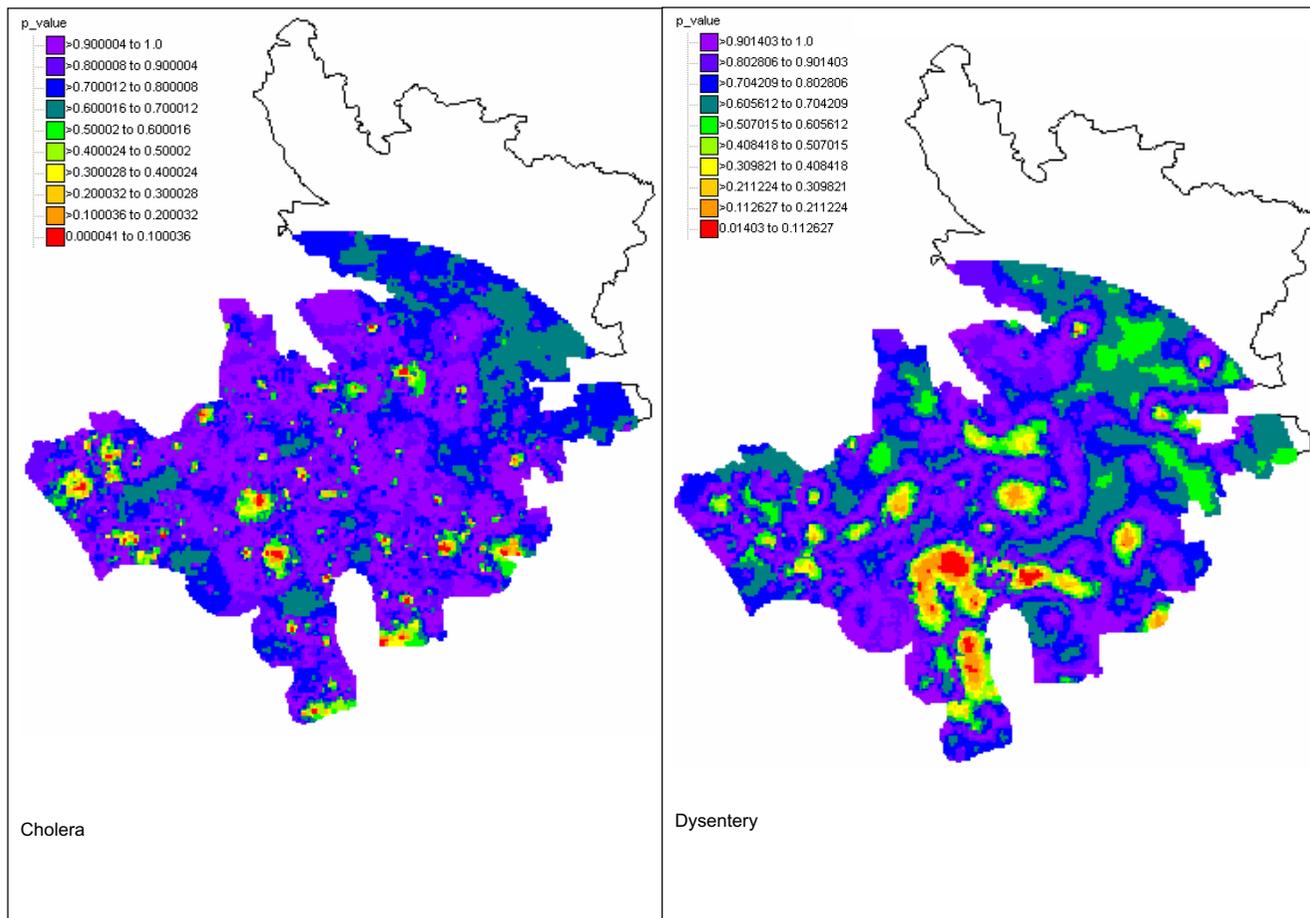
Figure 6
Maps of the p-values for the test of uniformity of local incidence rates for cholera and dysentery in Matlab 1994. At each location, the population-weighted mean of all bari rates is compared to the probability distribution of the local incidence rate which is fully characterized by the Poisson kriging estimate and variance. The lower the p-value, the more likely the local rate is significantly smaller or higher than the regional rate.

The maps of disease risk illustrated the better spatial continuity of rates recorded for dysentery relatively to cholera. Combined with the different location of high-incidence areas, these contrasted spatial patterns suggest that different mechanisms are responsible for the occurrence of these two diseases. Within the field of spatial epidemiology, knowledge about host, agent and environment is needed to get insight into the etiology of the disease and to develop and promote disease control measures [30]. This means that spatial epidemiological studies have to characterize spatial patterns of disease risk. Spatial variations of disease risk results in clustering of disease, and that may reflect the existence of specific subpopulations that are under higher or lower exposure to the disease. Our risk maps for cholera and dysentery may be useful for

identifying putative factors of increased disease risk and for helping health officials take remedial actions.

The high incidence rates for dysentery in the urban part of the study area can be attributed to food habit pattern given that habitants of the study area use their hands to eat. Those who live close to urban areas have easy access to restaurants which tend to ignore basic hygienic rules. Thus, eating meals that are not served hygienically at those places increases the risk of acquiring dysentery. Our dysentery risk map suggests the need of interventions for combating the threats of dysentery in the urban part of the study area. On the other hand, the isolated clusters of high cholera incidence may be attributed to the poor health behavior (e.g. hands are not washed systematically after

bathroom break and before eating, drinking water is not boiled or filtered, cooking utensils are washed with surface water, etc.) of the community. We can not also ignore the possibility of environmental niche of endemic cholera in those areas. Further investigations focusing on the areas of high incidence risk for cholera could shed some light on the factors that cause these local spikes in cholera incidence.

Acknowledgements

References

1. Kelsall J, Wakefield J: **Modeling spatial variation in disease risk: a geostatistical approach.** *Journal of the American Statistical Association* 2002, **97(459):**692-701.
2. Kulldroff M: **A spatial scan statistics.** *Communications in Statistics: Theory and Methods* 1997, **26:**1481-1496.
3. Bithell JF: **A classification of disease mapping methods.** *Stat Med* 2000, **19(17–18):**2203-2215.
4. Diggle PJ: **Overview of statistical methods for disease mapping and its relationship to cluster detection.** In In *Spatial Epidemiology: Methods and Applications* Edited by: Elliot P, Wakefield JC, Best NG, Briggs DJ. Oxford: Oxford University Press; 2000.
5. Lawson AB: **Disease map reconstruction.** *Stat Med* 2001, **20(14):**2183-204.
6. Lawson AB, Clark A: **Spatial mixture relative risk models applied to disease mapping.** *Stat Med* 2002, **21(3):**359-70.
7. Wakefield J, Salway R: **A statistical framework for ecological and aggregate studies.** *Journal of Royal Statistical Society, Series A* 2001, **164:**119-137.
8. Gunnlaugsson G, Angulo FJ, Einarsdottir J, Passa A, Tauxe RV: **Epidemic cholera in Guinea-Bissau: the challenge of preventing deaths in rural West Africa.** *Int J Infect Dis* 2000, **4(1):**8-13.
9. La Raja M: **Cholera-like diarrhoeal disease and rivers in rural Mozambique.** *Trop Doct* 2000, **30(1):**53-4.
10. Germani Y, Quilici ML, Glaziou P, Mattera D, Morvan J, Fournier JM: **Emergence of cholera in the Central African Republic.** *Eur J Clin Microbiol Infect Dis* 1998, **17(12):**888-890.
11. Glass RI, Becker S, Huq MI, Stoll BJ, Khan MU, Merson MH, Lee JV, Black RE: **Endemic cholera in rural Bangladesh, 1966–1980.** *Am J Epidemiol* 1982, **116(6):**959-70.
12. Burrows W: *Textbook of Microbiology* Philadelphia: Saunders; 1979.
13. Glass RI, Black RE: **The epidemiology of cholera.** In *Cholera* Edited by: Barua D, Greenough III W. New York: Plenum Medical Company; 1992:129-54.
14. World Health Organization: **Cholera today. A new look at an old disease.** *WHO Features* 1987, **107:**1-3.
15. Blake PA: **Cholera: a possible endemic focus in the United States.** *N Engl J Med* 1980, **302:**305-09.
16. Salmaso S, Greco D, Bonfiglio B, Castellani-Pastoris M, De Felip G, Bracciotti A, Sitzia G, Congiu A, Piu G, Angioni G, Barra L, Zampieri A, Baine WB: **Recurrence of Pelecypod-associated cholera in Sardinia.** *Lancet* 1980, **2(8204):**1124-1127.
17. Emch M: **Diarrheal disease risk in Matlab, Bangladesh.** *Social Science & Medicine* 1999, **49:**519-530.
18. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, Morris GJ Jr, Khan MNH, Siddique AK, Yunus M, Albert MJ, Sack DA, Colwell RR: **Critical Factors Influencing the Occurrence of Vibrio chol-**
19. erae in the Environment of Bangladesh. *Appl Envir Microbiol* 2005, **71(8):**4645-4654.
19. Webster R, Oliver MA, Munir KR, Mann JR: **Kriging the local risk of a rare disease from a register of diagnoses.** *Geographical Analysis* 1994, **26(2):**168-185.
20. Carrat F, Valleron AJ: **Epidemiologic mapping using the "kriging" method: application to an influenza-like illness epidemic in France.** *Am J Epidemiol* 1992, **135(11):**1293-1300.
21. Goovaerts P: **Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging.** *Int J Health Geogr* 2005, **4:**31.
22. Goovaerts P: **Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation.** *Int J Health Geogr* 2006, **5:**7.
23. Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C: **Geostatistical modelling of spatial distribution of Balenoptera physalus in the northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts.** *Ecological Modelling* 2006, **193:**615-628.
24. Rashid ER: *Geography of Bangladesh* University Press Limited, Dhaka, Bangladesh; 1991.
25. Hall AJ: *An evaluation of the effectiveness and impact of Bangladesh's Chandpur Flood Control and Irrigation Project. Doctoral Thesis* Department of Agriculture, University College of North Wales, Bangor, United Kingdom; 1988.
26. Ali M, Emch M, Ashley C, Streatfield PK: **Implementation of a medical geographic information system: concepts and uses.** *J Health Popul Nutr* 2001, **19(2):**100-110.
27. Koenig MA, Fauveau V, Chowdhury AI, Chakraborty J, Khan MA: **Maternal mortality in Matlab, Bangladesh: 1976–85.** *Studies in Family Planning* 1988, **19(2):**69-80.
28. D'Souza S: **A population laboratory for studying disease processes and mortality-the demographic surveillance system, Matlab, Bangladesh.** *Special Publication No. 13, ICDDR,B, Dhaka, Bangladesh* 1981.
29. Phillips JF, Simmons R, Koneig MA, Chakraborty J: **Determinants of reproductive change in a traditional society: evidence from Matlab, Bangladesh.** *Studies in Family Planning* 1988, **19(6):**313-334.
30. Berke O: **Exploratory spatial relative risk mapping.** *Preventive Veterinary Medicine* 2005, **71:**173-182.