

On Studying Spatial Patterns and Association Between these Patterns of Mortality and Prosperity in Malaysia

Faisal G. Khamis, Abdul Aziz Jemain, and Kamarulzaman Ibrahim
School of mathematics, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract: In this study, we examined the geographical distribution (the spatial structure) of mortality and prosperity in Malaysia. In addition we proposed an approach to investigate the association between clustering patterns of mortality and prosperity across different areas of the country. To characterize the geographic pattern of mortality and prosperity, three indicators (infant, neonatal, and stillbirth) were proposed for mortality and also three indicators (class1, class2 and class3) for occupation were proposed for prosperity. These indicators measure the level of mortality and prosperity factors for all 81 districts in peninsula Malaysia based on 1993 census data. Two statistics of spatial autocorrelation based on sharing boundary neighbours known as global and local Moran are used to investigate the global and local clustering respectively. We found that both mortality and prosperity varied significantly across the different districts. Also, we found many significant local clusters in both mortality (in the north, south and mid-west), and in prosperity (in the north and west). A significant association was found between mortality and prosperity based on the spatial correlation coefficient.

Key words: Autocorrelation, sharing boundary neighbors, mapping, global and local moran, monte carlo simulation

INTRODUCTION

More than 10 000 newborn babies die every day.^[12] Every year, it is estimated that under nutrition contributes to the deaths of about 5.6 million children under the age of five; 146 million children in the developing world are underweight and at increased risk of an early death.^[2] Infant mortality is considered, as a standard indicator of population health used through out the world; the rates of infant mortality can reflect levels of social and economic development, levels of care, and the effectiveness of preventive programs, as well as post-birth services to both mothers and their children.^[8] The importance of our goal or our purpose is followed from what Weeks^[13] stated: There are few things in the world more frightening and awesome than the responsibility for a newborn child-fragile and completely dependent on others for survival. Benach and Yasui^[9] analyzed the geographical pattern and the magnitude of the association between deprivation and mortality in Spain; they found that geographical gradient from north east to south west was shown by both mortality and deprivation.

The general pattern is that the higher the individual income the lower the risk of disease and mortality.^[16]

Their study in Rio de Janeiro, Brazil revealed that intra-city variations of the post-neonatal mortality rate are associated with geographic patterns of poverty, and that pregnancy in adolescence is strongly and contextually correlated with intra-neighborhood poverty clustering. Contextual effects operate when the health status of individual depends not only on their characteristics but also on the supra-individual effects associated with the area where they live or the social group to which they belong. To understand the linkages between those variables, investigations should focus on features of the areas rather than on the compositional characteristics of residents of the area, which can not fully describe the social environment in which people live.^[17] In Japan, area socioeconomic disadvantage is significantly related to higher mortality, especially on premature death^[8]. Researchers have suggested that not only the absolute standard of living but also the magnitude of the gap between the rich and the poor matter in terms of population health.^[19] These findings suggest that we should be looking at certain features of local areas, where in our case mortality and prosperity levels for different districts are considered. So, we studied the association of the geographic pattern of prosperity on the geographic pattern of babies' mortality distribution.

We used an approach to investigate the hypothesis that the clustering of prosperity affects the mortality of communities more compared to when prosperity is randomly scattered. We define three indicators for both mortality and prosperity factors that measure the level of mortality and prosperity factors among districts and analyze the effects of geographic prosperity clustering on the geographic mortality clustering in Malaysia, based on index of similarity. Spatial autocorrelation is the term used for the interdependence of the values of a variable over space.^[15] The purpose of spatial analysis is to identify pattern in geographic data and attempt to explain these patterns. Findings are expected to enhance mortality monitoring and policing capabilities across districts in Malaysia. Mortality mapping plays an important role in the monitoring of community's health.

Maps can reveal spatial patterns not previously recognized or suspected from the examination of a table of statistics and reveal high risk communities or problem areas.^[18] They stated that, at the relevant spatial locations, some covariables which are related to the disease distribution, such as social or behavioral measures can be used to describe the local population. These covariables may relate closely to the health status of the local community and so their inclusion in any analysis could help to assess more accurately the local population "at risk" structure. Often, these covariables are lifestyle or occupational indices which help to indicate, albeit indirectly, the expected incidence of disease. The inclusion of socioeconomic factors facilitates interpretation of the disease maps. In accordance to the inclusion of social economic factors such as level of income to describe the mapping of a disease, we consider prosperity factor as the covariable for mortality. This study investigated the existence of spatial clustering and clusters of districts with respect to mortality and prosperity factors and the association between the geographic patterns *globally and locally* of these two factors.

MATERIALS AND METHODS

Data: The data are collected from the department of statistics based on the census conducted in Malaysia-1993.^[5] For each of the (N=81) districts, indicators (observed variables) for the factors, are measured in terms of ratios and percentages. All indicators are transformed to normal distribution. We must construct on the basis of the prior concept or statistical analyses, which particular indicators load on each factor. More

precisely, we construct the following factors with their respective indicators:

Mortality: has three indicators which are: standardized infant mortality ratio, standardized neonatal mortality ratio, and standardized stillbirth mortality ratio. Infant mortality indicates the deaths under one year of age. Neonatal mortality refers to the deaths within 28 days after birth. Stillbirth occurs after 24 weeks of gestation.^[6] Standardized mortality ratio (SMR) allows comparison of the causes of death between population groups.^[7] Standardized ratio (SR) measures the extent to which each area has a number of cases more than would be expected (SR > 1) or less than expected (SR < 1) under the hypothesis of a random distribution of cases over the region.^[20] It is calculated as follows^[20,24]:

$$SMR_i = O_i/E_i, i = 1, 2, \dots, 81$$

$$E_i = \frac{\sum_{i=1}^{81} O_i}{\sum_{i=1}^{81} n_i} \times n_i$$

where SMR_i is standardized mortality ratio for i th district, O_i is observed number of deaths for i th district, E_i is expected number of deaths for i th district, and n_i is the number of live births (population at risk) for i th district.

Prosperity: it means level of economic development, represents the type of occupation status, which is grouped into three classes starting from top to bottom in the income and social level as follows: class1 includes professional, administrative and managerial workers; class 2 includes clerical workers and class3 includes sales, and service workers. All classes are measured in percentages. Low income status is considered one of the important factors to have poorer health than those with higher income status, as well as income provides necessities such as food and health care.^[1] When pregnant women are not adequately nourished, their babies are borne at low weights, putting their survival at risk.^[2] The babies of fathers in semi-routine occupations had infant mortality rates over 2.5 times higher than those of babies whose fathers were in higher professional occupations.^[3] Low levels of

occupational security often accompany poverty status and poverty can induce serious health risks including mortality.^[4]

Analysis: Our analysis involves eight steps. Step 1 was to calculate the standardized mortality ratio (data not shown; available upon request). Step 2 was to assess the distribution of all indicators. Several indicators suffer from non-normality; thus normal scores were used for transformation, which are a monotonic transformation of the original scores with same mean and standard deviation^[25] to produce more normal distributions, and also multivariate normal distribution were tested against those indicators. In step 3 we constructed mortality and prosperity factors using factor analysis. In step 4 we conducted visual inspection based on the quantified gradients for the two factors using quantiles, in particular comparing districts with high mortality and districts with low prosperity. Step 5 includes the calculation of global Moran's I for each factor to detect the global clustering and then local Moran's I_i for i th district to detect the local clusters. In step 6 we visually inspected the gradients of local Moran values for both factors using quantiles. Step 7 includes the calculation of bivariate spatial correlation between mortality factor and prosperity factor. In step 8, we investigated the significant of autocorrelation coefficients for both factors using permutation test; and investigated the significant of bivariate spatial correlation using Monte Carlo simulation.

Based on global Moran's I statistic we can test whether the geographical distribution of mortality and prosperity is random or not. We are interested in detecting and evaluating localized clusters using local Moran's I_i statistic. In regional data analysis, districts in close proximity to one another with similar values produce a spatial pattern indicative of positive spatial autocorrelation. Identifying groups of districts in close proximity to one another with high values is often of particular interest suggesting a "cluster" of elevated risk. Another goal in regional data analysis is identification of the spatial risk factors for the response of interest using choropleth mapping.

Factor analysis with a maximum likelihood (ML) approach^[10] is employed to identify mortality and prosperity factors; mortality and prosperity scores were computed for each district by computing factor scores using regression method:

$$\hat{\eta}_j = \hat{\Lambda}'S^{-1}y_j, j = 1, 2, \dots, N,$$

where;

$\hat{\eta}_j$ = factor score for j th district

$\hat{\Lambda}_{(3 \times 1)}$ = ML estimates of the factor loadings

$S_{(3 \times 3)}$ = sample covariance matrix

y_j = (3×1) vector of observations

and $N=81$. Those scores were then categorized into quantiles of certain interval, and then this interval is used for all maps in Fig. 1, using darker shades of gray to indicate increasing positive values for mortality and increasing negative values for prosperity. Such approaches enable qualitative evaluation of patterns of mortality and prosperity status. Districts are considered connected if they share a common border. With each pair of districts we associate a weight w_{ij} which is zero if $i = j$ or if the two districts are not spatially connected; otherwise, w_{ij} takes on a non-zero value (in this research $w_{ij} = 1$). Pearson's correlation coefficient between mortality and prosperity factors is calculated to study the association between them. All programs in this study are manipulated using S+6.2.1 software.

A choropleth map is used commonly to portray data collected for units, such as counties, districts, or states. To construct a choropleth map, data for enumeration units are typically grouped into classes and a gray tone is assigned to each class. Color is considered to be key to the development of good visualization tool for the purposes of quantitative data analysis, that is for helping the viewer to notice patterns in data.^[20] Although maps allow us to visually assess spatial pattern, they have two important limitations: their interpretation varies from person to person, and there is the possibility that a perceived pattern is actually the result of chance factors, and thus not meaningful. For these reasons, it makes sense to compute a numerical measure of spatial pattern, which can be accomplished using spatial autocorrelation.

Identification of global spatial clustering: The goal of a global index of spatial autocorrelation is to summarize the degree to which similar observations tend to occur near to each other. In our exploratory spatial analysis, we tested for spatial autocorrelation using standard normal deviates (z-values) of Moran's I under a normal assumption. Spatial autocorrelation is the measure of tendency for things that are alike to occur near one another in geographic space. The interpretation of the Moran statistic is as follows: if $I > E(I)$, then a district tends to be connected to districts

that have similar attribute values, and vice versa. Global clustering test is used to determine whether clustering is present throughout the study area, without determining statistical significance of local clusters.^[22] The autocorrelation coefficient can be used to test the null hypothesis of no autocorrelation versus the alternative of positive spatial autocorrelation:

- H_0 : No clustering exists (no spatial autocorrelation)
- H_1 : Clustering exists (positive spatial autocorrelation)

Moran's I is a weighted correlation coefficient used to detect departures from spatial randomness. It is used to determine whether neighboring areas are more similar than would be expected under the null hypothesis^[15]:

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} \eta_i \eta_j}{S_0 \sum_{i=1}^N \eta_i^2}$$

where;

- $N = 81$ is the number of districts,
- w_{ij} = weight denoting the strength of the connection between districts i and j ,
- η_i and η_j = estimated factor scores for mortality in i th and j th districts respectively,

and

$$S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}, i \neq j.$$

To compute Moran's I for prosperity, we used ξ instead of η in the above statistic. A significant positive value for Moran's I for a particular factor indicates positive spatial autocorrelation, showing that the overall pattern in the districts having a high/low level of mortality or prosperity similar to their neighboring districts. A significant negative value for Moran's I indicates negative spatial autocorrelation, showing that there are districts having a high/low level of mortality or prosperity unlike neighboring districts. Fundamentally, a major interest is whether mortality hot spots exist (a hot spot represents a grouping of incidents that are spatially clustered). To test the significance of global Moran's I we apply the z -

statistic, which follows a standard normal distribution, and it is calculated as follows^[24]:

$$z = \frac{I - E(I)}{\sqrt{\text{var}(I)}}$$

Converting counts to ratios to take account of population size differences is not, however, sufficient to ensure comparability of data values for the exploring purpose.^[20] Cressie^[21] determined that the square root of the number of live births times those factors have approximately equal variances. Our analysis suggests evidence of clustering if the test is significant but does not identify the locations of any particular clusters. Accordingly, local spatial statistic is advocated for identifying and assessing potential hot spots in this analysis.

Identification of local spatial clusters: A global index can suggest clustering but cannot identify individual clusters.^[24] Local Indicators of Spatial Associations (LISAs) measure the degree of spatial dependence to allow for the effects of neighborhood based on each district's associated value (in our case, mortality and prosperity factors), where neighborhood is defined according to some measure of proximity or contiguity. The main purpose of such indexes is to provide a local measure of similarity between each region's associated value and those of nearby regions. Anselin^[22] proposed the local Moran's I_i statistic to test for local autocorrelation. Local spatial clusters, sometimes referred to as hot spots, may be identified as those locations or sets of contiguous locations for which the local Moran's I_i is significant.^[22] He stated that the indication of local patterns of spatial association may be in line with a global indication, although this is not necessarily be the case. It is quite possible that the local pattern is an aberration that the global indicator would not pick up, or it may be that a few local patterns run in the opposite direction of the global spatial trend. Local values that are very different from the mean (or median) would indicate locations that contribute more than their expected share to the global statistic. These may be outliers or high leverage points and thus would invite closer scouting. Moran's I_i serves two purposes or provide two interpretations^[22]: First, it may be interpreted as indicator of hot spots (i.e., the assessment of significant local spatial clustering around an individual location). Second, it may be used to assess

the influence of individual locations on the magnitude of the global Moran statistics and to identify outliers. Moran's I_i for *ith* district may be defined as^[24]:

$$I_i = \frac{\eta_i}{s} \sum_{j=1}^N w_{ij} \frac{\eta_j}{s}, \quad i = 1, 2, \dots, 81$$

where, analogous to the global Moran's I , the η_i and η_j are the estimated factor scores for mortality, and s is the standard deviation of the factor scores. To compute Moran's I for prosperity, we used ξ instead of η in the above statistic. Local Moran statistic is used to test the null hypothesis of no clusters.

The typical output of a LISA analysis involves the values of the LISAs themselves, which are typically mapped to indicate areas with high values, suggesting stronger local correlation than others. LISA values may be due to aggregations of high values, aggregations of low values, or aggregations of moderate values. Thereby, high values of I_i suggest clusters of similar (but not necessarily large) values across several districts, and low values of I_i suggest an outlying cluster in a single district i (being different from most or all of its neighbors). We can map each district's I_i value to provide insight into the location of districts with comparatively high or low local association with their neighboring values. The application of statistical techniques to spatial data faces an important challenge, as expressed in the first law of geography: "everything is related to everything else, but closer things are more related than distant things".^[18] The quantitative expression of this principal is the effect of spatial dependence, i.e. when the observed values are spatially clustered, the samples are not independent. The obvious question after finding significant clusters of mortality is-why? Could this pattern be explained by the pattern of prosperity factor?

Bivariate spatial association: So far, we have presented only spatial method that quantifies the spatial structure of one factor at a time. There is much discussion about what is an appropriate measure of bivariate spatial association. Lee^[33] for example, develops an index L that combines Pearson's bivariate correlation with Moran's spatial autocorrelation measures. However the index does not seem interpretable without reference back to its three components.^[25] Spatial dependence or spatial clustering causes losing in the information that each observation carries. When N observations are made on a variable that is spatially dependent (and that dependence is

positive so that nearby values tend to be similar) the amount of information carried by the sample is less than the amount of information that would be carried if the N observations were independent, because a certain amount of the information carried by each observation is duplicated by other observations in the cluster. A general consequence of this is that the sampling variance of statistics is underestimated. As the level of spatial dependence increases the underestimation increases. The problem is that when spatial autocorrelation is present the variance of the sampling distribution of (e.g. Pearson correlation coefficient), which is a function of the number of pairs of observations, is underestimated. Spatial autocorrelation coefficients can be modified to estimate the spatial correlation between two variables.^[37]

$$I_{\eta\xi} = \frac{1}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (\eta_i - \bar{\eta})(\xi_j - \bar{\xi})}{\left[\sqrt{\sum_{i=1}^N (\eta_i - \bar{\eta})^2 / N} \right] \left[\sqrt{\sum_{j=1}^N (\xi_j - \bar{\xi})^2 / N} \right]}$$

where η and ξ are the mortality and prosperity factors respectively. Although the mathematics is quite straightforward, very few software packages offer the option of computing $I_{\eta\xi}$.^[37] Thus, we used programming with S-plus software to find the value of $I_{\eta\xi}$. To test the significance of $I_{\eta\xi}$ we apply z -statistic, which follows approximately standard normal distribution: $z = I_{\eta\xi} \sqrt{N-1}$.

RESULTS AND DISCUSSION

RESULTS: Descriptive statistics for all indicators considered in the study were calculated, and Pearson's correlation coefficients between all indicators were estimated and provided in Table 1. From factor analysis, mortality factor ranged from -2.61 to 2.58. The highest was district 13, followed by district 15 (2.09) and district 52 (1.87). The lowest was district 19, followed by district 22 (-2.11) and district 71 (-1.86). Prosperity factor ranged from -2.05 to 2.23. The highest was district 12, followed by district 80 (2.14) and district 79 (1.73). The lowest was district 73, followed by district 10 (-1.91), and district 56 (-1.86). The Pearson correlation coefficient between mortality and prosperity factors is found moderately negative (-.40),

Table 1: Descriptive statistics and Pearson correlation matrix after normalization

Indictor	Y ₁	Y ₂	Y ₃	X ₁	X ₂	X ₃	mean	st. dev.
tsimr, y ₁	1.00						1.07	.28
tsnmr, y ₂	.72**	1.00					1.03	.27
tssmr, y ₃	.38**	.27*	1.00				1.07	.40
tclass1, x ₁	-.41**	-.19	-.10	1.00			10.07	3.30
tclass2, x ₂	-.38**	-.17	-.27*	.84**	1.00		6.82	3.84
tclass3, x ₃	-.28*	-.13	-.13	.70**	.71**	1.00	18.36	4.98

** Correlation is significant at the .01 level (2-tailed)

* Correlation is significant at the .05 level (2-tailed)

and it is significant with ($p < .001$). Figure 1a and 1c show visual insight of mortality and prosperity factors respectively. The suggestion of spatial clustering of similar values that follows from a visual inspection of these maps is confirmed by a strong positive and significant global Moran's I of .32 for mortality factor, with an associated standard normal z-value of 4.79 ($p < .0001$); and .27 for prosperity factor, with an associated standard normal z-value of 4.11 ($p < .0001$). Thus, we reject the null hypothesis of no spatial autocorrelation. The results of local Moran's I_i values for mortality and prosperity factors are reported in Table 2. A positive value of I_i indicates spatial clustering of similar values (either high or low), and negative value indicates a clustering of dissimilar values (i.e. a location with high value surrounded by neighbours with low values, and vice versa), as in the interpretation of the global Moran's I.^[22] Since the (Fig. 1a, and 1c). Higher scores of mortality were mainly shown in the north and south east of the country (Fig. 1a). Low scores of prosperity were mainly shown in the same part of the country (Fig. 1c). We found that the bivariate spatial correlation between the two factors ($I_{\eta\epsilon} = -.24$) with significant value ($z = -2.15$). Table 3 shows many significant clusters for both factors as shown from the p-values.

Six significant clusters (13, 52, 56, 71, 73, and 79) were found to have high level in both mortality and prosperity. Possibly, the high level of mortality in these districts could probably be contributed by the low level of prosperity in some of their neighbors (or by the prosperity inequality among their neighbors) as shown in Fig. 1b and 1d. tsimr, tsnm, and tssmr are transformed standardized (infant, neonatal and

stillbirth) mortality ratio; tclass1, tclass2, and tclass3 are transformed (class1, class2 and class3) of occupation percentages. st.dev.=standard deviation.

DISCUSSION: We formulated two factors, mortality and prosperity using three indicators for each, and examined the global clustering and local clusters for each; then we examined the association between the spatial pattern of mortality and the spatial pattern of prosperity, allowing for the effects of neighboring districts that share the boundary with a particular district. Findings allow policy makers to better identify what types of resources are needed and precisely where they should be employed. The above framework proposed for analyzing the spatial pattern of mortality reveals some noteworthy findings. Our results from the negative Pearson and bivariate spatial correlation coefficients between the two factors, however support the assumption that the high mortality is associated with low prosperity. Fukuda *et al.*^[8] identified the association between mortality rates and socioeconomic factors among cities in Japan using factor analysis and multiple regression analysis; they found that mortality is positively associated with unemployment.

After rejecting the null hypothesis, concluding that there is some form of clustering, it is of course of interest to know the exact nature of the clustering process. Is it only global type clustering or are there hot-spot clusters? If the later, how many hot-spots are there and where are they located? Our analysis of the association between mortality and prosperity factors used exploratory tools such as descriptive tables and small area choropleth maps. Geographical distributions of mortality and prosperity in quantiles were examined

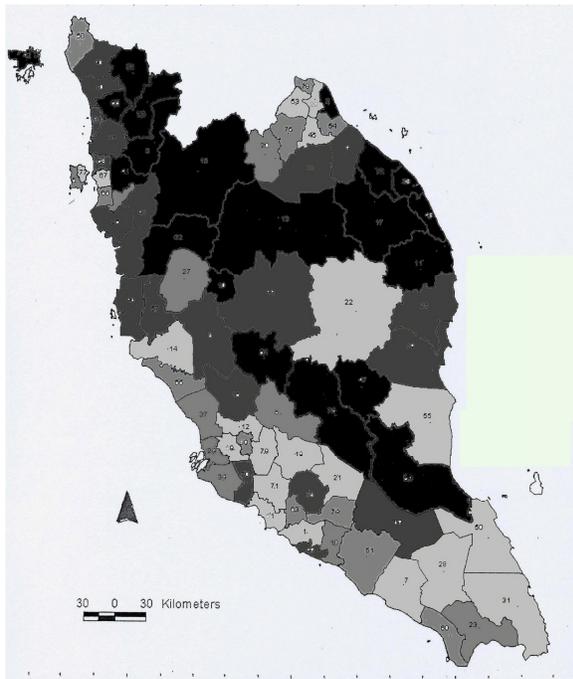


Fig. 1a: Mortality factor

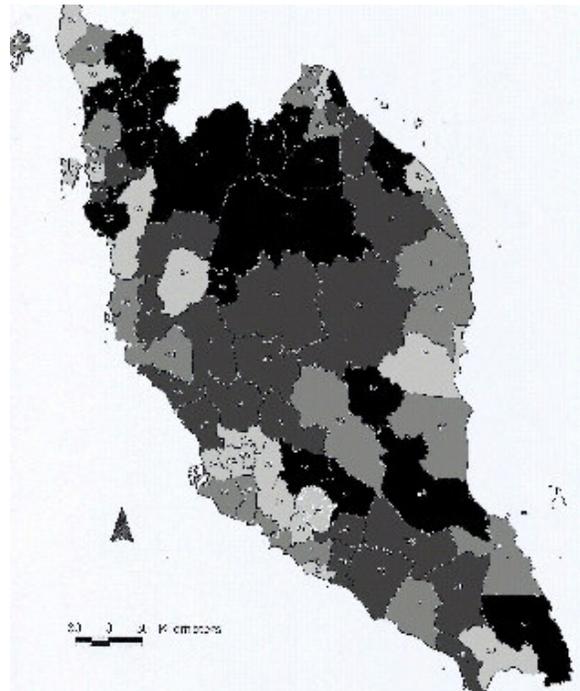


Fig. 1c: Prosperity factor

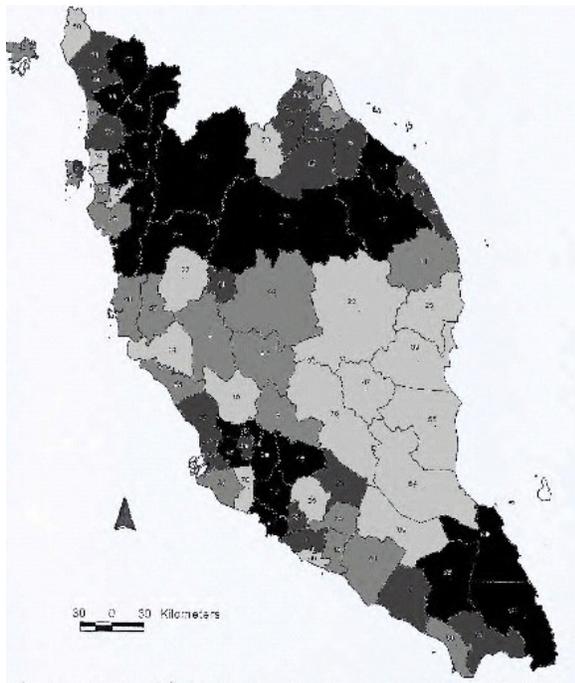


Fig. 1b: Local Moran values of mortality factor

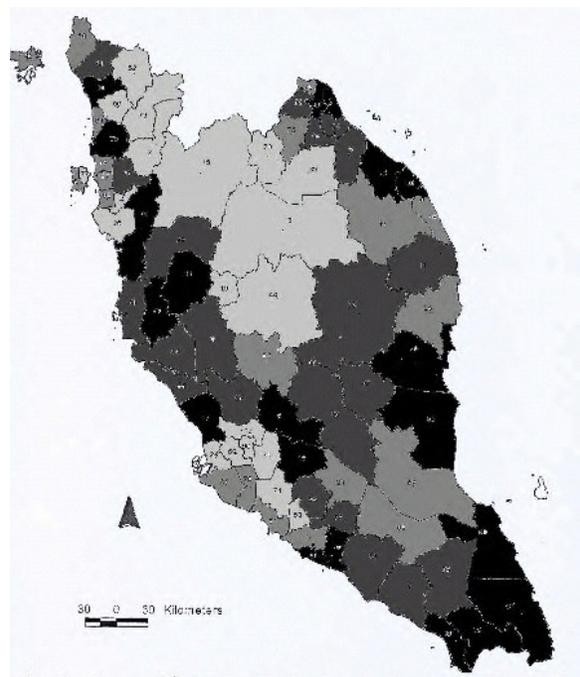


Fig. 1d: Local Moran values of prosperity factor

Quantiles	Quantiles	38	2.84	.042	-1.99	.900
(-2.61) – (-0.72)	(-19.26) - (-.04)	39	-1.12	.802	-.37	.649
(-0.65) – (-0.01)	(.01) - (.64)	40	1.57	.116	.06	.453
(0.02) – (0.66)	(.65) - (2.84)	41	3.42	.073	.23	.393
(0.70) – (2.58)	(2.89) - (13.90)	42	.20	.328	.26	.301
a	b	43	3.82	.052	-3.38	.937
-----	-----	44	.24	.363	2.11	.105
Quantiles	Quantiles	45	1.17	.222	-.01	.500
(-2.05) – (-0.65)	(-5.01) - (-.19)	46	.29	.367	.06	.462
(-0.62) – (-0.01)	(-.13) - (.23)	47	-.82	.721	-.13	.555
(0.00) – (0.63)	(.24) - (1.68)	48	2.02	.069	.50	.267
(0.64) – (2.23)	(1.72) - (14.82)	49	-.29	.690	-.19	.642
c	d	50	3.10	.028	-.21	.612
		51	.56	.271	.01	.473
		52	6.64	.001	4.79	.018
		53	1.62	.131	.01	.475
		54	.02	.477	-.09	.562
		55	-3.43	.974	-.30	.646
		56	5.07	.019	6.14	.010
		57	.19	.407	-.29	.620
		58	-.69	.790	.35	.284
		59	3.65	.053	14.82	.001
		60	.17	.415	-.42	.681
		61	3.21	.046	1.12	.170
		62	.14	.410	.24	.374
		63	1.72	.119	1.74	.116
		64	-.81	.733	1.31	.191
		65	.26	.369	.11	.430
		66	.16	.401	.37	.333
		67	-.06	.523	.91	.224
		68	-.04	.541	1.01	.204
		69	-.07	.528	.76	.246
		70	-1.02	.759	1.68	.150
		71	8.82	.002	4.22	.037
		72	3.45	.031	-.88	.792
		73	8.46	.002	9.90	.001
		74	.58	.317	.01	.518
		75	.94	.201	1.26	.155
		76	-3.31	.922	.00	.491
		77	1.55	.127	2.33	.080
		78	.01	.497	.49	.267
		79	6.66	.014	12.32	.001
		80	1.08	.168	12.75	.000
		81	.53	.280	.32	.343

Fig. 1: Choropleth maps show:
 a. mortality factor,
 b. local Moran values of mortality factor,
 c. prosperity factor and
 d. local Moran values of prosperity factor

Table 2: Local Moran's I_i for mortality and prosperity factors, and their corresponding p-values

Id	I_i for mortality	p-value	I_i for prosperity	p-value
1	1.60	.162	.37	.333
2	-.74	.797	-1.05	.836
3	8.70	.006	2.64	.087
4	-.14	.567	2.06	.096
5	.46	.219	1.56	.051
6	.38	.353	.06	.465
7	2.38	.065	-.01	.499
8	.42	.362	-.26	.606
9	.73	.259	.12	.416
10	2.19	.076	1.72	.119
11	.38	.319	.15	.382
12	2.89	.063	11.82	.000
13	8.72	.006	6.57	.012
14	-.57	.705	-.08	.546
15	13.90	.000	2.13	.113
16	-.58	.693	-.09	.548
17	6.38	.014	.93	.227
18	.37	.325	-.47	.695
19	6.29	.012	-1.78	.853
20	-3.02	.939	2.90	.052
21	1.30	.200	.45	.339
22	-19.26	.999	.14	.402
23	1.60	.107	-1.29	.852
24	.99	.189	2.20	.072
25	-.47	.693	1.04	.162
26	.07	.448	1.90	.098
27	-.67	.721	-4.33	.967
28	6.49	.012	.09	.417
29	.64	.283	-1.09	.793
30	.65	.248	-5.01	.980
31	5.27	.008	-.69	.740
32	8.87	.001	.04	.460
33	1.01	.211	3.45	.046
34	.27	.346	.36	.324
35	2.39	.097	-2.30	.905
36	-.77	.731	.02	.473
37	1.11	.172	-1.71	.852

visually using maps. Small area studies are a valuable tool to analyze and to pinpoint areas with higher mortality.^[9]

The levels of prosperity and mortality differed substantially within the country. A striking example was found in many districts (e.g. 13, 52, 73), which showed relatively high levels of mortality as shown in Fig. 1b. As causes of death were not studied, explanations of this spatial pattern are not straightforward. Whether material or individual circumstances cause mortality differences may be debated. Prosperity may be associated with mortality by reflection of high individual income which provides good medical care, high quality of food, and acceptable household conditions. As well as when the parents have high education, they can take care of their babies better

than those who are less educated. An important question is raised by these incident distributions. Are there clusters? Is there a discernable pattern? Is the high mortality clustering explained by the low prosperity clustering? It seems apparent from studying those Fig.s that spatial pattern assessment technique is necessary to evaluate this mortality and prosperity data.

The usual correlation coefficient statistics (such as Pearson) only test whether there is an association between two attributes by comparing values at the same location. Map comparison involves more than pair wise comparison between data recorded at the same locations; so even if relatively large values of two attributes (in our study mortality and prosperity) exist, it would still be indicative of an association if relatively large values on mortality and prosperity occupied locations that were close together in space (i.e. that are spatially correlated).^[20] This is because spatial units are arbitrary subdivisions of the study region and people may move around from one area to another; so they will be affected by prosperity levels in areas other than where they live (i.e. the level of mortality in *i*th district is thought to be influenced by the levels of prosperity not just in *i*th district but also in neighbouring districts). Several papers examining the relation between population mortality and income inequality seem to support the relative income hypothesis.^[29-32] They suggest that greater income inequality is associated with higher population mortality.

The result of autocorrelation coefficient is sensitive to the choice of neighbours and weights, so it may be desirable to run the autocorrelation under several different scenarios. Permutation distribution can be used to test the significance of the computed coefficient, so we used 1000 random permutations. We found that $p < .0001$ for both mortality and prosperity factors. Simulated data is useful for validating the results for such analysis. However, using Monte Carlo simulation, we simulate 999 random samples, 81 values for each sample, for both mortality and prosperity factors. These samples (999 matrices, each has two columns) are generated under bivariate normal distribution with mean zero and standard deviation 1 for both factors, and allowing for bivariate spatial correlation (-.24). The choice made for generating data under bivariate normality was made since the data do not show any departure from this assumption (p -value = .526). A significant p -value ($p < .0001$) was found for the bivariate spatial correlation.

CONCLUSION

We studied the clustering of mortality and prosperity separately, and the spatial association between them.

The analysis of spatial association between mortality and prosperity factors has shown that mortality is negatively correlated with prosperity based on Pearson and spatial correlation coefficients. Although, we cannot provide a causal relationship between the clustering patterns of mortality and prosperity, our results are conclusive in at least four aspects:

Firstly, Fig. 1a shows that high mortality is concentrated along the north-south axis, particularly in the central region and also several districts in the east coast of peninsula, for instance in the districts (52, 56, 73...etc). Fig. 1c shows that low prosperity is concentrated along the north-south axis, particularly also in the central region and also some districts in the coastal areas of peninsula, for instance in the districts (52, 56, 81...etc). Based on *visual* inspection, the patterns formed by those districts with highest ranking in mortality and those with lowest ranking in prosperity are nearly identical (Fig. 1a and 1c).

Secondly, many districts are not observed visually as hot spots for both mortality and prosperity factors as shown in Fig.s 1a and 1c respectively, but after considering the information of their neighbors (i.e., calculating local Moran's I_i values, and after represent them on the map), we can obviously see the patterns of hot spots. For instance, districts (3, 13, 15...etc) for mortality factor, and districts (12, 13, 33...etc) for prosperity factor, as shown in Fig.s 1b and 1d respectively.

Thirdly, our results show that levels of mortality and prosperity vary between districts. The clustering tendency shows that each of the factors, either mortality or prosperity, for each district can be spatially correlated with mortality or prosperity respectively in neighbouring districts based on global Moran index. The significant of bivariate spatial correlation and the visual inspection supports the hypothesis that the spatial pattern of prosperity factor can be associated negatively with the spatial pattern of mortality factor.

Fourthly, districts which possess neighbours with high degree of inequality in prosperity seem to show higher levels in mortality, for instance districts (13, 52, 56...etc). This is consistent with what Haining^[20] stated, the levels of such variable in area *i* is thought to be influenced by the levels of another variable not just in area *i* but also in neighbouring areas. This supports the hypothesis that the degree of variations in prosperity factor between these districts and their neighbors could influence mortality factor.

Many districts which exhibit high mortality status (Fig. 1b) are in the north-east, central and southern part of the peninsula; while many districts which exhibit high prosperity conditions (Fig. 1d) are generally found in the northern, central and southern part of the

peninsula. The analytical approach used here accurately delineates districts of high mortality, and permits policy makers to develop strategies in way that should minimize the difference between districts in mortality. Policy which pays attention to area characteristics will diminish mortality inequalities and consequently improve the health of the overall population.

REFERENCES

1. Hosseinpoor, A. R. and K. Mohammad, *et al.*, 2005. Socioeconomic inequality in infant mortality in Iran and across its provinces. *Bulletin of the World Health Organization*, 83: 837-844.
2. UNICEF, 2006. Progress for Children: A report card on nutrition. www.childinfo.org/eddb/malntrition/imdex.htm.
3. National statistics, 2003. In brief: population estimates for England and Wales. *Health Statistics Quarterly*, No.20: 1-69.
4. Aber, J. L. and N. G. Bennett, *et al.*, 1997. The effects of poverty on child health and development. *Annual review of public health*, 18: 463-483.
5. Department of statistics, 1995. Population and housing census of Malaysia 1993: General report of the population census, 2. Kuala Lumpur: Malaysia.
6. Hansell, A. and P. Aylin, 2000. "Routine data and health impact assessment-a review of epidemiological studies of socio-economic influence on health and evaluation of outcome indicators derived from routine health data for health impact assessment," www.phel.gov.uk/hiadocs/142_routine_data_and_HIA_a_review.pdf.
7. 2006. Public Health Annual Report 2004-2005: Children & Young People 'Being Healthy', www.rochdale.gov.uk/PublicServices/council.asp?URL=stratlnit.
8. Fukuda, Y., K. Nakamura and T. Takano, 2004. Wide range of socioeconomic factors associated with mortality among cities in Japan. *Health promotion international*, 19, No.2: 177-187.
9. Benach, J. and Y. Yasui, 1999. Geographical patterns of excess mortality in Spain explained by two indices of deprivation. *Journal of epidemiology and community Health*, 53, No.7: 423-431.
10. Johnson, R.A. and D.W. Wichern, 2002. *Applied multivariate statistical analysis*. 5th Ed. New Jersey: Prentice-Hall.
11. Fukuda, Y., K. Nakamura, and T. Takano, 2007. Higher mortality in areas of lower socioeconomic position measured by a single index of deprivation in Japan. *Public Health*, 121 (3): 163-173.
12. Martines, J. and V.K. Paul, *et al.*, 2005. Neonatal survival: a call for action. *Lancet*, 365: 1189-1197.
13. Weeks, J. R., 1992. *Population: An Introduction to concepts and Issues*, 5th. Ed. Wadsworth, Inc.
14. Lawson, A.B. and F.R. Williams, 2001. *An Introductory Guide to Disease Mapping*. New York: Wiley and sons.
15. Cliff, A.D. and J.K. Ord, 1981. *Spatial processes: Models and Applications*. London: Page Bros.
16. Szwarcwald, C.L., C.L.T. Andrade and I.B. Francisco 2002. Income inequality, residential poverty clustering and infant mortality: a study in Rio de Janeiro, Brazil. *Social science & medicine* 55: 2083-2092. www.elsevier.com/locate/socscimed.
17. Macintyre, S., S. Maciver and A. Sooman, 1993. Area, class, and health: Should we be focusing on places or people? *Journal of social policy*, 22: 213-234.
18. Tobler, W.R., 1979. Cellular geography, In: *Philosophy in Geography*. Holland: D Reidel Publishing Company.
19. Wilkinson, R.G., 1996. *Unhealthy societies: The afflictions of inequality*. London: Routledge.
20. Haining, R., 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge university press.
21. Cressie, N.A.C., 1993. *Statistics for spatial data*. Revised Ed. New York: Wiley and sons.
22. Anselin, L., 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis* 27, No.2: 93-115.
23. Lee, S-IL., 2001. Developing a bivariate spatial association measure: An integration of Pearson's r and Morans' I . *Journal of Geographical Systems*, 3: 369-385.
24. Waller, L.A. and C.A. Gotway, 2004. *Applied Spatial Statistics for Public Health Data*. New Jersey: Wiley and Sons.
25. Jöreskog, K., D. Sörbom, S. Toit, and M. Toit, 2000. *LISREL8: New Statistical Features*. 3rd ed. Lincolnwood: SSI.
26. Fortin, M.J. and M.R.T. Dale, 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge: Cambridge University Press.