

Correlates of being sick or injured in under five-year old children by district in Pakistan: A spatial analysis case study

Masood Ali Shaikh

Abstract

This case study demonstrates the use of population based, spatial analysis of public health indices in Pakistan. The data for this study were obtained from the Pakistan Bureau of Statistics website. Using district level data, for the spatial analysis of having being sick or injured, and their correlates in the under-five year old children. Percent under 5 year children, who either fell sick or injured during the past two-weeks by district, was used as an outcome variable in the final spatial regression model. While district level population density, average household size, literacy ratios for females were used as explanatory variables. As opposed to the final Ordinary-Least-Squares model, only population density was statistically significant in the spatial model. Limitations in terms of availability of current and regularly updated, attribute as well as geographic data in the country are underscored by the results of this case study.

Keywords: Spatial Analysis, Spatial Regression, Injury, Pakistan.

Introduction

Globally, over two thousand deaths occur daily owing to injuries in children and teenagers; based on 2008 joint estimate by the World Health Organization (WHO) and United Nations Children's Emergency Fund (UNICEF).¹ For the year 2011, WHO estimated that there were over 630,000 fatalities in under 15 year olds globally.² In Pakistan, several hospital-based studies have been published on causes and correlates of injuries in children.^{3,4} The Pakistan Social and Living Standards Measurement (PSLM) survey for the year 2014-15, is the most recent nationally representative survey reporting on the prevalence of being sick or injured in the country.⁵ PSLM provides these prevalence estimates for two age groups; under five-year olds and five-year and older individuals in the country, by district. For under five-year olds, the prevalence of being sick or injured during the last two weeks of interview, was reported as 12.44 percent by the PSLM.⁵

.....
Independent Consultant, Karachi.

Correspondence: Email: masoodali1@yahoo.com

In Pakistan, there are no studies on spatial analysis of injuries in children. Several studies in Europe and North America have also utilized Geographic Information Systems to conduct the spatial analysis of correlates and determinants of injuries in children and adults.^{6,7} This study is the first attempt to use district level data in Pakistan, for the spatial analysis of having being sick or injured, and their correlates in the under-five year olds; using the population census based correlates.

Methods and Results

The data for this study were obtained from the Pakistan Bureau of Statistics (PBS) website for the Pakistan Social and Living Standards Measurement (PSLM) survey for the year 2014-15, released in March, 2016.⁵ The PSLM survey data, providing district level population based representative estimates, are used for monitoring health and various other development plans and activities in the country. The survey methodology, detailed report, and tabular data are freely available on the PBS website. Briefly, a stratified two-stage sample design was adopted for the PSLM survey, and data were collected from October 2014 to June 2015; covering 78,635 households across the country, based on 1998 census that excludes the Federally Administered Tribal Areas (FATA). The district-wise 1998 Pakistan Census data, freely available at the PBS website as "District at Glance" reports was also used for all four provinces and Islamabad. These are the latest census figures available in the country.

The district level, PSLM 2014-15 and 1998 Census, data for all available districts were downloaded as hard copy tabular data and entered in Excel 2016 for analysis. For Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan provinces, PSLM district data were available for 37, 24, 25, and 28 districts respectively. The province of Punjab, includes the capital city/district Islamabad as well. While Pakistan Census 'district at glance' reports were available for 35, 17, 23, and 26 districts of Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan, respectively.⁸ The district level GIS shapefiles for Pakistan were obtained from the Humanitarian Data Exchange website of the United Nations Office for Coordination of Humanitarian Affairs (OCHA) website.⁹

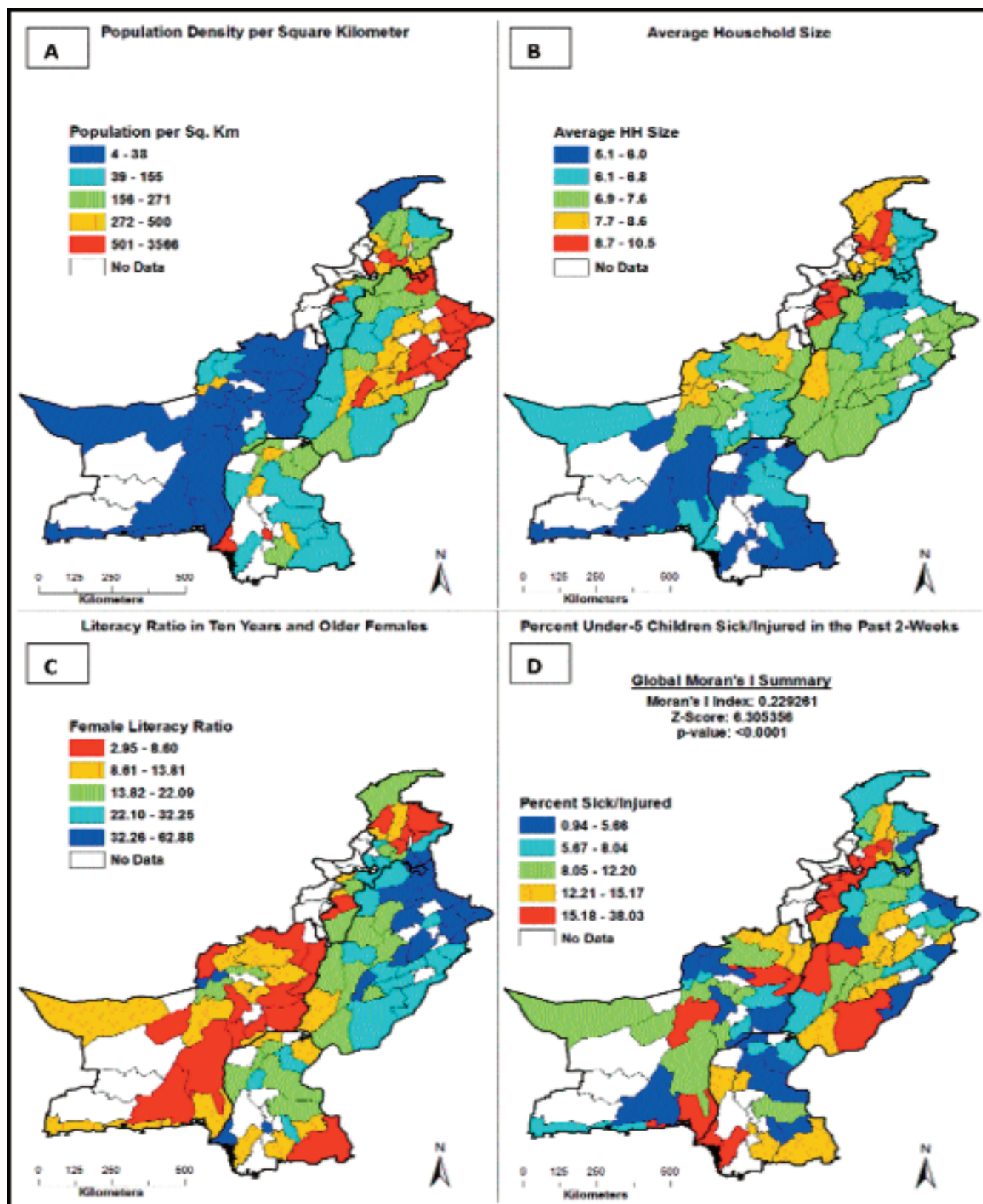


Figure-1: 'A' to 'C', quantile maps of three independent variables. 'D' map of local clustering and Global Moran's I value for the outcome variable.

Table-1: Ordinary least squares regression model output from GeoDa.

Summary of Output			
Number of Observations:		96	
Number of Variables:		4	
R-squared:		0.161073	
Adjusted R-squared:		0.133716	
F-statistic:		5.88795	
Prob(F-statistic):		0.00101047	
Independent Variables:		Coefficient	Probability
Constant		1.92284	0.67419
Population Density		0.0048606	0.00654
Average HH Size		1.44298	0.01942
Female Literacy		-0.131109	0.03549
Regression Diagnostics			
Multicollinearity Condition Number:		17.510885	
Test on Normality of Errors			
TEST	DF	VALUE	PROB
Jarque-Bera	2	99.6271	0.00000
Diagnostics for Heteroscedasticity			
Random Coefficients			
TEST	DF	VALUE	PROB
Breusch-Pagan test	3	12.5091	0.00583
Koenker-Bassett test	3	4.0463	0.25651
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	9	11.2895	0.25639
DIAGNOSTICS FOR SPATIAL DEPENDENCE - FOR WEIGHT MATRIX: Queen_Contiguity_1			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3768	5.5041	0.00000
Lagrange Multiplier (lag)	1	26.2360	0.00000
Robust LM (lag)	1	1.9545	0.16210
Lagrange Multiplier (error)	1	24.3434	0.00000
Robust LM (error)	1	0.0619	0.80352
Lagrange Multiplier (SARMA)	2	26.2979	0.00000
DIAGNOSTICS FOR SPATIAL DEPENDENCE - FOR WEIGHT MATRIX: Queen_Contiguity_2			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.2085	3.9765	0.00007
Lagrange Multiplier (lag)	1	9.5543	0.00199
Robust LM (lag)	1	0.2816	0.59563
Lagrange Multiplier (error)	1	11.7328	0.00061
Robust LM (error)	1	2.4601	0.11677
Lagrange Multiplier (SARMA)	2	12.0144	0.00246

Owing to some missing districts in all three data sources; cumulatively, for 96 districts all data were available. All these 96 districts were included for subsequent analysis reported in this study. Data on the percent under 5 year old children who either fell sick or injured during the past

Table-2: Selected spatial lag regression model output from GeoDa.

Summary of Output: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Spatial Weight:Queen_Contiguity_1

Lag coefficient (Rho):0.468206

R-squared:0.364971

Independent Variables:

W_Dependent-Variable0.4682060.00000

Constant-0.2025780.95854

Population Density0.003291340.02740

Average HH Size0.9592690.07074

Female Literacy-0.09295940.07818

Regression Diagnostics

Diagnosis for Heteroscedasticity

Random Coefficients

TESTDFVALUEPROB

Breusch-Pagan test310.18130.01709

Diagnosis for Spatial Dependence

Spatial Lag Dependencefor Weight Matrix: Queen_Contiguity_1

TESTDFVALUEPROB

Likelihood Ratio Test120.45700.00001

two-weeks of the survey interview, by district were obtained from the PSLM. While data on district level population density, average household size, literacy ratios for males/females/combined, average annual growth rate, percent of brick and mortar houses; percent housing units with electricity, piped water, and gas for cooking, were obtained from the PSB's 'District at Glance' reports.

Data analysis entailed spatial and non-spatial exploratory analysis, using ArcGIS Desktop 10.4, and GeoDa 1.8.14; followed by the ordinary least squares regression, spatial lag regression, and the spatial error regression models. The variable district-wise percent under 5 year olds who either fell sick or injured during the past two-weeks of the survey interview was used as a dependent variable; all other variables listed above were used as potentially explanatory/independent variables for selecting the final regression model.

Moran's I test with 'inverse distance' spatial conceptualization was run on the dependent variable (DV) of district-wise percent under 5 year age children who either fell sick or injured during the past two-weeks of the survey interview. Moran's index was 0.23 with a p value of <0.0001, suggesting statistically significant positive spatial correlation. Figure-1 'D' shows the map of local clustering for the dependent variable.

Next ordinary least squares regression model was developed by using the aspatial diagnostics to compare

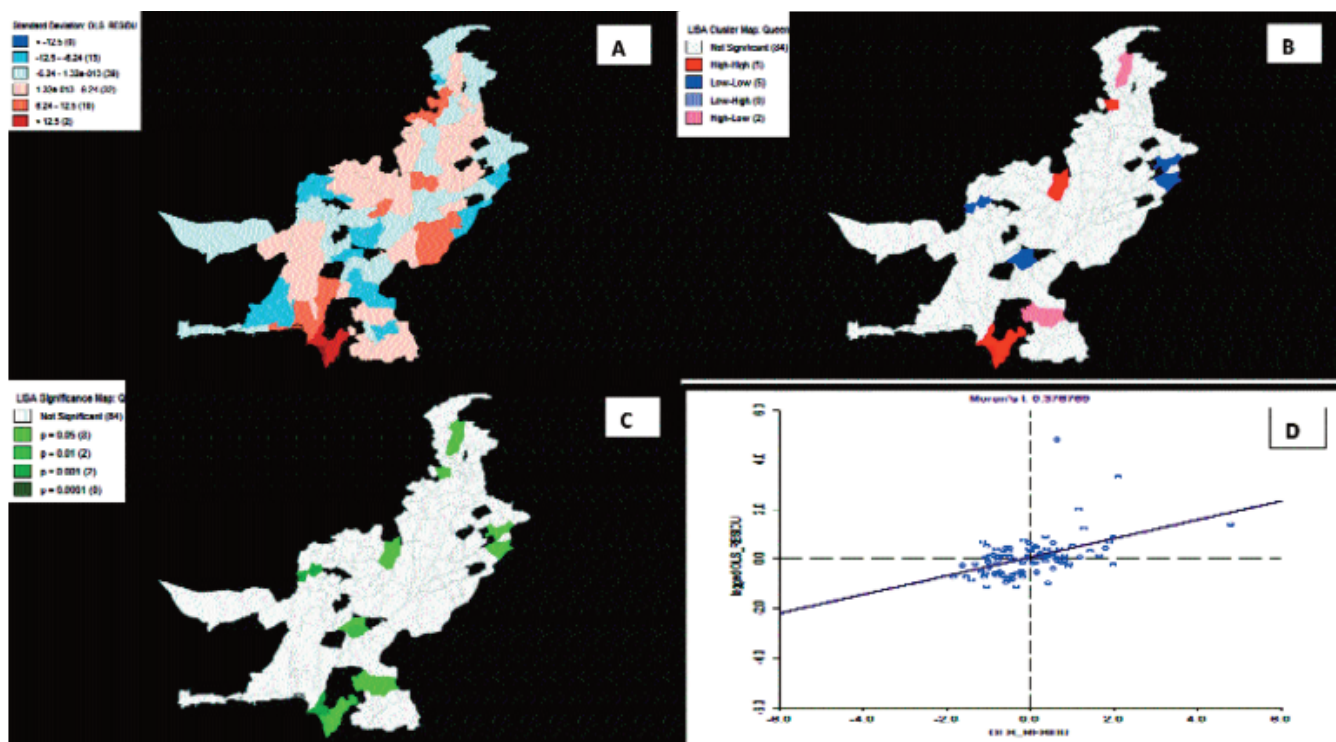


Figure-2: Map 'A' showing OLS regression model residuals standard deviation map. Map 'B' and 'C' showing LISA statistics. Part 'D' showing the Moran scatter plot.

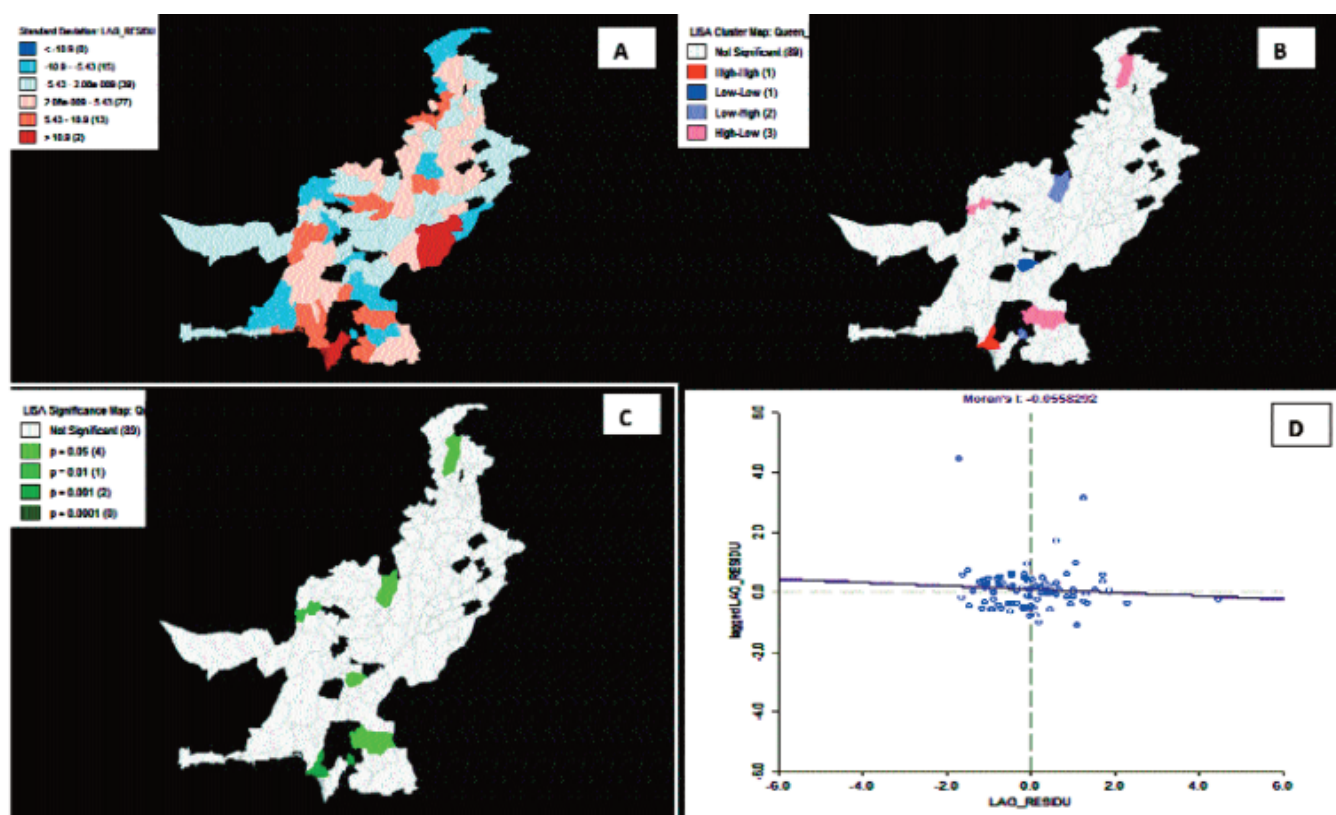


Figure-3: Map 'A' showing spatial lag regression model residuals standard deviation map. Map 'B' and 'C' showing LISA statistics. Part 'D' showing the Moran scatter plot.

different models i.e. log likelihood, Aikake information criterion, and the Schwartz criterion. The final model included three independent variables: population density, average household size, and literacy ratio in females aged 10 years and older. Figure-1, 'A' to 'C', shows three maps of the spatial distribution of independent variables. As ordinary least squares (OLS) regression output in Table-1 shows; all three independent variables were statistically significant at the level of <0.05 . Both, population density and average household size were positively associated with the percent sick/injured under-five children, while female literacy ratio was negatively associated with the dependent variable. This model's F-statistic value was statistically significant, with the adjusted R-squared value of 0.1337. Regarding regression diagnostics, the multicollinearity condition was 17.5. The normality of residual errors assessed by Jarque-Bera test indicated problems, as expected; since data are spatially correlated. Heteroskedasticity was tested three ways: Breusch-Pagan, Koenker-Bassett, and White tests; only Breusch-Pagan was statistically significant. Figure-2, map 'A' shows the OLS regression model residuals standard deviation; map 'B' and 'C' show results of the Local Indicator of Spatial Autocorrelation (LISA) statistics, while the Figure-2 'D' shows the Moran scatter plot, based on univariate local Moran's I test; Moran's Index was 0.376769 with p-value <0.0001 . The Moran scatter plot indicates that there is a significant global autocorrelation of the residuals. The LISA cluster map (B) indicates the quadrant of the Moran scatter plot that each significant observation falls into. Significant local clustering of like values is indicated by the bright shades of red and blue while significant spatial outliers are shown in pale colors. The significance map (C) shows the p-values on a test of spatial randomness.

As shown in Table-1, two Queen contiguity spatial weight matrixes (SWM) were created with order of contiguity 1 and 2. The Global Moran's I value, based on SWM with order of contiguity one and two, were 5.5041 and 3.9765, respectively; both were statistically highly significant at the p-value of <0.0001 . Lagrange Multiplier results for the first-order and second-order Queen contiguity spatial weight matrix favour using the Lag model for the SWM-queen-1 but favour the Error model when using the SWM-queen-2; as shown in table 1. Since in this analysis the focus was on how a dependent variable is spatially correlated across space, the Lag model was selected. Comparison of the spatial lag term between SWM-1 and SWM-2 reveals that it decreases. Therefore, adding additional influence of neighbouring districts reduces the effect of the Lag-DV on the focal district. Table-2 gives the output from the spatial lag model using first-order queen

contiguity weight matrix. The lag coefficient (Rho) was 0.365 and statistically significant.

As opposed to OLS model, only population density was statistically significant. And like the OLS model, the spatial lag model also violated the homoscedasticity assumption. However, the likelihood ratio test for spatial dependence was statistically significant; demonstrating evidence for the spatial lag dependence for the SWM used. Figure-3, map 'A' shows the spatial lag model residuals standard deviation; map 'B' and 'C' show results of the Local Indicator of Spatial Autocorrelation (LISA) statistics, while the figure 3-D shows the Moran scatter plot, based on univariate local Moran's I test; Moran's Index -0.0558292, with a p-value of 0.275. The Moran scatter plot indicates that there is no evidence of statistically significant global autocorrelation of the residuals. The LISA cluster map (B) indicates the quadrant of the Moran scatter plot that each significant observation falls into. Significant local clustering of like values is indicated by the bright shades of red and blue while significant spatial outliers are shown in pale colors. The significance map (C) shows the p-values on a test of spatial randomness.

Discussion

This is the first study looking at district-level spatial correlates of having being sick or injured in the past two weeks in the under five-year old population in Pakistan. All three independent variables i.e. population density, average household size and female literacy were found to be statistically significant in the ordinary least squares regression model. However, ordinary least squares model does not take into account the spatial nature of data. This was demonstrated by the Moran scatter plot indicating the statistically significant global autocorrelation of the residuals of the model. Hence violating the assumption of independently distributed residuals. Secondly, the homoscedasticity assumption was also violated as demonstrated by the statistical significance of Breusch-Pagan test.

The spatial lag model offered an improvement upon the ordinary least squares regression model. However, only population density was statistically significant at the 0.05 level, while other two independent variables were only significant at the 0.01 level. The lag coefficient (Rho) was statistically significant with a value of 0.468; stating that the dependent variable (DV) of under five-year olds having been sick or injured in the past two-weeks, in neighboring districts was associated with DV in the focal district, after accounting for between area differences of other independent variables of population density, average household size, and the female literacy rate. But

spatial lag model also violated the homoscedasticity assumption. Hence, results need to be interpreted with these caveats. Other limitations included the availability of census data: the dependent variable was for the year 2014-15, while all the dependent variables were gleaned from the last census conducted in the country i.e. 1998. Albeit these are big time differences; but the objective of this case study was to demonstrate the use of population based, spatial analysis of public health indices; using the example of under five-year olds having been sick or injured in the past two weeks, by district in Pakistan. Limitations in terms of availability of current and regularly updated, attribute as well as geographic data e.g. updated shapefiles for all administrative sub-divisions in the country are underscored by the results of this case study.

Disclaimer: None.

Conflict of Interest: None.

Funding Sources: None.

References

1. World Health Organization and United Nations Children's Fund. World report on child injury prevention. 2008. [Online] [Cited 2016 December 07]. Available from: URL: http://apps.who.int/iris/bitstream/10665/43851/1/9789241563574_eng.pdf.
2. World Health Organization. Child injuries. [Online] [Cited 2016 December 07]. Available from: URL: http://www.who.int/violence_injury_prevention/child/injury/en/.
3. Kanwal N, Chaudhry J, Amjad M, Zaheer M. Causes of childhood unintentional injuries in urban cities of Pakistan. *Pak J Med Res*. 2014; 53:63-6.
4. Faruque AV, Mateen Khan MA. Unintentional Injuries In Children: Are Our Homes Safe? *J Coll Physicians Surg Pak*. 2016; 26:445-6.
5. PSLM-2014-15 Pakistan Social and Living Standards Survey (2014-15) National / Provincial / District. Pakistan Bureau of Statistics, Statistics Division, Government of Pakistan. Islamabad. March 2016. [Online] [Cited 2016 May 12]. Available from: URL: http://www.pbs.gov.pk/sites/default/files//pslm/publications/PSLM_2014-15_National-Provincial-District_report.pdf.
6. Lawson F, Schuurman N, Amram O, Nathens AB. A geospatial analysis of the relationship between neighbourhood socioeconomic status and adult severe injury in Greater Vancouver. *Inj Prev*. 2015; 21: 260-5.
7. Goltsman D, Li Z, Bruce E, Connolly S, Harvey JG, Kennedy P, et al. Spatial analysis of pediatric burns shows geographical clustering of burns and 'hotspots' of risk factors in New South Wales, Australia. *Burns*. 2016; 42:754-62.
8. Population Census. Pakistan Bureau of Statistics, Statistics Division, Government of Pakistan. [Online] [Cited 2016 November 11]. Available from: URL: <http://www.pbs.gov.pk/pco-kpk-tables>.
9. Office for the Coordination of Humanitarian Affairs. Humanitarian Data Exchange. [Online] [Cited 2016 July 12]. Available from: URL: <https://data.humdata.org/dataset/44c2b2a4-b1cb-49d3-8299-0e544f1cab52>.