A Spatial and Bayesian approach to Infant Mortality in Puerto Rico: A study of risk differences, clustering and determinants

ALEXIS R. SANTOS-LOZADA, M.A.
The University of Texas at San Antonio

Introduction

Previous literature on the determinants of infant mortality in Puerto Rico is virtually inexistent. This is surprising since Puerto Rico exhibits high risk factors than other jurisdictions of the United States. According to the Center for Diseases Control and Prevention (2003) differences in Infant Mortality among Hispanics point to Puerto Ricans having poorest outcomes than other Hispanic Groups. Research has also indicated that Puerto Ricans in both the continental United States and Puerto Rico have the higher incidence of low birth weight and infant mortality (Becerra, et al., 1993). Cesareans have also been associated with higher risk of infant mortality, in case of Puerto Ricans cesarean rates in the island have gone up when compared to the same rates for Puerto Ricans in the United States (Center for Diseases Control and Prevention, 2006). Additional research conducted by the Center for Diseases Control and Prevention (2003) has pointed to Puerto Rican infant mortality is higher for those in the island when compared to those living in the continental United States. The pressing issue of the persistent high risk factors for infant mortality in Puerto Ricans deem necessary that we apply new methods to understand factors that influence infant mortality in the island. The objectives of this paper are two: to study spatial clustering of Infant Mortality in Puerto Rico and to study influences of covariates in a set of different specified models.

Data

Data for this analysis comes from the Puerto Rico Vital Statistics System. Infant Mortality rates come from the Annual Vital Statistics Report for Mortality (Departamento de

Salud, 2008) and data for total births comes from the Annual Vital Statistics Report for Births, Marriages and Divorces (Departamento de Salud, 2008b). Data to calculate county level socioeconomic proportions come from the 2006-2010 American Community Survey as published by the US Census Bureau (US Census, 2010).

Methods

The first step is to examine the spatial structure of the Standardizes Infant Mortality Ratios (SIMR) in Puerto Rico through exploratory spatial data analysis (ESDA). ESDA is a crucial step in a spatial statistics project (Chi & Zhu, 2008). The Global Moran's I statistic is used as an initial measure of spatial clustering of the outcome and all variables considered in this model. A high Moran's I indicates clustering of values (Voss, et al., 2006). To further explore this initial spatial structures this paper will also fit local version of the Moran's I, the Local Indicator of Spatial Autocorrelation (LISA) is used to obtain a measure of autocorrelation for each county or *municipio* (Colón-Lugo & Sparks, 2013). To support the existence of spatial associations, the spatial processes used in this paper incorporate a 15,000 simulations (Anselin, 1995; Colón-Lugo & Sparks, 2013). All ESDA is weighted using the Queen Rule for neighbor definition with 1 order of contiguity.

This paper will employ spatial statistic methods for analyzing the multiple issues at hand. For studying clustering on infant mortality at the island I will employ chi-square tests. The Potthoff-Whittinghill test for homogeneity of relative risks (Potthoff & Whittinghill, 1966; Potthoff & Whittinghill, 1966b). The hypothesis for the patterns of Infant Mortality is that infant mortality is not homogeneous across Puerto Rico.

The existence of heterogeneity in risk of infant mortality in Puerto Rico raises the question of the existence of clustering for this important issue in the island. For testing of clustering the Openshaw's Geographical Analysis Machine Method (Oppenshaw, et al., 1987) and the Kulldorff-Nargawalla Analysis (Kulldorff & Nargawalla, 1995) are employed to explore clustering of in the risk of infant mortality. This analysis will incorporate distance based weights where each county has 3 neighbors.

As one of the main objectives of this paper is to explore the county level determinants for infant mortality risk, I will explore the determinants of county level infant mortality risk by fitting various regression models. Several regression models will be fitted, these include: Ordinary Least Squares (OLS), Binomial, Negative Binomial, and a Hierarchical Model. The Ordinary Least Squares is a basic model which will be used to explore initial associations of variables with the Standardized Infant Mortality Ratios (SIRs). If the OLS is correctly defined it is possible to interpret the coefficients in substantive terms as showing the influence of each variable in the outcome. I will additionally fit three models to study the consistency of the effects of the variables across all of them. The additional models are: Negative Binomial Model, Binomial Model and a Binomial Hierarchical Model. All statistical processes were completed using R, ESDA was completed using GeoDa and the hierarchical regression model was fitted using the INLA library in the R statistical software. Finally, a Bayesian model is deemed appropriate considering the nature of the outcome and its distribution. JAGS (Just Another Gibbs Sampler) was used to fit Bayesian multilevel models to study the risk of infant mortality. When talking about the multilevel nested models fitted using JAGS the DIC will be used to choose which model explains more of the outcome variable.

Measurements

The outcome of this analysis is the Standardized Infant Mortality Ratio which is defined as the number of infant deaths over the number of expected infant deaths based on the island's 2008 Infant Mortality Rate 8.65 (per 1,000 births).

$$\text{SIMR} = \frac{\text{Infant Deaths 2008}}{\text{Expected Infant Deaths 2008}}$$

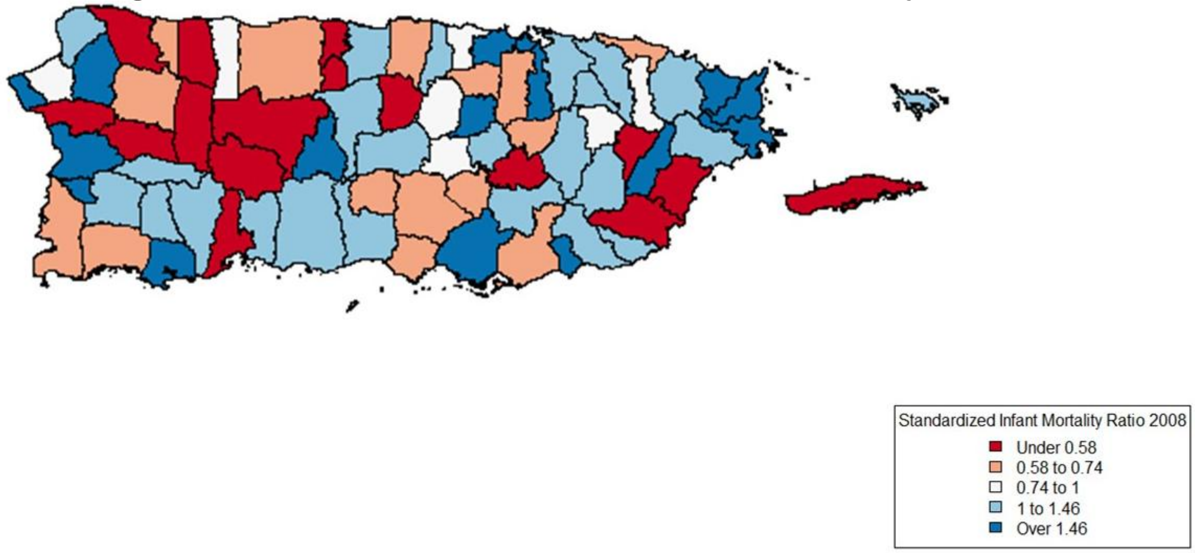
Covariates included in this analysis are poverty rates, which has been associated with high mortality outcomes for Puerto Ricans in the established literature (Sims, et al., 2007). Unemployment rate is also considered in this model, which is bound to be related to mortality and poverty rates as established in the literature. The analysis will also consider the proportion of women who have had a birth, which will also serve to approximate the effect of levels of fertility in the risk of infant mortality at the county level. Additionally, and highly correlated with poverty rate and unemployment rate is level of education. This model considers the effect of the proportion of persons without a High School Diploma in terms of the risk of infant mortality. I hypothesize that higher proportions of persons without a High School Diploma will lead to higher levels of risk of mortality.

Results

Descriptive Statistics and Exploratory Spatial Data Analysis

A visualization of the county level is presented in Figure 1. Higher ratios or risk is found in metro or metro adjacent areas. This is to say near the four major cities, SIRs are over 1, these cities are San Juan (North), Ponce (South), Mayaguez (West) and Fajardo (East). Relatively low SIRs are seed in Mountain Area and in the south-central area.

Figure 1: Standardized Incidence Ratio for Infant Mortality in Puerto Rico



Based on the initial analysis presented in Table 1, most of the variables included in this analysis show signs of the existence of a spatial effect. Interestingly enough the outcome variable of this study only has a marginally significant Moran's I. To further study spatial autocorrelation, Figure 1 presents clusters of concentration for all the variables used in this study. From Figure 2, it is clearly appreciated that local spatial autocorrelation exist to some degree for all the variables included in this study.

Table 1 Descriptive Statistics and Exploratory Spatial Data Analysis (ESDA) for outcome variable and covariates

Variables	Mean	Moran's I	p-value
Standardized Infant Mortality Rate	1.04	0.09*	0.07
Poverty Rate	0.50	0.53***	<0.01
Unemployment Rate	0.19	0.33***	<0.01
% Females with Births	0.04	0.09	0.38
% of Persons Age 25, no H.S. Diploma	0.36	0.49***	<0.01

* $p \leq 0.10$ ** $p < 0.05$ *** $p \leq 0.01$

Considering the evidence presented so far, a spatial effect is present in the data. I will further study clustering of the Standardized Infant Mortality Ratios using the tests discussed previously (Methods section). The Potthoff-Whittinghill is used to study the existence of heterogeneity in the SIMR, the results are presented in Table 2. The results of the chi-square test supports the first hypothesis of this paper, which is to say, something is happening in terms of spatial effects.

Table 2 Chi-Square test for Overdispersion, with parametric specifications and assuming a Poisson model (15,000 simulations)		
Outcome-Offset	Statistic	p-value
Infant Mortality-log(Expected Mortality)	126.90***	< 0.01
* p $\leq$ 0.10 ** p< 0.05 *** p $\leq$ 0.01		

Knowing about the existence of this spatial effect, this paper will expand on the initial clustering analysis done previously and will apply the clustering analysis that were also discussed previously.

The GAM Method for studying the clustering of SIMR show significant clusters (p-value $\leq$ 0.05) in the Fajardo-Luquillo Area, Toa Baja-Guaynabo, Guánica and Mayaguez-Hormigueros-Rincón. The results for the GAM Method are presented in Figure 3. The Spatial Scan Statistics Method presents supportive evidence to the existence of clustering of areas of high SIRs in the Fajardo-Luquillo area, these holds true for both specifications of the Spatial Scan Statistics.

Regression Analysis

After having completed the descriptive/explorative analysis, now I explore the determinants of risk of infant mortality in Puerto Rico by applying regression analysis.

Table 3 Ordinary Least Squares, Negative Binomial, Binomial, Hierarchical Binomial Regressions and Geographically Weighted Regression for Standardize Incidence Ratio for Infant Mortality in Puerto Rico 2008

	Ordinary Least Squares	**Negative Binomial Model**	**Binomial Model**	**Binomial Hierarchical Model**		
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Mean**	**95% LL**	**0.95% UL**
Intercept	-5.07***	-5.73***	-5.85***	**-6.33**	**-8.37**	**-4.37**
Poverty Rate	-4.24	-3.11	-2.47	-2.27	-7.98	3.47
Unemployment Rate	3.88	5.25**	5.05**	**4.75**	**0.40**	**9.26**
% Females with Birth	-34.02**	-45.14***	-45.13***	**-36.81**	**-63.82**	**-10.58**
% Age > 25 no H.S. Diploma	8.02*	5.55	5.00	5.33	-1.97	12.39
Model Dispersion	--	1.20	1.09	--		
*** p$\leq$ 0.01 ** p$\leq$ 0.05 * p$\leq$ 0.10 **For hierarchical model**: ***UL*** – Upper Limit ***LL*** – Lower Limit						

The initial OLS model results point to % of Females with Births being protection in relation to risk of infant mortality and to the proportion of 25 year old persons without a High School Diploma as a hazard with regards to the outcome variable (marginally significant). The non-significance of education through in the other three models, and the weak support it had in the initial OLS leads to not support of the hypothesis.

The second model, Negative Binomial, presents that Unemployment Rate is a hazard in terms of risk infant mortality and again the % of Females with Birth continues to have a protective effect in terms of risk of infant mortality at county level. This model has a dispersion of 1.20 which points to the existence of deviance in the model. The results from the third model (Binomial Model), seem to agree with the significance, direction and proportion of effects of the Unemployment Rate and the % of female headed households in terms of mortality risk. The dispersion of the model is 1.09, although it is smaller than the previous model, it continues to point to dispersion.

A hierarchical model was fitted using the INLA library in R. The hierarchically specified model fitted through a binomial distribution, presents results in terms of credible intervals. For this particular outcome (SMR), the coefficient effects that are credibly associated with the outcome are % females with births and the unemployment rate. The unemployment rate seems to have a positive effect in risk of infant mortality whereas the proportions of female with births continue to have a protective effect in terms of the outcome variable.

Bayesian approach to Infant Mortality in Puerto Rico

For starting the Bayesian approach to Infant Mortality, I calculate the credible intervals for Standardized Incidence Ratios for Infant Mortality in Puerto Rico. The credible intervals, similar to the clustering analysis discussed in the previous section points to credible (equivalent to significant) higher risk of mortality in the towns were clusters were identified (Figure 3). When the level of analysis is increased to 95% the credible higher levels are reduced from 8 to 4 counties. It can be observed that 8 counties had credible intervals, which are higher than the island's mortality rate. These credible intervals are presented in Table 4.

Table 4 Credible Intervals for Infant Mortality in Puerto Rico 2008					
90% Credible Intervals			**95% Credible Intervals**		
County	**Value**	**Probability**	**County**	**Value**	**Probability**
Culebra	0.038-0.249	0.965	Culebra	0.038-0.249	0.965
Fajardo	0.020-0.040	0.999	Fajardo	0.020-0.040	0.999
Guaynabo	0.012-0.041	0.929	Rincón	0.021-0.061	0.984
Luquillo	0.014-0.041	0.917	Guánica	0.028-0.062	1.00
Mayagüez	0.012-0.023	0.916			
Rincón	0.021-0.061	0.984			
Toa Baja	0.011-0.022	0.903			
Guánica	0.028-0.062	1.00			

To further study this interesting subject, I will fit a Poisson-Gamma Model for the SMR which will provide a better idea about the Risk of Infant Mortality at the county level. The results of this process are presented in Table 5.

Table 5 Credible Intervals for Infant Mortality Standardized Incidence Ratio in Puerto Rico 2008					
90% Credible Intervals			95% Credible Intervals		
County	BSMR(CI)	Probability	County	SMR(CI)	Probability
Fajardo	1.58(1.01-2.40)	0.977	Fajardo	1.58(1.01-2.40)	0.977

The Poisson-Gamma model results indicate that 17 counties have an SMR higher than 1.1. From these 17 counties, 1 county has a credible interval with higher risk of infant mortality (Table 5). This municipality is Fajardo and it was previously flagged as county with credible interval higher than the average and with clustering. Figure 4 present the exceedence probabilities and a Bayesian estimate of the Standardized Infant Mortality Ratio. The Bayesian estimates present credible intervals with *municipios* previously flagged as places with clustering, through the scan statistics and GAM Method, have credible intervals which point to higher levels of risk of infant mortality at the county level.

To further develop the Bayesian approach to the outcome variable, I used JAGS to fit Bayesian Hierarchical models to study the risk of Infant Mortality in Puerto Rico. The results of this models are presented in Figure 5. This models are controlling for the variation due to

Poverty Rate, Unemployment Rate, % of Females with Births and % Age > 25 with no High School Diploma. The regression diagnostics are presented in Table 6.

Table 6 Deviance Information Criterion for Multilevel and Non-Multilevel Poisson-Lognormal Bayesian Models for Infant Mortality in Puerto Rico

Model	DIC	Difference
Poisson-Lognormal	345.6	6
Poisson-Lognormal with Random Intercept	339.6	

Contrary to the conclusion drawn from just observing the maps of Risk, the inclusion of a Multilevel Element (Random Intercept) produced significant differences in the models. The results from Table 6 are consistent with this, as both models have a big difference in their DIC, the multilevel models should be considered when contemplating working or researching about Infant Mortality in Puerto Rico.

Figure 2: Univariate Local Autocorrelation (LISA) Cluster Maps

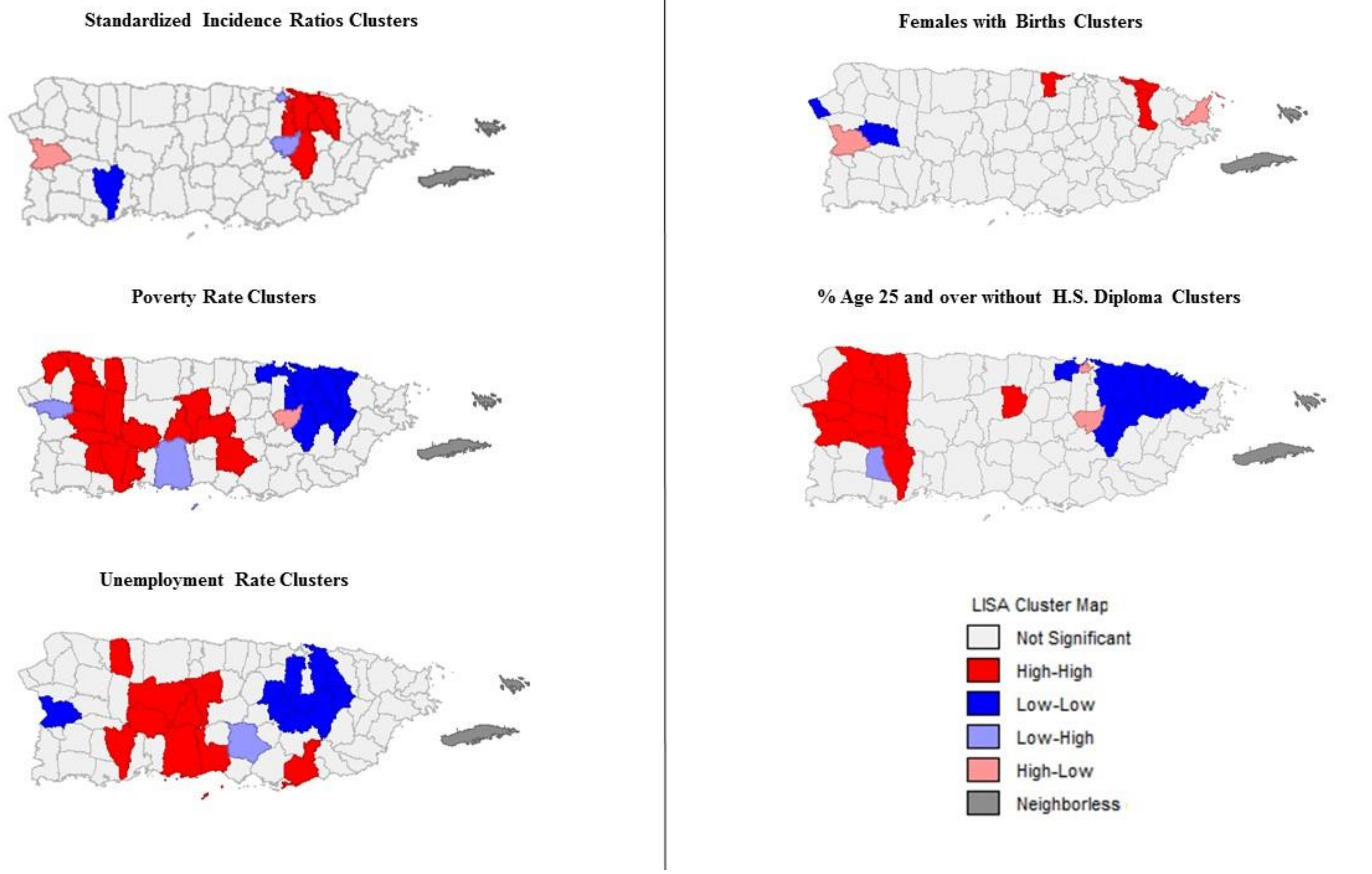


Figure 3: Clustering of Infant Mortality Risk in Puerto Rico

Significant Clusters using the GAM Method

Significant Clusters using the Spatial Scan Statistics I

Significant Clusters using the Spatial Scan Statistics II

Figure 4: Exceedence Probabilities and Bayesian Estimates for the SIMRs

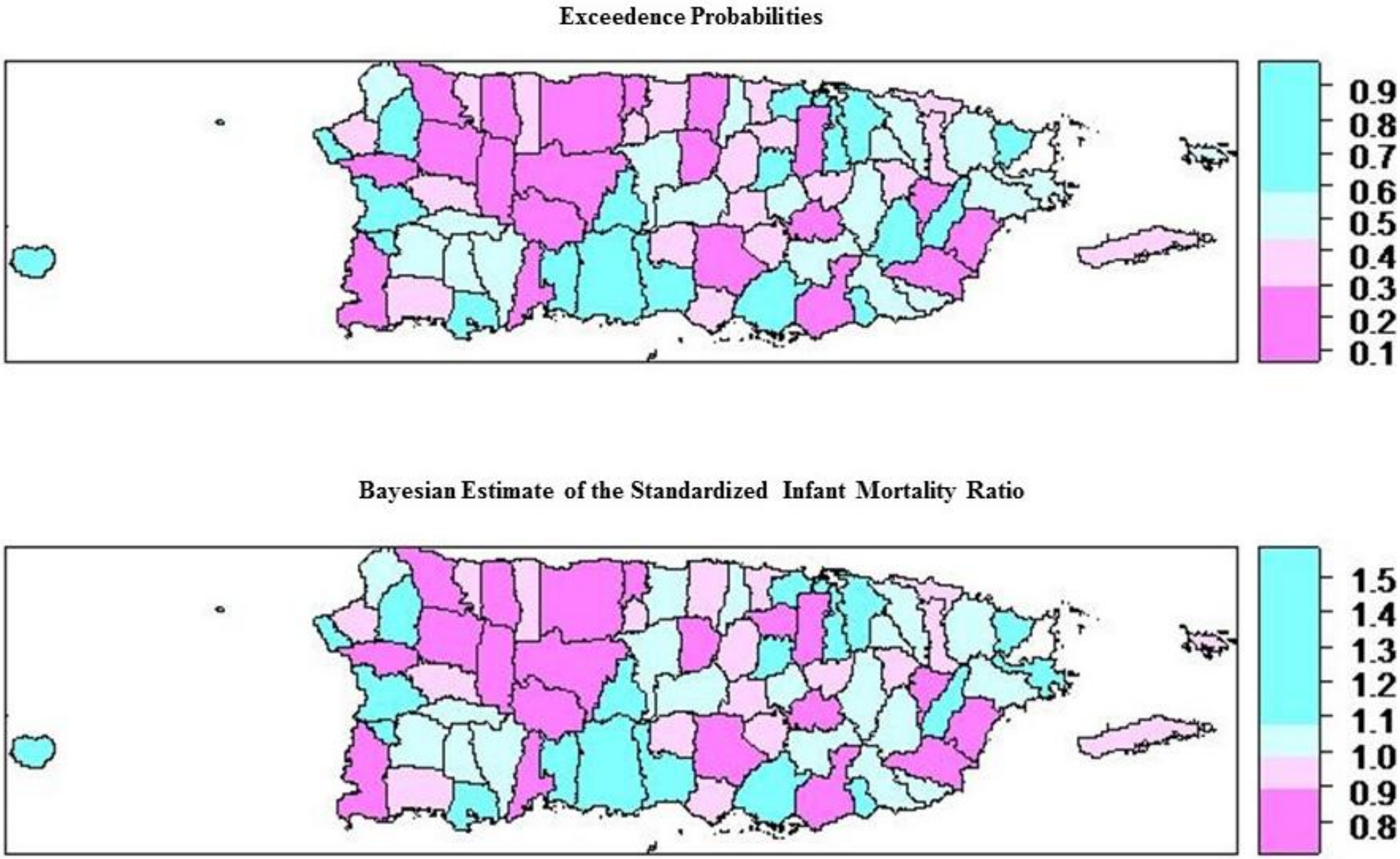
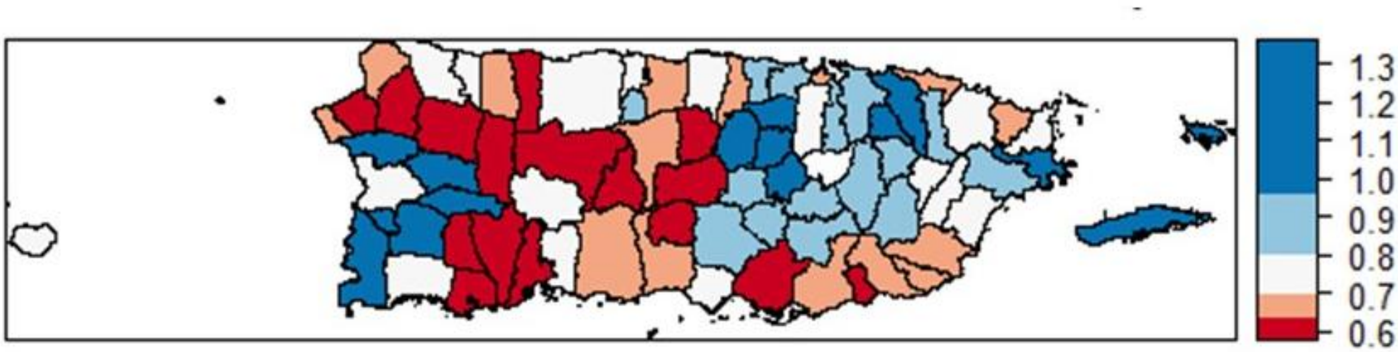
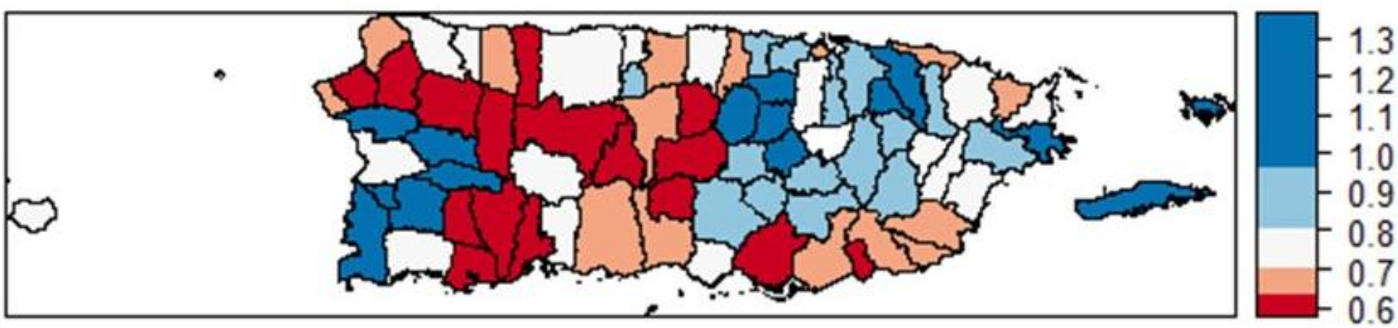


Figure 5: Mapping of Risk of Infant Mortality fitted through Bayesian Models

Bayesian Poisson-Log Normal Regression Model with no structure of Random Effects



Bayesian Poisson-Log Normal Regression Model with a Random Intercept



Limitations

This paper employs spatial analysis, in most of the cases the municipalities of Vieques and Culebra did not have neighbors because they do not share axis/vertices with any other municipality of the island. These two municipalities can be identified by being gray in all clustering analysis presented in Figure 2.

Discussion

This paper had two objectives. The first was to study spatial clustering of infant mortality in Puerto Rico. The second was to estimate the effect of the covariates considered on it in a set of competing models. The results revealed that risk of mortality in Puerto Rico has a spatial component that has not been addressed through the literature on infant mortality. This result supports the first hypothesis of this paper as the effect of space is patent through the spatial approach (Moran's I and Chi-Square test). Additionally there are signs of spatial clustering based in different methodologies which also support the notion of existence of a space effect in the outcome variable.

The results revealed that county level poverty does not affect the risk of mortality at that level. In the Negative Binomial, Binomial and the Binomial Hierarchical Model unemployment rate has a hazardous effect in increasing the risk of mortality at the county level. Higher proportions of females with birth have a protective effect in when associated with Mortality risk. Finally, higher proportions of persons aged 25 and over without a High School diploma has a hazardous effect when evaluating the risk of mortality using an OLS. The results of this exploratory analysis provide important information about the relations of the county level proportions when associated with the risk of mortality.

References

Anselin, L. (1995). Local indicators of Spatial Associations-LISA. Geographical Analysis 27, 93-115.

Becerra, J.E., Atrash, H.K., Perez, N. and Saliceti, J.A. (1993). Low birthweight and Infant Mortality in Puerto Rico. American Journal of Public Health 83(11), 1572-1576.

Center for Diseases Control and Prevention (2003). Infant Health Among Puerto Ricans --- Puerto Rico and U.S. Mainland, 1989-2000. Mortality Monitoring Weekly Report, 52(42), 1012-1016.

Center for Diseases Control and Prevention (2006). Rates of cesarean delivery among Puerto Rican women – Puerto Rico and the U.S. mainland, 1992-2002. Mortality Monitoring Weekly Report 55(3), 68-71.

Chi, G. and Zhu, J. (2008). Spatial Regression Models for Demographic Analysis. Population Research and Policy Review 27, 17-42.

Colón-Lugo, H. and Sparks, C. S. (2013), A Study of County Child Poverty Rates in Puerto Rico: Does Space Matter? Spatial Demography 1(1), 96-110.

Departamento de Salud (2008). Informe Annual de Estadisticas Vitales 2007-2008: Mortalidad Infantil, Materna y Fetal. Secretaría Auxiliar de Planificación y Desarrollo.

Departamento de Salud (2008). Informe Anual de Estadísticas Vitales 2007-2008: Nacimientos, Matrimonios y Divorcios. Secretaría Auxiliar de Planificación y Desarrollo.

Kulldorff, M. and Nargawalla, N. (1955). Spatial Disease Clusters: Detection and Inferences. Statistics in Medicine 14, 799-810.

Potthoff, R. F. and Whittinghill, M.(1966). Testing for Homogeneity: I. The Binomial and Multinomial Distributions. Biometrika 53, 167-182.

Potthoff, R. F. and Whittinghill, M.(1966b). Testing for Homogeneity: The Poisson Distribution. Biometrika 53, 183-190.

Sims, M., Sims, T., Bruce, M.A. (2007). Urban Poverty and Infant Mortality Rate Disparities. Journal of the National Medical Association 99(4), 349-356.

U.S. Census (2010). American Community Survey 2006-2010 County Summary File.