

MORTALITY AMONG YOUNG NICARAGUAN  
IMMIGRANTS TO COSTA RICA: AN APPLICATION  
OF GEOGRAPHICALLY WEIGHTED STATISTICAL  
REGRESSION

MORTALIDAD ENTRE LOS INMIGRANTES  
NICARAGÜENSES EN COSTA RICA: UNA  
APLICACIÓN DE LA REGRESIÓN GEOGRÁFICA  
PONDERADA

ROGER E. BONILLA\*      JUAN B. CHAVARRÍA†

*Received: 7/May/2013; Revised: 21/May/2014;*

*Accepted: 10/Jun/2014*

---

\*Escuela de Estadística, Universidad de Costa Rica, 2060 San José, Costa Rica. Fax: (506) 2511-6483, (506) 2511-6500. E-mail: roger.bonilla@ucr.ac.cr

†Misma dirección que/same address as R. Bonilla. E-mail: jchavarr@fce.ucr.ac.cr

### Abstract

This paper applies a geographically weighted statistical regression (GWR) model to young Nicaraguan immigrant homicides in Costa Rica during the period 1998-2008 and identifies possible covariates. The parameters for the GWR model are:

$$Y_i(g) = \beta_0(g) + \beta_1(g)x_1 + \beta_2(g)x_2 + \dots + \beta_k(g)x_k + \varepsilon,$$

which may be obtained from the solution of

$$\beta(g) = (X^T W(g) X)^{-1} X^T W(g) Y.$$

The GWR model is a more adequate model than the classic models, such as the *log-linear Poisson* model. In the GWR model, poverty was the most significant variable. The map of the estimators associated with the percentage of poor households suggests that the relationship between poverty and mortality by homicide for young Nicaraguan immigrants is stronger in the Caribbean region and neighboring zones. When the GWR model was run for homicides among young Costa Ricans, this effect was not observed, as it was among Nicaraguan immigrants.

**Keywords:** geographically weighted regression (GWR); spatial correlation; homicides; Costa Rica; immigration.

### Resumen

Este trabajo aplica un modelo de regresión estadística especial ponderada (GWR) a los homicidios de inmigrantes nicaragüenses jóvenes en Costa Rica en el período 1998-2008 e identifica sus posibles covariables. Los parámetros del modelo GWR

$$Y_i(g) = \beta_0(g) + \beta_1(g)x_1 + \beta_2(g)x_2 + \dots + \beta_k(g)x_k + \varepsilon,$$

$$\beta(g) = (X^T W(g) X)^{-1} X^T W(g) Y.$$

El modelo GWR es un modelo más adecuado con respecto a modelos clásicos como el log-lineal de Poisson. En el modelo GWR la variable pobreza resultó la más significativa. El mapa de los estimadores asociados con el porcentaje de hogares pobres sugiere que la relación entre la pobreza y la mortalidad por homicidios de jóvenes nicaragüenses es más fuerte en el Caribe y zonas aledañas. Cuando el modelo GWR se aplicó a los homicidios entre jóvenes costarricenses, este efecto no se observó.

**Palabras clave:** regresión estadística espacial ponderada (GWR); correlación espacial; homicidios; Costa Rica; inmigración.

**Mathematics Subject Classification:** 91B72, 62J99, 62J05.

## 1 Introduction

The geographically weighted statistical regression model (GWR) allows different relationships between the dependent variable and independent covariates existing in different geographic points  $(x, y)$  (municipalities) within a geographic space (Brunsdon et al. 1996). The GWR model produces a set of  $\beta$  estimators for a regression in each geographic unit: the value for the geographic unit plus the values for the neighboring municipalities, weighted by distance. In each regression, the observations are weighted by a function related to the distance of the geographic unit for which the regression is being carried out. The function used depends on an estimate of the bandwidth obtained by a Monte Carlo simulation procedure [7], [4].

The function for the weights  $W_j$  used in the GWR model has the following form [5]:

$$W_j = \exp\left(\frac{-d_j}{b^2}\right)$$

Where  $d_i$  is the Euclidean distance from the centroid of the  $i^{th}$  municipality and each one of the centroids of the other municipalities where the regression is being carried out so

$$\arg \min_b = \sum_{i=1}^n (y_i - y_{\neq i}(b))^2,$$

where  $y_{\neq i}(b)$  is the fitted value of the GWR model using a parameter  $b$  omitting the observation  $i$  from the calibration process.

In each Monte Carlo simulation the values are distributed randomly throughout the space and the GWR model is repeated. Once the bandwidth calibration process is completed, the GWR model uses the best-fit bandwidth to adjust the GWR model in each geographic unit.

In modeling count data, such as homicides among Nicaraguan immigrants, on the basis of traditional models, such as the Poisson log-linear model, there can be problems capturing all of the relevant sources of variation, because of spatial autocorrelation. The Poisson log-linear model assumes that the coefficients do not vary locally. The GWR model takes into account this spatial autocorrelation [6], [23].

### Homicides among young Nicaraguan immigrants to Costa Rica

A preliminary analysis of the core information for this study allowed us to identify that homicides are the predominant cause of death among young Nicaraguan

**Table 1:** Standardized mortality rates ( $\times 100\,000$ ) and deaths by causes of deaths among Nicaraguan immigrants and Costa Ricans. Costa Rica 1998–2008.

| Causes of deaths          | Costa Ricans |             | Nicaraguan Immigrants |            | Relative risk |
|---------------------------|--------------|-------------|-----------------------|------------|---------------|
|                           | Rate         | Deaths      | Rate                  | Deaths     |               |
| Infectious diseases       | 1,95         | 251         | 2,65                  | 35         | 1,4           |
| Cancer                    | 12,18        | 1564        | 9,62                  | 125        | 0,8           |
| Chronic respirat diseases | 1,66         | 212         | 1,490                 | 20         | 0,9           |
| Cardiovascular diseases   | 5,71         | 726         | 6,63                  | 86         | 1,2           |
| Diabetes                  | 0,62         | 77          | 0,45                  | 6          | 0,7           |
| Alcoholism                | 1,84         | 232         | 1,95                  | 25         | 1,1           |
| Motor vehicle accidents   | 20,02        | 2575        | 23,43                 | 311        | 1,2           |
| Other fatalities          | 8,82         | 1147        | 15,04                 | 194        | 1,7           |
| Suicides                  | 9,84         | 1265        | 10,64                 | 137        | 1,1           |
| <b>Homicides</b>          | <b>13,36</b> | <b>1578</b> | <b>24,72</b>          | <b>325</b> | <b>2,0</b>    |
| Other causes              | 15,65        | 2009        | 15,86                 | 205        | 1,0           |

Source: Instituto Nacional de Estadística y Censos, (2009).

immigrants from 15 to 34 years of age (Table 1). This coincides with the literature consulted ([13], [12], [19], [20], [22], [18]), which emphasizes that the main causes of death among immigrants are external causes, particularly homicides, although the relationship between migration and external causes of deaths is not clear.

In Costa Rica, up to this point, the only studies of mortality among Nicaraguan immigrants have been [12] and more recently [13], but they do so in a general context. The researchers find the same results as those found in other countries: high mortality from external causes (accidents and homicides) versus a low mortality from disease related deaths. Due to the age groups studied, homicides and accidents are the main causes of death among immigrants and Costa Ricans. A large proportion of the Nicaraguan immigration to Costa Rica consists of individuals in these age groups and their mortality continues without having been studied [12].

## 2 Methods and data

### 2.1 Mathematical notation

Define the classical linear model with  $k$  covariates with a sample of size  $n$  as

$$Y = \beta^T X + \varepsilon,$$

where  $Y = (Y_1, \dots, T_n)^T \in \mathbb{R}^n$ ,  $\beta = (\beta_0, \dots, \beta_k)^T \in \mathbb{R}^{k+1}$ ,

$$\begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{nk} \end{pmatrix}$$

and  $\varepsilon \in \mathbb{R}^n$  is some vector with  $E[\varepsilon|X] = 0$ , can be extended as follows:

$$Y_i(g) = \beta_0(g) + \beta_1(g)x_1 + \beta_2(g)x_2 + \dots + \beta_k(g)x_k + \varepsilon,$$

where  $(g) \in \mathbb{R}^2$  is a vector which indicates that the parameter will be estimated at a location whose geographic coordinates are given by the vector  $g$  [9], [8]. That model allows local variation among the estimators.

The parameters for a traditional linear regression model can be resolved by solving:

$$\beta = (X^T X)^{-1} X^T Y.$$

So that to estimate the parameters of a GWR model we have to use the following weighting scheme:

$$\beta(g) = (X^T W(g) X)^{-1} X^T W(g) Y.$$

The weighting factors are selected in such a way that the observations that are close to an estimation point have more influence on the results than those observations that are far from that point. The results from the parameter estimates can be presented on maps to study regional variations. By using GWR models, it is assumed that geographic autocorrelation exists.

## 2.2 Data and general procedures

The study population consists of young immigrants (from 15 to 34 years of age) that died in Costa Rica due to homicide between 1998 and 2008, as reported in the national vital statistics system available at the National Institute of Statistics and Census [15]. The geographic unit was the Costa Rican municipality and the geographic location of the municipality was the geometric centroid geographic coordinate according to the Lambert North cartography system [14].

The following variables based on an empirical model were included in the database:

- Response Variable: Number of young immigrant homicides in the municipality ( $y$ ) (years 1998 to 2008).

- Exposure Variable: Population of young Nicaraguan immigrants (census 2000). Factors (covariates predictors)

*Demographic.*

- *Percentage of Nicaraguan immigrants in the municipality ( $x_1$ ) (census 2000).*
- *Percentage of persons in the municipality between the ages of 15 and 34 (youth) ( $x_2$ ) (census 2000).*
- *Sex ratio of the municipality ( $x_3$ ) (census 2000).*

*Geographic.*

- *Percentage of households in the municipality classified as in the urban zone ( $x_4$ ) (census 2000).*

*Socio-economic characteristics.*

- *Percentage of households in the municipality with at least one unmet basic need (UBN) ( $x_5$ ) (census 2000).*
- *Percentage of the economically active population in the municipality working in each of the three economically productive sectors (primary, secondary, tertiary) ( $x_6, x_7, x_8$ ) (census 2000).*

The data were processed with the STATA statistical package [21], GWR module. The GeoDA package [2]<sup>1</sup> was also be used for the geographic statistical regression. Finally, the MapInfo package [16] was used for the map presentation.

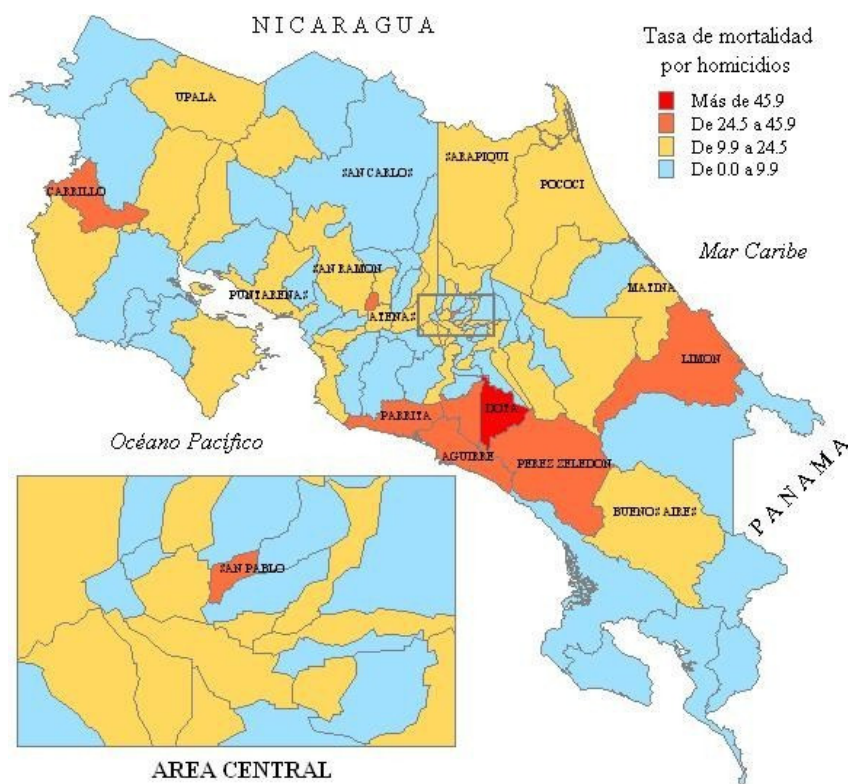
### 3 Results

Figure 1 presents the mortality rate from homicide among young immigrants on a percentile scale. According to the map, the highest homicide rates (90<sup>th</sup> percentile and higher) are found in the municipalities of the southern zone: Dota, Tarrazú, Parrita, Aguirre and Pérez Zeledón, as well as in the municipality of Limón (on the Caribbean coast), in Carrillo (Guanacaste), in Palmares, and San Pablo de Heredia, in the Central Valley. Moran's  $I$  was used as the statistics to evaluate global spatial autocorrelation [1]. Geographic autocorrelation of the homicide rates for young immigrants is significant (Moran's  $I = 0.0976$ ). The

---

<sup>1</sup><http://geodacenter.asu.edu/software/>

GWR model takes into account the spatial autocorrelation, since the model coefficients vary locally.



**Figure 1:** Mortality rate from homicide for young Nicaraguan immigrants (percentiles). Costa Rica 1998–2008. Legend: Red, more than 45.9; Orange, from 24.5 to 45.9 Yellow, from 9.9 to 24.5 Blue, from 0.0 to 9.9.

Table 2 presents the GWR model regression results, as well as those from the classic model. The test for bandwidth suggests that the GWR model offers a better description of the homicide rates among immigrants than a classic *Poisson log-linear* regression model with the same covariates ( $p < 0.000$ ). The best-fit bandwidth is 142.9 kilometers. This means that the GWR model will use the municipalities located within a distance of 142.9 kilometers to estimate the regression. In practical terms, this means that the GWR regression is better than the classical model of a log-linear Poisson regression, because in the other case, the bandwidth would be greater, covering the country as a whole [5]. Any bandwidth that is less than one-half of the maximum distance between a point of the periphery (borders) of Costa Rica and another, indicates that the GWR models

are more adequate than the classical Poisson log-linear regression models. Traditional models, such as the Poisson log-linear model, ignore the bandwidth. To have an idea of the dimension of the bandwidth in this case, it is approximately the distance between the capital, San José and the city of Limón (approximately 141 kilometers).

**Table 2:** Incidence rate ratio from the classic Poisson log-linear regression model and test for non-stationarity of the coefficients from the GWR model. The response variable is the number of deaths by homicide for Nicaraguan immigrant or Costa Rican youth by municipality in Costa Rica. Costa Rica 1998–2008.<sup>a</sup>

| Sociodemographic covariates                      | Classic model | GWR Model <sup>b</sup> |              |
|--|---------------|------------------------|--------------|
|  |               | Nicaraguan immigrants  | Costa Ricans |
| Percentage of immigrants ( $x_1$ )               | 0.96*         | 0.01                   | 0.00         |
| Sex Ratio ( $x_3$ )                              | 1.01*         | 0.01                   | 0.00         |
| Percentage Urban Population ( $x_4$ )            | 1.01*         | 0.01                   | 0.00         |
| Percentage Poor Households ( $x_5$ )             | 1.04**        | 0.01*                  | 0.01         |
| Percentage Adults (35+ years) ( $x_2$ )          | 1.05          | 0.02                   | 0.01         |
| % EAP <sup>c</sup> in Secondary Sector ( $x_7$ ) | 1.05          | 0.01                   | 0.01         |
| % EAP in Tertiary Sector ( $x_8$ )               | 1.05*         | 0.01                   | 0.02         |
| Number of cells ( $N$ )                          | 81            | 81                     | 81           |
| <i>Goodness of fit</i>                           |               |                        |              |
| $\chi^2$   | 92.10         |                        |              |
| Prob > $\chi^2$                                  | 0.08          |                        |              |
| Bandwidth  |               | 142986                 | 142986       |
| $p$ value  |               | 0.00                   | 0.00         |

<sup>a</sup>The exposure variable is the population of Nicaraguan immigrant or Costa Rican youths, as may be the case.

<sup>b</sup>Non-stationarity test for the coefficient. Value  $s_i$ . The test for significance of non-stationarity of the estimators is a test that indicates whether the variable is significant within the GWR model.

<sup>c</sup>EAP = Economically Active Population.

\*Significant at 5%.

\*\*Significant at 1%.

The significance test for non-stationarity for the estimators is a test indicating whether the variable is significant within the GWR model [5]. The test provided only one significant variable, which is the percentage of poor households. The





The map of the estimators associated with the percentage of poor households (Figure 2) suggests that the relationship between poverty and mortality by homicide for young Nicaraguan immigrants is stronger in the Caribbean region and neighboring zones, where the main center is the municipality of Matina. In the municipalities in the Central Valley, Sarapiquí and the Southern Zone, there is an intermediate relationship between poverty and mortality by homicide for young immigrants. In the rest of the country, this relationship is low, where its lowest point is in the municipalities of Golfito and Osa, close to the Panamanian border.

When the GWR model was run for homicides among young Costa Ricans, this effect was not observed, as it was among Nicaraguan immigrants (Table 2). In particular, the covariate, percentage poor households, was not significant at 5%. This indicates that the relationship between poverty and homicides interacts differently among Nicaraguan immigrants and Costa Ricans, since it affects the former more and this association is significant.

## 4 Final discussion

The results from this study demonstrate that the GWR model is a more adequate model for describing homicides among young Nicaraguan immigrants than classical models, such as Poisson log-linear regression since the GWR model incorporates spatial autocorrelation. GWR models allow calculate local estimators versus classical models that do not incorporate spatial autocorrelation. Up to what point are these results valid and reliable? One element that could lessen the validity of the results obtained has to do with aggregate statistics. The fact that the data are aggregated at the municipality level presents two serious problems: (1) *The Numerator Effect* and (2) the *Ecological Fallacy*. In regards to the Numerator Effect, in this paper, we have used the municipalities as our unit of analysis, since it is a geographic unit that guarantees a sufficient number of cases for the numerator. Ideally, we would have used smaller administrative units, such as the *districts* in Costa Rica. However, this was not possible, primarily because of the fact that there would have been numerous zeros in the numerators [11], [3]. For that reason, the municipality was chosen as the geographic unit for analysis. Secondly, in the analysis in this paper we run into the *Ecological Fallacy*, which is present in research using aggregated statistics. The fallacy consists of affirming that the individual has the characteristics of the statistical aggregate to which the individual belongs ([17], [10]). Another effect of the ecological fallacy is that some of the relationships, for example the inverse relationship between poverty and homicides, are difficult to explain upon aggregating the data. These relationships can only be observed at the individual level.

Notwithstanding, the results obtained are consistent with the literature that we found, complementing the studies by [12] and [13]. The findings from this study contribute to present useful evidence for an appropriate design for preventive public policies and social programs that would benefit the young immigrants.

## References

- [1] Anselin, L. (1995) "Local indicators of spatial association – LISA", *Geographical Analysis* **27**: 93–115.
- [2] Anselin, L.; Syabri, I.; Youngih, K. (2005) "GeoDa: an introduction to spatial data analysis", *Geographical Analysis* **38**(1): 5–22.
- [3] Bailey, L.; Vardulaki, B.L.; Langham, J.; Chandramohan, D. (2005) *Introduction to Epidemiology*. Open University Press, London.
- [4] Bowman, A.W. (1984) "An alternative method of cross-validation for the smoothing of density estimates", *Biometrika* **71**(2): 353–360.
- [5] Brunson, C.; Fotheringham, A.S.; Charlton, M.E. (1996) "Geographically weighted regression: a method for exploring spatial nonstationarity". *Geographical Analysis* **28**(4): 281–298.
- [6] Clayton, D.G.; Bernardinelli, L.; Montomoli, G. (1993) "Spatial correlation in ecological analysis", *Int. J. Epidemiol.* **22**(6): 1193–1202.
- [7] Cleveland, W.S. (1979) "Robust locally weighted regression and smoothing scatterplots", *Journal of the American Statistical Association* **74**(368): 829–836.
- [8] Fotheringham, A.S.; Brunson, C.; Charlton, M.E. (2000) *Quantitative Geography*. Sage, London.
- [9] Fotheringham, A.S.; Brunson, C.; Charlton, M.E. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.
- [10] Greenland, S.; Robins, J. (1994) "Invited commentary: ecologic studies—biases, misconceptions, and counterexamples", *American Journal of Epidemiology* **139**(8): 747–760.
- [11] Hennekens, C.H.; Buring, J.E.; Mayren, S.L. (1987) *Epidemiology in Medicine*. Little Brown & Company, Boston.

- [12] Herring, A; Bonilla, R.; Borland, R.; Hill, K. (2008) “Patrones diferenciales de mortalidad entre inmigrantes nicaragüenses y residentes nativos de Costa Rica”, *Población y Salud en Mesoamérica* **6**(1): article 2, 20 pp.
- [13] Herring, A.; Bonilla, R. (2009) “Inmigrantes nicaragüenses en Costa Rica: estado y utilización de servicios de salud”, *Población y Salud en Mesoamérica* **7**(1): article 4, 20 pp.
- [14] Instituto Geográfico Nacional de Costa Rica (IGN) (1984) *Proyección Costa Rica Lambert Norte*. San José, Costa Rica.
- [15] Instituto Nacional de Estadística y Censos (2009) “Estadísticas vitales”. En: <http://censos.ccp.ucr.ac.cr>, consultado 15/05/2009.
- [16] MapInfo Corporation (1985) *MapInfo Professional*. MapInfo Corporation, New York.
- [17] Robinson, W.S. (1950) “Ecological correlations and the behavior of individuals”, *International Journal of Epidemiology* **38**(2): 337–341.
- [18] Sharma, R.D.; Michalowski, M.; Verma, R.B. (1990) “Mortality differentials among immigrant populations in Canada”, *International Migration* **28**(4): 443–450.
- [19] Singh, G.K.; Miller, B.A. (2004) “Health, life expectancy and mortality patterns among immigrant populations in the United States” *Can. J. Public Health* **95**(3): 114–121.
- [20] Sorenson, S.; Shen, H. (1999) “Mortality among young immigrants to California: injury compared to disease deaths”. *Journal of Immigrant Health* **1**(1): 41–47.
- [21] StataCorp (2005) *STATA Package*. Stata Corporation, College Station TX.
- [22] Trovato, F. (1992) “Violent and accidental mortality among four immigrant groups in Canada, 1970-1972”, *Biodemography and Social Biology* **39**(1-2): 82–101.
- [23] Wolpert, R.L.; Ickstadt, K. (1998) “Poisson/gamma random field models for spatial statistics”, *Biometrika* **85**(2): 251–267.