Exploring the Association between Low Birth Weight and Exposure to Lead Using Geographically Weighted Odds Ratio

Extended Abstract

Jerome Yang[1], Hong Xia[1]

1. Introduction

Low birth weight (LBW) is closely related with fetal and neonatal mortality and morbidity, inhibited growth and cognitive development, and chronic diseases later in life. In this study we examined the association between low birth weight and maternal exposure to chemicals, which are listed in the National Ambient Air Quality Standards in Massachusetts. To characterize the spatial non-stationionarity of this association, we used Geographically Weighted Odds Ratio (GWOR) as an improved local statistic.

Previous studies have applied various methods to evaluate the relationship between low birth weight and exposure to air pollutants. Aguilera et al. (2009) applied linear regression models to assess the relationship between prenatal air pollution exposure and birth weight. Wilhelm and Ritz (2005) employed logistic regression model to explore the association of CO and particulate matter (PM) with adverse birth outcomes. Besides linear and logistic regression, Maisonet et al. (2001) also used odds ratio to analyse the relationship between low birth weight and exposures to ambient air pollution. Slama et al. (2007) used Poisson regression to estimate the prevalence ratios (PR) of birth weight lower than 3,000 g due to traffic related atmospheric pollutants.

In these studies, using odds ratio or linear, logistic and Poisson regressions implies that every sample is considered independently. To consider the previously ignored spatial autocorrelation and non-stationarity, Fotheringham et al (2002) developed Geographically Weighted Regression (GWR), which uses different parameters in the linear regression model to characterize the non-stationarity of relationship between variables. In this study, however, we explored the association between two binary variables. We employed GWOR, a localized spatial association index, to replace the existing methods.

2. Data

The US Environmental Protection Agency (EPA) Risk-Screening Environmental Indicators (RSEI) Model provided us with the relative levels of exposure in every 810*810m cell. This model estimates people's exposure to chemicals based on the total release amount from the EPA Toxics Release

[1] Clark University, Worcester, Massachusetts, USA

Inventory (TRI) program. Three results of 591 chemicals are all available: the pound-based, hazard-based and risk-related scores (Table 1). Both the hazard-based and risk-related scores were employed in this study.

Table 1. The results of EPA RSEI

Results	Calculation
Risk-related results	Surrogate Dose x Toxicity Weight x Population
Hazard-based results	Pounds x Toxicity Weight
Pounds-based results	TRI Pounds released

The LBW data, provided by the Massachusetts Department of Public Health (MassDPH), recorded 623,844 births from 2000 to 2007 with variables including infant's birth weight, sex and plurality as well as the mother's age, marital status, race, education, gestational age and diseases. The variable types include numerical, ordinal and binary. The geographic locations of all birth cases were geocoded by the MassDPH. This data is available down to the census block level. Primary analysis (Table 4. Refer to Appendix) indicates some variables including plurality, marital status, gestational age, number of cigarettes during pregnancy and other diseases have stronger associations with LBW.

3. Study Process

We first explored the spatial heterogeneity of LBW and exposure to lead in the entire Massachusetts. They are represented by the ratio of LBW cases and the RSEI scores respectively in every census block. The latter, originally a raster image, was converted to a vector feature class by averaging the cell values within each polygon. The global and local Moran's I tools provided by GeoDa were used to explore the spatial pattern of LBW ratio and exposure to lead (both the hazard-based and risk-related scores) as well as examine the stationarity of their associations. The results showed that the associations are spatially non-stationary (Figure 1 a-d).

We proposed the use of GWOR to characterize this non-stationarity. Due to the limitation of computational capacity, in this phase we only focused on the cases of 2007 in Springfield, Massachusetts. Within this study area, the birth weight of each case was converted to a binary variable using 2,500 grams as the breakpoint (LBW versus normal). In the same way, exposure to lead was converted using the median (high exposure versus low exposure). After conversion, those variables became binary and their relationships could be explored with GWOR.

4. Geographically Weighted Odds Ratio

For two binary variables, a and b, p_{11}, p_{10}, p_{01} and p_{00} are the probability of 4 types of outcome combination (Table 2).

Table 2. Four types of outcome combination of two binary variables

	$P(b=1)$	$P(b=0)$
$P(a=1)$	p_{11}	p_{10}
$P(a=0)$	p_{01}	p_{00}

And the Odds Ratio is $OR = \frac{p_{11} \cdot p_{00}}{p_{10} \cdot p_{01}}$

In the non-stationary spatial analysis context, each polygon within the study area had an odds ratio calculated from the cases nearby. Derived from Brunsdon et al (2002), for any polygon i with one of its neighbor cases j, the Geographically Weighted Odds Ratio (GWOR) is:

$$GWOR_i = \frac{p_{11i} \cdot p_{00i}}{p_{10i} \cdot p_{01}i}$$

where $p_{11i} = \frac{\sum_j x_{11j} \cdot w_{ij}}{\sum_j w_{ij}}$, $p_{10i} = \frac{\sum_j x_{10j} \cdot w_{ij}}{\sum_j w_{ij}}$, $p_{01i} = \frac{\sum_j x_{01j} \cdot w_{ij}}{\sum_j w_{ij}}$, $p_{00i} = \frac{\sum_j x_{00j} \cdot w_{ij}}{\sum_j w_{ij}}$

$x_{11j}, x_{10j}, x_{01j}$ and x_{00j} are an alternative to show 4 types of outcome combination of the case j. Their relationships with variables a and b are shown in Table 3.

Table 3. The conversion from variables a and b to $x_{11j}, x_{10j}, x_{01j}$ and x_{00j}

j	(a_j, b_j)	x_{11j}	x_{10j}	x_{01j}	x_{00j}
1	(1, 1)	1	0	0	0
2	(1, 0)	0	1	0	0
3	(0, 1)	0	0	1	0
4	(0, 0)	0	0	0	1

w_{ij} weights the outcome of case j in the calculation for polygon i. It is a function of d_{ij}, the distance between polygon i and case j. Quite a few choices are available for this function. Two commonly used weighting functions are Inverse Distance Weighting (IDW) and the bisquare kernel smoother. They are computed as:

$$w_{ij} = \frac{1}{{d_{ij}}^p} \quad p > 0$$

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h}\right)^2\right]^2 & if\ d_{ij} < h \\ 0 & otherwise \end{cases}$$

where p, the power of IDW, could be any positive real number and h, and the maximum neighborhood distance from polygon I or bandwidth, could be either a positive constant or a function of number of neighbors. The choice of weighting function replies on the nature of a spatial phenomenon. In our case, the bisquare kernel smoother was applied with various fixed bandwidths. The effect of changing the value p is discussed in the result section.

5. Result

The results of GWOR show a regionalized distribution of GWOR. In several sites within the study area, high GWORs are observed. This implies there are stronger associations between LBW and exposure to lead in these locations than in other locations. The GWOR maps also indicate the influence of changing bandwidth. The greater bandwidths we chose, the greater amount of neighbors were taken into account. This accounts for the variation of odds ratio among the bandwidths 632, 1256 and 1881 meters. The 632-meter results of hazard-based and risk related scores both contain many zero polygons.

These are the census blocks with either no birth or the cases that odds ratios cannot be calculated. When bandwidths are greater, however, less polygons show the value zero.

There are several limitations in our analysis. First, the exclusion of the cases outside the study area resulted in the lack of neighbors along the boundary. The distinctive high GWORs along the edge in Figure 2c display this problem. In addition, zero has multiple meanings in our analysis. It could be either no birth or those blocks where the odds ratio cannot be calculated. Excluding these blocks from analysis would influence the continuity of features. Furthermore, knowing some variables are also associated with LBW, the GWORs adjusted by those variables might be more meaningful.

6. Conclusion

Rather than summarizing the association with only one single index, GWOR is able to characterize the non-stationary of odds ratios across a study area. It not only provides an overview of the places where the association is stronger, but also takes the similarity of neighborhoods into account. Using this approach, we are able to know the areas with stronger association between LBW and exposure to lead in Springfield.

In this phase, we only developed the calculation of GWOR and applied it within one of the 351 towns and cities in Massachusetts. We plan to improve computational capacity in order to explore the huge dataset of the entire state. We will also involve other geographically weighted statistics, such as geographically weighted logistic regression and geographically weighted Poisson regression to understand the relationship between LBW and other categorical and ordinal variables.

7. Reference

Aguilera, I., Guxens, M., Garcia-Esteban, R., Corbella, T., Nieuwenhuijsen, M. J., Foradada, C. M., & Sunyer, J. (2009). Association between GIS-based exposure to urban air pollution during pregnancy and birth weight in the INMA Sabadell Cohort. *Environmental health perspectives*, *117*(8), 1322-7. doi:10.1289/ehp.0800256

Brunsdon, C., Fotheringham, S., & Charlton, M. (2002). Geographically Weighted Local Statistics Applied to Binary Data. GIScience, 38-50.

Fotheringham, A.S., Brunsdon, C., & Charlton, M.E. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester: Wiley.

Maisonet, M., Bush, T. J., Correa, a, & Jaakkola, J. J. (2001). Relation between ambient air pollution and low birth weight in the Northeastern United States. *Environmental health perspectives*, *109 Suppl* (June), 351-6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1240552&tool=pmcentrez&rendertypr=abstract

Slama, R., Morgenstern, V., Cyrys, J., Zutavern, A., Herbarth, O., Wichmann, H.-E., & Heinrich, J. (2007). Traffic-related atmospheric pollutants levels during pregnancy and offspring’s term birth weight: a study relying on a land-use regression exposure model. *Environmental health perspectives*, *115*(9), 1283-92. doi:10.1289/ehp.10047

Wilhelm, M., & Ritz, B. (2005). Local Variations in CO and Particulate Air Pollution and Adverse Birth Outcomes in Los Angeles County, California, USA. Environmental Health Perspectives, 113(9), 1212-1221. doi:10.1289/ehp.7751

8. Appendix

Table 4. Primary Analysis of the LBW dataset

Variable Explanation	Value	Descriptive Analysis		Odds Ratio with LBW
		Sum	Percentage	
Sex	**1=male**	**318904**	**51.12%**	**1**
	2=female	304936	48.88%	1.166518
	3=unknown	4	0.00%	N/A
Plurality	**1**	**595584**	**95.47%**	**1**
	2	26679	4.28%	18.7847
	3	1514	0.24%	270.6037
	4	66	0.01%	570.7418
	9=unknown	1	0.00%	N/A
Mother's Age	12-17	12560	2.01%	1.646147
	18-24	121162	19.42%	1.171772
	25-30	**187271**	**30.02%**	**1**
	31-35	193471	31.01%	1.005817
	36-40	93362	14.97%	1.244766
	41-45	1549	0.25%	3.279283
	46-50	144	0.02%	0.056494
	51+	12	0.00%	0
	99=unknown	6	0.00%	20.33786
Marital Status	**1=Married**	**441823**	**70.82%**	**1**
	2=Unmarried	180737	28.97%	1.396648
	3=Previously Married	1258	0.20%	1.401363
	9=unknown	26	0.00%	4.578592
Mother's Race	**1=white**	**449952**	**72.13%**	**1**
	2=black	49078	7.87%	1.825045
	3=asian/pacific islander	41914	6.72%	1.144626
	4=american indian	1246	0.20%	1.194977
	5=other	80758	12.95%	1.267274
	8=refused	507	0.08%	1.047503
	9=unknown	389	0.06%	1.447371
Mother college years completed	0	225179	36.10%	1.384859
	1-4	306117	49.07%	1.100146
	5-8	**86418**	**13.85%**	**1**
	9+	5208	0.83%	1.111955
	88=Refused	922	0.15%	1.72983
Clinical estimate of gestational age	0-20	358	0.06%	30961.54
	21-25	1965	0.31%	86647.44
	26-30	4881	0.78%	8193.3
	31-35	25909	4.15%	721.0473
	36	19787	3.17%	116.8261
	37	40390	6.47%	45.4915
	38	89553	14.36%	12.62909
	39	159246	25.53%	3.956601
	40	199581	31.99%	2.936979
	41	70849	11.36%	1.030176
	42	**7028**	**1.13%**	**1**
	43	204	0.03%	2.678913
	44	18	0.00%	33.65385

	45	7	0.00%	44.87179
	46+	60	0.01%	14.17004
	99=Unknown	4008	0.64%	27.23911
Kessnes index	1=adequate	488216	78.26%	0.823709
	2=intermediate	107714	17.27%	1.026562
	3=inadequate	**19592**	**3.14%**	**1**
	4=unknown	6496	1.04%	2.133199
	5=no prenatal care	1826	0.29%	4.298884
Kotlchuck index	**0=unknown**	**488216**	**78.26%**	**1**
	1=inadequate	107714	17.27%	1.246268
	2= Intermediate	19592	3.14%	1.214021
	3=Adequate Basic	6496	1.04%	2.589748
	4=Adequate Intensive	1826	0.29%	5.218935
Cigarettes during pregnancy	**False**	**571968**	**91.68%**	**1**
	True	51434	8.24%	1.68512
	(Unknown)	442	0.07%	3.021218
Maternal weight gained or lost (lb.)	0-	2415	0.39%	2.494678
	0	4047	0.65%	2.976705
	1-5	8145	1.31%	2.332416
	6-10	18975	3.04%	2.553748
	11-15	33009	5.29%	2.091897
	16-20	72456	11.61%	1.561975
	21-25	106523	17.08%	1.120656
	26-30	**131348**	**21.05%**	**1**
	31-35	92022	14.75%	0.883847
	36-40	67970	10.90%	0.900767
	41-45	32908	5.28%	0.881928
	46-50	23154	3.71%	0.943029
	51-55	9997	1.60%	0.858257
	56-60	7149	1.15%	1.114943
	61-100	6949	1.11%	1.137204
	101+	227	0.04%	1.584083
	(Unknown)	6485	1.04%	3.666505
	(Wrong Format)	65	0.01%	1.230064
Cardiac disease (risk factor for this pregnancy)	1=Yes	4226	0.68%	1.87778
	2=No	**615681**	**98.69%**	**1**
	9=Unknown	3936	0.63%	2.105597
	(blank)	1	0.00%	N/A
Diabetes, gestational	1=Yes	23023	3.69%	1.38486
	2=No	**596884**	**95.68%**	**1**
	9=Unknown	3936	0.63%	2.122963
	(blank)	1	0.00%	N/A
Eclampsia	1=Yes	7009	1.12%	7.975885
	2=No	**612898**	**98.25%**	**1**
	9=Unknown	3936	0.63%	2.204481
	(blank)	1	0.00%	N/A
Hydramnios/ oligohydramnios	1=Yes	12546	2.01%	3.804089
	2=No	**607361**	**97.36%**	**1**
	9=Unknown	3936	0.63%	2.19313
	(blank)	1	0.00%	N/A
Hemoglobinopathy	1=Yes	305	0.05%	1.400165
	2=No	**619602**	**99.32%**	**1**
	9=Unknown	3936	0.63%	2.094229
	(blank)	1	0.00%	N/A
Hypertension, chronic	1=Yes	7410	1.19%	2.969602
	2=No	**612497**	**98.18%**	**1**
	9=Unknown	3936	0.63%	2.136962
	(blank)	1	0.00%	N/A
Hypertension, pregnancy related	1=Yes	21337	3.42%	2.722845
	2=No	**598570**	**95.95%**	**1**
	9=Unknown	3936	0.63%	2.204846
	(blank)	1	0.00%	N/A
Incomplete cervix	1=Yes	3909	0.63%	8.9645

	2=No	**615998**	**98.74%**	**1**
	9=Unknown	3936	0.63%	2.16053
	(blank)	1	0.00%	N/A
Lupus erythematosus	1=Yes	446	0.07%	3.308914
	2=No	**619461**	**99.30%**	**1**
	9=Unknown	3936	0.63%	2.096797
	(blank)	1	0.00%	N/A
Previous infant 4000+ grams	1=Yes	4296	0.69%	0.133062
	2=No	**615611**	**98.68%**	**1**
	9=Unknown	3936	0.63%	2.080371
	(blank)	1	0.00%	N/A
Previous preterm infant	1=Yes	5452	0.87%	3.307573
	2=No	**614455**	**98.49%**	**1**
	9=Unknown	3936	0.63%	2.13017
	(blank)	1	0.00%	N/A
Renal disease	1=Yes	2105	0.34%	2.064534
	2=No	**617802**	**99.03%**	**1**
	9=Unknown	3936	0.63%	2.100842
	(blank)	1	0.00%	N/A
Sickle cell disease	1=Yes	493	0.08%	2.435121
	2=No	**619414**	**99.29%**	**1**
	9=Unknown	3936	0.63%	2.095988
	(blank)	1	0.00%	N/A
Uterine bleeding	1=Yes	4231	0.68%	3.22503
	2=No	**615676**	**98.69%**	**1**
	9=Unknown	3936	0.63%	2.12115
	(blank)	1	0.00%	N/A
Death Birth	**FALSE**	**620989**	**99.54%**	**1**
	TRUE	2855	0.46%	36.02023
Total		**623844**	**100.00%**	**574975**

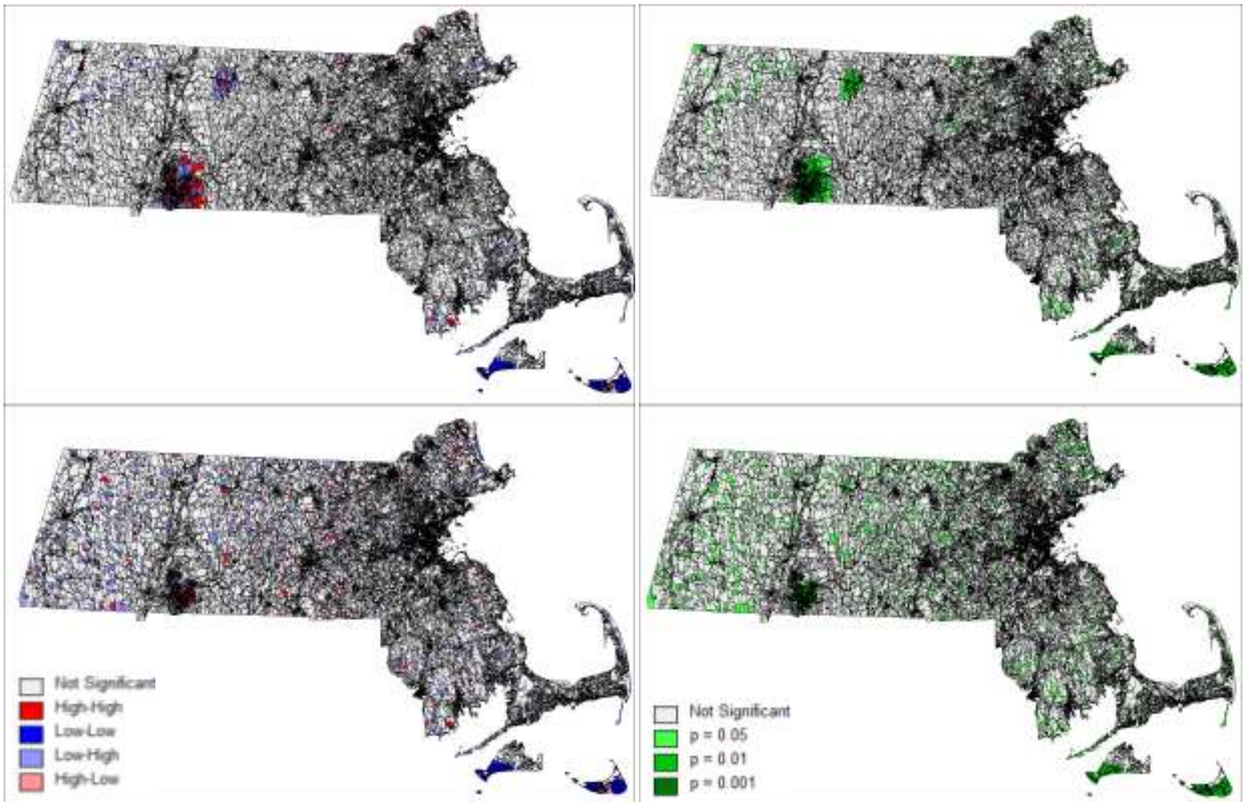

Figure 1. (a) (Top-left) Bivariate Indicators of spatial association (LISA) values of Hazard-Based Score and LBW Ratio. (b) (Top-right) Bivariate LISA significance of Hazard-Based Score and LBW Ratio. (c) (Bottom-left) Bivariate LISA values of risk-related score and LBW ratio. (d) (Bottom-right) Bivariate LISA significance of risk-related score and LBW ratio.

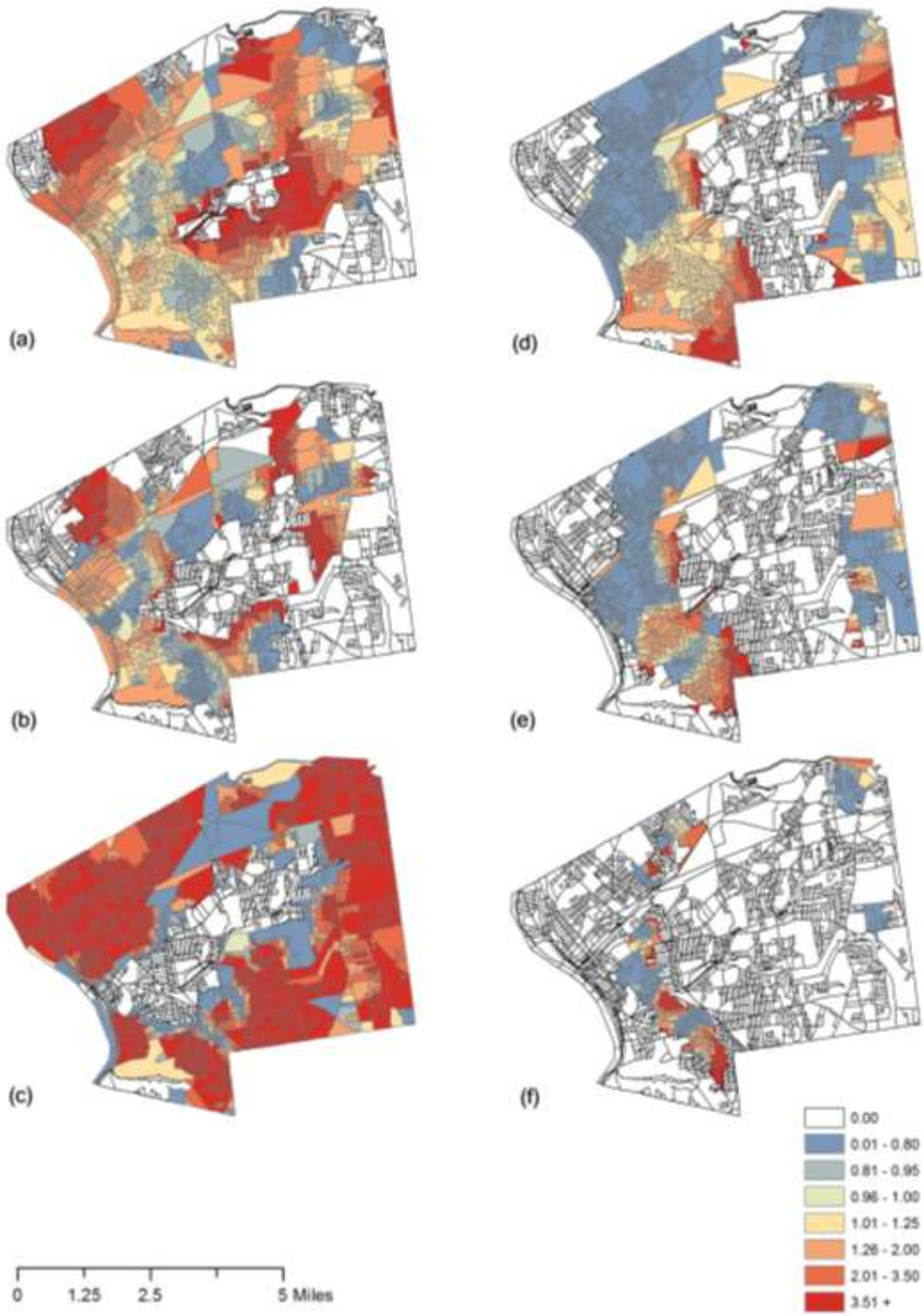

Figure 2. (a) The GWOR of LBW and hazard-based exposure score with bandwidth 632 meters. (b) The GWOR of LBW and hazard-based exposure score with bandwidth 1,256 meters. (c) The GWOR of LBW and hazard-based exposure score with bandwidth 1,881 meters. (d) The GWOR of LBW and risk-related exposure score with bandwidth 632 meters. (e) The GWOR of LBW and risk-related exposure score with bandwidth 1,256 meters. (f) The GWOR of LBW and risk-related exposure score with bandwidth 1,881 meters.