

When homogeneity meets heterogeneity: the geographically weighted regression with spatial lag approach to prenatal care utilisation

Carla Shoff¹, Vivian Yi-Ju Chen², Tse-Chuan Yang³

¹*Office of Information Products and Data Analytics, Centers for Medicare and Medicaid Services, Baltimore, USA;* ²*Department of Statistics, Tamkang University, New Taipei City, Taiwan;* ³*Department of Sociology, Center for Social and Demographic Analysis, University at Albany, SUNY, New York, USA*

Abstract. Using geographically weighted regression (GWR), a recent study by Shoff and colleagues (2012) investigated the place-specific risk factors for prenatal care utilisation in the United States of America (USA) and found that most of the relationships between late or no prenatal care and its determinants are spatially heterogeneous. However, the GWR approach may be subject to the confounding effect of spatial homogeneity. The goal of this study was to address this concern by including both spatial homogeneity and heterogeneity into the analysis. Specifically, we employed an analytic framework where a spatially lagged (SL) effect of the dependent variable is incorporated into the GWR model, which is called GWR-SL. Using this framework, we found evidence to argue that spatial homogeneity is neglected in the study by Shoff et al. (2012) and that the results change after considering the SL effect of prenatal care utilisation. The GWR-SL approach allowed us to gain a place-specific understanding of prenatal care utilisation in USA counties. In addition, we compared the GWR-SL results with the results of conventional approaches (i.e., ordinary least squares and spatial lag models) and found that GWR-SL is the preferred modelling approach. The new findings help us to better estimate how the predictors are associated with prenatal care utilisation across space, and determine whether and how the level of prenatal care utilisation in neighbouring counties matters.

Keywords: prenatal care, geographically weighted regression, spatial non-stationarity, United States of America.

Introduction

Prenatal care consists of a series of clinical visits and services that ideally begin before conception and extend throughout the ante-partum period. They are designed to promote the health and well-being of the woman, her fetus and their family (Daniels et al., 2006; Ong, 2008). In the United States of America (USA), it is recommended that women begin receiving prenatal care during the first trimester of pregnancy (US Department of Health and Human Services, 2011). When a woman receives prenatal care throughout the pregnancy, the prenatal care is more likely to be effective (US Department of Health and Human Services, 2011). Delays or not receiving prenatal care at all could result in a lost opportunity to manage health complications (e.g. diagnose and treat medical problems) and facilitate health promotion (e.g. receive health behaviour advice or referrals for other

resources), both of which could positively influence the pregnancy outcome (Cook et al., 1999; Luecken et al., 2009; Ayoola et al., 2010; Johnson et al., 2010).

Studies reporting the effectiveness of prenatal care have been somewhat mixed (Goss et al., 1997; Hueston et al., 2003; Walford et al., 2011); however, most agree that prenatal care is an important facilitator of infant and maternal health (Mikhail and Curry, 1999; Johnson et al., 2007; Epstein et al., 2009). For example, when a woman receives adequate and timely prenatal care, this decreases her risk for maternal morbidity and mortality (Misra and Guyer, 1998; Harper et al., 2003; Schrag et al., 2003), while delayed prenatal care is a significant risk factor for poor birth outcomes including higher rates of infant mortality, low birth weight and prematurity (Vintzileos et al., 2002a, 2002b, 2002c, 2003).

Recent studies have identified several factors that are associated with late or no prenatal care utilisation. These factors include socio-demographic characteristics including single marital status (Hueston et al., 2008), minority race/ethnicity (Nothnagle et al., 2000; Frisbie et al., 2001; Alexander et al., 2002; Kuo et al., 2008), lower socioeconomic status (e.g. poverty, unemployment or low levels of education) (Epstein et al., 2009; Schempf and Strobino, 2009; Ayoola et al.,

Corresponding author:

Carla Shoff

Office of Information Products and Data Analytics
Centers for Medicare and Medicaid Services
7500 Security Boulevard, Baltimore, MD 21244, USA
Tel. +1 410 786-2186; Fax: +1 410 786-5515
E-mail: carla.shoff@cms.hhs.gov

2010; Sunil et al., 2010), and no health insurance coverage (Chandler, 2002; Egerter et al., 2002; Gavin et al., 2007; Rosenberg et al., 2007). Another major contributor to the lack of prenatal care utilisation among women is the declining number of obstetrics-gynaecology (Ob-Gyn) physicians (Aved et al., 1993; Davis et al., 2004).

Since research has shown that late or no prenatal care is associated with poor maternal health (Misra and Guyer, 1998; Harper et al., 2003; Schrag et al., 2003) and birth outcomes (Chang et al., 2003; Taylor et al., 2005), identifying the factors that are associated with late or no prenatal care is essential. Recently, there is a trend in health research to question the analytic assumption that risk factors impose effects on health outcomes universally across space (Yang et al., 2009, 2013; Chen et al., 2010; Shoff and Yang, 2012; Yang and Matthews, 2012). This assumption may not hold for various reasons (Ricketts and Holmes, 2007). For example, the size of the aggregate unit varies greatly, the local practice patterns may alter the local associations between the outcome and independent variables, the differences in attitudes, knowledge and cultural beliefs of health care across space may lead to divergent health care seeking behaviours. This trend echoes the emphasis of studies on the impacts of locality and residential environments on health (MacIntyre et al., 1993; Casagrande et al., 2009; Matthews et al., 2009). Following this trend, a recent study (Shoff et al., 2012) investigated the place-specific risk factors for prenatal care utilisation in the USA and found that most of the relationships between late or no prenatal care and its determinants are spatially varying. As concluded by Shoff et al. (2012), the conventional analytic approach, namely global modelling, may result in a model misspecification (Fotheringham, 1997), and the variables in the global modelling may not well explain the spatial distribution of late or no prenatal care in the USA. From a policy-making viewpoint, a global model is useful to summarise how health outcomes are shaped. However, it does not provide detailed information for effective and efficient interventions (Yang and Matthews, 2012) and thus, including the place-specific perspective into prenatal care research would better identify local risks and may further improve maternal and infant health.

Shoff et al. (2012) used geographically weighted regression (GWR) (Fotheringham et al., 2002) to explore the place-specific risk factors for late or no prenatal care utilisation. The literature suggests that GWR contributes to incorporating the local analytic perspective into health research, although it is focused

on tackling spatial heterogeneity and does not explicitly address spatial homogeneity. It should be emphasised that spatial heterogeneity and spatial homogeneity are unique, not mutually exclusive, features of spatial data (Fotheringham, 2009). Contrarily, spatial heterogeneity and homogeneity may inform and/or affect each other (Fotheringham, 2009), and the literature suggests that both features could be considered in the analysis simultaneously (Brunsdon et al., 1998). Despite the possibility of addressing spatial heterogeneity and homogeneity at the same time, few empirical studies have adopted this analytic approach to explore the determinants of prenatal care utilisation.

One important message that could be drawn from the discussion above is that using GWR to explore the place-specific determinants for prenatal care utilisation may still be subject to the confounding effect of spatial homogeneity. We have employed an analytic framework, where a spatially lagged (SL) effect of the dependent variable (that captures spatial homogeneity) is incorporated into the GWR model (called GWR-SL) to a USA prenatal care utilisation county-level data set. The goal of this study is to fill this gap by including both spatial homogeneity and heterogeneity into the analysis, aiming to answer the following research questions: (i) is the prenatal care utilisation of a place positively related to the levels of prenatal care utilisation in neighbouring areas? (explicitly, we will test if the SL effect improves our understanding of the spatial distribution of prenatal care utilisation); (ii) do the SL effects vary across space?; and (iii) does controlling spatial homogeneity in a GWR framework alter the findings reported in our previous research?

Materials and methods

Data sources and measures

For this study, data on counties in the continental USA ($N = 3,106$) in and around the year 2000 were compiled from multiple data sources. Three-year (1999-2001) county average data was extracted from the Office on Women's Health Quick Health Data Online (<http://www.womenshealth.gov/statistics/government-in-action/>) to construct the dependent variable, *late or no prenatal care* (National Center for Health Statistics, 1999-2001). The dependent variable was measured as the percentage of women who received prenatal care during their second or third trimester of pregnancy or did not receive prenatal care at all. This is a county-level percentage based on the woman's county of residence,

not the county in which her infant was born. It should be noted that the original data from the National Center for Health Statistics are from individuals, but due to the confidentiality issue researchers are only permitted to aggregate individuals into counties and use the county as the analytic unit.

Data on the racial, ethnic and nativity composition of the county come from the 2000 USA decennial Census Summary Files 1 and 3 (US Census Bureau 2000a). The racial composition variables include the percentage of the female population aged 15-44 years, who identify themselves with one race: *percentage of black females of childbearing age*, *percentage of American Indian/Alaskan Native (AIAN) females of childbearing age* and the *percentage of Asian females of childbearing age*. As for the ethnic composition of the county, we include the *percentage of Hispanic females of childbearing age*, which is measured as the percentage of the female population aged 15-44 years who identify themselves as Hispanic; this includes Hispanics who report any racial category. A measure of nativity status composition, *percentage foreign born*, is included in the analysis. This variable is measured as the percentage of the total population who reported that they were a USA citizen by naturalisation or were not a citizen of the USA.

In order to capture the socioeconomic status of the female population in the county, a composite measure called *female socioeconomic disadvantage* was created. This composite measure was created in order to avoid problems with multicollinearity of the three variables that measure socioeconomic status that were highly correlated¹: percentage of females in poverty (percentage of females who are in poverty out of the total female population for whom the poverty status was determined), percentage of females with less than a high school education (percentage of females who are 25 years of age or older with less than a high school education), and the percentage of females unemployed (percentage of the female population 16 years of age or older who are in the labour force but are unemployed). Principal components analysis was used to create the *female socioeconomic disadvantage* variable using the regression method. The data that were used to create

these measures were also downloaded from the Office on Women's Health Quick Health Data Online.

County-level measures of health insurance coverage and healthcare providers were also included in the analyses. Specifically, the percentage of the total population who do not have insurance coverage, *percentage uninsured*, was extracted from the US Census Bureau's Small Area Health Insurance Estimates Program (US Census Bureau, 2000b). The total number of Ob-Gyn physicians in a county divided by the female population aged 15-44 years and multiplied by 100,000 is the *Ob-Gyn per 100,000 females of childbearing age* variable. The *Area Resource File* was used to extract the number of Ob-Gyn physicians per county in the year 2000 (US Department of Health and Human Services, 2008).

Analytic strategy

The methodological contribution of this study is to employ the GWR-SL analytic framework for analysis of prenatal data. The founders of GWR implied that spatial autoregressive modelling could be integrated into GWR (Brunsdon et al., 1998). One of the dominant spatial autoregressive regression models is the spatial lag model (Ward and Gleditsch, 2008) where a spatial lagged effect of the dependent variable is included into the analysis. It has been argued that the spatial lag model could capture the spatial homogeneity embedded in spatial data and it has been widely used in previous research (Sparks and Sparks, 2010; Yang et al., 2011; Sparks et al., 2012). To the best of our knowledge, Páez et al. (2002) first demonstrated that a spatial lag effect could be fused into a Gaussian-based GWR and be estimated with the maximum likelihood (ML) approach.

What distinguishes our work from that by Páez et al. (2002) is that we use the instrumental variable (IV) approach to estimate our GWR-SL model. In contrast to the ML estimation, IV estimation does not require the normality assumption and could be used to facilitate causal inferences and to control for unmeasured errors (Greenland, 2000; Hernán and Robins, 2006). Let y_i be the 3-year averaged percentage of late or no prenatal care in county i , $i = 1, 2, \dots, N$, and y the vector ($N \times 1$) containing y_i . Denote the covariates of i^{th} county in a vector containing a constant and the aforementioned eight county-level characteristics of interest as x_i . Our analytic model could be expressed as:

$$y_i = \rho(u_i, v_i) w_i y + x_i' \beta(u_i, v_i) + \varepsilon_i$$

(equation 1)

¹The three socioeconomic status variables were highly correlated and significant at the $P < 0.001$ level: the percentage of females in poverty and the percentage of females with less than a high school degree (0.72), the percentage of females in poverty and the percentage of females who are unemployed (0.69), and the percentage of females with less than high school degree and percentage of females who are unemployed (0.52).

where ε_i is a random error term, (u_i, v_i) the location coordinates (centroid of the county), w_i the i^{th} row ($1 \times N$) of a known spatial contiguity matrix W ($N \times N$) specifying spatial neighbours (first order queen) and $\rho(u_i, v_i)$ the spatial lag dependence between county-level percentages of late or no prenatal care at location (u_i, v_i) determined by w_i . The interpretation of ρ is similar to that in the spatial lag model. For a given location, say j , ρ indicates the relationship between j 's dependent variable (i.e. percentage of late or no prenatal care) and the dependent variable of j 's neighbours defined by the contiguity matrix. A positive and significant ρ means the positive spatial autocorrelation, whereas a negative ρ refers to the negative spatial autocorrelation (Anselin, 1988). The notation $\beta(u_i, v_i)$ represents the vector of (local) regression coefficients associated with x_i at location (u_i, v_i) .

The regression parameters of model (1) are estimated by adopting the technique of two-stage least squares (2SLS) (Kelejian and Prucha, 1998; Kim et al., 2010) on a point-wise basis using kernel-based methods. Specifically, for a given location (u_0, v_0) in the studied region, the β s and ρ are locally estimated to minimise the following function:

$$\sum_{i=1}^n [y_i - \rho(u_0, v_0) w_i y - x_i' \beta(u_0, v_0) - z_i' \delta(u_0, v_0)]^2 K\left(\frac{d_{i0}}{h}\right) \quad (\text{equation 2})$$

where K is a given kernel function with h being the bandwidth and d_{i0} the distance between (u_0, v_0) and the i^{th} designed location (u_i, v_i) . The quantity $\delta(u_0, v_0)$ is the coefficient vector associated with the instrumental variables z_i that is used to construct a proxy (fitted values) in the estimation process for model endogeneity posed by the SL dependent variable $w_i y$. When constructing the instruments, we followed the suggestion by past studies that for a matrix X containing x_i , $z_i = w_i X$, is the standard candidate used in spatial 2SLS. See Chen and Yang (2013) and Huang and Long (2011) for a more detailed discussion of the estimation theory.

In this study, statistical inference for model parameters is established by using bootstrapping techniques. A Monte Carlo approach similar to that of Brunson et al. (1998) was applied to test for the variability of individual regression coefficients. A full presentation of these two methods can be found in Chen and Yang (2013).

Comparison of GWR-SL with other approaches

As the GWR-SL approach is novel particularly for prenatal care research, we would like to compare this

approach to other conventional spatial analysis methods, including Bayesian conditional autoregressive (CAR) model and simultaneous autoregressive (SAR) model. Several features that distinguish GWR-SL from CAR and SAR are discussed below.

For the analysis of spatial areal unit data, the Bayesian approach is an alternative technique for estimation of the ecological relationships between outcomes of interest and regional covariates. In essence, Bayesian methods, orientated by the work of Besag et al. (1991), decomposed the area-specific random effect across geographical units into two components, spatially structured and non-structured random effect, and analysed them with a hierarchical perspective. The spatially structured random effects are typically captured with a prior from CAR models that consider spatial dependence (clustering effect) through specification of the spatial structure underlying the data (Besag et al., 1991). The non-structured random effects are often taken to account for the heterogeneity among regions. This Bayesian approach has become widely used in various fields, especially in disease mapping studies.

The GWR-SL model used in this paper is a non-Bayesian approach that attempts to simultaneously capture both spatial dependence (spatial homogeneity) and spatial heterogeneity. The intent of modelling seems to be similar to that of the Bayesian approach, but essentially in a different context. Before further elaboration on the GWR-SL approach, we must include a discussion of SAR modelling as it is widely used in health research using spatial analysis techniques. The SAR modelling was developed as part of spatial econometrics (LeSage and Pace, 2009) and it assumes that the dependent variable in a specific area/unit is a function of (or related to) the dependent variables in neighbouring areas/units. This provides an overall estimate of how the neighbours' dependent variables affect the dependent variable in a certain area, which is also known as the spatial lag effect in spatial analysis (Ward et al., 2008). This spatial lag effect in SAR is obtained with the whole spatial structure underlying the data and it provides thus a single parameter estimate. Similar to the CAR model mentioned above, the spatial structure underlying the data plays a critical role in the estimation process and the spatial lag effect is a single (i.e. global) estimate, rather than a location-specific statistic.

Five features distinguish our GWR-SL from CAR and SAR modelling. First, GWR-SL generalises the spatial lag effect into the local spatial modelling context. Explicitly, the spatial lag effect in SAR is an overall estimate, but the spatial lag effect in GWR-SL is for

each observation (i.e. location-specific). The “local” spatial lag effects allow researchers to better explore the pattern of where the neighbours matter. Second, while the spatially structure error terms in CAR modelling are location-specific and aim to capture spatial dependence and heterogeneity (Besag et al., 1991), they are essentially “variations” that cannot be explained by the independent variables and cannot really be interpreted. By contrast, our local spatial lag effects are drawn from the dependent variable and it can thus be understood how the neighbours’ values of a dependent variable affect the dependent variable in an area. Our GWR-SL assumes that the error terms are randomly distributed and unrelated to spatial structure. Third, both the SAR and the CAR approaches were developed under the global spatial modelling framework, while GWR-SL was proposed within the local spatial modelling family. In other words, the associations estimated with SAR or CAR between the dependent and the independent variables are for all observations in the data; however, GWR-SL focuses on “each” observation and its neighbours defined by spatial adjacency or kernel density functions and the estimated associations, therefore, are specific to each observation. Fourth, following the previous point, GWR-SL will generate abundant parameter estimates (i.e. coefficients) and provide detailed information on how one location interacts with its neighbouring areas. Nonetheless, the SAR and CAR approaches only yield the global coefficients, which cannot be used to visualise the spatial patterns. Finally,

the spatial heterogeneity addressed in GWR-SL model is expressed by varying regression coefficients, which is a type of observed spatial heterogeneity (Anselin 1988, 2008). By contrast, CAR models handle the unobserved spatial heterogeneity that lead to heteroscedasticity error terms due to missing variables or other forms of model specifications.

Results

Global model results

The ordinary least squares (OLS) and spatial lag regression model results are provided in Table 1. These are the global modelling results. This model explained 35% of the total variance among the dependent variable, the percent late or no prenatal care, and has an Akaike Information Criterion (AIC) of 19,617. In contrast, the spatial lag regression model explains 48% of the total variance in percent late or no prenatal care and has an AIC of 19,070, indicating that it is statistically preferable to the OLS regression model.

The SL dependent variable (ρ) captures the association between the prenatal care utilisation of a county and the levels of prenatal care utilisation in neighbouring counties. The spatial lag regression modelling suggested that if the level of late or no prenatal care utilisation in neighbouring counties increases by 1%, the percentage of late or no prenatal care utilisation of a specific county would increase by roughly 0.5%. After including ρ into the model, each of the

Table 1. OLS and spatial lag regression models predicting the percentage of women who received late or no prenatal care (global regression models) N = 3,106 counties.

	OLS regression		Spatial lag regression	
	Estimate	Std. error	Estimate	Std. error
Intercept	7.22***	0.41	3.15***	0.48
Percent Black females of childbearing age	0.21***	0.04	0.14***	0.03
Percent AIAN females of childbearing age	0.89***	0.08	0.71***	0.07
Percent Asian females of childbearing age	-2.26***	0.35	-2.06***	0.32
Percent Hispanic females of childbearing age	-0.03	0.07	-0.16**	0.06
Percent foreign born	0.30***	0.04	0.30***	0.04
Female socioeconomic disadvantage	0.00	0.00	0.39**	0.15
Percent uninsured	0.58***	0.03	0.34***	0.03
Ob-Gyn per 100,000 females of childbearing age	-0.01*	0.00	-0.00	0.00
ρ (spatially lagged dependent variable)			0.48***	0.02
Adjusted R-square		0.35		0.48
Akaike information criterion		19,617		19,070

*P <0.05; **P <0.01; ***P <0.001.

racial/ethnic and nativity status composition measures was significantly associated with late or no prenatal care. A 1% point increase in the Black and AIAN female of childbearing age population was associated with a 0.14% and 0.71% increase in the percentage of women who receive late or no prenatal care. While increases in the Asian and Hispanic female of childbearing age populations were associated with a 2.06% and 0.16% decrease in the percentage of women receiving prenatal care after their first trimester of pregnancy or not receiving any prenatal care at all. The significant relationship between the percentage of Hispanic females of childbearing age and percentage late or no prenatal care was not identified in the OLS model. This relationship may not have been identified, because the spatial structure of the data was not taken into account in the regression model, which can result in the estimated regression coefficients to be biased, inconsistent or inefficient (Ward and Gleditsch, 2008; LeSage and Pace, 2009). A 1% increase in the foreign born population is associated with a 0.30% increase in the percentage of women who receive late or no prenatal care.

The global model results also show that a one unit increase in female socioeconomic disadvantage is associated with a 0.39% increase in the percentage of women with late or no prenatal care. This is another relationship that was not identified using the OLS regression approach. A 1% increase in the uninsured population resulted in a 0.34% increase in the percentage of women who received prenatal care after the first trimester of pregnancy or did not receive prenatal care at all. After including the SL dependent variable into the model, the negative relationship between the rate of Ob-Gyn physicians in the county and late or no prenatal care was no longer statistically significant.

Local model results

The GWR and GWR-SL results are provided in Table 2. These are the local model 5-number parameter summary results. The 5-number summary of parameter estimates defines the extent of the variability in the parameter estimates based on the minimum, lower quartile, median, upper quartile and maximum local parameter

Table 2. Geographically weighted regression and geographically weighted regression-spatial lag 5-number parameter summary results (N = 3,106 counties).

	Geographically weighted regression					
	Minimum	Lower quartile	Median	Upper quartile	Maximum	Monte Carlo
Intercept	-17.32	2.95	7.66	13.43	25.72	Non-Stationary
Percent Black females of childbearing age	-16.43	-0.06	0.43	0.82	18.43	Non-Stationary
Percent AIAN females of childbearing age	-96.06	-2.57	-0.04	0.98	34.00	Non-Stationary
Percent Asian females of childbearing age	-30.35	-6.38	-2.86	-0.07	23.58	Non-Stationary
Percent Hispanic females of childbearing age	-11.07	-1.26	0.05	1.85	14.40	Non-Stationary
Percent foreign born	-3.19	-0.20	0.35	0.80	3.38	Non-Stationary
Female socioeconomic disadvantage	-5.38	-0.38	0.63	2.28	6.61	Non-Stationary
Percent uninsured	-1.20	0.20	0.54	0.98	3.73	Non-Stationary
Ob-Gyn per 100,000 females of childbearing age	-0.08	-0.02	-0.01	0.01	0.10	Stationary
Rho (spatially lagged dependent variable)	-	-	-	-	-	-
	Geographically weighted regression spatial lag					
	Minimum	Lower quartile	Median	Upper quartile	Maximum	Monte Carlo
Intercept	-21.67	-1.31	2.63	7.61	22.75	Non-Stationary
Percent Black females of childbearing age	-13.85	0.10	0.37	0.74	12.60	Stationary
Percent AIAN females of childbearing age	-82.10	-1.63	-0.05	0.72	18.78	Stationary
Percent Asian females of childbearing age	-24.67	-5.66	-2.74	-0.19	21.22	Stationary
Percent Hispanic females of childbearing age	-6.73	-1.11	0.00	1.53	11.31	Non-Stationary
Percent foreign born	-2.11	-0.13	0.31	0.74	3.34	Non-Stationary
Female socioeconomic disadvantage	-4.88	-0.05	1.18	2.27	5.70	Non-Stationary
Percent uninsured	-0.93	0.13	0.40	0.89	3.33	Stationary
Ob-Gyn per 100,000 females of childbearing age	-0.07	-0.01	-0.00	0.01	0.07	Non-Stationary
Rho (spatially lagged dependent variable)	-1.04	0.20	0.39	0.56	1.46	Non-Stationary

estimates reported in the GWR and GWR-SL models (Matthews and Yang, 2012). The Monte Carlo tests indicated the associations between the independent and dependent variable are non-stationary across space for each of the associations in the GWR model with the exception of the rate of Ob-Gyn physicians in the county. However, in the GWR-SL model, the relationship between the rate of Ob-Gyn physicians and late or no prenatal care is non-stationary, in addition to the relationships between percent Hispanic females of child-bearing age, percent foreign born, female socioeconomic disadvantage and the SL dependent variable and late or no prenatal care.

The output from the GWR-SL models was used to generate surfaces for each model parameter and each surface depicts the spatial variation of a relationship with the outcome variable (Matthews and Yang, 2012). Here we use a continuous bivariate colour scheme for the parameter estimate surface where the significant positive parameter estimates are displayed in shades of blue and green and the non-significant areas are masked out (Matthews and Yang, 2012).

The local parameter estimates of the GWR-SL model results are displayed in Figs. 1 and 2. The first map in Fig. 1 shows the local parameter estimates of the SL dependent variable. The coloured areas of the map represent the counties where neighbouring counties prenatal care utilisation has a significant effect on the prenatal care utilisation of that county. For the counties shaded blue, the higher the percentage of late or no prenatal care utilisation in the neighbouring counties, the higher the percentage of late or no prenatal care utilisation in that county. Recalling our previous discussion, this relationship is captured by the local ρ and most of these associations are positive spatial autocorrelations, which was found in the majority of the significant areas. For example, if the level of late or no prenatal care utilisation in neighbouring counties increases by 1%, the percentage of late or no prenatal care utilisation of a specific county would increase by roughly 1.4%. However, in some counties located in Iowa, if the percentage of late or no prenatal care utilisation in the neighbouring counties increases by 1%, the percentage of late or no prenatal

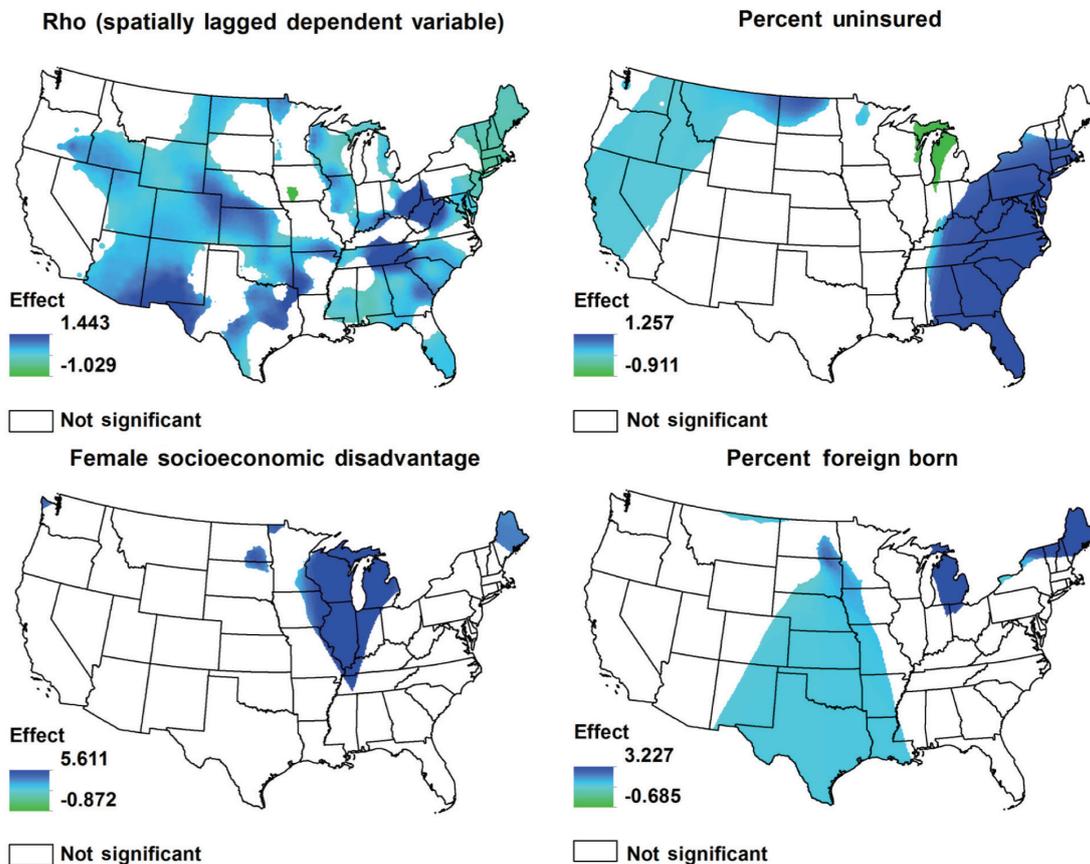


Fig. 1. Map of GWR-SL estimates for the spatially lagged dependent variable, percent uninsured, female socioeconomic disadvantage and percent foreign born (significant areas at $P < 0.05$ level).

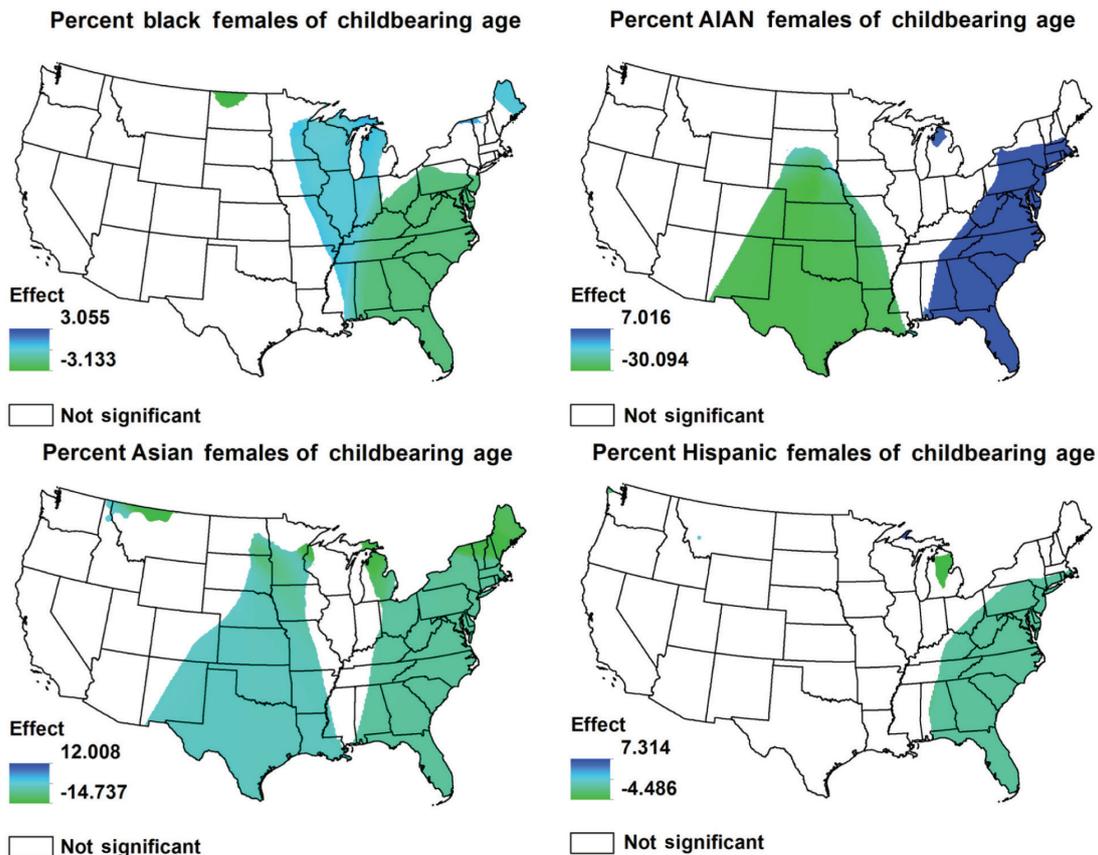


Fig. 2. Map of GWR-SL estimates for percent Black females of childbearing age, percent American Indian/Alaskan Native females of childbearing age, percent Asian females of childbearing age and percent Hispanic females of childbearing age (significant areas at $P < 0.05$ level).

care utilisation in those counties decreases by approximately 1%.

Figs. 1 and 2 include the results of the GWR-SL regression estimates for the parameters that mapped in the paper by Shoff et al. (2012) using the method proposed by Matthews and Yang (2012). One notable difference between the GWR and GWR-SL results is that for each of the parameters, the range of the estimates (minimum to maximum) is not as wide in the GWR-SL model results as it is for the GWR model results (see Table 2), and the maps reveal different significant spatial patterns (see Figs. 1 and 2 in this paper for the GWR-SL model results and Figs. 2 and 3 in Shoff et al. (2012) for the GWR model results).

Discussion

The GWR-SL approach allowed us to gain a better understanding of prenatal care utilisation in USA counties improving on more traditional approaches in a number of ways. The traditional OLS regression approach allowed us to determine which variables were significant predictors of late or no prenatal care

utilisation; however, these results may be biased because they did not take into account the spatial nature of the data. This was made clear from the comparison between the results of the spatial lag regression model and those of the OLS regression model. The spatial lag model revealed that the percentage of Hispanic females of childbearing age and the female socioeconomic disadvantage were in fact significant predictors of late or no prenatal care utilisation. This model also showed that the Ob-Gyn per 100,000 females of childbearing age was not a significant predictor of late or no prenatal care utilisation. The spatial lag regression model was an improvement over the traditional OLS regression approach, because it allowed us to determine that neighbouring counties prenatal care utilisation is a significant predictor of a county's prenatal care utilisation. The downfall was that we could not determine which counties are influenced by their neighbouring counties as the estimates were for the overall associations between prenatal care utilisation and the explanatory covariates. This approach also did not allow us to determine whether the relationship between late or no prenatal care utili-

sation and its predictors are constant over space. A GWR approach has been used in previous research and the findings reveal that the relationship between prenatal care utilisation and its predictors vary significantly across space, with the exception of the relationship between the rate of Ob-Gyn physicians in the county and the percentage of late or no prenatal care (Shoff et al., 2012). Nevertheless, this approach still does not explicitly address the issue of which counties prenatal utilisation is associated with their neighbouring counties' prenatal care utilisation. Herein lies the strength of the GWR-SL approach; it allows us to take into account spatial heterogeneity by allowing the relationships to vary across space *and* spatial homogeneity by controlling for the neighbouring counties prenatal care utilisation.

The GWR-SL approach was beneficial for this study and allowed us to compare with the results in Shoff et al. (2012). First, we were able to understand how the predictors of prenatal care utilisation vary across space and determine whether and how the level of prenatal care utilisation in neighbouring counties matters, which was absent in previous work. Second, after accounting for the percentage of late or no prenatal care utilisation in neighbouring counties, the GWR-SL approach helped us to further clarify the relationship between the rate of Ob-Gyn physicians and the percentage of late or no prenatal care. Specifically, without considering the level of prenatal care utilisation in neighbouring areas (ordinary GWR approach), higher percent of Ob-Gyn physicians seemed to impose a spatially universal impact on prenatal care utilisation. Nonetheless, this universal association was found to be a consequence of ignoring spatial homogeneity and is essentially spatially non-stationary. Similarly, some of the relationships that were non-stationary in the GWR results were in fact not (the relationships between the racial composition of the county and late or no prenatal care, as well as percentage uninsured and late or no prenatal care). After including the SL dependent variable into the model (GWR-SL approach), the significant spatial patterns of the parameter estimates have changed.

The maps presented here are essential, because they identify areas where policies can be implemented to improve prenatal care utilisation across the USA. For example, the results reveal for which counties prenatal care utilisation can be improved by increasing the prenatal care utilisation in neighbouring counties. As discussed above, previous research found that a lack of health insurance coverage is a significant risk factor for not receiving prenatal care on time or even at all

(Chandler, 2002; Egerter et al., 2002; Gavin et al., 2007; Rosenberg et al., 2007). However, the GWR-SL model results show the areas where increasing the health insurance coverage can improve the prenatal care utilisation. As shown in Fig. 1, counties located in the Mid-Atlantic, southeast, west and northwest regions are areas where the higher the percentage of the population who are uninsured the higher the percentage of late or no prenatal care. These are the areas that can be targeted to increase health insurance coverage which will, in turn, increase prenatal care utilisation. Implementing these place specific policies is essential if we are going to meet the goal of *Healthy People 2020* of increasing the percentage of pregnant women who receive prenatal care in the first trimester from 70.8% to 77.9% (US Department of Health and Human Services, 2011).

While the methodological approach used in this study was beneficial for gaining a better understanding of prenatal care utilisation in the USA, it poses some limitations. First, as this research uses data that have been aggregated to a particular geographic unit (i.e. county) the modifiable areal unit problem (MAUP) could be an issue (Openshaw, 1984; Raghavan et al., 2013). Second, like other ecological studies, the findings presented here are subject to the ecological fallacy, which refers to the error of deriving conclusions about individuals based on the results of aggregate-level analyses (Piantadosi et al., 1988). Third, the local spatially lagged effects (ρ) are defined by the spatial contiguity matrix, but there is no agreement on how to define "contiguity". We defined neighbours as those who share the same boundary or vertex (first-order) and using a higher order may alter the findings and conclusions.

Acknowledgements

We received support from the Population Research Institute (Penn State; R24-HD41025), the Center for Social and Demographic Analysis (University at Albany; R24-HD044943), and the National Science Council of Taiwan (NSC102-2119-M-032-001).

References

- Alexander GR, Kogan MD, Nabukera S, 2002. Racial differences in prenatal care use in the United States: are disparities decreasing? *Am J Public Health* 92, 1970-1975.
- Anselin L, 1988. *Spatial econometrics: methods and models*. Kluwer Academic Publishers, 284 pp.
- Anselin L, Le Gallo J, Jayet H, 2008. Spatial panel econometrics. In: *The econometrics of panel data, fundamentals and*

- recent developments in theory and practice (third edition). Matyas L, Sevestre P (eds). New York: Springer, 625-662 pp.
- Aved BM, Irwin MM, Cummings LS, Findeisen N, 1993. Barriers to prenatal care for low-income women. *West J Med* 158, 493-498.
- Ayoola AB, Nettleman MD, Stommel M, Canady RB, 2010. Time of pregnancy recognition and prenatal care use: a population-based study in the United States. *Birth* 37, 37-43.
- Besag J, York J, Mollie A, 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1-59.
- Brunsdon C, Fotheringham AS, Charlton M, 1998. Spatial non-stationarity and autoregressive models. *Environ Plann A* 30, 957-974.
- Casagrande SS, Whitt-Glover MC, Lancaster KJ, Odoms-Young AM, Gary TL, 2009. Built environment and health behaviors among African Americans: a systematic review. *Am J Prev Med* 36, 174-181.
- Chandler D, 2002. Late entry into prenatal care in a rural setting. *J Midwifery Womens Health* 47, 28-34.
- Chang SC, O'Brien KO, Nathanson MS, Mancini J, Witter FR, 2003. Characteristics and risk factors for adverse birth outcomes in pregnant black adolescents. *J Pediatr* 143, 250-257.
- Chen VYJ, Wu PC, Yang TC, Su HJ, 2010. Examining the non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. *Sci Total Environ* 408, 2042-2049.
- Chen VYJ, Yang TC, 2013. Estimation and inference of geographically weighted autoregressive lag models. Working paper.
- Cook CA, Selig KL, Wedge BJ, Gohn-Baube EA, 1999. Access barriers and the use of prenatal care by low-income, inner-city women. *Soc Work* 44, 129-139.
- Daniels P, Noe GF, Mayberry R, 2006. Barriers to prenatal care among black women of low socioeconomic status. *Am J Health Behav* 30, 188-198.
- Davis K, Baksh L, Bloebaum L, Streeter N, Rolfs B, 2004. Barriers to adequate prenatal care in Utah. *PRAMS Perspectives* 1, 1-8.
- Egerter S, Braveman P, Marchi K, 2002. Timing of insurance coverage and use of prenatal care among low-income women. *Am J Public Health* 92, 423-427.
- Epstein B, Grant T, Schiff M, Kasehagen L, 2009. Does rural residence affect access to prenatal care in Oregon? *J Rural Health* 25, 150-157.
- Fotheringham AS, 1997. Trends in quantitative methods I: Stressing the local. *Prog Hum Geogr* 21, 88-96.
- Fotheringham AS, 2009. "The problem of spatial autocorrelation" and local spatial statistics. *Geogr Anal* 41, 398-403.
- Fotheringham AS, Brunsdon C, Charlton ME, 2002. Geographically weighted regression: the analysis of spatially varying relationships. New York: John Wiley and Sons, 269 pp.
- Frisbie WP, Echevarria S, Hummer RA, 2001. Prenatal care utilization among non-hispanic whites, African Americans, and Mexican Americans. *Matern Child Health J* 5, 21-33.
- Gavin NI, Adams EK, Manning WG, Raskind-Hood C, Urato M, 2007. The impact of welfare reform on insurance coverage before pregnancy and the timing of prenatal care initiation. *Health Serv Res* 42, 1564-1588.
- Goss GL, Lee K, Koshar J, Heilemann MS, Stinson J, 1997. More does not mean better: prenatal visits and pregnancy outcome in the hispanic population. *Public Health Nurs* 14, 183-188.
- Greenland S, 2000. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 29, 722-729.
- Harper MA, Byington RP, Espeland MA, Naughton M, Meyer R, Lane K, 2003. Pregnancy-related death and health care services. *Obstet Gynecol* 102, 273-278.
- Hernán MA, Robins JM, 2006. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 17, 360-372.
- Huang Y, Long ZH, 2011. 2SLS estimation for the GWR-SL model. 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), December 16-18, Changchun, China.
- Hueston WJ, Geesey ME, Diaz V, 2008. Prenatal care initiation among pregnant teens in the United States: an analysis over 25 years. *J Adolesc Health* 42, 243-248.
- Hueston WJ, Gilbert GE, Davis L, Sturgill V, 2003. Delayed prenatal care and the risk of low birth weight delivery. *J Commun Healthc* 28, 199-208.
- Johnson AA, Hatcher BJ, El-Khorazaty MN, Milligan RA, Bhaskar B, Rodan MF, Richards L, Wingrove BK, Laryea HA, 2007. Determinants of inadequate prenatal care utilization by African American women. *J Health Care Poor Underserved* 18, 620-636.
- Johnson PJ, Call KT, Blewett LA, 2010. The importance of geographic data aggregation in assessing disparities in American Indian prenatal care. *Am J Public Health* 100, 122-128.
- Kelejian HH, Prucha IR, 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J Real Estate Financ* 17, 99-121.
- Kim SG, Cho SH, Lambert DM, Roberts RK, 2010. Measuring the value of air quality: application of the spatial hedonic model. *Air Qual Atmos Health* 3, 41-51.
- Kuo TM, Gavin NI, Adams EK, Ayadi MF, 2008. Racial disparities in medicaid enrollment and prenatal care initiation among pregnant teens in Florida: comparisons between 1995 and 2001. *Med Care* 46, 1079-1085.
- LeSage JP, Pace RK, 2009. Introduction to spatial econometrics, Chapman and Hall/CRC, 374 pp.
- Luecken LJ, Purdom CL, Howe R, 2009. Prenatal care initiation in low-income hispanic women: risk and protective factors.

- Am J Health Behav 33, 264-275.
- MacIntyre S, Maciver S, Sooman A, 1993. Area, class and health: should we be focusing on places or people? *J Soc Policy* 22, 213-234.
- Matthews SA, Moudon AV, Daniel M, 2009. Using geographic information systems (GIS) for enhancing research relevant to policy on diet, physical activity, and weight. *Am J Prev Med* 36, S171-S176.
- Matthews SA, Yang TC, 2012. Mapping the results of local statistics: using geographically weighted regression. *Demogr Res* 26, 151-166.
- Mikhail BI, Curry MA, 1999. Perceived impediments to prenatal care among low-income women. *West J Nurs Res* 21, 335-355.
- Misra DP, Guyer B, 1998. Benefits and limitations of prenatal care. *JAMA* 279, 1661-1662.
- National Center for Health Statistics, 1999-2001. National vital statistics system detail natality files. Edited by US Department of Health and Human Services. Washington DC: The Office on Women's Health Quick Health Data Online.
- Nothnagle M, Marchi K, Egerter S, Braveman P, 2000. Risk factors for late or no prenatal care following medicaid expansions in California. *Matern Child Health J* 4, 251-259.
- Ong S, 2008. Guidelines for perinatal care. *Obstet Gynecol* 10, 207-207.
- Openshaw S, 1984. Ecological fallacies and the analysis of areal census data. *Environ Plann A* 16, 17-31.
- Páez A, Uchida T, Miyamoto K, 2002. A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environ Plann A* 34, 883-904.
- Piantadosi S, Byar DP, Green SB, 1988. The ecological fallacy. *Am J Epidemiol* 127, 893-904.
- Raghavan RK, Brenner KM, Harrington Jr. JA, Higgins JJ, Harkin KR, 2013. Spatial scale effects in environmental risk-factor modelling for diseases. *Geospat Health* 7, 169-182.
- Ricketts TC, Holmes GM, 2007. Mortality and physician supply: does region hold the key to the paradox? *Health Serv Res* 42, 2233-2251.
- Rosenberg D, Handler A, Rankin KM, Zimbeck M, Adams EK, 2007. Prenatal care initiation among very low-income women in the aftermath of welfare reform: does pre-pregnancy medicaid coverage make a difference? *Matern Child Health J* 11, 11-17.
- Schempf AH, Strobino DM, 2009. Drug use and limited prenatal care: an examination of responsible barriers. *Am J Obstet Gynecol* 200, 1-10.
- Schrag SJ, Arnold KE, Mohle-Boetani JC, Lynfield R, Zell ER, Stefonek K, Noga H, Craig AS, Sanza LT, Smith G, 2003. Prenatal screening for infectious diseases and opportunities for prevention. *Obstet Gynecol* 102, 753-760.
- Shoff C, Yang TC, 2012. Spatially varying predictors of teenage birth rates among counties in the United States. *Demogr Res* 27, 377-418.
- Shoff C, Yang TC, Matthews SA, 2012. What has geography got to do with it? Using GWR to explore place-specific associations with prenatal care utilization. *GeoJournal* 77, 331-341.
- Sparks JP, Sparks CS, 2010. An application of spatially autoregressive models to the study of us county mortality rates. *Popul Space Place* 16, 465-481.
- Sparks JP, Sparks CS, Campbell JJA, 2012. An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the US. *GeoJournal* 1-17.
- Sunil TS, Spears WD, Hook L, Castillo J, Torres C, 2010. Initiation of and barriers to prenatal care use among low-income women in San Antonio, Texas. *Matern Child Health J* 14, 133-140.
- Taylor CR, Alexander GR, Hepworth JT, 2005. Clustering of us women receiving no prenatal care: differences in pregnancy outcomes and implications for targeting interventions. *Matern Child Health J* 9, 125-133.
- US Census Bureau, 2000a. Census 2000 summary file 1 and 3, detailed tables.
- US Census Bureau, 2000b. Health insurance coverage status by age for counties and states. Suitland: Small Area Health Insurance Estimates.
- US Department of Health and Human Services, 2008. Area resource file. Rockville: Health Resources and Services Administration, Bureau of Health Professions.
- US Department of Health and Human Services, 2011. Healthy people 2020: Maternal, infant, and child health objectives. Washington DC.
- Vintzileos AM, Ananth CV, Smulian JC, Scorza WE, 2003. The impact of prenatal care on preterm births among twin gestations in the United States, 1989-2000. *Am J Obstet Gynecol* 189, 818-823.
- Vintzileos AM, Ananth CV, Smulian JC, Scorza WE, Knuppel RA, 2002a. The impact of prenatal care in the United States on preterm births in the presence and absence of antenatal high-risk conditions. *Am J Obstet Gynecol* 187, 1254-1257.
- Vintzileos AM, Ananth CV, Smulian JC, Scorza WE, Knuppel RA, 2002b. The impact of prenatal care on neonatal deaths in the presence and absence of antenatal high-risk conditions. *Am J Obstet Gynecol* 186, 1011-1016.
- Vintzileos AM, Ananth CV, Smulian JC, Scorza WE, Knuppel RA, 2002c. Prenatal care and black-white fetal death disparity in the United States: heterogeneity by high-risk conditions. *Obstet Gynecol* 99, 483.
- Walford HH, Trinh S, Wiencrot A, Lu MC, 2011. What is the role of prenatal care in reducing racial and ethnic disparities in pregnancy outcomes? In: *Reducing racial/ethnic disparities in reproductive and perinatal outcomes: the evidence from population-based interventions*. Handler A, Kennelly J, Peacock N

- (eds). New York: Springer, 461 pp.
- Ward MD, Gleditsch KS, 2008. Spatial regression models, Sage Publications, California, 112 pp.
- Yang TC, Matthews SA, 2012. Understanding the non-stationary associations between distrust of the health care system, health conditions, and self-rated health in the elderly: a geographically weighted regression approach. *Health Place* 18, 576-585.
- Yang TC, Shoff C, Matthews SA, 2013. Examining the spatially non-stationary associations between the second demographic transition and infant mortality: a poisson GWR approach. *Spatial Demog* 1, 18-41.
- Yang TC, Wu PC, Chen VYJ, Su HJ, 2009. Cold surge: a sudden and spatially varying threat to health? *Sci Total Environ* 407, 3421-3424.
- Yang TC, Jensen L, Haran M, 2011. Social capital and human mortality: explaining the rural paradox with county-level mortality data. *Rural Sociol* 76, 347-374.