Applied Geography 62 (2015) 191–200

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Geographically-weighted regression analysis of percentage of late-stage prostate cancer diagnosis in Florida

Pierre Goovaerts [a,*], Hong Xiao [b], Georges Adunlin [b], Askal Ali [b], Fei Tan [c], Clement K. Gwede [d], Youjie Huang [e]

[a] BioMedware, Inc., Ann Arbor, MI, USA
[b] University of Florida, Gainesville, FL, USA
[c] Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
[d] Moffitt Cancer Center, Tampa, FL, USA
[e] Florida Department of Health, Tallahassee, FL, USA

A R T I C L E   I N F O

Article history:
Available online

Keywords:
Logistic regression
Cancer
Contextual factors
Spatial statistics

A B S T R A C T

This study assessed spatial context and the local impacts of putative factors on the proportion of prostate cancer diagnosed at late-stages in Florida during the period 2001–2007. A logistic regression was performed aspatially and by geographically-weighted regression (GWR) at the nodes of a 5 km spacing grid overlaid over Florida and using all the cancer cases within a radius of 125 km of each node. Variables associated significantly with high percentages of late-stage prostate cancer included having comorbidities, smoking, being Black and living in census tracts with farmhouses. Having private or public insurance, being married or diagnosed in a for-profit facility, as well as living in census tracts with high household income reduced significantly this likelihood. Geographically-weighted regression allowed the identification of areas where the local odds ratio is significantly different from the ratio estimated using aspatial regression (State-level). For example, the local odds ratios for the comorbidity covariates were significantly smaller than the State-level odds ratio in Tallahassee and Pensacola, while they were significantly larger in Palm Beach. This emphasizes the need for local strategies and cancer control interventions to reduce the percentage of prostate cancer diagnosed at late-stages and ultimately eliminate health disparities.

Introduction

Prostate cancer (PCa) is the most common solid malignancy and the second leading cause of cancer-related death for American men. It has been estimated that there will be 233,000 new cases and 29,480 deaths from this disease in the United States (US) in 2014 (American Cancer Society, 2014). The State of Florida ranks second, behind California, for both incidence (16,590 estimated new cases) and mortality (2170 estimated deaths) from PCa in 2014 (American Cancer Society, 2014).

Difference in individual and contextual factors—including age, race, socioeconomic status (SES), comorbidity, geographic location, and access to health care, display striking disparities across geographic regions with respect to the incidence, mortality and percent of PCa diagnosed at late-stages. For instance, PCa incidence rates are approximately 70 percent higher for African Americans than for Caucasians and death rate are twice as high for African Americans as for any other racial/ethnic group (American Cancer Society, 2013). An important factor associated with high percentage of late-stage PCa is the presence and severity of comorbidity. Comorbidity is the co-occurrence of one or more diseases or disorders in an individual (Bartsch et al., 1992; Siu, Lau, Tam, & Shiu, 2002). Comorbidity reflects the aggregate effect of all clinical conditions a patient might have, excluding the disease of primary interest (Arcangeli, Smith, Ratliff, & Catalona, 1997). A growing body of evidence supports the association of PCa risk with farming, due to exposure to toxic chemicals, especially pesticides (Alavanja et al., 2003; Meyer, Coker, Sanderson, & Symanski, 2007; Settimi, Masina,

* Corresponding author. BioMedware, Inc, 121 W. Washington St., 4th floor-TBC, Ann Arbor, MI 48104, USA.
E-mail addresses: goovaerts@biomedware.com (P. Goovaerts), hxiao18@ufl.edu (H. Xiao), george.adunlin@gmail.com (G. Adunlin), askalaali@gmail.com (A. Ali), ftan@math.iupui.edu (F. Tan), clement.gwede@moffitt.org (C.K. Gwede), YH2010FL@gmail.com (Y. Huang).

http://dx.doi.org/10.1016/j.apgeog.2015.04.018
0143-6228/

Andrion, & Axelson, 2003). Geographical disparities in percent of late-stage PCa have been associated with poor access to primary health care, lack of health insurance and difference in coverage (Mandelblatt, Yabroff, & Kerner, 1999; Mullins, Blatt, Gbarayor, Yang, & Baquet, 2005; Roetzheim et al., 1999; Talcott et al., 2007).

Studies of geographic variations have made important contributions to our understanding of how geography, individual and contextual factors jointly shape the distribution of PCa incidence and percent of late-stage PCa. To our knowledge, no PCa studies have explicitly investigated spatial heterogeneity in individual and contextual factor differences across counties in the State of Florida.

The study of the correlation between health data and risk factors is traditionally performed using global or aspatial regression, with the implicit assumption that the impact of covariates is constant across the study area. This assumption is likely unrealistic for large states such as Florida that display substantial geographic variation in demographic, social, economic, and environmental conditions. To account for the non-stationarity of relationships in space, aspatial regression can be supplemented with geographically-weighted regression (GWR), whereby the regression model is fitted within local windows selected by the user so as to include enough observations. Each observation (i.e. prostate cancer case whose residence falls within that window) is weighted according to its proximity to the center of the window (Fotheringham, Brunsdon, & Charlton, 2003). Local regression coefficients and associated statistics (i.e. proportion of variance explained, odds ratio) can then be mapped to visualize how the explanatory power of covariates changes spatially (Cardozo, García-Palomares, & Gutiérrez, 2012; Mennis, 2006; Su, Xiao, & Zhang, 2012).

A study by Goovaerts introduced the first application of GWR to the analysis of health disparities, with a study of PCa mortality across the United States (Goovaerts, 2005). Wheeler & Tiefelsdorf used GWR to explore local relationships between bladder cancer mortality rates at the state economic area (a group of similar counties) level and two explanatory variables: population density (proxy for environmental and behavioral differences) and lung cancer mortality rates (proxy for the risk factor smoking) (Wheeler & Tiefelsdorf, 2005). The same year, Nakaya and colleagues developed a geographically weighted Poisson regression approach to conduct ecological regression (i.e. study of relationship between aggregated data) in space, with an application to the relationship between working-age death in the Tokyo metropolitan area and socio-economic factors (Nakaya, Fotheringham, Brunsdon, & Charlton, 2005). Since then, geographically-weighted regression has been increasingly applied to the analysis of local relationships between health outcomes and putative factors (Chen & Truong, 2012; Chen, Wu, Yang, & Su, 2010; Chi, Grigsby-Toussaint, Bradford, & Choi, 2013; Shoff, Chen, & Yang, 2014; Yang & Matthews, 2012). Most studies have however been conducted using aggregated health data, such as county-level rates, and application to PCa data has been sparse. This study aims to conduct regression analysis in a spatial context to assess the local impacts of individual and contextual factors on percent of late-stage PCa in Florida.

Materials and methods

Data and data sources

The analysis was conducted on 39,374 cases aged 40 or older that were diagnosed with PCa in the State of Florida between 10/1/2001 and 12/31/2007. Data were obtained from four different sources and at three different spatial scales (individual, census tract, county). First, individual-level data were acquired from the Florida Cancer Data System (FCDS) housed at the University of Miami. The FCDS was established as the state central cancer registry in 1981, and is the largest single population-based cancer incidence registry in the nation (Florida Department of Health, 2014). The FCDS has been part of the Centers for Disease Control and Prevention National Program of Cancer Registries since 1996. The FCDS collects information on patient demographics, residence, prostate tumor characteristics and other data such as tobacco use and primary payer of health insurance.

Second, diagnoses data were obtained from the Florida Agency for Health Care and Administration (AHCA). AHCA maintains two databases (Hospital Patient Discharge Data and Ambulatory Outpatient Data) on all patient encounters within hospitals and freestanding ambulatory surgical and radiation therapy centers in Florida. Comorbidity was computed following the Elixhauser Index method (Elixhauser, Steiner, Harris, & Coffey, 1998) based on diagnoses information from AHCA. The study used a total of 45 conditions, including 29 from the Elixhauser Index plus 16 additional conditions based on clinical characteristics of the study population. The method used to come up with these conditions is explained in greater details in another publication (Xiao et al., 2013).

Third, data on socio-demographic and environmental characteristics were extracted at the census tract level from the U.S. Census Bureau (Census 2000, Summary File-3) public use files for the State of Florida.

Fourth, health provider information by county was obtained from the Florida Department of Health Division of Medical Quality Assurance to calculate provider to case ratios. Specifically, the number of primary health providers and urologists was divided by the number of prostate cancer diagnoses for each county during 2001–2007. This measure was used to capture provider availability.

Statistical analysis

The relationship between percent of PCa diagnosed at late-stages and putative factors was modeled using logistic regression. The dependent variable is an indicator variable taking a value of 1 if the patient was diagnosed late, and zero otherwise. Covariates include age, race, marital status, smoking, type of health insurance (uninsured, public or private insurance) and facilities (for-profit and not-for-profit) where diagnosis was made, presence of no comorbidity, 1 to 2 comorbidities, more than 2 comorbidities, census-tract median household income and presence of farmhouse, year of diagnosis, county-level provider-to-case ratios.

The regression model was fitted using two different methods: 1) the traditional approach that ignores the coordinates of the observations (aspatial regression), and 2) geographically-weighted regression (GWR) that fits a local regression model at the 5970 nodes of a 5 km spacing grid overlaid over Florida and using all the cancer cases within a radius of 125 km of each node (Fig. 1). A grid spacing of 5 km provided enough resolution to map local spatial patterns while keeping the number of regression models (5970) low enough to be computationally feasible. Details of the procedure can be found in Fotheringham et al. (2003), Cardozo et al. (2012), Chen and Truong (2012), Chi et al. (2013), Su et al. (2012), Wang, Zhang, and Li (2013). In particular, a description of geographically-weighted logistic regression is provided by Rodrigues, de la Riva, and Fotheringham (2014).

The window size for GWR had to be large enough to include, for each grid node, all levels of each categorical covariate so that logistic regression could be performed. This condition was satisfied by a radius of 125 km, which also ensured that at all but seven grid nodes a minimum of 1000 observations was available for regression; see the spatial distribution of number of observations in Fig. 1. The average and median numbers of cases within each window are 7237 and 7,598, respectively. The use of a constant window size was