

Identifying counties vulnerable to diabetes from obesity prevalence in the United States: a spatiotemporal analysis

Xiao Li,¹ Amanda Staudt,² Lung-Chang Chien³

¹*Department of Biostatistics, University of Texas Health Science Center at Houston (UTHealth), School of Public Health, Houston, TX;* ²*Department of Epidemiology, Human Genetics & Environmental Sciences, University of Texas Health Science Center at Houston (UTHealth), School of Public Health in San Antonio, San Antonio, TX;* ³*Department of Biostatistics, University of Texas Health Science Center at Houston (UTHealth), School of Public Health in San Antonio, San Antonio, TX, USA*

Abstract

Clinical and epidemiological research has reported a strong association between diabetes and obesity. However, whether increased diabetes prevalence is more likely to appear in areas with increased obesity

prevalence has not been thoroughly investigated in the United States (US). The Bayesian structured additive regression model was applied to identify whether counties with higher obesity prevalence are more likely clustered in specific regions in 48 contiguous US states. Prevalence data adopted the small area estimate from the Behavioral Risk Factor Surveillance System. Confounding variables like socioeconomic status adopted data were from the American Community Survey. This study reveals that an increased percentage of relative risk of diabetes was more likely to appear in Southeast, Northeast, Central and South regions. Of counties vulnerable to diabetes, 36.8% had low obesity prevalence, and most of them were located in the Southeast, Central, and South regions. The geographic distribution of counties vulnerable to diabetes expanded to the Southwest, West and Northern regions when obesity prevalence increased. This study also discloses that 7.4% of counties had the largest average in predicted diabetes prevalence compared to the other counties. Their average diabetes prevalence escalated from 8.7% in 2004 to 11.2% in 2011. This study not only identifies counties vulnerable to diabetes due to obesity, but also distinguishes counties in terms of different levels of vulnerability to diabetes. The findings can provide the possibility of establishing targeted surveillance systems to raise awareness of diabetes in those counties.

Correspondence: Lung-Chang Chien, Department of Biostatistics, University of Texas Health Science Center at Houston (UTHealth), School of Public Health in San Antonio, 7411 John Smith Drive, Suite 1100, 78229 San Antonio, TX, USA.

Tel: +1.210.276.9051 - Fax: +1.210.276.9028.

E-mail: Lung-Chang.Chien@uth.tmc.edu

Key words: Diabetes prevalence; Obesity prevalence; Structured additive regression.

Acknowledgements: this study was supported by Dr. Chien's start-up funds at University of Texas School of Public Health. Funding for this project was provided to Amanda Staudt by the Southwest Center for Occupational and Environmental Health (SWCOEH), a NIOSH Education and Research Center, and awardee of Grant No. 5T42OH008421 from the National Institute for Occupational and Environmental Health (NIOSH)/Centers for Disease Control and Prevention.

Conflict of interest: the authors declare no potential conflict of interest.

Contributions: L-CC and XL designed and coordinated this project; AS and L-CC collected data and did literature review; L-CC and XL performed spatiotemporal analyses; XL, L-CC and AS wrote, reviewed, and edited the manuscript.

Note: all findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention, Atlanta, GA, USA. Part of this paper was presented at the 2016 APHA annual meeting, 2016 Nov 1-4, Chicago, IL, USA.

Received for publication: 9 December 2015.

Revision received: 12 July 2016.

Accepted for publication: 15 July 2016.

©Copyright X. Li et al., 2016
Licensee PAGEPress, Italy
Geospatial Health 2016; 11:439
doi:10.4081/gh.2016.439

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Introduction

Over the past three decades, the prevalence of diagnosed diabetes in the US increased from 2.5 to 6.9% (CDC, 2014a). Consequently, diabetes-related morbidity has dramatically increased over time in the United States (US). The number of hospital discharges among those diagnosed with diabetes has nearly doubled with 2.8 million discharges in 1988 and 5.5 million discharges in 2009 (CDC, 2014b). In addition, diabetes has led to increases in emergency department visits, self-reported heart disease or stroke, visual impairment, lower extremity conditions, end-stage renal disease, and other causes of morbidity (CDC, 2014b). Obesity and diabetes are both epidemic in the US, and diabetes has long been linked to obesity (Dansinger, 2014). Starting in the 1990s, the rate of obesity in the US increased dramatically. The age-adjusted obesity prevalence among adults was 15.6% in 1995, 19.8% in 2000, and 23.7% in 2005 (Blanck *et al.*, 2006). The current prevalence of obesity is 34.9% among adults and 17% among children (CDC, 2014c). In 2010, the majority (84.7%) of diabetics in the US was overweight or obese and over half of diabetics (56.9%) were obese (CDC, 2014b). The prevalence of obesity and diabetes vary



across US states. The highest recorded prevalence of obesity in the southern region of the US was 30.2%. In Mississippi and West Virginia it was over 35%, and no state in the US had a prevalence of obesity lower than 20% (CDC, 2014c). The *diabetes belt*, which is where the prevalence of diabetes is concentrated, includes 644 counties mostly located in the southern states, and the belt includes the entire state of Mississippi (CDC, 2011). Sullivan *et al.* (2005) conducted a study in the US on the prevalence of diabetes and found obesity to be strongly associated with diabetes. Similarly, a California study conducted in Alameda County, found that obesity and overweight were mediators of Type 2 diabetes (Maty *et al.*, 2005). There may be a possibility of a geographic disparity in the association of obesity and diabetes. However, whether increased diabetes prevalence is more likely to appear in areas with increased obesity prevalence has not been thoroughly investigated in the US. The number of spatial analyses conducted on obesity and diabetes is limited. Research on urbanisation as a potential risk factor for diabetes has led to mixed results. In Asia, those living in urban areas were found to be more likely to have diabetes (Mohan *et al.*, 2008; Ning *et al.*, 2009; Katulanda *et al.*, 2013), while in Greece, rural populations had a significantly increased prevalence of diabetes (Melidonis *et al.*, 2006). A study in Canada did not find a difference in the relative risk (RR) of diabetes based on a rural-urban comparison, but a difference was found based on region (Foulds *et al.*, 2012). A few studies have been conducted on the spatial relationship between diabetes and obesity in the US. In 2007, counties in the top two quintiles in both obesity and diabetes prevalence were located in the South and the Appalachian regions (Gregg *et al.*, 2009). Congdon used a multilevel method using 2007 data to confirm that the influence of geographic variation existed on joint weight and diabetes status (Congdon, 2010). Other studies have reported geographic clustering based on diabetes and obesity status, but these studies were limited to small regions in the US (Schlundt *et al.*, 2006; Laraia *et al.*, 2014; Zheutlin *et al.*, 2014). These results provide evidence that a more comprehensive analysis that uses a spatio-temporal approach should be implemented with a longer time period.

Preliminary research reveals that areas in the US may have higher joint obesity and diabetes risks (Congdon, 2010), but those findings lack advanced analyses providing reliable and enhanced evidence for people who live in areas vulnerable to diabetes due to obesity. Hence, the research purpose of this study is to investigate if diabetes is spatially correlated with obesity at the county level, distinguish spatial clusters, and identify counties vulnerable to diabetes due to obesity. We address three research questions in this study: i) whether obesity prevalence is spatially correlated to diabetes prevalence; ii) whether an increase in obesity prevalence changes the geographic distribution of diabetes at the county level; and iii) whether counties vulnerable to diabetes can be clustered. The ultimate goal is to apply disease-risk mapping for carrying out spatial association and variation between diabetes prevalence and obesity prevalence in the US.

Materials and Methods

Data source

This study contains three data sources: the Behavioral Risk Factor Surveillance System (BRFSS) (<http://www.cdc.gov/brfss>), the American Community Survey (ACS) (US Census Bureau, 2014), and the Cartographic Boundary Files (<https://www.census.gov/geo/maps-data/data/tiger-cart-boundary.html>). The BRFSS data has been collected by individual state health departments under the direction of US

Centers for Diseases and Prevention since 1984. It is the largest nationwide telephone survey, and it annually gathers data on individual's characteristics, risk behaviours, living status, and health conditions. Using a Bayesian multilevel small area estimating method, BRFSS calculates and publishes data for public-use on age-gender-race adjusted prevalence for diabetes, obesity, and physical inactivity in each county (Congdon and Lloyd, 2010). BRFSS has been approved by Human Research Review Boards from the Department of Health in each state. Selected participants need to sign an informed consent for some specific questions. Detailed information about the BRFSS survey design, full-text questionnaires, and data collection can be found at its website (<http://www.cdc.gov/brfss>).

The data of ACS are collected by the largest decennial survey, administered by the US Census Bureau since 2005 to provide the most current and detailed information about population, social, housing and economic conditions for states and local areas. The ACS has been widely used by federal/state/local agencies, nongovernmental organizations, educators, business, and journalists (US Census Bureau, 2014). Every year the ACS publishes 1-year, 3-year, and 5-year estimates for socioeconomic status (SES) factors, and this study adopted the 5-year estimate (2007-2011) to reduce the amount of missing data in any county. In addition, the cartographic boundary files are maintained and stored by the US Census Bureau's geographic database.

Study area

This study only considered the 48 contiguous states with a total of 3109 counties. Each county has at least one neighbouring county. We selected county as the geographic unit in this research because the smallest geographic data collected by BRFSS is at the county level. For the purpose of facilitating explanations of the spatial pattern variations, we divided up the 3109 counties into nine regions using the US climate regions defined by the National Climate Data Center (Appendix 1).

Statistical data analysis

Firstly, we applied Moran's I statistics to measure and test the spatial autocorrelation of age-adjusted diabetes prevalence and age-adjusted obesity prevalence at the county level for each year. Then, we built two models to examine spatial association between diabetes prevalence and obesity prevalence:

Model 1:

$$\log(DM_{it}) = \alpha + b_0 + \beta X_i + f_i + OB_{it} \times f_{spat} \quad (\text{eq. 1})$$

where DM_{it} and OB_{it} are the adjusted diabetes prevalence and the adjusted obesity prevalence at county i ($i=1, 2, \dots, 3109$) and calendar year t ($t=1, 2, \dots, 8$) from 2004 to 2011, respectively. The notations α and b_0 represent a fixed intercept and a random intercept, respectively. The covariate vector X_i with a dimension of 16×1 contains SES confounding factors obtained from 5-year estimates of ACS, including health insurance coverage percentage, high school or higher education percentage, male percentage, elders (≥ 65 year-old) percentage, non-Hispanic white percentage, non-Hispanic black percentage, Latino/Hispanic percentage, average age, median family income per 1000 US dollars, poverty percentage, adjusted physical inactivity prevalence and five occupational percentages (management/business/science/art, service, sales/office, natural resources/construction/maintenance, and production/transportation/material moving). The notation f_i is a nonlinear time smoother to control for temporal autoregressive correlations. This model can assess the spatial heterogeneity of obesity prevalence on diabetes prevalence across counties. The spatial func-

tion can be regarded as a geographic weight to evaluate how obesity prevalence contributes to diabetes prevalence in each county differently, along with considering complex spatial influence related to obesity prevalence. We used the intrinsic conditional autoregressive prior for the spatial function to capture the neighbourhood structure. Spatial clusters of diabetes prevalence due to obesity prevalence can be first demonstrated by Model 1. For the purpose of spatially comparing different levels of adjusted obesity prevalence on diabetes, we categorised obesity prevalence into quartiles (low, median-low, median-high, and high levels), and modified Model 1 to fulfil the study purpose:

Model 2:

$$\log(DM_{it}) = \alpha + b_0 + \beta X_i + f_t + f_{\text{spat.L}} + I_{ML}^{OB} \times f_{\text{spat.ML}} + I_{MH}^{OB} \times f_{\text{spat.MH}} + I_H^{OB} \times f_{\text{spat.H}} \quad (\text{eq. 2})$$

where three indicator variables (I_{ML}^{OB} , I_{MH}^{OB} , I_H^{OB}) along with spatial functions ($f_{\text{spat.ML}}$, $f_{\text{spat.MH}}$, $f_{\text{spat.H}}$) represented counties with median-low, median-high, or high level of obesity prevalence, respectively. It is noted that the reference level is low obesity prevalence level, which is not needed and can be presented by a single spatial function $f_{\text{spat.L}}$. All the other notations and symbols in Model 2 are identical to those used in Model 1. Both models were implemented by the structured additive regression model with a fully Bayesian inference using Markov chain Monte Carlo simulations (Brezger and Lang, 2006; Fahrmeir and Lang, 2001). The estimated coefficients and statistical significances of parameters in all linear terms were determined by posterior means and 95% confidence intervals (CIs). The nonlinear smoother, f_t , was estimated by a B-spline function with a second-order random walk (Lang and Brezger, 2004). We applied Markov random fields (Kindermann and Snell, 1980) with a conditional autoregressive prior to estimate all spatial functions. The estimated spatial function can be used to calculate the increased percentage of RR (RR%) for diabetes for every 1% increase in obesity prevalence in each county. The 95% CI was also used to determine the spatial significance of each estimate in a spatial function. We defined spatial vulnerability as those counties that have a RR% significantly greater than 0. In particular, Model 2 can conduct a spatial significance in each spatial function, representing that each county can be identified to have a spatial vulnerability in each level of obesity prevalence. Hence, we counted the number of spatial vulnerability in each county in Model 2, where counties with 0 spatial vulnerability indicate no impact on diabetes prevalence from each level of obesity prevalence, while counties with four spatial vulnerabilities

indicates each level of obesity prevalence significantly affect diabetes prevalence. Then, we defined five vulnerable levels (definite, higher, moderate, lower, and least) for counties from four spatial vulnerabilities to no spatial vulnerability in Model 2. Counties with definite vulnerable level represent residents in any of the four obesity levels who are vulnerable to diabetes, and so on. To verify discrimination between the five vulnerable levels among all counties, we applied the analysis of variance to compare predicted diabetes prevalence among five vulnerable levels. This study also performed a sensitivity analysis to examine the robustness of spatial estimates by using different hyper-parameters of the prior of the spatial variance in both models. Both data cleaning, management and summary were accomplished by SAS v9.3 (SAS Institute Inc., Cary, NC, USA). Spatial analysis was carried out by BayesX software version 2.1 (Brezger *et al.*, 2005). The significance of multiple comparisons was decided by $P < 0.05$.

Results

The average adjusted diabetes prevalence and average adjusted obesity prevalence were unevenly distributed across the nation (Figure 1). The average adjusted diabetes prevalence ranged from 3.90 per 100,000 people to 16.03 per 100,000 people, a higher rate appeared in the southeastern region. The spatial distribution of the average annual adjusted obesity prevalence resembles the spatial pattern of average annual adjusted diabetes prevalence, ranging from 13.00 to 43.03 per 100,000

Table 1. Spatial autocorrelations of the annual adjusted diabetes and obesity prevalence at the county level in the US, 2004-2011.

Year	Diabetes prevalence		Obesity prevalence	
	Moran's <i>I</i>	P	Moran's <i>I</i>	P
2004	0.4868	<0.0001	0.4185	<0.0001
2005	0.4822	<0.0001	0.4212	<0.0001
2006	0.5122	<0.0001	0.4039	<0.0001
2007	0.5090	<0.0001	0.3975	<0.0001
2008	0.5222	<0.0001	0.4053	<0.0001
2009	0.4793	<0.0001	0.3655	<0.0001
2010	0.4565	<0.0001	0.3577	<0.0001
2011	0.4337	<0.0001	0.3363	<0.0001

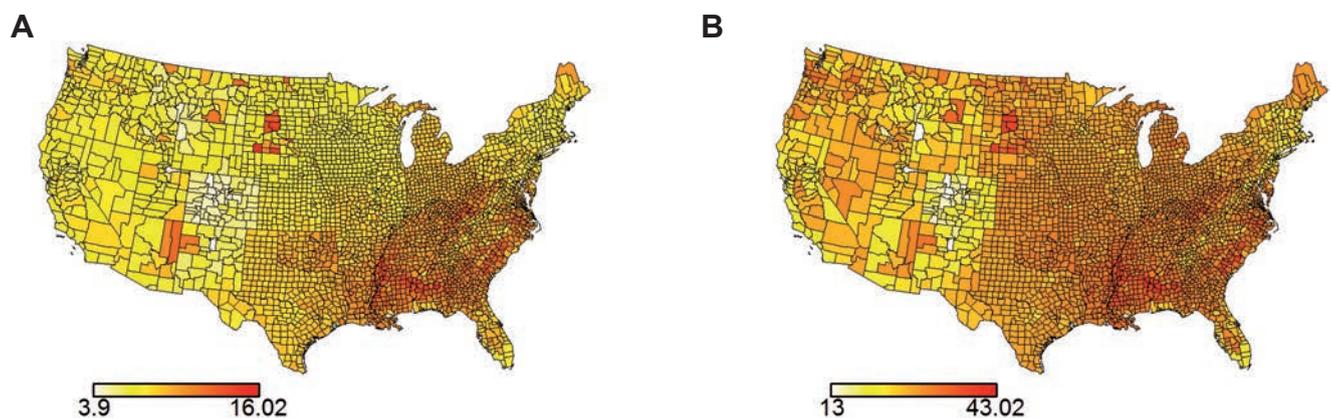


Figure 1. Spatial distribution of the average adjusted diabetes (A) and obesity (B) prevalence per 100,000 people in the US.

populations. The Moran's *I* statistics indicate that there is an existing significant spatial autocorrelation of average adjusted diabetes prevalence and average adjusted obesity prevalence in the 3109 US counties each year (Table 1).

The results derived from Model 1 are shown in Figure 2, which reveals an uneven distribution in the influence of obesity prevalence on diabetes in the US. As obesity prevalence increased 1%, the increased percentage of RR for diabetes ranged from -0.11% (95% CI=-0.38, -0.15) in Lake County, Colorado to 2.08% (95% CI=1.89, 2.26) in Boone County, West Virginia. A higher increased RR% for diabetes was more likely to appear in Southeast, Northeast, Central and South regions. The geographic distribution of the increased RR% varied across the US, and the significance map reveals that most counties had a significantly increased RR% greater than 0%.

After analysing and comparing the spatial patterns of the four quartile levels of adjusted obesity prevalence, 36.83% of counties with a low level of adjusted obesity prevalence were found to be vulnerable to diabetes. Those counties were most prominent in Southeast, Central and South regions as shown in Figure 3A. When obesity prevalence increased to the median-low level, the highest RR% of diabetes increased 7.05% (95% CI=-0.77, 15.24) in San Juan County, Washington. Regions had more counties with a significantly increased RR% for diabetes when adjusted obesity prevalence increased to the median-low level, except in the Southeast, South and East North Central regions. In particular, a great amount of counties in Southwest and West regions become vulnerable to diabetes as shown in Figure 3B. When adjusted obesity prevalence elevated to the median-high level, the significance map in Figure 3C indicates that counties vulnerable to diabetes expanded to East North Central, West North Central and Northwest regions. When adjusted obesity prevalence increased to the median-high level, counties vulnerable to diabetes expanded to East North Central region. When adjusted obesity prevalence elevated to the high level, the greatest RR% increment for diabetes was in Union County, Florida at 23.65% (95% CI=-17.56, -29.44). Additionally, more counties vulnerable to diabetes appeared in Central and West North Central regions.

The geographic distribution of the five vulnerable levels shown in Figure 4 explains a spatial cluster of adjusted diabetes prevalence as 230 counties (7.40%) were attributed to the definite vulnerable level. In these counties, suffering due to high diabetes prevalence escalated

from 8.74% in 2004 to 11.18% in 2011. Most of the counties in definite vulnerable level were concentrated in Central and Southeast regions. Moreover, multiple comparisons resulted in significant differences among vulnerable levels by at least 0.58% (moderate level *vs* lower level), while the difference between definite level and higher level was only 0.13% (P=0.0629; 95% CI=-0.01, 0.27) and not statistically significant (Table 2). When spatial estimates in Model 1 and Model 2 were plotted by adjusted physical inactivity prevalence *vs* without physical inactivity prevalence, distribution was along the 45° line (Figure 5). This reflects that our results, in terms of spatial functions, were robust and not sensitive. The sampling trace also shows that main estimated parameters reached convergence (Appendix 2).

Discussion

The findings of this study contribute to the known geographical differences of diabetes prevalence across the US and add critical information to the body of knowledge on diabetes prevalence. The county-level

Table 2. Multiple comparisons of predicted diabetes prevalence among the five vulnerable levels.

Comparison	Mean difference	95% CI*	P
Definite <i>vs</i> higher	0.13	-0.01, 0.27	0.0629
Definite <i>vs</i> moderate	1.33	1.18, 1.47	<0.0001
Definite <i>vs</i> lower	1.90	1.76, 2.04	<0.0001
Definite <i>vs</i> least	2.94	2.80, 3.08	<0.0001
Higher <i>vs</i> moderate	1.19	1.05, 1.33	<0.0001
Higher <i>vs</i> lower	1.77	1.63, 1.91	<0.0001
Higher <i>vs</i> least	2.81	2.67, 2.95	<0.0001
Moderate <i>vs</i> lower	0.58	0.44, 0.72	<0.0001
Moderate <i>vs</i> least	1.62	1.48, 1.76	<0.0001
Lower <i>vs</i> least	1.04	0.90, 1.18	<0.0001

CI, confidence interval.

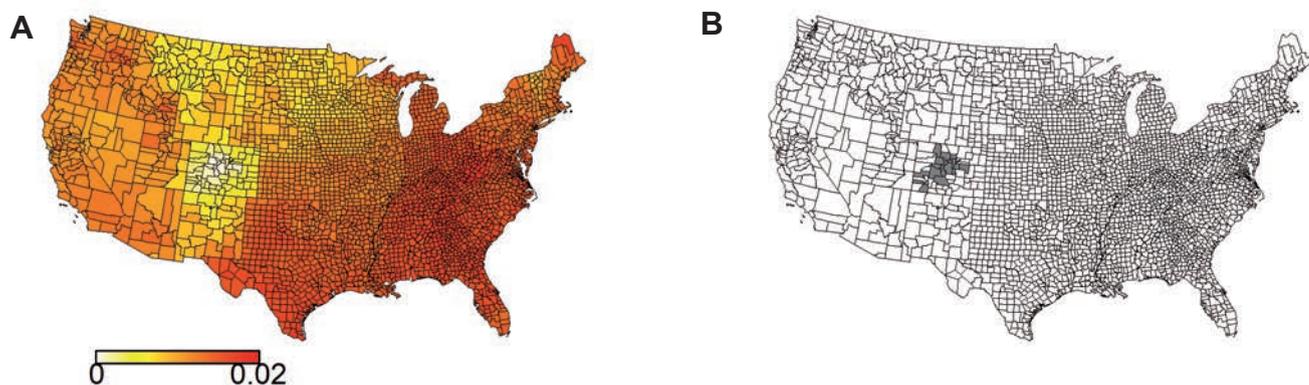
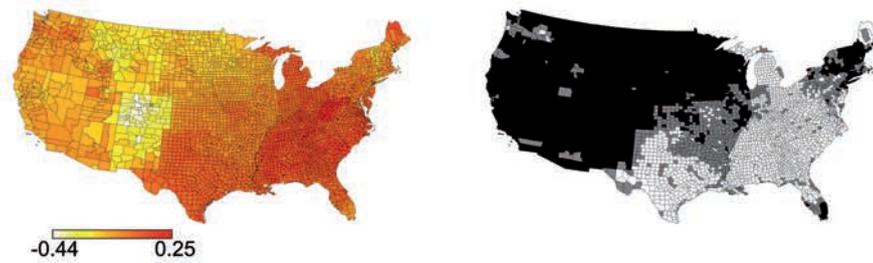
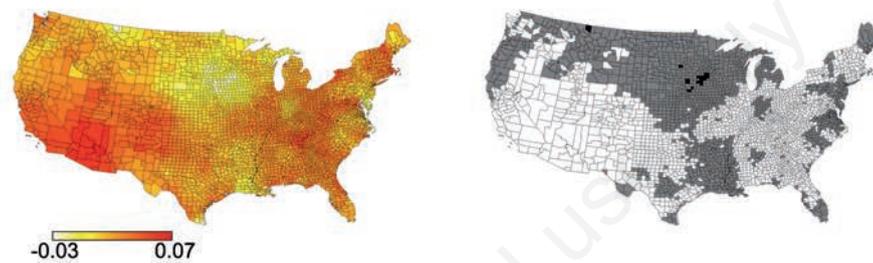


Figure 2. Spatial association between diabetes and obesity. A) Spatial map of the increased percentage of relative risk (RR) due to adjusted obesity prevalence; B) significance map, where counties shaded by white colour had a significantly positive increased percentage of RR, black colour had a significantly negative increased percentage of RR, and grey counties had a non-significant increased percentage of RR.

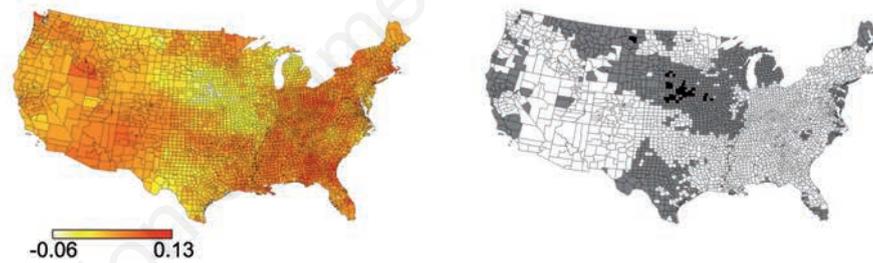
A



B



C



D

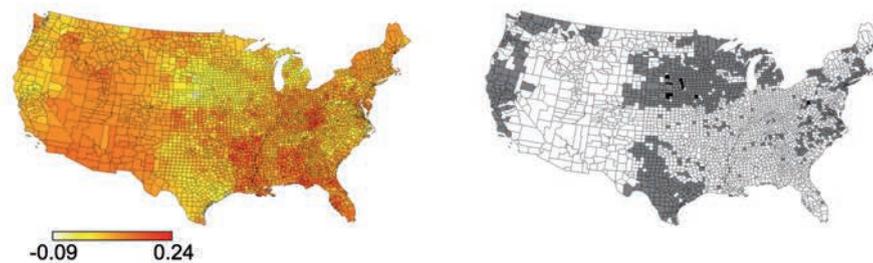


Figure 3. Spatial association of diabetes prevalence by levels of adjusted obesity prevalence shown with significance maps. On the left are spatial maps for diabetes prevalence; on the right are significance maps where counties shaded by white colour had a significantly increased relative risk% (RR%), black colour had a significantly decreased RR%, and grey counties had a non-significant RR%. A) Low level, B) low-median level, C) median-high level, and D) high level of adjusted obesity prevalence.

analysis determined that counties vulnerable to diabetes were more likely to be clustered in Southeast, Central and South regions because of higher adjusted obesity prevalence. As obesity prevalence level increased, counties vulnerable to diabetes expanded to Northeast, East North Central, West North Central and Northwest regions. The study was able to visualise the county-level spatial heterogeneity relationship between obesity prevalence and diabetes prevalence across states, to monitor counties with high risk of diabetes, and to quantify geographical disparity, explaining the level of variation of diabetes risk.

The spatial impact of adjusted diabetes prevalence discovered in this study was elaborated by the geographic variation of adjusted obesity prevalence, which has not been statistically proven in previous studies. For example, diabetes prevalence was only previously investigated to have significant spatial autocorrelations and an association with $PM_{2.5}$ at the county-level in the US (Pearson *et al.*, 2010; Chien *et al.*, 2015). Advanced evidence concluded that neighbourhood characteristics related to greater affluence, occupation, and education are associated with higher Type 1 diabetes risk (Liese *et al.*, 2012). Spatial clustering analysis also revealed significant county-level diabetes prevalence in the US after adjusting for socio-demographic and built environment-related variables (Hipp and Chalise, 2015). In addition, previous studies have shown a spatial variation in diabetes incidence. After controlling for population density, SES, remoteness and ethnicity, researchers showed the risk of diabetes incidence in Western Australia varied with latitude (Ball *et al.*, 2014). Similar results were reported in Finland as the incidence rate of diabetes was higher in rural areas as compared to urban areas (Voutilainen *et al.*, 2015).

Counties vulnerable to diabetes in the Southeast, Central and South regions were also researched in previous studies. In particular, southern Texas counties were investigated to have higher rates of obesity and diabetes than the rest of the state and the nation, with nearly one-third of population classified as obese and approximately one in nine were diagnosed with diabetes (Ramirez *et al.*, 2008). In Nashville, Tennessee, the geographic distribution of obesity, diabetes, health behaviour, and environmental characteristics was clustered and identi-

fied high vulnerability to diabetes and obesity in terms of census tract (Schlundt *et al.*, 2006). In Ohio and South Carolina, evidence of the presence of a local variation in Type 1 and Type 2 diabetes mellitus incidence was reported, which is important for future surveillance efforts for diabetes (Liese *et al.*, 2010). Our study is consistent with previous results, and we provide solid statistical evidence, in terms of spatial heterogeneity and vulnerability of diabetes, to strengthen our results.

The CDC defined the diabetes belt as a geographic region consisting of 644 counties in 15 southern states with an estimated prevalence of diagnosed diabetes greater than 11% (Barker *et al.*, 2011), and some counties in the diabetes belt appear again in the definite and higher vulnerable levels defined by this study. Specifically, our findings reveal that counties in these two vulnerable levels had higher levels of suffer-

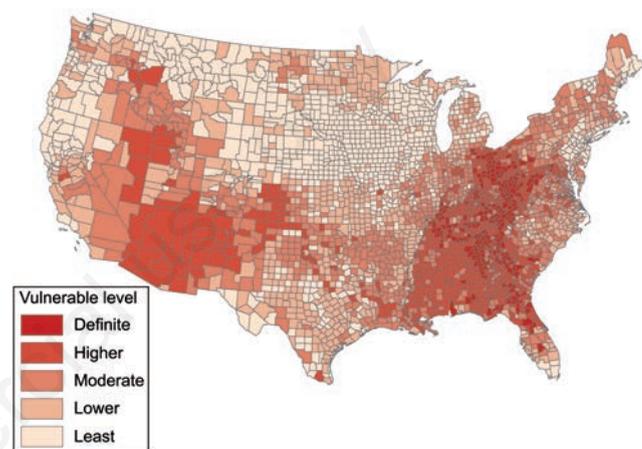


Figure 4. Geographic distribution of five vulnerable levels to diabetes in the US.

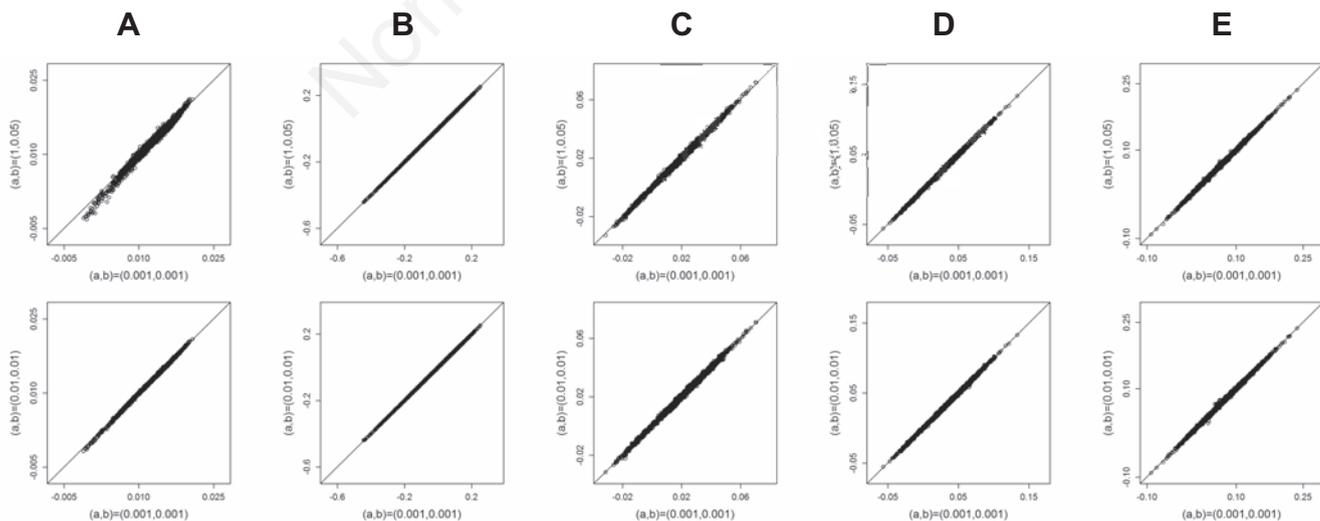


Figure 5. Scatter plots of spatial function estimates with different hyper-parameters (a, b) in the prior of spatial variance. A) Estimate of f_{spat} in Model 1; B-E) estimates of $(f_{spat-L}, f_{spat-ML}, f_{spat-MH}, f_{spat-H})$ in Model 2. The upper plots show (a, b) = (0.001, 0.001) vs (a, b) = (1, 0.05); the lower plots show (a, b) = (0.001, 0.001) vs (a, b) = (0.01, 0.01).

ing due to high diabetes prevalence linked to obesity than the other counties. This differs from the CDC report because people living in the diabetes belt had a lower odds ratio of obesity than those living in the rest of the US. Thus, we calculated the prevalence of predicted diabetes based on our models and proved that the five vulnerable levels can be distinguished from each other. Moreover, we also presented increasing trends of predicted diabetes prevalence in five vulnerable levels and concluded that counties in definite vulnerable level had the largest increment annually. Therefore, we believe that our findings are compelling as we present a solid analysis with a longer study period. More importantly, a cluster of counties with a higher vulnerable level was detected in Southwest region, which had not been discussed previously. There is a need to do further investigations in that area. Comprehensively speaking, we suggest that the diabetes belt should be reconsidered, in terms of obesity, and some counties, which are not located in the diabetes belt, should be also under surveillance.

A new finding of this study is the geographic expansion of counties vulnerable to diabetes when obesity prevalence increased, especially in Northern US (Figure 3B and D). Eid (2011) addressed the prevalence of obesity and diabetes mellitus in South Dakota, and elaborated on some of the mechanisms of association between obesity and diabetes mellitus. In a southeastern Wisconsin population, excess weight gain during childhood was a risk factor for early manifestation of Type 1 diabetes mellitus (Evertsen *et al.*, 2009). In Pennsylvania, researchers found the burden of obesity and diabetes is extensive and growing (Garcia-Dominic *et al.*, 2014). The reason of the expansion of diabetes prevalence due to high obesity prevalence in Northern US is still not well understood from previous studies, so it is necessary to conduct further work in counties vulnerable to diabetes, especially in West North Central and East North Central regions.

Some limitations of this study need to be considered when interpreting these findings. First, the data from BRFSS does not include people younger than 18; thus, the spatial obesity impact on diabetes may be not accurate due to missing childhood obesity information. Second, our quantified findings can be only explained in terms of county, while other geographic units, such as ZIP code or census tract, may produce findings. Third, formal medical records, such as clinic visits and hospitalizations, were not considered in the self-reported survey. Lastly, the questionnaire in BRFSS does not specify Type 1, Type 2 and gestational diabetes, so the impacts on obesity cannot be differentiated.

Conclusions

This study determined a significant variation of spatial pattern for diabetes in terms of the geographic variation of obesity prevalence, and identified geographical clusters of diabetes prevalence in terms of four quartiles of obesity prevalence. Counties in Central, South and Southeast regions are more likely to be vulnerable to diabetes, even with a low prevalence of obesity. In addition, as obesity prevalence increased to higher levels, the geographic distribution of counties vulnerable to diabetes tended to expand to the Northern US regions. This study highlighted the importance of surveillance efforts for diabetes with small area estimates. Future research should focus on development of interventions and prevention methods in those areas where people are vulnerable to diabetes.

References

- Ball SJ, Haynes A, Jacoby P, Pereira G, Miller LJ, Bower C, Davis EA, 2014. Spatial and temporal variation in type 1 diabetes incidence in Western Australia from 1991 to 2010: increased risk at higher latitudes and over time. *Health Place* 28:194-204.
- Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ, 2011. Geographic distribution of diagnosed diabetes in the US: a diabetes belt. *Am J Prev Med* 40:434-9.
- Blanck HM DW, Galuska DA, Gillespie C, Hamre R, Khan K, Serdula MK, Ford ES, Garvin WS, Mokdad AH, Densmore D, 2006. State-specific prevalence of obesity among adults - United States, 2005. *MMWR Morb Mortal Wkly Rep* 55:985-8.
- Brezger A, Kneib T, Lang S, 2005. BayesX: analysing Bayesian structured additive regression models. *J Stat Softw* 14:1-22.
- Brezger A, Lang S, 2006. Generalized structured additive regression based on Bayesian P-splines. *Comput Stat Data An* 50:967-91.
- CDC, 2011. CDC identifies diabetes belt. Available from: <http://www.cdc.gov/diabetes/pdfs/data/diabetesbelt.pdf>
- CDC, 2014a. Crude and age-adjusted rate per 100 of civilian, noninstitutionalized population with diagnosed diabetes, United States, 1980-2011. Available from: <http://www.cdc.gov/diabetes/statistics/prev/national/figage.htm>
- CDC, 2014b. Diabetes home. Available from: <http://www.cdc.gov/diabetes/data/national.html>
- CDC, 2014c. Overweight and obesity. Available from: <http://www.cdc.gov/obesity/index.html>
- Chien LC, Alamgir H, Yu HL, 2015. Spatial vulnerability of fine particulate matter relative to the prevalence of diabetes in the United States. *Sci Total Environ* 508:136-44.
- Congdon P, 2010. A multilevel model for comorbid outcomes: obesity and diabetes in the US. *Int J Environ Res Public Health* 7:333-52.
- Congdon P, Lloyd P, 2010. Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. *J Data Sci* 8:235-52.
- Dansinger M, 2014. What increases my risk of diabetes? Available from: <http://www.webmd.com/diabetes/risk-factors-for-diabetes>
- Eid WE, 2011. Obesity and type 2 diabetes mellitus in South Dakota: focused insight into prevalence, physiology and treatment. *S D Med Spec No*:68-73.
- Evertsen J, Alemzadeh R, Wang X, 2009. Increasing incidence of pediatric type 1 diabetes mellitus in Southeastern Wisconsin: relationship with body weight at diagnosis. *PLoS ONE* 4:e6873.
- Fahrmeir L, Lang S, 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. *J Roy Stat Soc C-App* 50:201-20.
- Foulds HJ, Bredin SS, Warburton DE, 2012. An evaluation of the physical activity and health status of British Columbian aboriginal populations. *Appl Physiol Nutr Metab* 37:127-37.
- Garcia-Dominic O, Lengerich EJ, Camacho F, Gallant NR, Wray LA, Ahern F, Bogdan G, Weinberg G, Ulbrecht JS, 2014. Prevalence of diabetes and associated obesity in Pennsylvania adults, 1995-2010. *Prev Chronic Dis* 11:E111.
- Gregg EW, Cadwell BL, Burrows NR, Barker LE, Thompson TJ, Geiss L, Pan L, 2009. Estimated county-level prevalence of diabetes and obesity - United States, 2007. *MMWR Morb Mortal Wkly Rep* 58:1259-63.
- Hipp JA, Chalise N, 2015. Spatial analysis and correlates of county-level diabetes prevalence, 2009-2010. *Prev Chronic Dis* 12:E08.
- Katulanda P, Jayawardena R, Ranasinghe P, Rezvi Sheriff MH, Matthews DR, 2013. Physical activity patterns and correlates



- among adults from a developing country: the Sri Lanka diabetes and cardiovascular study. *Public Health Nutr* 16:1684-92.
- Kindermann R, Snell JL, 1980. Markov random fields and their applications. *Contemporary mathematics*. Vol 1. American Mathematical Society, Providence, RI, USA.
- Lang S, Brezger A, 2004. Bayesian P-splines. *J Comput Graph Stat* 13:183-212.
- Laraia BA, Blanchard SD, Karter AJ, Jones-Smith JC, Warton M, Kersten E, Jerrett M, Moffet HH, Adler N, Schillinger D, Kelly M, 2014. Spatial pattern of body mass index among adults in the diabetes study of Northern California (DISTANCE). *Int J Health Geogr* 13:48.
- Liese AD, Lawson A, Song H-R, Hibbert JD, Porter DE, Nichols M, Lamichhane AP, Dabelea D, Mayer-Davis EJ, Standiford D, Liu L, Hamman RF, D'Agostino RB Jr, 2010. Evaluating geographic variation in type 1 and type 2 diabetes mellitus incidence in youth in four US regions. *Health Place* 16:547-56.
- Liese AD, Puett RC, Lamichhane AP, Nichols MD, Dabelea D, Lawson AB, Porter DE, Hibbert JD, D'Agostino RB Jr, Mayer-Davis EJ, 2012. Neighborhood level risk factors for type 1 diabetes in youth: the SEARCH case-control study. *Int J Health Geogr* 11:1.
- Maty SC, Everson-Rose SA, Haan MN, Raghunathan TE, Kaplan GA, 2005. Education, income, occupation, and the 34-year incidence (1965-99) of Type 2 diabetes in the Alameda County Study. *Int J Epidemiol* 34:1274-81.
- Melidonis AM, Tournis SM, Kompoti MK, Lentzas IL, Roussou VR, Iraklianos SL, Michail IM, Mariolis AM, 2006. Increased prevalence of diabetes mellitus in a rural Greek population. *Rural Remote Health* 6:534.
- Mohan V, Mathur P, Deepa R, Deepa M, Shukla DK, Menon GR, Anand K, Desai NG, Joshi PP, Mahanta J, Thankappan KR, Shah B, 2008. Urban rural differences in prevalence of self-reported diabetes in India – the WHO-ICMR Indian NCD risk factor surveillance. *Diabetes Res Clin Pr* 80:159-68.
- Ning F, Pang ZC, Dong YH, Gao WG, Nan HR, Wang SJ, Zhang L, Ren J, Tuomilehto J, Hammar N, Malmberg K, Andersson SW, Qiao Q, Qingdao Diabetes Survey Group, 2009. Risk factors associated with the dramatic increase in the prevalence of diabetes in the adult Chinese population in Qingdao, China. *Diabetic Med* 26:855-63.
- Pearson JF, Bachireddy C, Shyamprasad S, Goldfine AB, Brownstein JS, 2010. Association between fine particulate matter and diabetes prevalence in the US. *Diabetes Care* 33:2196-201.
- Ramirez AG, Thompson IM, Vela L, 2008. *The South Texas health status review*. Springer, Amsterdam, The Netherlands.
- Schlundt DG, Hargreaves MK, McClellan L, 2006. Geographic clustering of obesity, diabetes, and hypertension in Nashville, Tennessee. *J Ambul Care Manage* 29:125-32.
- Sullivan PW, Morrato EH, Ghushchyan V, Wyatt HR, Hill JO, 2005. Obesity, inactivity, and the prevalence of diabetes and diabetes-related cardiovascular comorbidities in the US, 2000-2002. *Diabetes Care* 28:1599-603.
- US Census Bureau, 2014. American community survey information guide. Available from: <https://www.census.gov/programs-surveys/acs/about/information-guide.html>
- Voutilainen A, Hartikainen S, Sherwood PR, Taipale H, Tolppanen AM, Vehvilainen-Julkunen K, 2015. Associations across spatial patterns of disease incidences, socio-demographics, and land use in Finland 1991-2010. *Scand J Public Health* 43:356-63.
- Zheutlin AR, Adar SD, Park SK, 2014. Carbon dioxide emissions and change in prevalence of obesity and diabetes in the United States: an ecological study. *Environ Int* 73:111-6.