Spatial Clusters of County-Level Diagnosed Diabetes and Associated Risk Factors in the United States

Sundar S. Shrestha[*], Karen A. Kirtland, Theodore J. Thompson, Lawrence Barker, Edward W. Gregg and Linda Geiss

Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract: *Introduction*: We examined whether spatial clusters of county-level diagnosed diabetes prevalence exist in the United States and whether socioeconomic and diabetes risk factors were associated with these clusters.

Materials and Methods: We used estimated county-level age-adjusted data on diagnosed diabetes prevalence for adults in 3109 counties in the United States (2007 data). We identified four types of diabetes clusters based on spatial autocorrelations: high-prevalence counties with high-prevalence neighbors (High-High), low-prevalence counties with low-prevalence neighbors (Low-Low), low-prevalence counties with high-prevalence neighbors (Low-High), and high-prevalence counties with low-prevalence neighbors (High-Low). We then estimated relative risks for clusters being associated with several socioeconomic and diabetesrisk factors.

Results: Diabetes prevalence in 1551 counties was spatially associated (p<0.05) with prevalence in neighboring counties. The rate of obesity, physical inactivity, poverty, and the proportion of non-Hispanic blacks were associated with a county being in a High-High cluster versus being a non-cluster county (7% to 36% greater risk) or in a Low-Low cluster (13% to 67% greater risk). The percentage of non-Hispanic blacks was associated with a 7% greater risk for being in a Low-High cluster. The rate of physical inactivity and the percentage of Hispanics or non-Hispanic American Indians were associated with being in a High-Low cluster (5% to 21% greater risk).

Discussion: Distinct spatial clusters of diabetes prevalence exist in the United States. Strong association between diabetes clusters and socioeconomic and other diabetes risk factors suggests that interventions might be tailored according to the prevalence of modifiable factors in specific counties.

Keywords: County-level, diabetes cluster, diabetes risk factors, geographic cluster, socioeconomic factors, spatial analysis.

INTRODUCTION

Diabetes contributes to blindness, renal disease, amputation, cardiovascular disease, and disability. Therefore, it poses substantial health impact and economic costs for the United States. In 2010, 8.3% (25.8 million) of the U.S. population had diabetes, diagnosed or not [1]. In 2008, 6.1 % of the non-institutionalized U.S. civilian population had diagnosed diabetes, more than twice the prevalence of diagnosed diabetes in 1980 [2]. Along with the increased prevalence of diabetes, the economic cost of diabetes in the United States has increased from $98.2 billion in 1997 to $174 billion in 2007 [3, 4].

Diabetes incidence and prevalence in the United States is well understood at the state level [5, 6]. The prevalence of diabetes also varies at smaller geographical scales, such as the county (usually the smallest unit of local government in the United States) or census tract levels [7-9]. Accordingly, national- and state-level estimates of diabetes prevalence could conceal local pockets of high or low prevalence, and

therefore make it difficult for policymakers to efficiently and effectively plan diabetes prevention or control interventions [7, 10]. County-level estimates of the prevalence of diagnosed diabetes among adults exist, both raw and age-adjusted to the 2000 standardized population [11]. These estimates provide the best available basis for understanding county-level spatial patterns of diabetes prevalence and inequalities in diabetes burden.

Recent advances in spatial analytical methods, such as exploratory spatial data analysis implemented in GeoDa Software [12, 13], provide the capability to uncover spatial clusters. These methods have been used for obesity, physical activity, cancer, the pattern for prescribing cardiovascular drugs, and sexually transmitted diseases [14-19]. However, few studies of diabetes spatial clustering have been published [8, 9, 20]. We are aware of no studies on spatial clustering of diabetes prevalence and its correlates using U.S. county-level data, except Barker *et al*. [21]. Barker *et al*. identified a U.S. diabetes belt—a regional pattern of counties with high prevalence (11.0% or more) of diagnosed diabetes. While that paper studied differences in risk factors between the diabetes belt and the rest of the United States, it examined neither negatively associated spatial clusters nor smaller groupings of high-prevalence counties outside the larger belt.

Address correspondence to this author at the U.S. Centers for Disease Control and Prevention, Division of Diabetes Translation, 4770 Buford Highway, MS K-10, Atlanta, GA 30341, USA; Tel: (770) 488-1148; Fax: (770) 488-5966; E-mail: SShrestha@cdc.gov

In this study, we identified county-level diabetes spatial clusters based on spatial autocorrelations of county-level diabetes prevalence. We then estimated the extent to which identified diabetes spatial clusters were associated with selected socioeconomic and diabetes risk factors, aggregated at the county level. Identifying spatial patterns of diabetes prevalence clusters could help us further understand potential factors governing the spatial heterogeneity in the distribution of populations at high or low risk for diabetes [22]. Knowing this would aid in translating factors governing a low prevalence of diabetes from low-prevalence clusters to high-prevalence clusters.

MATERIALS AND METHODS

Data Source and Variables

The outcome variable was the diabetes spatial clusters we identified (described below) using 2007 data on age-adjusted prevalence of diagnosed diabetes among adults (aged ≥20 years) from the 3109 counties in the 48 contiguous states and the District of Columbia [11]. The 2007 county-level estimates of diabetes prevalence among adults were derived by applying small area estimation techniques to self-reported Behavioral Risk Factor Surveillance System (BRFSS) data for 2006, 2007, and 2008, and U.S. Census data for 3141 counties in the United States (this number includes noncontiguous U.S. counties). BRFSS determined persons with diagnosed diabetes by asking, "Has a doctor ever told you that you have diabetes?" Women with gestational diabetes were excluded. Data from the 2000 U.S. Census were used to age-adjust the results by population [10]. The prevalence estimates included diagnosed diabetes (type 1 or type 2), but do not distinguish diabetes by type. We did not include counties in Alaska and Hawaii because the significance level for these counties would not be computed correctly when using contiguity-based spatial weights [http://geodacenter.asu.edu/node/402#lisaisle, Accessed (August 2, 2011)].

To assess the factors associated with diabetes spatial clusters, we considered the following county-level socioeconomic and risk factors: the estimated percentage of people of all ages living in poverty (2007 data) [23]; the percentage of residents of different races/ethnicities: non-Hispanic white, non-Hispanic black, Hispanic, non-Hispanic American Indian, non-Hispanic Asian, and other races/ethnicities (2007 data) [24]; the age-adjusted estimates of the percentage of obese adults (i.e., with a body mass index [BMI] of ≥30 kg/m^2 based on self-reported height and weight) (2007 data) [11]; the age-adjusted self-reported percentage of adults who engaged in no leisure time physical activity (2007 data) [11]; the percentage of county population (≥20 years) that was female (2007 data) [25]; whether the county was in a metropolitan statistical area (MSA), based on 2003 MSA classification (hereafter referred to as the county's "metropolitan status") [26]; and percentage of adults (aged 18-64 years) in all income groups in the county who were uninsured (2007 data) [27].

ANALYSIS

Exploratory Spatial Data Analysis

We mapped age-adjusted estimates of diabetes prevalence, then visually identified clusters, and computed global and local Moran I statistics, which allowed us to test for overall spatial autocorrelation in diabetes prevalence and identify statistically significant diabetes clusters and outliers on the U.S. county-level map.

Estimating the Global Spatial Association

To evaluate overall spatial association in diabetes prevalence across the counties, we computed the univariate global Moran's I statistic, as:

$$I = \left(\frac{1}{\sum_i \sum_j w_{ij}} \right) \left(\frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2 / n} \right)$$

where x_i is the age-adjusted diabetes prevalence of the ith county; x_j is the age-adjusted diabetes prevalence of the jth county; $\bar{x}$ is the county-level mean diabetes prevalence; n is the number of counties; and w_{ij} is the first-order queen contiguity spatial weight matrix used to define the immediately spatially contiguous neighbors, in which county j is a neighbor of county i if counties i and j share a boundary.

We standardized the weights so that rows summed to 1 [13]. Thus, the Moran's I statistic takes values between -1 and 1 and can be interpreted much like a Pearson's correlation. To test the statistical significance of Moran's I statistic, a p-value was calculated by comparing the observed statistic to a reference distribution generated by 999 randomized permutations of the observed data. The null hypothesis was that age-adjusted county-level diabetes prevalence was spatially independent. Rejection of hypothesis suggests the existence of a spatial association between diabetes prevalence among neighboring counties, indicating the existence of spatial clusters. However, it does not indicate where clusters are located nor does it indicate the types of spatial clusters [12].

Identifying Diabetes Spatial Clusters

Diabetes spatial clusters were identified based on a significant spatial autocorrelation between the diabetes prevalence of a county and its immediate neighboring counties. We computed the local spatial autocorrelation using local Moran's I statistics, also referred to as Local Indicators of Spatial Association (LISA), using the formula

$$I_i = \left(\frac{(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2 / n} \right)$$

in which I_i is the Moran's I statistics of the ith county and other variables in the equation are as defined previously. Statistical significance was tested using the same approach as used for global Moran's I. Based on LISA, U.S. counties were grouped into four types of statistically significant diabetes spatial cluster counties and one group of statistically insignificant counties (i.e., non-cluster counties) [28]. The identified diabetes spatial cluster counties were: 1) *High-High* spatial clusters: counties with a high diabetes prevalence (county-level mean of >8.6%) that were positively spatially correlated with high diabetes prevalence neighbors; 2) *Low-Low* spatial clusters: counties with low diabetes prevalence (county-level mean of <8.6%) that were

positively spatially correlated with low-prevalence neighbors; 3) *Low-High* spatial clusters: counties with low diabetes prevalence that were negatively spatially correlated with high-prevalence neighbors; and 4) *High-Low* spatial clusters: counties with high diabetes prevalence that were negatively spatially correlated with low-prevalence neighbors; and 5) Non-cluster counties, in which the diabetes prevalence of a countywas not statistically correlated with the diabetes prevalence of its neighbors. We used GeoDa 0.95 software [13] to estimate global and local Moran's I statistics and spatially dependent diabetes prevalence patterns.

Associating Clusters with Correlates

Following the identification of diabetes spatial clusters, we described the characteristics of counties in clusters and non-cluster counties by computing the means and standard errors of variables (See Table **1**). We compared means:

i.　Between High-High and Low-Low cluster counties to assess the extent to which the risk factor prevalence of these two types of clusters with concordant counties (i.e., in which neighboring counties all have the same level of diabetes prevalence) was different.

ii.　Between High-High and High-Low cluster counties and between Low-Low and Low-High cluster counties to assess whether the mean risk factor prevalence of counties differed from that of the same type of county when their neighbors had a different

prevalence.

iii.　Between counties in High-High clusters and high-prevalence non-cluster counties and between counties in Low-Low clusters and low-prevalence non-cluster counties to assess whether or not the risk factor prevalence of a county differed from that of the same type of county depending on whether or not that county's diabetes prevalence was correlated with its neighboring counties.

We used a multinomial logistic regression model to examine the association between the four types of clusters identified and the socioeconomic and diabetes risk factors, adjusting for gender, health insurance coverage, and metropolitan status of the counties. We presented estimated results as the relative risk for a county being in a cluster versus being a non-cluster county. We also estimated the relative risk for a county being in a High-High cluster relative to being in a Low-Low cluster. This enabled us to compare results between these two concordant clusters, but represent substantially unequal diabetes prevalence. We used STATA v.11 for these analyses.

RESULTS

Diabetes Prevalence Map

The quintiles map of diabetes prevalence (Fig. **1**) shows that counties in the top quintile (diabetes prevalence ≥10.5%) were concentrated in the South—most of the counties of Alabama, Georgia, Mississippi, and considerable proportions
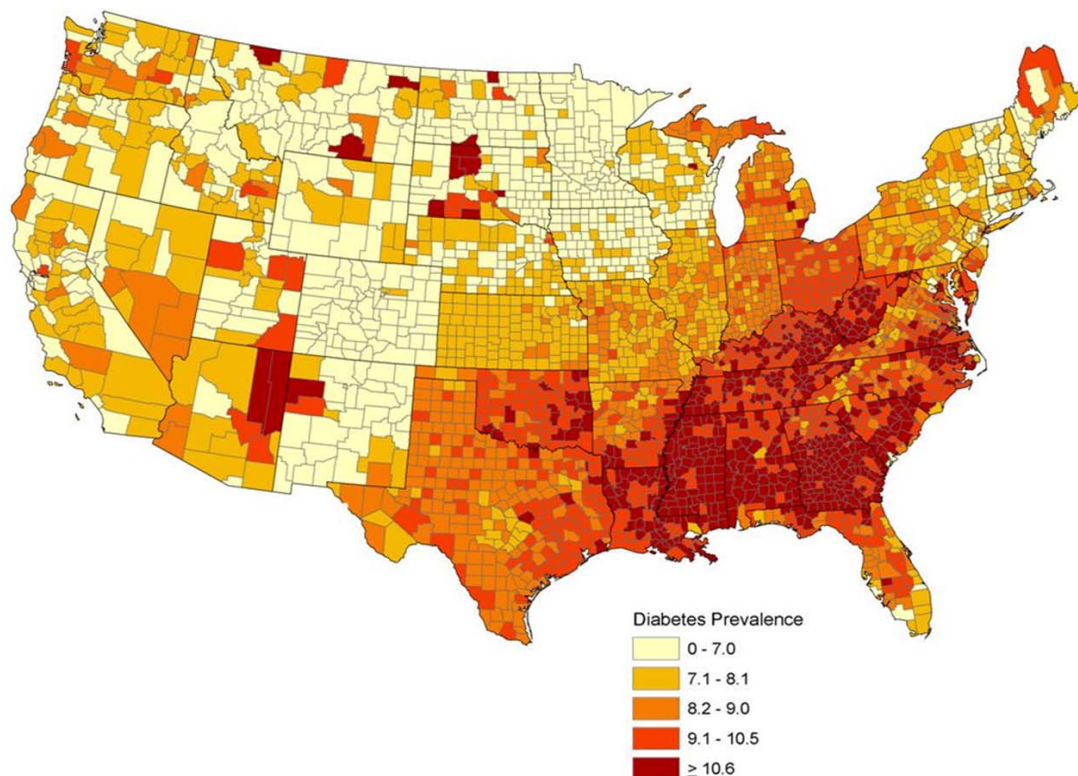

Fig. (1). Quintiles of age-adjusted county-level diagnosed diabetes prevalence (%), 2007.

of the counties in Louisiana, Kentucky, South Carolina, and West Virginia. Counties in the bottom quintile (diabetes prevalence <7.1%) were concentrated in Colorado, Iowa, Minnesota, Montana, New Mexico, North Dakota, South Dakota, Wisconsin, and Wyoming.

Diabetes Spatial Clusters

The formal test of global spatial associations showed that county-level age-adjusted diabetes prevalence was spatially correlated (Moran's I=0.78, p<0.05). This provided evidence of the spatial clustering of like diabetes prevalence. Of the total 3109 counties, 717 (23.1%) were in High-High clusters; 786 (25.3%) were in Low-Low clusters; 22 (0.7%) were in High-Low clusters and 26 (0.8%) were in Low-High clusters (Fig. **2**). The remaining 1558 (50.1%) were non-cluster counties.

The cluster counties were summarized for this report by identifying the U.S. census regions (Northeast, Midwest, South, and West) in which they are located. High-High cluster counties were located mainly in the South and some in the Midwest (Fig. **2**). Low-Low cluster counties were located mainly in the West, with some in the Midwest and the Northeast (Fig. **2**). Low-High cluster counties were scattered throughout the South, the Northeast, and the West (Fig. **2**). More than half (55%) of Low-High cluster counties were located in metropolitan statistical areas [26]. High-Low cluster counties were scattered around the West, the Midwest, and the South (Fig. **2**). The majority of High-Low clusters included all or part of areas designated as American Indian reservations.

Descriptive Results

The mean diabetes prevalence in non-cluster counties was lower than the average in all counties (p<0.05). However, this difference was very small (0.1%) (Table **1**). In the High-High cluster counties, the mean diabetes prevalence was significantly higher than in the Low-Low cluster counties. While the High-Low cluster counties had significantly lower mean diabetes prevalence than High-High cluster counties the Low-High cluster counties had significantly higher mean diabetes prevalence than Low-Low cluster counties. The mean diabetes prevalence in a county in a High-High or Low-Low cluster and those in their corresponding cluster neighbors were similar. However, the mean diabetes prevalence rate in the Low-High cluster counties was lower than in their in neighboring counties and in the High-Low cluster counties, it was higher than in the neighbor counties. Low-High or High-Low cluster counties were classified based on rates of diabetes prevalence that were obviously significantly different from those of their neighbors, so we did not test for statistical significance.

The mean poverty rate was significantly greater for counties in High-High clusters than for those in Low-Low clusters (Table **1**). The percentage of non-Hispanic whites was significantly higher for counties in Low-Low clusters than for those in High-High clusters (Table **1**). In contrast, the percentage of non-Hispanic blacks was significantly higher for counties in High-High clusters than for those in Low-Low clusters. The percentages of Hispanics, non-Hispanic Asians, and non-Hispanic American Indians were significantly higher counties in Low-Low clusters than for
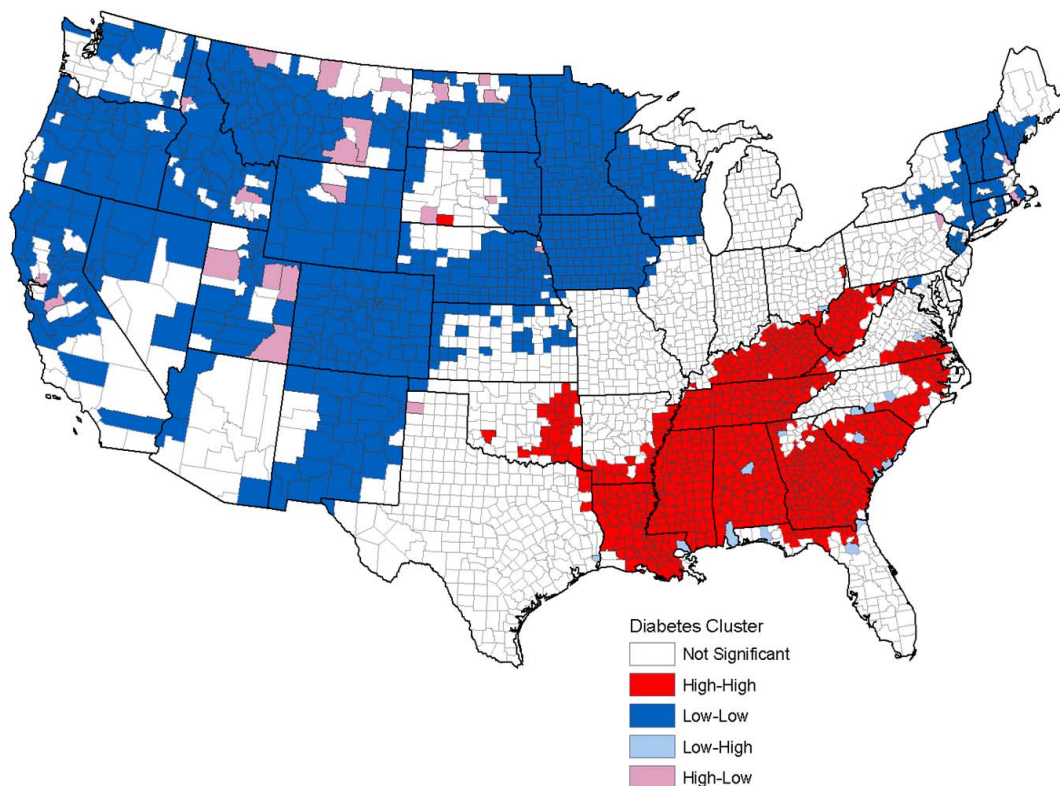

Fig. (2). Spatial clusters of age-adjusted county-level diagnosed diabetes prevalence, 2007. Note: White colored counties were spatially not correlated with neighboring counties at a 5% level.

Table 1.   Characteristics of U.S. Counties: By Diabetes Spatial Clusters and Non-Cluster Type, and in Overall, 2007

Variables	Cluster Counties (n=1551)				[a]Non-Cluster Counties (n=1558)			All
	HH **(n=717)**	**HL** **(n=26)**	**LH** **(n=22)**	**LL** **(n=786)**	**High Prev.** **(n=636)**	**Low Prev.** **(n=922)**	**All** **(n=1558)**	**Counties** **(n=3109)**
Age-adjusted diabetes (%)	11.16[H]	10.21[H]	8.02[L]	6.66[L]	9.65	7.86	8.59	8.71
	(0.04)	(0.37)	(0.20)	(0.03)	(0.03)	(0.21)	(0.03)	(0.03)
Age-adjusted diabetes of neighbor (%)	11.11[H]	6.93[H]	8.68[L]	6.70[L]	9.16	8.27	8.63	8.70
	(0.03)	(0.07)	(0.90)	(0.02)	(0.03)	(0.20)	(0.02)	(0.03)
Age-adjusted obesity (%)	31.65[H]	30.12[H]	25.14	25.71	29.05	27.19	27.95	28.24
	(0.11)	(0.89)	(0.72)	(0.13)	(0.84)	(0.09)	(0.07)	(0.07)
Age-adjusted physical inactivity (%)	31.2[H]	27.52[H]	22.87	21.72	27.29	24.48	25.62	25.92
	(0.14)	(0.63)	(0.83)	(0.14)	(0.12)	(0.11)	(0.09)	(0.09)
Poverty: All ages (%)	20.44	20.83	11.90	11.93	15.82	13.13	14.23	15.12
	(0.24)	(2.28)	(1.04)	(0.15)	(0.23)	(0.16)	(0.14)	(0.11)
Race/Ethnicity (%):								
NH_White	73.17[H]	59.38[H]	82.92[L]	89.27[L]	77.30	85.98	82.43	81.83
	(0.72)	(5.21)	(1.91)	(0.46)	(0.73)	(0.51)	(0.44)	(0.32)
NH_Black	22.27[H]	2.92[H]	10.13[L]	1.02[L]	10.01	3.19	5.97	8.48
	(0.72)	(1.32)	(1.54)	(0.07)	(0.45)	(0.15)	(0.22)	(0.24)
Hispanic	2.35[H]	8.51[H]	4.20	6.65	8.66	7.96	8.24	6.45
	(0.09)	(2.18)	(0.69)	(0.40)	(0.56)	(0.45)	(0.35)	(0.21)
NH_American Indian	0.94[H]	26.40[H]	0.49	1.25	2.15	0.78	1.34	1.43
	(0.13)	(5.66)	(0.10)	(0.10)	(0.33)	(0.05)	(0.14)	(0.10)
NH_Asian/Pacific islander	0.53[H]	1.43[H]	1.45	1.02	0.97	1.27	1.14	0.98
	(0.02)	(0.58)	(0.35)	(0.07)	(0.06)	(0.08)	(0.05)	(0.03)
Other races/ethnicity	0.74[H]	1.35[H]	0.80	0.79	0.91	0.83	0.87	0.82
	(0.03)	(0.13)	(0.07)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)
Female (%)	51.52	50.38	51.47[L]	50.22[L]	50.74	50.97	50.88	50.85
	(0.11)	(0.68)	(0.21)	(0.08)	(0.13)	(0.06)	(0.06)	(0.05)
Without health insurance (%)	21.07	19.72	20.93[L]	18.25[L]	21.85	19.89	20.69	20.16
	(0.17)	(1.09)	(0.99)	(0.21)	(0.29)	(0.22)	(0.18)	(0.11)
Metro (yes=1)	0.34	0.27	0.55[L]	0.25[L]	0.42	0.39	0.40	0.35
	(0.02)	(0.09)	(0.11)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)

Notes:
HH= High-High cluster counties; HL=High-Low cluster counties; LH=Low-High cluster counties; LL=Low-Low cluster counties.
[a]Non-cluster counties are those with non-significant spatial association with their neighbors.
NH= Non-Hispanic. Other races/ethnicity=NH-Native American and Hawaiian and those with ≥2 ethnic backgrounds.
Metro=located in metropolitan statistical area based on 2003 criteria.
All the means between High-High and Low-Low clusters; between High-High cluster and high-prevalence non-cluster counties; and between Low-Low cluster and low-prevalence non-cluster counties are statistically different (p<0.05).
[H]= means between High-High and High-Low clusters are statistically different (p<0.05).
[L]= means between Low-Low and Low-High clusters are statistically different (p<0.05).
Values in parentheses are standard errors.

those in High-High clusters. A high percentage of Hispanics and non-Hispanic American Indians were in High-Low cluster counties; a low percentage of non-Hispanic Asians were in Low-High or High-Low cluster counties (Table **1**).

Both the age-adjusted obesity prevalence and age-adjusted mean percentage of the population that was physically inactive were significantly higher in High-High cluster counties than in Low-Low cluster counties (Table **1**).

Table 2. **Relative Risks for Counties Being Located in a Specific Type of Diabetes Spatial Clusters Relative to Being a "Non-Cluster" County and Being Located in High-High Versus Low-Low Cluster**

Variables	Diabetes Cluster versus [a]Non-Cluster Counties				[c]HH versus LL
	[b]HH	[b]HL	[b]LH	[b]LL	
1	2	3	4	5	6
Age-adjusted obesity (%)	1.22*** (1.12–1.33)	0.96 (0.78–1.18)	0.71*** (0.58–0.86)	0.85*** (0.80–0.89)	1.45*** (1.31–1.59)
Age-adjusted physically inactivity (%)	1.36*** (1.29–1.43)	1.21* (1.03–1.43)	0.96 (0.80–1.15)	0.82*** (0.78–0.85)	1.67*** (1.56–1.78)
Poverty: All ages (%)	1.09*** (1.05–1.12)	0.86* (0.77–0.97)	0.94 (0.82–1.07)	0.96* (0.93–0.99)	1.13*** (1.08–1.19)
Race/Ethnicity (%):					
NH_Black	1.07*** (1.05–1.08)	0.94 (0.84–1.06)	1.07** (1.02–1.12)	0.67*** (0.62–0.72)	1.59*** (1.48–1.71)
Hispanic	0.82*** (0.78–0.86)	1.05* (1.01–1.10)	0.86* (0.76–0.98)	1.01 (0.99–1.02)	0.82*** (0.78–0.86)
NH_American Indian	0.94*** (0.91–0.97)	1.14*** (1.09–1.19)	0.82 (0.42–1.61)	1.03 (1.00–1.06)	0.91*** (0.87–0.95)
NH_Asian/Pacific Islander	1.02 (0.84–1.23)	1.13 (0.96–1.32)	0.88 (0.63–1.22)	1.05 (0.97–1.14)	0.97 (0.78–1.19)
Other races/ethnicity	0.91 (0.76–1.08)	1.07 (0.76–1.51)	1.33 (0.33–5.46)	0.67*** (0.52–0.85)	1.36* (1.01–1.83)
Female (%)	1.07* (1.01–1.12)	0.83* (0.71–0.98)	1.12 (0.85–1.46)	0.76*** (0.72–0.82)	1.40*** (1.29–1.52)
Without health insurance (%)	1.09*** (1.06–1.13)	0.92 (0.84–1.01)	1.09 (0.99–1.19)	0.93*** (0.91–0.96)	1.17*** (1.13–1.22)
Metro (yes=1)	2.21*** (1.59–3.06)	0.98 (0.29–3.32)	1.00 (0.34–2.91)	0.45*** (0.33–0.60)	4.94*** (3.19–7.63)
LR Chi^2	1827***				1827***

Notes:
HH= High-High cluster counties; HL=High-Low cluster counties; LH=Low-High cluster counties; LL=Low-Low cluster counties.
[a]Non-cluster counties are those with non-significant spatial association with their neighbors.
NH= Non-Hispanic. Other races/ethnicities=NH-Native American and Hawaiian and that with ≥2 ethnic background. Metro=metropolitan statistical area based on 2003 criteria.
[b]Estimates are risk ratios relative to "non-cluster" counties.
[c]Risk ratios for High-High cluster counties relative to Low-Low cluster counties.
Values in parentheses are 95% confidence intervals.
*=p<0.05, **=p<0.01, ***=p<0.001.

Multivariate Results

Table **2** presents the multinomial logistic regression-based relative risks for the diabetes spatial clusters associated with socioeconomic and diabetes risk factors. Both the age-adjusted obesity and age-adjusted physical inactivity rates were associated with increased risk (22% to 36%) of the county being a High-High cluster versus being a non-cluster county, or versus being in a Low-Low cluster (45% to 67% higher). In contrast, these factors were associated with decreased risk (15% to 18%) of a county being in a Low-Low cluster compared with being a non-cluster county. While the age-adjusted obesity rate was associated with 29% lower risk of a county being in a Low-

High cluster, the age-adjusted physical inactivity rate was associated with a 21% higher risk of being in a High-Low cluster compared with being a non-cluster county.

The county-level poverty rate was associated with an average 9% higher relative risk for a county being in a High-High cluster versus being a non-cluster county or a 13% higher relative risk when compared with being in a Low-Low cluster. The relative risk was 4% lower for a county being in a Low-Low cluster versus being a non-cluster county.

Compared with non-Hispanic whites, the percentage of non-Hispanic blacks in a county was associated with a 7% higher risk of that county being in a High-High cluster

versus being a non-cluster county or versus 59% higher risk of being in a Low-Low cluster. However, it was associated with a 33% lower risk of being in a Low-Low cluster, and a 7% higher risk of being in a Low-High cluster, compared with being a non-cluster county. Both the percentage of Hispanics or non-Hispanic American Indians in a county was associated with a lower risk (6% to 18%) for the county being in a High-High cluster versus being a non-cluster county versus being in a Low-Low cluster (9% to 18%). While both of these were associated with increased risk of being in a High-Low cluster, the percentage of Hispanics was associated with a 14% lower risk of being in a Low-High cluster and the percentage of non-Hispanic American Indians was associated with a 14% higher risk of being a High-Low cluster county. Other factors such as a county's percentage of females, the percentage of uninsured adults (aged 18-64 years), and metropolitan status were positively associated with a greater risk of a county being in a High-High cluster versus being in a Low-Low cluster or being a non-cluster county.

DISCUSSION

We found that, based on 2007 data, age-adjusted county-level diabetes prevalence in the United States was spatially associated and that significant diabetes clusters existed, a finding consistent with those in previous studies in Canada [9, 20]. Counties in which diabetes prevalence was higher and had similar neighboring counties were concentrated in the South, to a large extent corresponding with the location of the diabetes belt [21]; counties in which diabetes prevalence was lower and had similar neighboring counties were concentrated in the West, Midwest, and Northeast. About half of all counties in the study were not significantly spatially clustered by diabetes prevalence (non-cluster counties). The average diabetes prevalence in those counties relative to all counties was different, but the difference was negligible in magnitude.

Counties in spatial clusters typically stood out in diabetes prevalence relative to the average for all counties or for non-cluster counties. For instance, counties in High-High clusters were characterized by a higher prevalence of diabetes and most associated factors (e.g., poverty level) than in high-prevalence non-cluster counties. Counties in Low-Low clusters were characterized by lower prevalence of diabetes and associated factors than in low-prevalence non-cluster counties. This highlights the importance of identifying clusters as typical of groups of counties that may need strategies quite different from those for non-cluster counties to address their diabetes burden.

Our results showed that diabetes spatial clusters and inequality in diabetes burden were associated with socioeconomic correlates (poverty and race/ethnicity) and type 2 diabetes risk factors (obesity and physical inactivity), even after accounting for gender proportions, insurance coverage status, and metropolitan status of counties. The positive association between the rate of poverty in a county and the type of diabetes cluster in which it is located (Low-Low versus High-High) was consistent with previous findings based on individual data [20, 29-31]. The association between poverty and diabetes might be causal. Poverty may worsen health outcomes [32] because of limited

access to healthy food, recreational facilities, and other elements contributing to a healthier lifestyle [29]. Poorer health outcomes for populations in socioeconomically deprived areas could also be driven by the limited access to health care, lack of social support, and social disorder [31, 33].

The associated risk for a county being in a cluster was likely confounded by the racial/ethnic disparity between the South and the other regions. We tried to control for census regions, but because the High-High cluster counties were mainly concentrated in the South, the High-High cluster was collinear with that region. A small-area study of diabetes prevalence demonstrated that spatial variation in diabetes exists even after adjusting for the proportion of females and ethnicity [34]. The racial/ethnic differences in diabetes prevalence in a cluster may, in part, be attributable to differences in modifiable socioeconomic and environmental health risk conditions of the places in which they live. One study found no difference in the prevalence of diabetes between African Americans and whites who lived in places with similar socioeconomic and environmental factors [35]. The high rates of obesity, physical inactivity, or both of counties in High-High clusters, compared with those in Low-Low clusters, could also be associated with lower socioeconomic status and relatively limited environmental resources for physical activity, both of which contribute to both diabetes and obesity [16, 31, 36].

Information on risk and socioeconomic factors associated with diabetes prevalence clusters from this study can help policymakers when considering where to implement prevention programs and develop community and clinical services. The study also provided information about risk factors and the socioeconomic status of potential target audiences. For example, High-High spatial clusters identified swaths of the country in which diabetes was common. Interventions to prevent complications from diabetes might be needed in these areas. High-Low clusters serve to remind policymakers that a county may be in need of intervention even when its neighbors' needs are less severe. Low-High clusters identify pockets within regions in which efforts to prevent diabetes have perhaps been successful or have lowered risk factors for diabetes. Low-Low clusters identify broad areas of the country where diabetes is less common; perhaps there are lessons from these areas that can be applied elsewhere such as in counties in High-High or High-Low clusters. All of this knowledge, along with information on the influence of modifiable factors, moves us toward better preventing diabetes and its complications.

Our study has several limitations. First, the county-level diabetes prevalence estimates were modeled from survey data, and we did not account for the survey sampling uncertainty or the biases and limitations of the survey. Second, we did not consider changes over time and therefore do not know how rapidly the diabetes prevalence in our spatial clusters could change. Third, we did not account for the movement of people between counties in our estimates of county-level prevalence of diabetes. It is not known what percentage of the residents developed diabetes while residing in another county. Fourth, we could only account for diagnosed diabetes. It is not known if and how the rates of undiagnosed diabetes spatially vary. Further, because the

county-level diabetes prevalence data were not available by diabetes type but only collectively as "diagnosed diabetes" we could not separately assess the results between type1 and type 2 diabetes. However, as type 2 diabetes accounts for about 95% of diabetes cases [37], the location of clusters and their association with socioeconomic variables and risk factors probably reflects data mostly from people with type 2 diabetes. These locations and associations may be quite different for people with type 1 diabetes. A study of youth aged 10-19 years showed the existence of spatial clustering of type 1 and type 2 diabetes prevalence at the census tract level but provided no evidence of joint clustering of type 1 and type 2 diabetes prevalence [8]. Fifth, risk estimates from the negatively associated clusters (High-Low or Low-High) may be subject to lower statistical power because of the relatively smaller number of counties that fall into those clusters. This should be taken into consideration when interpreting the results of these clusters. Sixth, physical inactivity measures were only for leisure time inactivity; non-leisure time physical activity was not accounted for. The prevalence of non-leisure time physical activity should differ geographically (e.g., would likely be much greater in farming or mining areas). Sixth, our analysis was subject to the limitations of any ecological analysis: county- or cluster-level relationships do not necessarily apply to all individuals [38]. Lastly, we did not carry out census region-specific analyses, which mighthave provided better insight into the relationship between spatial patterns and the factors that underlie them at the regional level. Also, analyses at smaller geographical scales, such as census tracts might have provided better insight into the prevalence of diabetes and its risk factors.

We found distinct spatial clusters of diabetes prevalence in the United States that were associated with socioeconomic correlates (poverty and race/ethnicity) and diabetes risk factors (obesity and physical inactivity). These findings highlight the critical role that these nonclinical factors play in reducing diabetes burden and suggest that interventions aimed at reducing diabetes burden could be tailored to the modifiable socioeconomic and risk factors of specific counties. The methods applied here may be used for both discovering and tracking spatial clusters of diabetes burden across space and time as additional years of county-level data become available. Further, although our study used data only from the United States, the same methods could be applied in other countries to identify unique spatial patterns of diabetes prevalence.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention.

REFERENCES

[1]   Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (Accessed April 5, 2011). Available at: http://www.cdc.gov/diabetes/pubs/factsheet11.htm

[2]   Centers for Disease Control and Prevention. Crude and Age-Adjusted Percentage of Civilian, Noninstitutionalized Population with Diagnosed Diabetes, United States, 1980-2009. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (Accessed September 21, 2011). Available at: www.cdc.gov/diabetes/statistics/prev/national/figage. htm.

[3]   American Diabetes Association. Economic costs of diabetes in the U.S. in 2007. Diabetes Care 2008; 31: 596-15.

[4]   Ray NF, Thamer M, Gardner E, Chan JK, Kahn R, Assoc AD. Economic consequences of diabetes mellitus in the US in 1997. Diabetes Care 1998; 21(2): 296-09.

[5]   Danaei G, Friedman AB, Oza S, Murray CJ, Ezzati M. Diabetes prevalence and diagnosis in US states: analysis of health surveys. Popul Health Metr 2009; 7: 16.

[6]   Barker L, Tierney E, Lanza A, Kirtland K. Selected risk factors' contribution to state-level incidence of diagnosed diabetes, 2005-2007. Open Diabetes J 2011; 4(1): 123-30.

[7]   Gregg EW, Kirtland KA, Cadwell BL, et al. Estimated county-level prevalence of diabetes and obesity—United States, 2007. Morb Mortal Wkly Rep 2009; 58(45): 1259-63.

[8]   Liese AD, Lawson A, Song HR, et al. Evaluating geographic variation in type 1 and type 2 diabetes mellitus incidence in youth in four US regions. Health Place 2010; 16(3): 547-56.

[9]   Tompkins JW, Luginaah IN, Booth GL, Harris SB. The geography of diabetes in London, Canada. The need for local level policy for prevention and management. Int J Environ Res Public Health 2010; 7(5): 2407-22.

[10]   Cadwell BL, Thompson TJ, Boyle JP, Barker LE. Bayesian Small Area Estimates of Diabetes Prevalence by U.S. County, 2005. J Data Sci 2010; 8: 173-88.

[11]   Centers for disease Control and Prevention: National Diabetes Surveillance System. (Accessed 8/11/2010). Available at: www.cdc.gov/diabetes/statistics/index.htm.

[12]   Anselin L, Sridharan S, Gholston S. Using exploratory spatial data analysis to leverage social indicator databases. The discovery of interesting patterns. Soc Indicators Res 2007; 82(2): 287-09.

[13]   Anselin L, Syabri I, Kho Y. GeoDa: An introduction to spatial data analysis. Geogr Anal 2006; 38(1): 5-22.

[14]   Goovaerts P. Geostatistical analysis of county-level lung cancer mortality rates in the Southeastern United States. Geogr Anal 2010; 42(1): 32-52.

[15]   Loughnan ME, Nicholls N, Tapper NJ. Demographic, seasonal, and spatial differences in acute myocardial infarction admissions to hospital in Melbourne Australia. Int J Health Geogr 2008; 7: 42.

[16]   Michimi A, Wimberly MC. Spatial patterns of obesity and associated risk factors in the conterminous U.S. Am J Prev Med 2010; 39: 1-12.

[17]   Owusu-Edusei K Jr, Owens CJ. Monitoring county-level chlamydia incidence in Texas, 2004-2005. application of empirical Bayesian smoothing and Exploratory Spatial Data Analysis (ESDA) methods. Int J Health Geogr 2009; 8:12.

[18]   Schuurman N, Peters PA, Oliver LN. Are obesity and physical activity clustered? A spatial analysis linked to residential density. Obesity (Silver Spring) 2009; 17(12): 2202-09.

[19]   Chen DR, Wen TH. Elucidating the changing socio-spatial dynamics of neighborhood effects on adult obesity risk in Taiwan from 2001 to 2005. Health Place 2010; 16(6): 1248-58.

[20]   Green C, Hoppa RD, Young TK, Blanchard JF. Geographic analysis of diabetes prevalence in an urban area. Soc Sci Med 2003; 57(3): 551-60.

[21]   Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ. Geographic distribution of diagnosed diabetes in the U.S.: a diabetes belt. Am J Prev Med 2011; 40: 434-39.

[22]   Yiannakoulias N. Using population attributable risk to understand geographic disease clusters. Health Place 2009; 15(4): 1142-48.

[23]   US Census Bureau. Small Area Income and Poverty Estimates. Internet researched data: December 2008. (Accessed August 11,

2010). Available at: http://www.census.gov/did/www/saipe/data/statecounty/data/2007.html.

[24]   Inter-university Consortium for Political and Social Research. County Characteristics, 2000-2007 [United States] [Computer file]. ICPSR20660-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-01-24. doi:10.3886/ICPSR20660. 2008.

[25]   U.S. Census Bureau. Population estimates. Releaseddate: September 2011. (Accessed June 26, 2012). Available at: www.census.gov/popest/data/intercensal/county/files/CO-EST00INT-AGESEX-5YR.csv.

[26]   U.S. Department of Agriculture. County typology codes. Updated date: March 26, 2009. (Accessed August 11, 2010). Available at: www.ers.usda.gov/Data/TypologyCodes/.

[27]   U.S. Census Bureau: 2007 Health Insurance Coverage Status for Counties and States: Data Sets. (Accessed June 26, 2012). Available at: www.census.gov/did/www/sahie/data/2007/dataset.html.

[28]   Anselin L. Local Indicators of Spatial Association—LISA. Geogr Anal 1995; 27(2): 93-115.

[29]   Krishnan S, Cozier YC, Rosenberg L, Palmer JR. Socioeconomic status and incidence of type 2 diabetes: results from the Black Women's Health Study. Am J Epidemiol 2010; 171: 564-70.

[30]   Robbins JM, Vaccarino V, Zhang H, Kasl SV. Socioeconomic status and type 2 diabetes in African American and non-Hispanic white women and men: evidence from the Third National Health and Nutrition Examination Survey. Am J Public Health 2001; 91(1): 76-83.

[31]   Barker L, Crespo R, Gerzoff RB, Denham S, Shrewsberry M, Cornelius-Averhart D. Residence in a distressed county in Appalachia as a risk factor for diabetes, Behavioral Risk Factor Surveillance System, 2006-2007. Prev Chronic Dis 2010; 7(5): A104.

[32]   Fang R, Kmetic A, Millar J, Drasic L. Disparities in chronic disease among Canada's low-income populations. Prev Chronic Dis 2009; 6(4): A115.

[33]   Lopez-De Fede A, Stewart JE, Battersby SE, Remington KC, Hardin JW, Mayfield-Smith K. Diabetes and the socioeconomic and built environment: geovisualization of disease prevalence and potential contextual associations using ring maps. Int J Health Geogr 2011; 10: 18.

[34]   Congdon P, Lloyd P. Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. J Data Sci 2010; 8: 235-52.

[35]   LaVeist TA, Thorpe RJ, Galarraga JE, Bower KM, Gary-Webb TL. Environmental and socio-economic factors as contributors to racial disparities in diabetes prevalence. J Gen Intern Med 2009; 24(10): 1144-48.

[36]   Ford ES, Mokdad AH, Giles WH, Galuska DA, Serdula MK. Geographic variation in the prevalence of obesity, diabetes, and obesity-related behaviors. Obes Res 2005; 13(1): 118-22.

[37]   Centers for Disease Control and Prevention. National diabetes fact sheet: general information and national estimates on diabetes in the United States, 2005. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (Accessed April 5, 2011). Available at: http://www.cdc.gov/diabetes/pubs/factsheet05.htm.

[38]   Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. Int J Epidemiol 2001; 30(6): 1343-50.