KANSAS JOURNAL of MEDICINE

A Geo-Stratified Analysis of Associations Between Socio-Economic Factors and Diabetes Risk

Michelle L. Redmond, Ph.D., M.S.[1], Milan Bimali, Ph.D.[2], Elizabeth Ablah, Ph.D., MPH[1], Paigton Mayes, Ph.D., MPH[1], Katelyn Dugan, MS-3[1]

[1]University of Kansas School of Medicine-Wichita, Wichita, KS Department of Population Health

[2]University of Arkansas for Medical Services, Little Rock, AR Department of Biostatistics

Received Aug. 23, 2021; Accepted for publication March 23, 2022; Published online May 17, 2022
https://doi.org/10.17161/kjm.vol15.15799

ABSTRACT

Introduction. In 2019, diabetes was the seventh leading cause of death in the United States. The association between diabetes risk and socio-economic factors in the U.S. has been examined primarily at the national level; little is known about this association at the regional level. This study examined and compared the association between diabetes risk and previously established socio-economic factors across four geographic regions (South, Midwest, West, and Northwest).

Methods. This study analyzed the 2014 Behavioral Risk Factor Surveillance System (BRFSS) data stratified by four geographic regions of the U.S. The risk estimates of diabetes associated with previously established socio-economic factors, as well as diabetes prevalence, were compared across four geographic regions.

Results. There was marked variation in association between diabetes risk and previously established risk factors across the four geographic regions. In the South, rural residency was associated with increased diabetes risk, whereas in the other geographic regions rural residency had a protective effect. In the South, the diabetes risk for males was 22% higher compared to females, whereas the risk for males was 41% higher than females in the Northeast. Independently, age had the strongest discriminative ability to distinguish between a person with diabetes and a person without diabetes, whereas ethnicity, race, and sex had the weakest discriminative abilities.

Conclusions. These findings suggested a higher prevalence of diabetes by race/ethnicity (non-Hispanic Black and Hispanic) and income across all four regions. Rural residency was highest in the South, but protective in other regions. Overall, age and income provided the highest predictive ability for diabetes risk. This study highlighted differences in diabetes prevalence in association between previously established socio-economic variables and diabetes risk across four geographic regions. These findings could help public health professionals and policy makers in understanding the dynamic relationship between diabetes and risk factors at the regional level.

Kans J Med 2022;15:175-183

INTRODUCTION

In 2018, 34.2 million people were estimated to have diabetes mellitus (diabetes), and another 7.3 million were estimated to live with undiagnosed diabetes.[1] In 2019, type II diabetes mellitus was the eighth leading cause of death in the U.S., with more than 84,000 deaths.[2] Diabetes is associated with increased risk of a wide variety of diseases and health complications, such as cardiovascular disease, kidney disease, tuberculosis, obesity, ophthalmic disorders, nephrology complications, and periodontal disease.[3-8]

The economic toll of diabetes is equally catastrophic. In the U.S., the total estimated cost of diabetes in 2017 was $327 billion, 72% of which was accounted for by direct health care expenditures and 28% was associated with reduced productivity.[9] On average, the health care cost was estimated to increase by 230% for someone with diabetes. The increased cost associated with diabetes had more of a severe impact on low- and middle-income families. Considering the tremendous clinical and financial burden incurred, the increasing trend of diabetes prevalence, and the fact that diabetes has been viewed as a largely preventable disease, studies aimed at identifying the determinants of diabetes have become more important than ever.

Multiple demographic factors have been associated with the risk of diabetes, including gender, age, income, education, race, ethnicity, and rural residency.[10-12] Specifically, being male, non-Hispanic Black, Hispanic, low-income, having less education, and being older all put one at a higher risk of having diabetes.[10-16] However, understanding geographic disparities around diabetes is needed. Rural residents have a higher prevalence of diabetes compared to urban residents.[17-22] O'Connor and Wellenuis found disparities exist between rural and urban residents, with rural residents more likely to receive a diagnosis of diabetes mellitus compared to urban residents.[17] However, after controlling for risk factors, the prevalence of diabetes diminished for rural residents. Barker and colleagues[21] suggested the southern region of the U.S. is the "diabetes belt", meaning it had a higher prevalence of diabetes and most often was linked to non-Hispanic Blacks who were obese and led a sedentary lifestyle.[21] Additional studies have examined the differences in how rural residents receive treatment and manage their diabetes compared to urban residents.[22] Stark differences were discovered that left rural residents with a disadvantage because of several factors, such as being overweight, having less access to a primary care physician (e.g., living in a Health Professional Shortage Area), and cost of care.

There is a dearth of literature on the prevalence of diabetes based on geographical region within the U.S. Although several studies have addressed the prevalence of diabetes in terms of socio-economic variables, little is known about the association of diabetes risk by geographic region.[10,23] Voeks et al.[23] suggested there are regional differences in terms of diabetes prevalence within selected southern states. How diabetes prevalence differs across geographical regions of the U.S. is poorly understood. Therefore, the purpose of this study was to examine diabetes prevalence in regions across the U.S. and describe relationships between diabetes risk and previously established socio-economic determinants of diabetes by region.

METHODS

Dataset and Study Design. The Behavioral Risk Factor Surveillance System (BRFSS) survey is a collaborative effort between the states, participating U.S. territories, and the U.S. Centers for Disease Control and Prevention (CDC).[24] The BRFSS is an ongoing surveillance system designed to measure behavioral risk factors for non-institutionalized adults in the U.S. The BRFSS survey is administered through landline or cellular telephone. The landline telephone survey involved data collection from a randomly selected adult in each household. Among the cellular telephone users, information was collected from participants who resided in private residences or college housing. Information collected during the interview included demographics, preventive health practices, and risk behaviors.[24] This study used data collected during the 2014 survey cycle from four U.S. geographical regions: South, Midwest, West, and Northwest. A stratified analysis for each geographic region was conducted (Figure 1). The BRFSS inclusion criteria included U.S. residents 18 years or older who owned a landline telephone or cellular telephone.

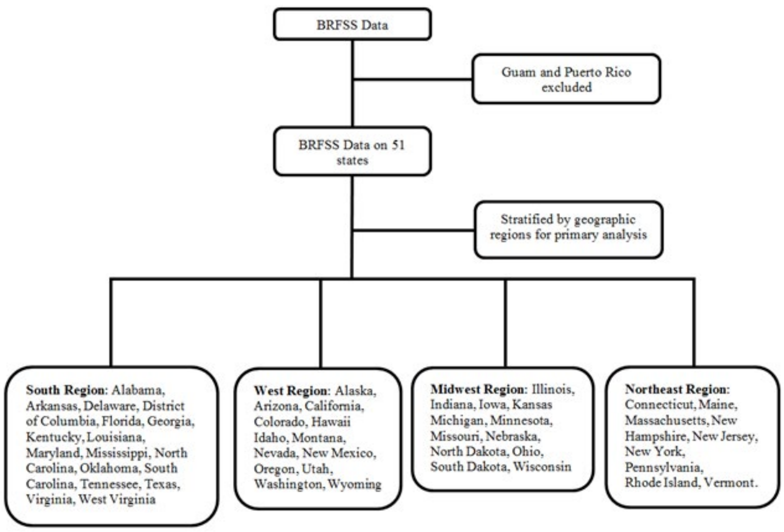

Figure 1. Study flowchart stratification.

Variables. The primary outcome of interest was if the respondent had ever been told by a doctor that they had diabetes. Although there are differences between type 1 and type 2 diabetes, the BRFSS database did not assess the type of diabetes; thus, the term "diabetes" referred to both types. Those who reported not having been diagnosed with diabetes or having been diagnosed only during pregnancy or with pre-diabetes or borderline diabetes were categorized as not having diabetes. Participants who refused to respond, were not asked, responded as being unsure, or left the question blank were excluded from analysis.

Based on residency, respondents were classified as rural residents if they reported their residency was not in a metropolitan statistical area (MSA). MSAs are defined by the U.S. Office of Management and Budget, and this definition has been applied by the U.S. Census Bureau for data collection.[25] Based on household income, participants were classified into one of the three categories: less than $25,000, between $25,000 and $50,000, and at least $50,000. Regarding race, participants were categorized as non-Hispanic White, non-Hispanic Black, and "other" (if race was reported as Asian, Native Hawaiian or other Pacific Islander, or other race). Regarding age, participants were categorized as 34 or younger, between 35 and 44, between 45 and 54, between 55 and 64, and 65 or older. To determine education level, participants were stratified into one of four groups: "did not graduate high school", "graduated high school", "attended college or technical school", and "graduated from college or technical school".[25]

Statistical Analysis. The association between diabetes risk and previously established socio-economic determinants of diabetes was explored by geographic region through modeling the relative risk ratio of probabilities in lieu of the more commonly reported odds ratio modeling. Zou's modified Poisson regression was used to obtain relative risk estimates.[26] Unlike odds ratios, relative risk estimates are more intuitive and comprehensible as they directly compare the probabilities of two mutually exclusive events: the presence and absence of diabetes.[27,28]

All variables included in this study were categorical. Descriptive statistics were reported as frequency (percentage). The association between diabetes status and each categorical variable was examined initially using Pearson's Chi-Square test of independence. Overall, diabetes prevalence was reported at each level of categorical variable. Zou's modified Poisson regression approach was used to obtain risk of diabetes associated with each variable.[26] Two-way interactions among variables were tested and adjusted for when the interaction was significant. Additionally, concordance index (C-index) was used to assess the predictive ability of a fitted model as well as the independent predictive ability of each variable. Reference categories for covariates were urban residency, female, income greater than $50,000, non-Hispanic ethnicity, non-Hispanic White, age younger than 34 years, and graduate from college or technical school. The analyses were performed using SAS 9.4 statistical software for Windows®. Statistical significance was based on two-sided tests, assuming a type I error rate of 0.05.

RESULTS

Descriptive Statistics. In the South (n = 71,441), 70% of respondents lived in an urban or semi-urban region; 61% were females; 43% had income greater than $50,000; 94% were non-Hispanic, 78% were non-Hispanic Whites, 14% were 34 years or younger, and 34% graduated from college or technical school. The prevalence of self-reported diabetes was higher in rural regions (20%) than in urban regions (17%); slightly higher among males (16%) than females (15%); highest among those with an income of less than $25,000 (22%) and least among those with incomes greater than $50,000 (10%); higher among Hispanics (16%) than non-Hispanics (14%); highest among non-Hispanic Blacks (21%) and least among "other" race (13%); highest among those 65 years or older (23%) and least among those 34 years or younger (2%); highest among those who graduated from high school but did not attend college or technical school (24%) and least among those who graduated from college or technical school (11%). The Chi-Square test of association suggested significant associations among all the covariates and diabetes status (Table 1).

In the West (n = 54,830), 63% of respondents lived in an urban or semi-urban region; 56% were females; 47% had income greater than $50,000; 89% were non-Hispanic; 86% were non-Hispanic Whites; 16% were 34 years old or younger; 38% graduated from college or technical school. The prevalence of self-reported diabetes was similar in rural and urban regions (13%); slightly higher among males (12%) than females (11%); highest among those with an income of less than $25,000 (16%) and least among those with incomes greater than

$50,000 (8%); higher among Hispanics (14%) than non-Hispanics (11%); highest among non-Hispanic Blacks and "other" race (14%) and least among non-Hispanic Whites (11%); highest among those who are at least 65 years old (18%) and least among those 34 years old or younger (2%); highest among those who did not graduate high school (18%) and least among those who graduated from college or technical school (8%). Gender, income category, ethnicity, race, age category, and education level were associated independently with diabetes status (Table 2). The association among rural residency and diabetes status was not significant.

In the Midwest (n = 62,443), 57% of respondents lived in an urban or semi-urban region; 57% were females; 46% had income greater than $50,000; 97% were non-Hispanic; 91% were non-Hispanic Whites; 15% were 34 years old or younger; 34% graduated from college or technical school (Table 3). The prevalence of self-reported diabetes was similar in rural and urban regions (15%); slightly higher among males (13%) than females (12%); highest among those with an income of less than $25,000 (19%) and least among those with incomes greater than $50,000 (8%); higher among Hispanics (13%) than non-Hispanics (11%); highest among non-Hispanic Blacks (20%) and least among non-Hispanic Whites (12%); highest among those with aged 65 years or older (20%) and least among those that are 34 years old or younger (2%); highest among those who did not graduate high school (20%) and least among those who graduated from college or technical school (9%). Gender, income category, ethnicity, race, age category, and education level were associated independently with diabetes status (Table 3). The association among rural residency and diabetes status was not significant.

Table 1. Prevalence of self-reported diabetes by socio-economic variables in southern region of the United States: BRFFS, 2014.

Variable	Response Options	Count (%)[1]	Prevalence (%)[2]	p Value[3]
Rural status	Urban	63,513 (70)	10,719 (17)	< 0.01
	Rural	27,790 (30)	5,491 (20)	
Gender	Male	53,383 (39)	8,499 (16)	< 0.01
	Female	82,838 (61)	12,646 (15)	
Income	< $25,000	36,120 (32)	7,908 (22)	< 0.01
	≥ $25,000 and ≤ $50,000	28,672 (25)	4,631 (16)	
	> $50,000	47,932 (43)	4,878 (10)	
Ethnicity	Not Hispanic	126,498 (94)	19,733 (16)	< 0.01
	Hispanic	8,482 (6)	1,196 (14)	
Race	Non-Hispanic White	103,488 (78)	14,809 (14)	< 0.01
	Non-Hispanic Black	23,768 (18)	5,107 (21)	
	Other	5,613 (4)	734 (13)	
Age	≤ 34	18,544 (14)	373 (2)	< 0.01
	≥ 35 and ≤ 44	15,294 (11)	906 (6)	
	≥ 45 and ≤ 54	22,460 (16)	2,649 (12)	
	≥ 55 and ≥ 65	30,854 (23)	5,977 (19)	
	≥ 65	49,069 (36)	11,240 (23)	
Education	Did not graduate high school	14,367 (11)	3,438 (24)	< 0.01
	Graduated high school	39,594 (29)	7,007 (18)	
	Attended college or technical school	35,121 (26)	5,583 (16)	
	Graduated from college or technical school	46,123 (34)	4,971 (11)	

[1]Count (%) is the frequency counts and percentage for each level of variable.
[2]Prevalence (%) represents prevalence of self-reported diabetes for each level of categorical.
[3]p value is based on Chi-Square test of association between diabetes status and the variables.

Table 2. Prevalence of self-reported diabetes by socio-economic variables in western region of the United States: BRFFS, 2014.

Variable	Response Options	Count (%)[1]	Prevalence (%)[2]	p Value[3]
Rural status	Urban	42,445 (63)	5,458 (13)	0.20
	Rural	25,345 (37)	3,350 (13)	
Gender	Male	48,416 (44)	5,780 (12)	< 0.01
	Female	62,747 (56)	6,643 (11)	
Income	< $25,000	25,485 (27)	4,140 (16)	< 0.01
	≥ $25,000 and ≤ $50,000	24,308 (26)	2,985 (12)	
	> $50,000	45,335 (47)	3,439 (8)	

Table 2. Prevalence of self-reported diabetes by socio-economic variables in western region of the United States: BRFFS, 2014. *cont.*

Variable	Response Options	Count (%)[1]	Prevalence (%)[2]	p Value[3]
Ethnicity	Not Hispanic	9,7248 (89)	10,474 (11)	< 0.01
	Hispanic	12,563 (11)	1,769 (14)	
Race	Non-Hispanic White	92,977 (86)	9,870 (11)	< 0.01
	Non-Hispanic Black	2,333 (2)	359 (14)	
	Other	12,978 (12)	1,835 (14)	
Age	LE 34	17,627 (16)	279 (2)	< 0.01
	≥ 35 and ≤ 44	13,948 (13)	611 (4)	
	≥ 45 and ≤ 54	17,715 (16)	1,597 (9)	
	≥ 55 and ≥ 65	24,609 (22)	3,234 (13)	
	≥ 65	37,264 (33)	6,702 (18)	
Education	Did not graduate high school	7,707 (7)	1,410 (18)	< 0.01
	Graduated high school	27,775 (25)	3,535 (13)	
	Attended college or technical school	32,470 (30)	3,864 (12)	
	Graduated from college or technical school	42,130 (38)	3,496 (8)	

[1]Count (%) is the frequency counts and percentage for each level of variable.
[2]Prevalence (%) represents prevalence of self-reported diabetes for each level of categorical.
[3]p value is based on Chi-Square test of association between diabetes status and the variables.

Table 3. Prevalence of self-reported diabetes by socio-economic variables in midwestern region of the United States: BRFFS, 2014.

Variable	Response Options	Count (%)[1]	Prevalence (%)[2]	p Value[3]
Rural status	Urban	43,277 (57)	6,375 (15)	0.74
	Rural	32,916 (43)	4,817 (15)	
Gender	Male	53,620 (43)	7,105 (13)	< 0.01
	Female	72,386 (57)	8,745 (12)	
Income	< $25,000	27,968 (26)	5,322 (19)	< 0.01
	≥ $25,000 and ≤ $50,000	30,436 (28)	4,106 (13)	
	> $50,000	50,035 (46)	4,109 (8)	
Ethnicity	Not Hispanic	121,334 (97)	15,301 (13)	< 0.01
	Hispanic	3,781 (3)	407 (11)	
Race	Non-Hispanic White	113,442 (91)	13,692 (12)	< 0.01
	Non-Hispanic Black	5,759 (5)	1,165 (20)	
	Other	4,937 (4)	765 (15)	
Age	≤ 34	18,899 (15)	305 (2)	< 0.01
	≥ 35 and ≤ 44	14,533 (12)	651 (4)	
	≥ 45 and ≤ 54	20,933 (17)	1,963 (9)	
	≥ 55 and ≥ 65	28,836 (23)	4,220 (15)	
	≥ 65	42,805 (34)	8,711 (20)	
Education	Did not graduate high school	8,033 (6)	1,590 (20)	< 0.01
	Graduated high school	38,967 (31)	5,816 (15)	
	Attended college or technical school	36,289 (29)	4,601 (13)	
	Graduated from college or technical school	42,027 (34)	3,746 (9)	

[1]Count (%) is the frequency counts and percentage for each level of variable.
[2]Prevalence (%) represents prevalence of self-reported diabetes for each level of categorical.
[3]p value is based on Chi-Square test of association between diabetes status and the variables.

In the Northeast (n = 46,571), 81% of respondents lived in an urban or semi-urban region; 59% were females; 50% had income greater than $50,000; 94% were non-Hispanic; 88% were non-Hispanic Whites; 13% were 34 years or younger; 42% graduated from college or technical school (Table 4). The prevalence of self-reported diabetes was slightly higher in urban regions (14%) than rural regions (13%); higher among males (14%) than females (11%); highest among those with an income of less than $25,000 (19%) and least among those with incomes greater than $50,000 (8%); higher among Hispanics (14%) than those who identified as non-Hispanic (12%); highest among non-Hispanic Blacks (18%) and least among non-Hispanic Whites and "other" race (12%); highest among those 65 years or older (20%) and least among those aged 34 years or younger (4%); highest among those who did not graduate high school (22%) and least among those who graduated from college or technical school (8%). Gender, income category, ethnicity, race, age category, and education level were associated independently with diabetes status (Table 4). The association among rural residency and diabetes status was moderately significant (p = 0.02).

Relative Risk Modeling. For each of the geographic regions, the interaction between income and age, and income and education level were significant. The relative risk modeling results presented in Table 5 provide risk estimates after adjusting for the interactions. The interpretation of risk estimates associated with any covariate in the subsequent paragraphs implicitly assumed values of all covariates, besides the covariate whose risk was being interpreted, were held fixed.

Table 4. Prevalence of self-reported diabetes by socio-economic variables in northeastern region of the United States: BRFFS, 2014.

Variable	Response Options	Count (%)[1]	Prevalence (%)[2]	p Value[3]
Rural status	Urban	46,877 (81)	6,595 (14)	0.02
	Rural	10,957 (19)	1,447 (13)	
Gender	Male	3,4284 (41)	4,652 (14)	< 0.01
	Female	48,484 (59)	5,535 (11)	
Income	< $25,000	18,161 (26)	3,537 (19)	< 0.01
	≥ $25,000 and ≤ $50,000	16,752 (24)	2,328 (14)	
	> $50,000	34,849 (50)	2,725 (8)	
Ethnicity	Not Hispanic	76,955 (94)	9,355 (12)	< 0.01
	Hispanic	4,878 (6)	691 (14)	
Race	Non-Hispanic White	70,751 (88)	8,351 (12)	< 0.01
	Non-Hispanic Black	5,552 (7)	992 (18)	
	Other	4,117 (5)	496 (12)	
Age	≤ 34	10,782 (13)	202 (2)	< 0.01
	≥ 35 and ≤ 44	9,441 (11)	402 (4)	
	≥ 45 and ≤ 54	14,934 (18)	1,303 (9)	
	≥ 55 and ≥ 65	19,748 (24)	2,768 (14)	
	≥ 65	27,863 (34)	5,512 (20)	
Education	Did not graduate high school	5,396 (7)	1,173 (22)	< 0.01
	Graduated high school	22,585 (28)	3,533 (16)	
	Attended college or technical school	19,699 (24)	2,553 (13)	
	Graduated from college or technical school	34,190 (42)	2,811 (8)	

[1]Count (%) is the frequency counts and percentage for each level of variable.
[2]Prevalence (%) represents prevalence of self-reported diabetes for each level of categorical.
[3]p value is based on Chi-Square test of association between diabetes status and the variables.

Table 5. Associations between diabetes and socio-economic variables, BRFSS, 2014.

Variables	Variable Category*	Variable Category*	Relative Risk (95% CI)			
			South (n = 71,441)	West (n = 54,830)	Midwest (n = 62,443)	Northeast (n = 46,571)
Rural status	Rural		1.09 (1.06 - 1.13)	0.96 (0.92 - 1)	0.95 (0.91 - 0.98)	0.91 (0.86 - 0.96)
	Urban		Reference category	Reference category	Reference category	Reference category
Gender	Male		1.22 (1.18 - 1.25)	1.29 (1.24 - 1.35)	1.30 (1.25 - 1.35)	1.41 (1.35 - 1.47)
	Female		Reference category	Reference category	Reference category	Reference category
Income	< $25,000		3.69 (2.28 - 5.97)	2.40 (1.43 - 4.02)	1.99 (1.17 - 3.37)	2.81 (1.62 - 4.88)
	≥ $25,000 and ≤ $50,000		1.92 (1.08 - 3.39)	1.28 (0.68 - 2.41)	1.19 (0.64 - 2.21)	1.76 (0.93 - 3.33)
	> $50,000		Reference category	Reference category	Reference category	Reference category
Ethnicity	Hispanic		1.26 (1.16 - 1.36)	1.34 (1.25 - 1.44)	1.08 (0.92 - 1.25)	1.29 (1.16 - 1.43)
	Not Hispanic		Reference category	Reference category	Reference category	Reference category
Race	Non-Hispanic Black		1.54 (1.49 - 1.60)	1.72 (1.53 - 1.95)	1.63 (1.52 - 1.74)	1.51 (1.40 - 1.63)
	Other		1.25 (1.14 - 1.36)	1.48 (1.39 - 1.57)	1.70 (1.56 - 1.86)	1.30 (1.16 - 1.46)
	Non-Hispanic White		Reference category	Reference category	Reference category	Reference category
Age	≥ 35 and ≤ 44		3.00 (1.93 - 4.68)	1.48 (0.93 - 2.37)	1.70 (1.10 - 2.64)	2.34 (1.45 - 3.76)
	≥ 45 and ≤ 54		6.00 (3.93 - 9.17)	3.93 (2.57 - 5.99)	4.32 (2.89 - 6.46)	3.94 (2.51 - 6.18)
	≥ 55 and ≤ 65		11.37 (7.48 - 17.27)	7.13 (4.72 - 10.78)	7.80 (5.26 - 11.56)	7.70 (4.95 - 11.97)
	≥ 65		15.06 (9.92 - 22.86)	10.59 (7.03 - 15.97)	11.76 (7.95 - 17.41)	11.47 (7.39 - 17.79)
	≤ 34		Reference category	Reference category	Reference category	Reference category
Education	Attended college or technical school		1.44 (1.35 - 1.55)	1.48 (1.36 - 1.62)	1.33 (1.23 - 1.45)	1.44 (1.3 - 1.59)
	Did not graduate high school		1.81 (1.48 - 2.21)	1.83 (1.37 - 2.45)	1.72 (1.34 - 2.21)	1.96 (1.48 - 2.58)
	Graduated high school		1.48 (1.36 - 1.60)	1.49 (1.34 - 1.66)	1.33 (1.22 - 1.46)	1.54 (1.39 - 1.71)
	Graduated from college or technical school		Reference category	Reference category	Reference category	Reference category
Income *age	≥ $25,000 and ≤ $50,000	≥ 35 and ≤ 44	1.10 (0.59 - 2.04)	2.16 (1.06 - 4.38)	2.11 (1.07 - 4.17)	0.85 (0.41 - 1.78)
	≥ $25,000 and ≤ $50,000	≥ 45 and ≤ 54	0.95 (0.53 - 1.69)	1.47 (0.77 - 2.82)	1.43 (0.76 - 2.70)	1.14 (0.59 - 2.20)
	≥ $25,000 and ≤ $50,000	≥ 55 and ≤ 65	0.8 (0.45 - 1.42)	1.19 (0.63 - 2.24)	1.3 (0.70 - 2.42)	0.98 (0.52 - 1.86)
	≥ $25,000 and ≤ $50,000	≥ 65	0.71 (0.4 - 1.25)	1.09 (0.58 - 2.04)	1.12 (0.60 - 2.07)	0.84 (0.45 - 1.59)
	< $25,000	≥ 35 and ≤ 44	0.92 (0.55 - 1.54)	1.64 (0.91 - 2.96)	1.90 (1.05 - 3.44)	0.95 (0.51 - 1.77)
	< $25,000	≥ 45 and ≤ 54	0.79 (0.48 - 1.28)	1.30 (0.77 - 2.21)	1.68 (0.98 - 2.88)	1.16 (0.66 - 2.04)

Table 5. Associations between diabetes and socio-economic variables, BRFSS, 2014. *cont.*

Variables	Variable Category*	Variable Category*	Relative Risk (95% CI)			
			South (n = 71,441)	West (n = 54,830)	Midwest (n = 62,443)	Northeast (n = 46,571)
Income *age	< $25,000	≥ 55 and ≤ 65	0.58 (0.36 - 0.93)	0.88 (0.53 - 1.48)	1.23 (0.73 - 2.09)	0.80 (0.46 - 1.39)
	< $25,000	≥ 65	0.45 (0.28 - 0.73)	0.73 (0.44 - 1.21)	0.87 (0.52 - 1.47)	0.60 (0.35 - 1.03)
Income * Education	≥ $25,000 and ≤ $50,000	Attended college or technical school	0.82 (0.74 - 0.91)	0.86 (0.76 - 0.98)	0.86 (0.76 - 0.97)	0.81 (0.7 - 0.94)
	≥ $25,000 and ≤ $50,000	Did not graduate high school	0.67 (0.53 - 0.85)	0.71 (0.51 - 1.00)	0.84 (0.63 - 1.12)	0.70 (0.5 - 0.97)
	≥ $25,000 and ≤ $50,000	Graduated high school	0.77 (0.68 - 0.86)	0.80 (0.69 - 0.93)	0.85 (0.75 - 0.96)	0.80 (0.69 - 0.92)
	< $25,000	Attended college or technical school	0.78 (0.7 - 0.87)	0.80 (0.7 - 0.92)	0.82 (0.72 - 0.93)	0.87 (0.75 - 1.02)
	< $25,000	Did not graduate high school	0.67 (0.54 - 0.83)	0.76 (0.56 - 1.04)	0.69 (0.53 - 0.91)	0.76 (0.56 - 1.03)
	< $25,000	Graduated high school	0.75 (0.67 - 0.84)	0.77 (0.66 - 0.89)	0.79 (0.7 - 0.91)	0.86 (0.74 - 1.00)

n = Number of observations used in analysis.
Reference Group: Female (Sex) - Urban (Rural) - Greater than 50,000 (Income Status) - Not Hispanic (Ethnicity) -Less than 34 years (Age category) -and graduated from college or technical school (Education)
Results that are statistically significant are highlighted in yellow.

Equipoise exists with respect to risk estimates associated with rural residency as both protective and harmful effects of rural residency were observed across the four geographic regions. Rural residents had a 9% higher diabetes risk in the South. However, in the other regions, rural residency appeared to have a protective effect. Diabetes risk among rural residents was 5% lower in the Midwest, 9% lower in the Northeast, and 4% lower in the West. Compared to females in the following regions, the diabetes risk among males was 41% higher in the Northeast, 30% higher in the Midwest, 29% higher in the West, and 22% higher in the South. Hispanic Americans were at consistently higher diabetes risk across all four geographic regions. Compared to non-Hispanic Whites in the following regions, the risk of diabetes for Hispanics was 34% higher in the West, 29% higher in the Northeast, 26% in the South, and 8% higher in the Midwest. Similarly, compared to non-Hispanic Whites in the following regions, the risk of diabetes for non-Hispanic Blacks was 72% higher in the West, 63% higher in the Midwest, 54% higher in the South, and 51% higher in the Northeast.

Owing to interaction between age and income, and between age and education, the increased risk estimates for each stratification of age, income, and education were not directly interpretable. For the sake of presentation, the risk estimates of an individual were computed with "an income between $25,000 and $50,000, age between 45 and 54 years, and graduated high school". Compared to the reference category (income greater than $50,000, 34 years or younger, and graduated from college or technical school), this risk estimate was 12.5 times higher in the South, 9 times higher in the Northeast, 8.8 times higher in the West, and 8.3 times higher in the Midwest.

Additional Analysis. The discriminative (predictive) ability of the model in correctly classifying respondents with diabetes and those without diabetes was estimated by C-index. The C-index for the fitted models ranged from 0.72 (for the South) to 0.73 (for the West, Midwest, Northeast; Figure 2). Traditionally, the threshold of 0.8 has been used to identify models with strong predictive abilities.[29] The fitted models are thus moderately strong in terms of their predictive abilities. The predictive ability of each variable also was examined independently. Age had the greatest predictive ability across all geographical regions, followed by income (Figure 2). Ethnicity, sex, and race had the least predictive ability with C-indices marginally above 0.5, thus performing no better than a random coin flip.

DISCUSSION

This study examined associations between socio-economic factors and diabetes risk across the four geographic regions in the U.S. Unlike previous studies, these associations were examined while considering the possible interactions among socio-economic factors. The independent predictive ability of each of these factors also were evaluated. It was seen that the magnitude of the effect of these factors on the diabetes risks varied markedly across the four geographic regions; and for rural residency the direction of the effect in the South (increased risk) and the rest of the geographic regions (protective effect) were in the opposite direction. These results suggested the relationship between socio-economic factors and diabetes risk could differ significantly across the four geographic regions. These findings were similar to Barker and colleagues who found there to be a diabetes belt in the Southern region of the U.S. linked to non-Hispanic Black residents.[22]

Other studies have found similar results related to the Southern and Midwestern region of the U.S., where a significant prevalence of diabetes and metabolic syndrome exists.[30-31] In this study, a significant difference in self-reported prevalence among rural and urban regions was found in the South, with rural regions reporting a higher prevalence of diabetes and non-Hispanic Blacks having a higher risk of diabetes. Even with regional differences when looking at relative risk of diabetes, the interaction between income and education is the strongest predictor of diabetes prevalence. More tools need to be implemented to lower the prevalence of diabetes among low income and those with a lower educational attainment to improve areas such as a "diabetes belt". Myers and colleagues also investigated community level factors to better understand their results.[31] Myers found the "diabetes belt" region lacked sufficient recreational opportunities, more so than economic. What is unclear is if the county level factors were impacted by the community economic factor rather than the individual. Understanding environmental determinants that contribute to diabetes at a county and city level would be beneficial.

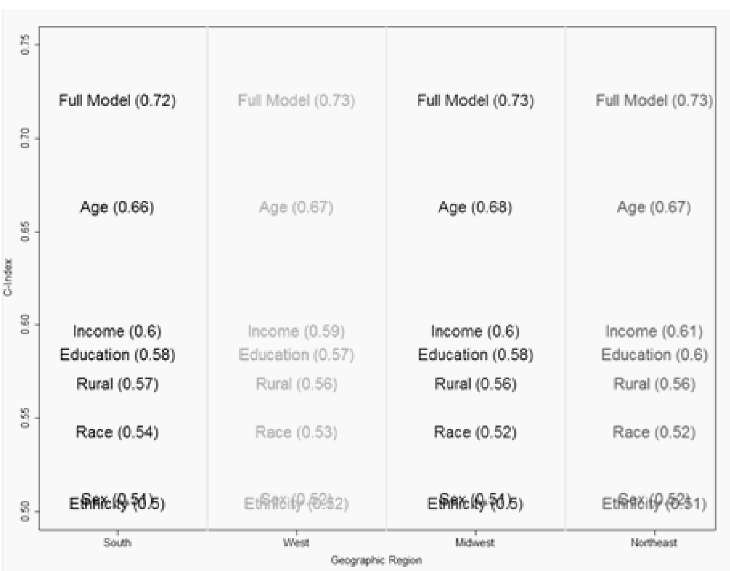

Figure 2. Discriminative (predictive) ability of diabetes classification by geographic region by estimated C-index.

There were also protective factors to consider in this work. Rural residency can be a protective factor in the Midwest, West, and Northeast, but not the South. Interestingly, a respondent's age and income were stronger predictors of diabetes risk. This study did not find race, ethnicity, or sex to be a dependable predictor of diabetes risk. This was important because of the overwhelming disparity in the prevalence of diabetes for specific racial/ethnic groups (non-Hispanic Black, Native American, Hispanic).

Limitations. The study had several limitations that were primarily attributable to the nature of the study design and the collected data. Limitations such as recall bias, inclusion of participants with access to phone services, missing information, lack of distinction between type 1 and type 2 diabetes, and biases associated with self-reporting (i.e., social desirability bias) limit the generalizability of our findings. The focus of our work was socio-economic factors, and influential variables (i.e., physical activity, food and beverage consumption, tobacco use) have not been included which likely would affect the risk estimates. Finally, retrospective studies can only identify associations and make it impossible to infer causality between the variables and diabetes risk.

CONCLUSIONS

This study highlighted novel findings, primarily, the variation in effect of socio-economic factors on diabetes risk across four geographic regions. This suggested that the dynamics between diabetes and the risk factors examined in the study differed by geographic region. The concordance index suggested that although variables could be associated significantly, their predictive ability may only be modest, essentially affirming the statistical adage that strongly associated variables may not be strongly predictive. There is a great need to develop a model with stronger predictive ability. Owing to interplay between social, economic, environmental, and genetic factors in establishing diabetes risk, the authors believe a strongly predictive model must incorporate individual and community-level information at the genetic, social, economic, and environmental levels. In conclusion, this study highlighted the differences in diabetes prevalence and the association between previously established socio-economic variables and diabetes risk across the four geographic regions of the U.S.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number K01HL135472. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1 U.S. Centers for Disease Control and Prevention. National Diabetes Statistics Report 2020. Estimates of Diabetes and its Burden in the United States. Retrieved from: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf.

2 U.S. Centers for Disease Control and Prevention. Mortality in the United States, 2020. 2021. https://www.cdc.gov/nchs/data/databriefs/db427.pdf.

3 Stojceva-Taneva O, Otovic NE, Taneva B. Prevalence of diabetes mellitus in patients with chronic kidney disease. Open Access Maced J Med Sci 2016; 4(1):79-82. PMID: 27275335.

4 Llambés F, Arias-Herrera S, Caffesse R. Relationship between diabetes and periodontal infection. World J Diabetes 2015; 6(7):927-935. PMID: 26185600.

5 Lai S. Chronic kidney disease and diabetes-a potential causal link. EBioMedicine 2016; 6:10-11. PMID: 27211535.

6 Bek T. Eye complications in type 2 diabetes mellitus. Ugeskr Laeger 2012; 174(37):2147-2150. PMID: 22971294.

7 Bhupathiraju SN, Hu FB. Epidemiology of obesity and diabetes and their cardiovascular complications. Circ Res 2016; 118(11):1723-1735. PMID: 27230638.

8 Calvin D, Watley S. Diabetes and hearing loss among underserved populations. Nurs Clin North Am 2015; 50(3):449-456. PMID: 26333602.

9 American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. Diabetes Care 2013; 36(4):1033-1046. PMID: 23468086.

10 Ablah E, Dong F, Cupertino AP, Konda K, Johnston JA, Collins T. Prevalence of diabetes and pre-diabetes in Kansas. Ethn Dis 2013; 23(4):415-420. PMID: 24392602.

11 Unwin N, Whiting D, Roglic G. Social determinants of diabetes and challenges of prevention. Lancet 2010; 375(9733):2204-2205. PMID: 20609961.

12 Saydah S, Lochner K. Socioeconomic status and risk of diabetes-related mortality in the U.S. Public Health Rep 2010; 125(3):377-388. PMID: 20433032.

[13] Rabi DM, Edwards AL, Southern DA. Association of socio-economic status with diabetes prevalence and utilization of diabetes care services. BMC Health Serv Res 2016; 6:124-130. PMID: 17018153.
[14] Tang M, Chen Y, Krewski D. Gender-related differences in the association between socioeconomic status and self-reported diabetes. Int J Epidemiol 2003; 32(3):381-385. PMID: 12777423.
[15] Bird Y, Lemstra M, Rogers M, Moraros J. The relationship between socioeconomic status/income and prevalence of diabetes and associated conditions: A cross-sectional population-based study in Saskatchewan, Canada. Int J Equity Health 2015; 14:93-101. PMID: 26458543.
[16] Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: A systematic review and meta-analysis. Int J Epidemiol 2011; 40(3):804-818. PMID: 21335614.
[17] O'Connor A, Wellenius G. Rural-urban disparities in the prevalence of diabetes and coronary heart disease. Public Health 2012; 126(10):813-820. PMID: 22922043.
[18] Krishna S, Gillespie KN, McBride TM. Diabetes burden and access to preventive care in the rural United States. J Rural Health 2010; 26(1):3-11. PMID: 20105262.
[19] Mohan V, Mathur P, Deepa R et al. Urban rural differences in prevalence of self-reported diabetes in India - the WHO-ICMR Indian NCD risk factor surveillance. Diabetes Res Clin Pract 2008; 80(1):159-168. PMID: 18237817.
[20] Chiwanga FS, Njelekela MA, Diamond MB, et al. Urban and rural prevalence of diabetes and pre-diabetes and risk factors associated with diabetes in Tanzania and Uganda. Glob Health Action 2016; 9:31440. PMID: 27221531.
[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(4):434-439. PMID: 21406277.
[22] Hale NL, Bennett KJ, Probst JC. Diabetes care and outcomes: Disparities across rural America. J Community Health 2010; 35(4):365-374. PMID: 20386968.
[23] Voeks JH, McClure LA, Go RC. Regional differences in diabetes as a possible contributor to the geographic disparity in stroke mortality: The reasons for geographic and racial differences in stroke study. Stroke 2008; 39(6):1675-1680. PMID: 18388336.
[24] U.S. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Survey Data, 2014 Codebook Report. Department of Health and Human Services, Centers for Disease Control and Prevention. 2014. https://www.cdc.gov/brfss/index.html. Accessed: September 3, 2021.
[25] U.S. Office of Management and Budget: 2010 Standards for delineating metropolitan and micropolitan statistical areas. 2010. https://www.govinfo.gov/content/pkg/FR-2010-06-28/pdf/2010-15605.pdf. Accessed: September 3, 2021.
[26] Zou G. A modified Poisson regression approach to prospective studies with binary data. Am K Epidemiol 2004; 159(7):702-706. PMID: 15033648.
[27] Lee J. Odds ratio or relative risk for cross-sectional data? Int J of Epidemol 1994; 23(1):201-203. PMID: 8194918.
[28] Schmidt CO, Kohlmann T. When to use the odds ratio or the relative risk? Int J Public Health 2008; 53(3):165-167. PMID: 19127890.
[29] Hosmer DW, Lemeshow S. Applied Logistic Regression. Second Edition. New York: John Wiley & Sons, 2000. ISBN: 9780471356325.
[30] Gurka MJ, Filipp SL, DeBoer MD. Geographical variation in the prevalence of obesity, metabolic syndrome, and diabetes among US adults. Nutr Diabetes 2018; 8(1):14. PMID: 29549249.
[31] Myers CA, Slack T, Broyles ST, Heymsfield SB, Church TS, Martin CK. Diabetes prevalence is associated with different community factors in the diabetes belt versus the rest of the United States. Obesity (Silver Spring) 2017; 25(2):452-459. PMID: 28009108.

Keywords: diabetes mellitus, relative risk, health surveys, socioeconomic factors, prevalence