The Community Explorer: Informing Policy with County-Level Data

Lopez, Claude and Butler, Brittney

Milken Institute, Milken Institute

February 2021

The Community Explorer: Informing Policy with County-Level Data

Claude Lopez, PhD, and Brittney Butler

Milken Institute

Corresponding author: Claude Lopez clopez@milkeninstitute.org

Contents

In A Nutshell 3

Introduction 5

Data and Methodology 6

- Data 6
- Methodology 7

Community Profiles and Chronic Diseases 9

- Community Profile One 10
- Community Profile Two 12
- Community Profile Three 14
- Community Profile Four 16
- Community Profile Five 19
- Community Profile Six 21
- Community Profile Seven 23
- Community Profile Eight 25

Community Profiles to Inform Policy 27

Appendix: 29

- Appendix 1: Variable Definitions 29
- Appendix 2: Categories of Characteristics, Statistical Summary 30

In A Nutshell

The COVID-19 pandemic highlights the relationship between the high prevalence of some chronic diseases and a population's ability to fight the new threat. It underscores the importance of understanding the populations' behavioral, demographic, economic, and social features most at risk. Yet, most of the current narrative on health inequality focus on one factor, usually race or gender, at the national level.

This report proposes a new approach to investigate US health disparities that focuses on understanding populations' specificities before looking at their health profile. It first identifies the US's different populations or communities based on their behavioral, demographic, economic, and social profiles. Then it links these profiles to chronic disease prevalence rates.

The community explorer presents the eight profiles that account for a combination of factors when describing the populations. They can be summarized as follows:

- **Community Profile One** represents 38 percent of the US population and is the most ethnically diverse community with the highest income level. It is a highly educated cohort that resides in large metro areas.
- **Community Profile Two** represents 25 percent of the US population and consists of highly educated, economically prosperous, mostly White counties in metro areas.
- **Community Profile Three** represents 12 percent of the US population and has the largest Black population, whose income is the lowest income of all the profiles. These counties are primarily concentrated in the Southeast.
- **Community Profile Four** represents 7 percent of the US population and has the largest White population. This population reports the lowest income of all the profiles. It encompasses mostly rural counties in the East North Central, and Northeast regions.
- **Community Five** represents 6 percent of the US population and consists of predominantly White counties whose economies depend mostly on manufacturing and are located around the Midwest region.
- **Community Profile Six** represents 5 percent of the US population. It captures the youngest cohort of the profiles, with the largest Hispanic population and the lowest education level, access to healthy food, and health insurance. The counties are concentrated in the West and South-Central regions.
- **Community Profile Seven** represents 4 percent of the US population. Its cohort is the oldest of the profiles and consists of mostly White, elderly retirement communities.

- **Community Profile Eight** represents 3 percent of the US population and is the most rural cohort, consisting of an older White population with the most limited access to healthy food. The counties are mostly in the north part of the West, Midwest, and Northeast regions.

These community profiles' health outputs link the health differences across the US to the prevailing behavioral, demographic, economic, and social profiles of the population.

Our novel approach sorts the information of 26 behavioral, demographic, economic, and social factors across 3,192 US counties into eight community profiles. Ultimately, it leverages and makes sense of county-level information to create a dataset that can inform local and national policies.

This data-driven method informs policy issues using community profiles as reference groups and highlights similarities across the US counties, even when they are non-neighbors. It identifies : (i) what factors matter depending on the community profile and the health issues, (ii) whether the policy that focuses on influencing the relevant factors should be at the local, regional, or national level, and (iii) refined policy benchmarks to monitor the impact of the policy.

When it comes to public health policy, our findings advocate for coordinated efforts between national and local authorities with community partners such as health-care professionals, business and community leaders, schools, and child-care facilities.

.

Introduction

Countless reports and papers explain how behavioral, demographic, economic, and social factors impact health disparities.[1] However, most of them estimate the relationship between these factors based on pre-established models and use national-level data. This report proposes to (1) use an agnostic approach to recognize the interactions between these factors at the county level and (2) identify patterns across these interactions and then sort them into county-level specific profiles. The Milken Institute Community Explorer provides a geographic visualization of these profiles.

We then calculate the prevalence rate of the 10 most common chronic diseases for each community profile. The rates vary across communities, yet three community profiles report the highest prevalence rates. They represent 25 percent of the US population, equally split between the community profile with the largest Black population and two community profiles with primarily White populations. Our analysis confirms that health inequalities are associated with a combination of factors, including race, income level, single parenthood, pollution, access to healthy food, and city size. More importantly, it shows that these factors are combined differently across community profiles, allowing to link health output to behavioral, demographic, economic, and social profiles. We also identify which factor explains the most the change in the prevalence rates for each community profile.

By grouping US counties into community profiles that share behavioral, demographic, economic, and social features and providing their geographic location, we highlight similarities across the US counties, even when they are non-neighbors. We create a new data-driven method to inform policy issues using community profiles as reference groups. This approach leverages the refined understanding of local characteristics to inform policy: from its geographic scope to the factors it should target when influencing health outcomes.

1. See LaVeist (2005), Smedly et al. (2003), and Roux (2012), among others.

The approach identifies : (i) what factors matter depending on the community profile and the health issues, (ii) whether the policy that focuses on influencing the relevant factors should be at the local, regional, or national level, and (iii) refined policy benchmarks to monitor the impact of the policy. Ultimately, when it comes to public health policy, we advocate for coordinated efforts between national and local authorities with community partners such as health-care professionals, business and community leaders, schools, and child-care facilities.

.

Data and Methodology

Data

Our county-level data combine behavioral, demographic, economic, and social factors and the prevalence of the 10 most common chronic diseases: arthritis, cancer, chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), diabetes, hyperlipidemia (HLD), hypertension (HTN), ischemic heart disease (IHD), obesity, and stroke. Building the dataset required merging information from the following sources:

- Behavioral Risk Factor Surveillance System (BRFSS) surveys for the prevalence of chronic conditions at the county level
- Centers for Disease Control and Prevention (CDC) Diabetes Surveillance System database for county-level diabetes prevalence
- Centers for Medicare & Medicaid Services database on county-level chronic conditions
- Robert Wood Johnson Foundation's County Health Rankings dataset for county-level socioeconomic indicators
- United Health Foundation's America's Health Rankings for state-level chronic disease prevalence
- Census Bureau's American Community Survey for demographic and race-based income and poverty measures

- Bureau of Labor Statistics data on unemployment at the county level
- Department of Agriculture database on county typology (manufacturing sector dependence and retirement destination indicators)[2]

Methodology

Merging different data sets increases the amount of information and the number of dimensions considered in the analysis. Yet, too many dimensions challenge the ability to draw meaningful, policy-relevant inferences. To address this concern, we combine two data reduction methods in a three-step strategy that summarizes the population's information:

(i) First, at the factor level, by identifying the underlying relationship between the behavioral, economic, and social factors, we combine the factors into categories. We identified seven categories that sort 26 factors.[3]

(ii) Then, at the county level, we apply machine learning techniques to these categories across the US counties, reducing the 3,192 counties considered into eight community profiles.

(iii) Finally, we use the community profiles to estimate the relationship between the disease prevalence rates and the categories of factors defined in (i).

More specifically, the steps are as follows:

1. **Categories of Factors**: We group the behavioral, demographic, economic, and social factors in categories that capture the underlying trend of their combined effect, using Exploratory Factor Analysis.[4] This technique reduces the number of observable factors to fewer latent factors that are meaningful underlying constructs. The estimation identifies seven latent factors, or categories, that best describe the community profiles. The factors are defined in the appendix. The categories are as follows:[5]

2. Some rural counties with small population sizes have very imprecise direct estimates of prevalence. We use the modified James-Stein (1961) method; that is, we shrink the county prevalence estimate significantly more toward the state-level estimate, a more reliable measure.
3. We initially consider a larger number of factors and then drop the ones that do not improve the ability of the underlying constructs, or category, to account for the total variance.
4. Factor analysis, one of the most common inter-dependency techniques, is used when the relevant set of variables shows a systematic inter-dependence and the objective is to determine the latent factors that create a commonality.
5. The Exploratory Factor Analysis identified the combination of the factors in each category. We named the categories.

- **Age-dependency factors**: percentage of the population above age 18 and the percentage of the population under age 65;
- **Behavioral and social factors**: excessive drinking, smoking, some post-secondary education, single-parent households, and unemployment;
- **Black population factor**: percentage of the population that is Black;
- **Economic factors**: average income for the Black population, the White population, and the entire population, and percentages of Black and Hispanic populations experiencing poverty;
- **Hispanic or White population factors**: percentage of the population that is Hispanic or White, and percentage of adults without health insurance;
- **Physical environment factors**: level of pollution, limited access to healthy food for the low-income population, and reliance on manufacturing activity; and
- **Urban-rural factors**: housing concerns, population density, metropolitan area, rural area, violent crime rate, and the number of fast-food establishments per 100,000 people.

2. **Community Profiles**: To understand the data's hidden structure, especially because we do not know how counties' characteristics relate to one another, we use an unsupervised machine learning technique called hierarchical clustering. It uses the categories of factors to identify shared characteristics across counties and classifies them into coherent groups.[6] The clustering analysis results in the eight community profiles, discussed in the next section.

3. **Community Profiles and Chronic Diseases**: Finally, we estimate the strength of the relation between the chronic diseases' prevalence and the categories of factors for each community profile. First, we regress each disease's prevalence on the seven categories. We obtain R^2, which measures how much the categories included in the regression

6. We use Tibshirani et al.' (2001) gap statistics to identify the optimal number of groups.

explain the prevalence rate variance. Then, we use the relative importance estimation to identify which category contributes the most to R^2.[7] Such analyses aim to partition explained variance among the multiple categories to understand better the role played by each one in the regression. Johnson and Lebreton (2004) define "relative importance as the proportionate contribution each predictor makes to R^2, considering both the unique contribution of each predictor by itself and its incremental contribution when combined with the other predictors."

Community Profiles and Chronic Diseases

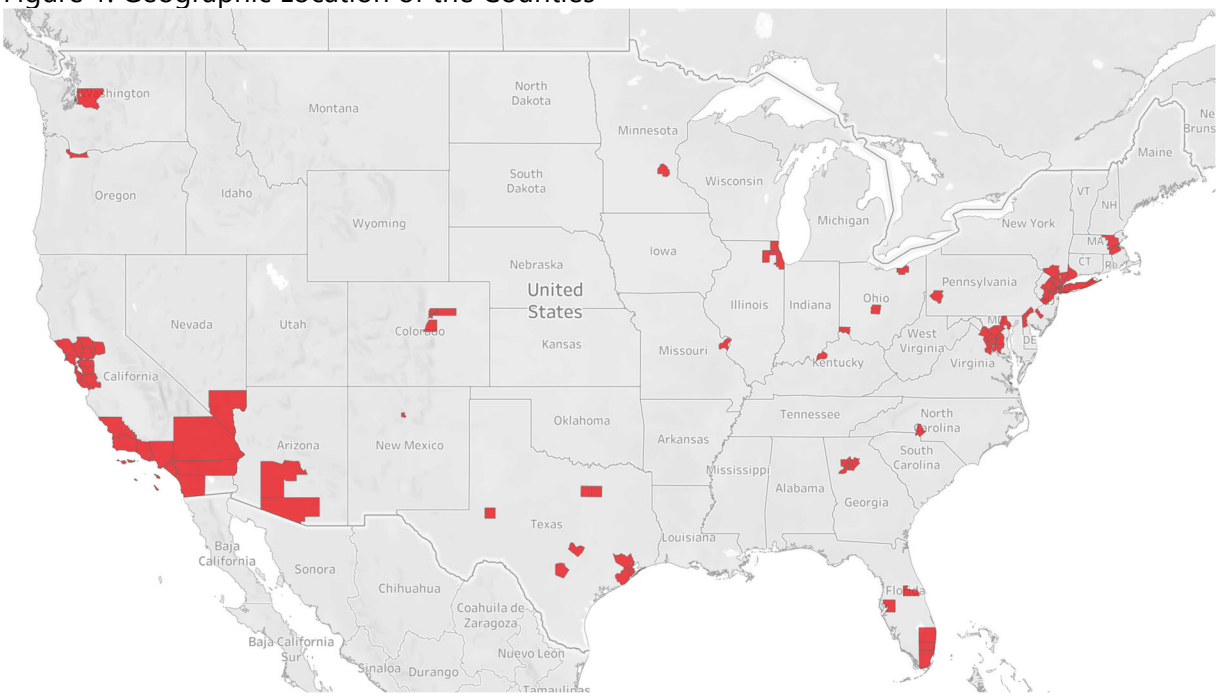
This section describes each community profile, using a map of the counties, the descriptive statistics reported in Appendix B, the chronic disease prevalence, and the outcome of the relative importance analysis.[8]

7. We use Grömping (2006, 2007) to calculate the relative importance.

8. Part of the analysis relies on the value of R^2. While R^2 provides an incomplete assessment of the relationship between the factors and the chronic diseases within a specific community, it allows us to contrast the different communities' features.

Community Profile One

Figure 1: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture

Community Profile One consists of large metropolitan counties that are, on average, the most ethnically diverse (Black, 14 percent; Hispanic, 23 percent; and White, 49 percent) and have the highest incomes of the eight community profiles. About 38 percent of the total US population resides in its 111 counties.

This community has the highest cancer prevalence (9 percent) and the lowest COPD (9 percent), diabetes (9 percent), and obesity (26 percent) rates among the eight communities.

The seven categories explain greater than 50 percent of the prevalence of two chronic diseases: obesity (68 percent) and CKD (61 percent). For both diseases, behavioral and social factors are essential in explaining the regression's good fit. These factors are important for six of the ten diseases, with R^2 ranging from 32 percent to 68 percent.

Compared to the other community profiles, the behavioral and social factors in this community are characterized by one of the lowest unemployment rates (3.6 percent) and

the lowest smoking rate (13.3 percent). Yet, it has the highest excessive alcohol consumption rate (19.4 percent). It also has the highest percentage of the population age 22-44 with some secondary education (70.5 percent), and the percentage of children living in a single-parent household is among the highest (30.7 percent).

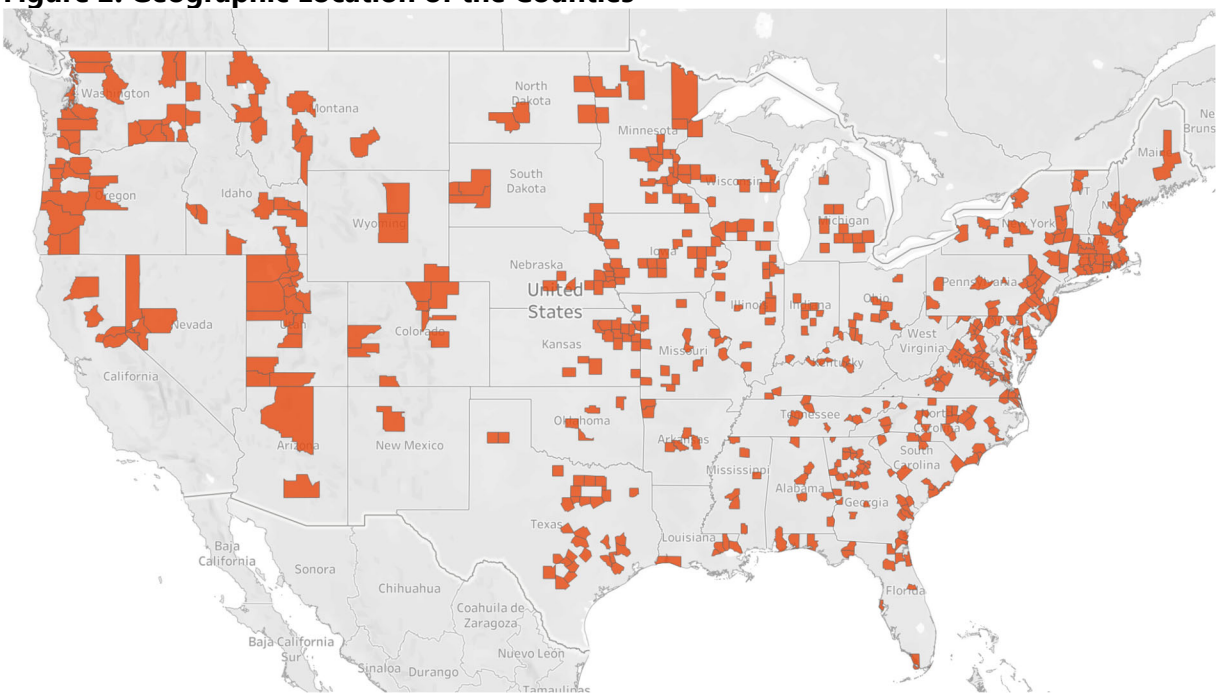
Table 1: Prevalence of Chronic Disease and Main Factors for Community Profile One

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	31 (33)	4 (4)	Physical Environment	39
Cancer	9 (8)	1 (1)	Age Dependency	47
Chronic Kidney Disease	24 (20)	10 (8)	Behavioral and Social	61
Chronic Obstructive Pulmonary Disease	9 (13)	-1 (1)	Behavioral and Social	41
Diabetes	9 (13)	1 (-2)	Behavioral and Social	44
Hyperlipidemia	40 (38)	-4 (-2)	Behavioral and Social	40
Hypertension	55 (57)	0 (-3)	Black Population	45
Ischemic Heart Disease	26 (33)	-4 (-3)	Behavioral and Social	32
Obesity	26 (32)	1 (2)	Behavioral and Social	68
Stroke	4 (2)	0 (1)	Black Population	37

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Two

Figure 2: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Two consists of economically prosperous and mostly White counties (79.2 percent) in metropolitan areas. About 25 percent of the US population resides in these 541 counties.

The prevalence of chronic diseases in this community follows the national average. The seven categories explain greater than 50 percent of Hypertension's prevalence (52 percent), and the physical environment factors are the most important in explaining R^2. These factors are important for five of the ten diseases whose R^2 ranges from 12 percent to 52 percent. Behavioral and social factors are important for three other diseases whose R^2 is close to 50 percent: diabetes (40 percent), obesity (44 percent), and COPD (47 percent).

The physical environment factors in this community are characterized by a higher-than-average level of pollution (an average of 9.2 polluted days compared to a US average of 9.1), better-than-average access to healthy food (6 percent of the low-income population does not have access to a grocery store compared to 8.3 percent for the US), and almost no reliance on the manufacturing industry.

This community's behavioral and social factors are characterized by the lowest unemployment rates (3.5 percent) of all communities. Excessive alcohol consumption is the second highest (19.2 percent), as is the percentage of the population age 22-44 with some college degree (67.8 percent). The percentage of children living in a single-parent household is below the US average (27.8 percent compared to 32.7 percent).

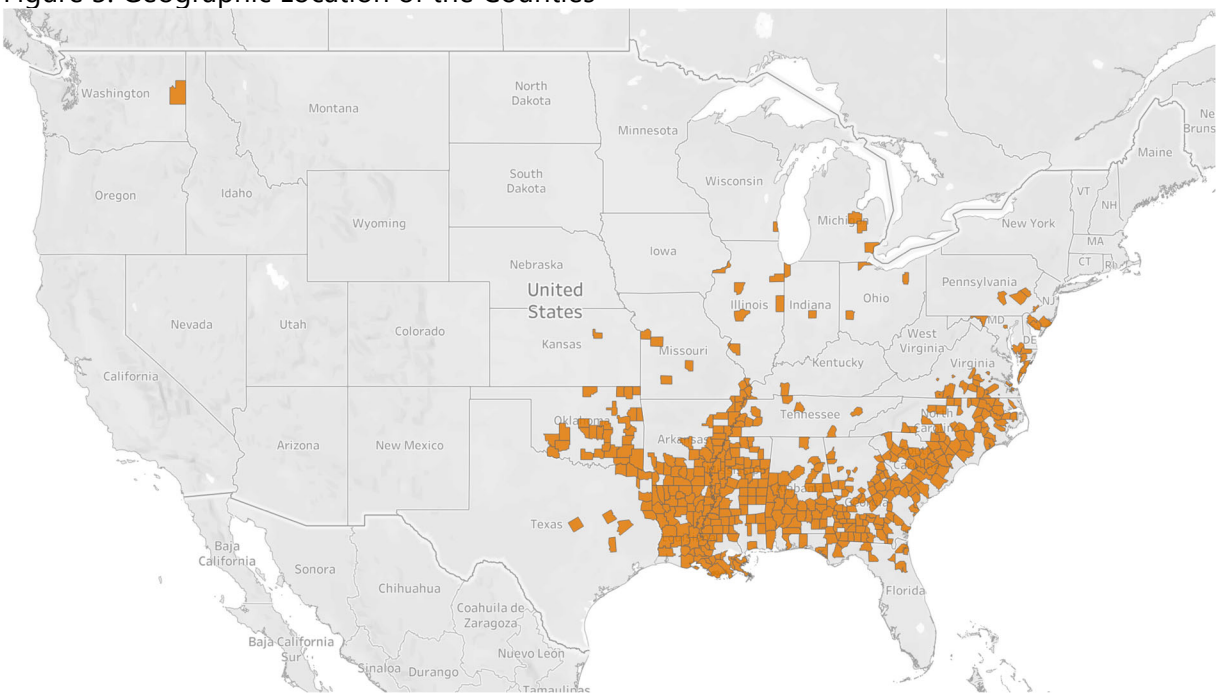
Table 2: Prevalence of Chronic Disease and Main Factors for Community Profile Two

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	32 (33)	6 (4)	Physical Environment	32
Cancer	8 (8)	0 (1)	Age Dependency	43
Chronic Kidney Disease	22 (20)	10 (8)	Physical Environment	39
Chronic Obstructive Pulmonary Disease	11 (13)	0 (1)	Behavioral and Social	47
Diabetes	10 (12)	1 (-2)	Behavioral and Social	40
Hyperlipidemia	38 (38)	-3 (-2)	Physical Environment	12
Hypertension	55 (57)	2 (-3)	Physical Environment	52
Ischemic Heart Disease	25 (27)	-2 (-3)	Physical Environment	26
Obesity	31 (32)	3 (2)	Behavioral and Social	44
Stroke	3 (3)	-1 (1)	Black Population	39

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Three

Figure 3: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Three consists of economically disadvantaged counties with the largest Black population (34.8 percent) among the eight communities. It has the lowest Black average household income ($18,193.47) and total average household income ($16,767.64). Its population has the highest rates of smoking (20.4 percent), unemployment (4.9 percent), single-parent households (46.3 percent), violent crime (455.96/100,000), and Hispanic poverty (33 percent). These 451 counties are primarily concentrated in the southeast region and account for 12 percent of the US population.

This community has the highest prevalence rates for most chronic diseases: arthritis (36 percent), CKD (26 percent), diabetes (15 percent), HLD (41 percent), HTN (65 percent), IHD (29 percent), and obesity (37 percent). However, the seven categories have a limited explanatory power on their prevalence: All the R^2 are less than 50 percent.[9] In this

[9] The low R^2 could also be the result of a lack of variance for each category across the counties of this community. However, the descriptive statistics reported in Appendis B show that is not the case.

community with the largest Black population, the Black population factor is important for COPD (R^2 of 43 percent) and IHD (R^2 of 32 percent).

This last point emphasizes the primary purpose of the factors selected: to help identify communities' different profiles, contrasting one from another. These factors are good proxies to synthesize complex differences across the US population. Yet, they may not be as useful to explain behaviors within each community: They narrow down the dimensions of interest and guide the focus of the community-specific analysis.

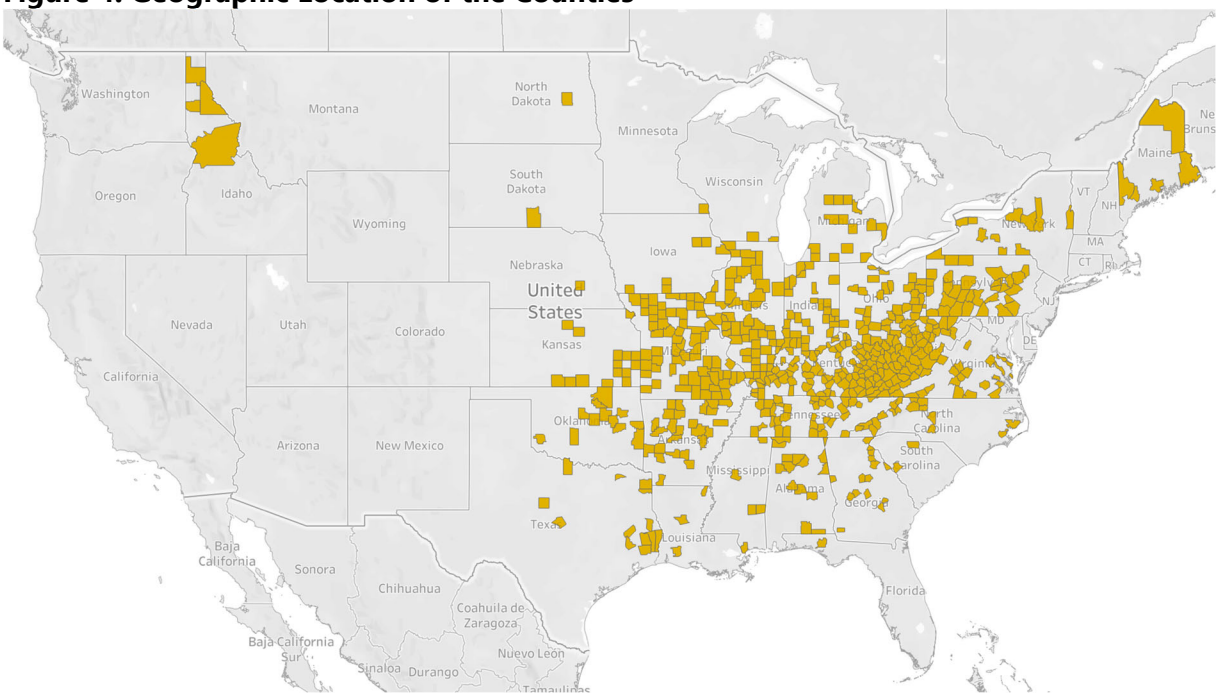
Table 3: Prevalence of Chronic Disease and Main Factors for Community Profile Three

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2(%)
Arthritis	36 (33)	10 (4)	Urban-Rural	15
Cancer	8 (8)	1 (1)	Urban-Rural	33
Chronic Kidney Disease	26 (20)	12 (8)	Urban-Rural	11
Chronic Obstructive Pulmonary Disease	13 (13)	0 (1)	Black Population	43
Diabetes	15 (12)	2 (-2)	Urban-Rural	24
Hyperlipidemia	41 (38)	1 (-2)	Physical Environment	19
Hypertension	65 (57)	4 (-3)	Behavioral and Social	29
Ischemic Heart Disease	29 (27)	-1 (-3)	Black Population	32
Obesity	29 (32)	3 (2)	Behavioral and Social	17
Stroke	4 (3)	0 (1)	Behavioral and Social	19

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Four

Figure 4: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Four consists of counties with, on average, the largest White population (88.9 percent) and the lowest average income ($46,611.17, compared to $57,265.00 for the US) of all the profiles. These are predominantly rural counties (70.5 percent) with the second-highest unemployment rate (4.6 percent) and a less-educated workforce (53 percent of the population age 22-44 with some secondary education compared to 58 percent for the US). Out of all eight communities, this one has the highest pollution level (on average 10.1 days per year) and the highest Black poverty level (35 percent). In contrast, the Black population represents, on average, only 5.1 percent of the community population, compared to 9.7 percent of the US population. It also has the lowest Hispanic population rate (3.03 percent) and White household income ($46,611). These 580 counties are mostly located in the Northeast Central and Northeast regions and account for 7 percent of the US population.

This community has the highest prevalence of arthritis (35 percent), COPD (16 percent), and HLD (41 percent) and among the highest rates of cardiovascular-related chronic diseases (29 percent for IHD and 60 percent for HTN), CKD (24 percent), diabetes (14 percent), and obesity (35 percent).

The seven categories explain greater than 50 percent of COPD prevalence (R^2 is 51 percent). Behavioral and social factors are the most important factors in explaining R^2. These factors are important for seven of the ten diseases whose R^2 ranges from 14 percent to 51 percent.

As discussed previously, this community's behavioral and social factors are characterized by the highest unemployment rates of all communities and one of the lowest education levels. Further, the smoking rate is the second highest (19.9 percent).

Table 4: Prevalence of Chronic Disease and Main Factors for Community Profile Four

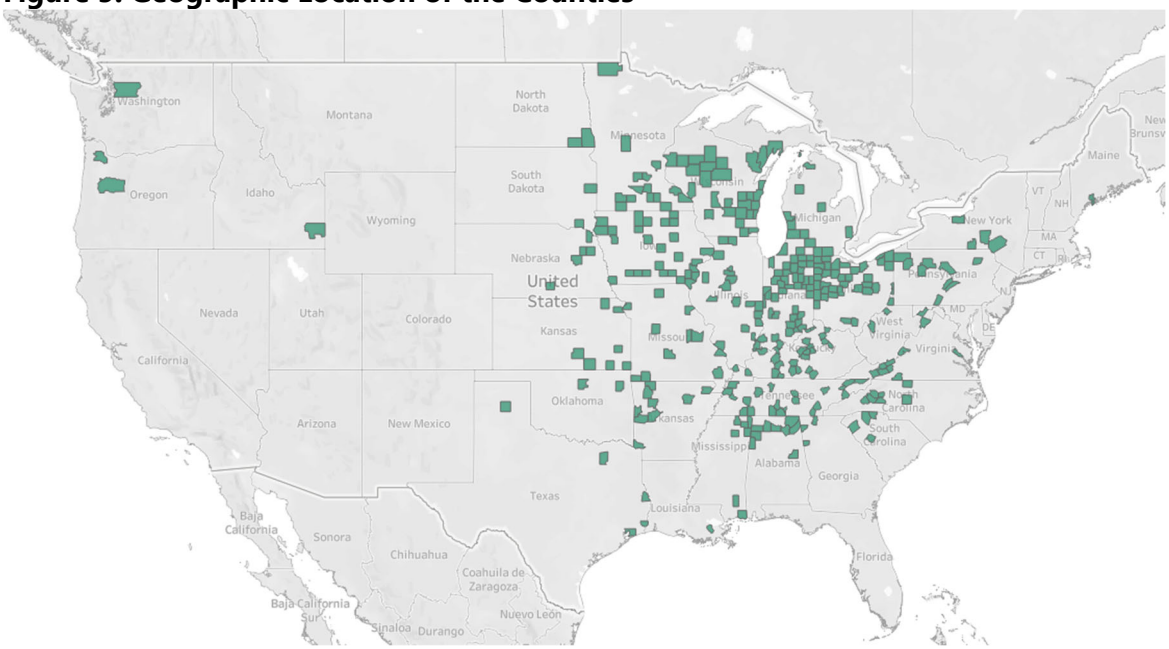
Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	35 (33)	7 (4)	Behavioral and Social	14
Cancer	7 (8)	0 (1)	Behavioral and Social	37
Chronic Kidney Disease	24 (20)	12 (8)	Physical Environment	24
Chronic Obstructive Pulmonary Disease	16 (13)	1 (1)	Behavioral and Social	51
Diabetes	14 (12)	3 (-2)	Behavioral and Social	18
Hyperlipidemia	41 (38)	-1 (-2)	Physical Environment	31
Hypertension	60 (57)	3 (-3)	Behavioral and Social	32

Ischemic Heart Disease	29 (27)	-3 (-3)	Behavioral and Social	23
Obesity	35 (32)	3 (2)	Behavioral and Social	14
Stroke	4 (3)	4 (1)	Physical Environment	16

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Five

Figure 5: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Five consists of predominantly white (88.2 percent) counties with the highest dependence on manufacturing employment and the lowest percentage of uninsured (an average of 10.5 percent compared to 14 percent for the US). The 334 counties are located mainly in the Midwest region and account for 6 percent of the US population.

This community has among the highest rates of CKD (24 percent), COPD (14 percent), HTN (58 percent), and HLD (40 percent). The seven categories explain greater than 50 percent of the prevalence of HTN (53 percent) and COPD (61 percent). In both cases, behavioral and social factors are the most important factors in explaining the regression fit. These factors are important for six of the ten diseases whose R^2 ranges from 13 percent to 61 percent. This community's behavioral and social factors are characterized by a below-average unemployment rate (3.8 percent compared to 4.1 percent) and the number of single-parent households (30.1 percent compared to 32.7 percent). However, it has a higher than

average smoking rate (18.2 percent compared to 17.2 percent) and excessive drinking (18.1 percent compared to 17.5 percent).

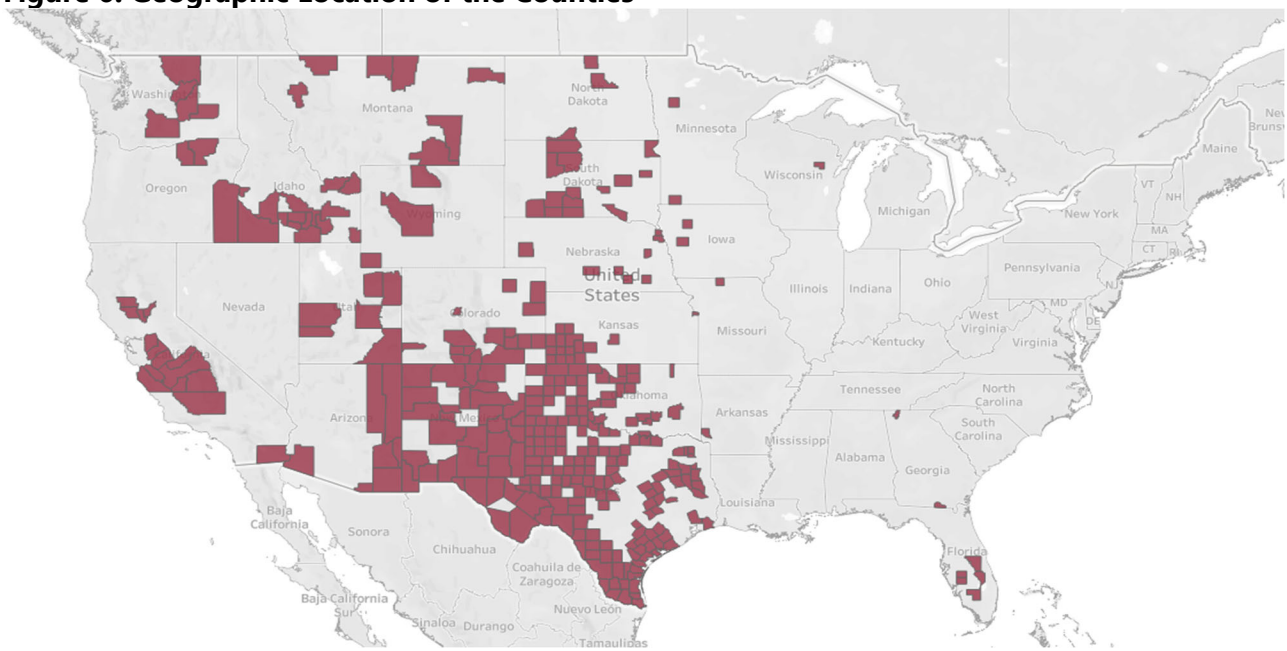
Table 5: Prevalence of Chronic Disease and Main Factors for Community Profile Five

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	33 (33)	6 (4)	Physical Environment	43
Cancer	7 (8)	0 (1)	Physical Environment	24
Chronic Kidney Disease	24 (20)	11 (8)	Behavioral and Social	28
Chronic Obstructive Pulmonary Disease	14 (13)	1 (1)	Behavioral and Social	61
Diabetes	12 (12)	1 (-2)	Behavioral and Social	32
Hyperlipidemia	40 (38)	-1 (-2)	Physical Environment	23
Hypertension	58 (57)	3 (-3)	Behavioral and Social	53
Ischemic Heart Disease	28 (27)	-2 (-3)	Behavioral and Social	39
Obesity	34 (32)	3 (2)	Behavioral and Social	13
Stroke	3 (3)	-1 (1)	Physical Environment	29

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Six

Figure 6: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Six includes counties with, on average, the largest Hispanic population (36.1 percent compared to 24.4 percent for the US) and the youngest (25.9 percent is under the age of 18, compared to 22.1 percent for the US). It has the highest number of uninsured (22.4 percent of adults do not have health insurance, compared to 14 percent for the US), the lowest level of education (49.3 percent have some post-secondary education, compared to 58 percent for the US), and the least access to healthy food (14.7 percent of the low-income population does not have access to a grocery store, compared to 8.3 percent for the US). The 343 counties account for 5.4 percent of the US population. They are concentrated in the West and Southwest-Central regions.

The prevalence of chronic diseases in this community is the lowest for arthritis (28 percent) and cancer (6 percent). The seven behavioral, demographic, economic, and social factors have a limited explanatory power on their prevalence: All the R^2 are less than 50 percent. Yet, physical environment factors are important for five of the ten diseases whose R^2 ranges from 22 percent to 42 percent.

As stated previously, this community's physical environment factors are characterized by the least access to healthy food. Further, the level of pollution is one of the lowest (7.7 polluted days per year compared to 9.1 for the US).

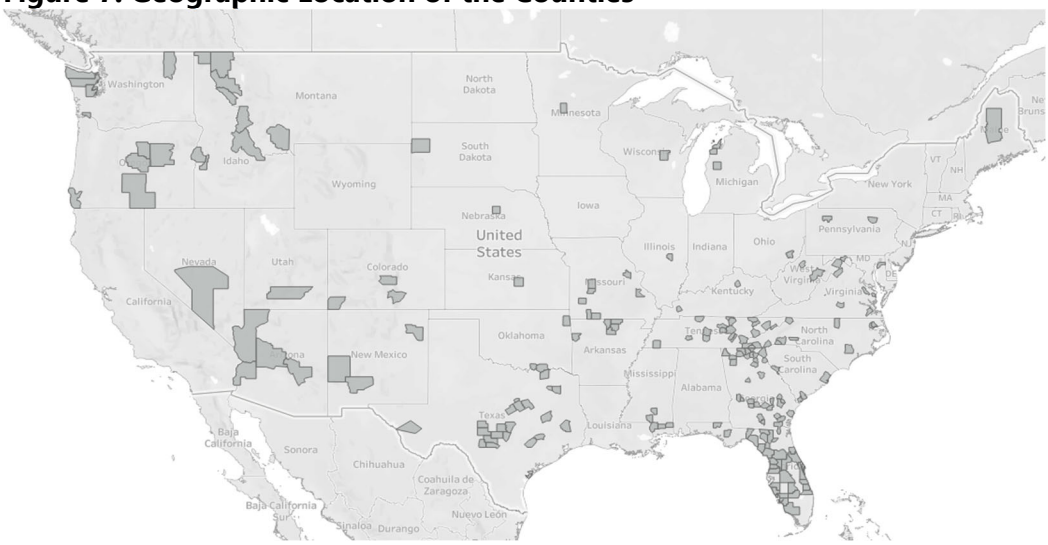
Table 6: Prevalence of Chronic Disease and Main Factors for Community Profile Six

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	28 (33)	5 (4)	Physical Environment	32
Cancer	6 (8)	-1 (1)	Physical Environment	25
Chronic Kidney Disease (CKD)	23 (20)	11 (8)	Hispanic or White Population	45
Chronic Obstructive Pulmonary Disease (COPD)	12 (13)	0 (1)	Physical Environment	22
Diabetes	11 (12)	1 (-2)	Behavioral and Social	14
Hyperlipidemia (HLD)	36 (38)	0 (-2)	Hispanic or White Population	39
Hypertension (HTN)	55 (57)	3 (-3)	Physical Environment	42
Ischemic Heart Disease (IHD)	28 (27)	-3 (-3)	Hispanic or White Population	34
Obesity	32 (32)	2 (2)	Age Dependency	29
Stroke	3 (3)	-1 (1)	Physical Environment	36

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Seven

Figure 7: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Seven is the oldest cohort (24.3 percent of the population is over 65, compared to 18.9 for the US) and is mostly White (78.5 percent). Its 198 counties represent 3.62 percent of the US population.

The prevalence of chronic diseases in this community follows the national average. The seven behavioral, demographic, economic, and social factors explain more than 50 percent of the prevalence of HTN (57 percent) and cancer (54 percent). The main factors are the physical environment and urban-rural. The physical environment factors are important for five of the ten diseases whose R^2 ranges from 31 percent to 57 percent.

The physical environment factors in this community are characterized by a level of pollution below the US average (8.8 polluted days per year compared to 9.1 for the US) and access to healthy food almost inline with the US average (8.8 percent of the low-income population does not have access to a grocery store, compared to 9.1 percent for the US).

The urban-rural factors in this community are characterized by a population density per county much lower than the US average (on average 61,152.3 compared to 197,568.6 for

the US) and the second-highest number of fast-food locations (525.2 per 100,000 compared to 370.6 per 100,000 for the US).

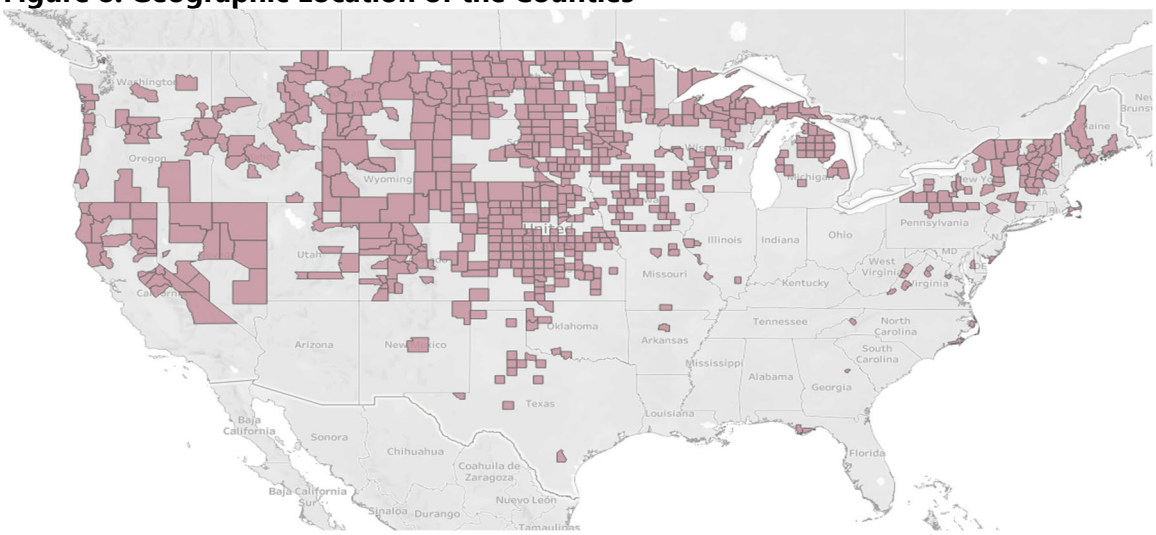
Table 7: Prevalence of Chronic Disease and Main Factors for Community Profile Seven

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2(%)
Arthritis	34 (33)	7 (4)	Physical Environment	36
Cancer	8 (8)	1 (1)	Urban-rural	54
Chronic Kidney Disease	23 (20)	11 (8)	Physical Environment	49
Chronic Obstructive Pulmonary Disease	14 (13)	1 (1)	Behavioral and Social	49
Diabetes	13 (12)	2 (-2)	Behavioral and Social	20
Hyperlipidemia	40 (38)	-2 (-2)	Physical Environment	41
Hypertension	57 (57)	2 (-3)	Physical Environment	57
Ischemic Heart Disease	27 (27)	-3 (-3)	Physical Environment	31
Obesity	32 (32)	2 (2)	Age-dependency	39
Stroke	4 (3)	0 (1)	Black Population	27

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Eight

Figure 8: Geographic Location of the Counties



Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profile Eight regroups the most rural (76 percent), second oldest (22.3 percent age 65 and older), and predominantly White (87.5 percent) cohort with the most limited access to healthy food (11.1 percent of the low-income population has limited access to a grocery store, while there are on average 1,322.6 fast-food locations per 100,000 habitants, compared to 8.3 percent and 370.6 for the US, respectively). Finally, it has the lowest violent crime rate (166 per 100,000 compared to 370.8 for the US) and the least polluted environment (6.9 pollution days per year compared to 9.1 for the US). The 634 counties in this community account for 3 percent of the US population.

The prevalence of chronic diseases in this community is among the lowest in the US, except for obesity (31 percent). The seven behavioral, demographic, economic, and social factors have a limited explanatory power on their prevalence: All the R^2 are less than 50 percent. Yet, physical environment factors are important for six of the ten diseases whose R^2 ranges from 14 percent to 39 percent.

As stated previously, this community's physical environment factors are characterized by the least access to healthy food and the least polluted environment. These counties' rural

location may make access to grocery stores less critical than urban settings in order to have access to a healthy diet.

Table 8: Prevalence of Chronic Disease and Main Factors for Community Profile Eight

Chronic Disease	Prevalence % (National %)		Significant Factors	
	2018	Change since 2009	Main Factor	R^2 (%)
Arthritis	30 (33)	5 (4)	Physical Environment	14
Cancer	7 (8)	-1 (1)	Hispanic or White Population	14
Chronic Kidney Disease	19 (20)	8 (8)	Physical Environment	18
Chronic Obstructive Pulmonary Disease	11 (13)	1 (1)	Behavioral and Social	26
Diabetes	10 (12)	1 (-2)	Behavioral and Social	18
Hyperlipidemia	31 (38)	-3 (-2)	Physical Environment	39
Hypertension	48 (57)	0 (2)	Physical Environment	31
Ischemic Heart Disease	23 (27)	-3 (-3)	Physical Environment	26
Obesity	31 (32)	2 (2)	Behavioral and Social	22
Stroke	3 (3)	0 (1)	Physical Environment	16

Source: Authors' calculations using the BRFSS, CDC's Diabetes Surveillance System, Centers for Medicare & Medicaid Services, Robert Wood Johnson Foundation's County Health Rankings, United Health Foundation's America's Health Rankings, Census Bureau's American Community Survey, Bureau of Labor Statistics, Department of Agriculture.

Community Profiles to Inform Policy

The health outputs of these community profiles link the health differences across US to the prevailing behavioral, demographic, economic, and social profiles of the population.

Let us look at the average prevalence rates for the 10 most common chronic diseases in the US for each community profile. They vary considerably across communities. Three profiles report the highest rates for most chronic diseases: Profiles Three, Four, and Five. They account for 25 percent of the US population. Furthermore, profile Three has the largest Black population (35 percent), while profiles Four and Five's population is more than 88 percent white. Profiles Three and Four have the two lowest incomes among all communities, the two highest unemployment rates. In contrast, profile Five has a relatively low unemployment rate and the lowest percentage of population without health insurance. Appendix B provides more detailed information, highlighting that factors such as single parenthood, pollution, access to healthy food, and city size also differ across these three profiles.

Results reported in Tables 3, 4, and 5 show that the behavioral and social factors have the most influence on disease prevalence rates across the three community profiles. When it is not the case, the most influential factors are community-profiles specific.

These communities cover a large part of the US, from Wisconsin, Iowa, Missouri, Arkansas, Oklahoma, and Louisiana to Pennsylvania, Virginia, North and South Carolina, Georgia, and Florida. Often the three profiles are present in one state.

By grouping US counties into community profiles that share behavioral, demographic, economic, and social features and providing their geographic location, this new approach highlights similarities across the US counties, even when they are non-neighbors. It leverages the refined understanding of local characteristics to inform policy: from its geographic scope to the factors it should target when influencing health outcomes.

The community profiles bridge the gap between local data and national trends by identifying similar populations across counties. Allowing for a unique interaction among the behavioral, demographic, economic, and social factors within each profile leads to three main benefits for policymakers.

1. **No forced factors interactions:** machine learning techniques allow us to process an extensive amount of information and group the factors that could influence that population's health outcome. Standard econometrics approaches require hypotheses on how health determinants should interact as they cannot process the same amount of information.
2. **Peer-Counties Benchmarking:** this pragmatic approach provides refined benchmarks to policymakers and policy implementors: for each community profile, the factors and corresponding health outputs serve as reference values and information for the community counties. These benchmarks allow comparisons among counties with relatively similar features. It provides meaningful benchmarks for assessing the impact of policy across and within community profiles.
3. **Policy geographic scope:** the factors' importance for a specific health issue across several community profiles call for policy initiatives at the national level. In contrast, if only one or a few community profiles report such a relation between the factors and the health condition, then the policy initiatives should be more local or a combination of local and national levels.

Finally, the community approach suggested is not limited to the health determinants and conditions used in this analysis. The community profiles build on populations' characteristics relevant to any policy issues that may have a regional or local component.

.

Appendix:

Appendix 1: Variable Definitions

Age-Dependency	
Over 65 (%)	Persons over 65 years and over, percent
Under 18 (%)	Persons under 18 years, percent
Behavioral and Social	
Excessive Drinking (%)	Percentage of adults reporting binge or heavy drinking
Single-Parent Households (%)	Percentage of children that live in a household headed by a single parent
Smoking (%)	Percentage of adults who are current smokers
Some College (%)	Percentage of adults ages 25-44 with some post-secondary education, such as enrollment in vocational/technical schools, junior colleges, or four-year colleges. It includes individuals who pursued education following high school but did not receive a degree as well as those who attained degrees
Unemployment Rate (%)	Percentage of population ages 16 and older unemployed but seeking work
Black Population Factors	
Black (%)	Percentage of population that is Black or Black alone
Economic Factors	
Average Household Income ($)	Average household income in US dollars of entire population
Black Average Household Income ($)	Average household income in US dollars of Black population
Black Poverty Rate (%)	Percentage of Black population that are experiencing poverty
Hispanic Poverty Rate (%)	Percentage of Hispanic population that are experiencing poverty
White Average Household Income ($)	Average household income in US dollars of White population
Hispanic or White Population Factors	
Hispanic (%)	Percentage of population that is of Hispanic origin
Uninsured Adults (%)	Percentage of adults under age 65 without health insurance
White (%)	Percentage of population that is White alone
Physical Environment Factors	
Average Polluted Days (#)	Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5)
Limited Access to Healthy Food (%)	Percentage of population who are low-income and do not live close to a grocery store
Manufacturing (%)	23 percent or more of average annual labor and proprietors' earnings derived from manufacturing or 16 percent of total employment during 2010-12
Urban-Rural factors	
Fast Food Locations per 100,000	Number of fast food locations per 100,000 population
Rural (%)	Rural-urban continuum code definition 4-9, 88, and 99
Metro (%)	Rural-urban continuum code definition 1-3
Population (#)	Total population
Severe Housing Cost (%)	Percentage of households with at least one of four housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities
Violent Crime Rate (#)	Number of reported violent crime offenses per 100,000 population

Appendix 2: Categories of Characteristics, Statistical Summary

Appendix B: Prevalence and Statistical Summary of Relevant Variables

	Factors	Age-Dependency		Behavioral and Social					Black Pop.	Economic					Hispanic or White Pop.		
Community Profiles		Over 65 (%)	Under 18 (%)	Excessive Drinking (%)	Single-Parent Households (%)	Smoking (%)	Some College (%)	Unemployment Rate (%)	Black (%)	Average Household Income ($)	Black Average Household Income ($)	Black Poverty Rate (%)	Hispanic Poverty Rate (%)	White Average Household Income ($)	Hispanic (%)	Uninsured Adults (%)	White (%)
1	Mean	14.56	22.35	19.44	30.69	13.26	70.49	3.62	14.4	61,379.69	60,143.4	18	17	87,658.32	22.98	10.95	48.98
	(Std Dev.)	(2.84)	(2.7)	(-2.53)	(-8.79)	(2.58)	(8.56)	(0.84)	(12.12)	(15,009.9)	(25,413.67)	(8)	(6)	(21,256.35)	(13.66)	(5.38)	(15.42)
2	Mean	16.72	22.24	19.19	27.76	15.56	67.75	3.54	7.06	52,537.35	47,248.26	23	20	65,822.84	8.25	10.93	79.22
	(Std Dev.)	(3.87)	(3.41)	(3.02)	(7.05)	(2.56)	(8.69)	(0.87)	(8.19)	(16,767.84)	(18,193.47)	(13)	(9)	(13,628.79)	(7.76)	(4.71)	(12.12)
3	Mean	17.71	22.43	14.89	46.27	20.35	51.57	4.91	34.76	38,152.07	28,989.94	33	33	50,512.78	5.61	16.45	55.01
	(Std Dev.)	(3.05)	(2.46)	(2.78)	(10.61)	(2.57)	(10.61)	(1.24)	(18.1)	(14,436.5)	(7,191.44)	(9)	(19)	(8,577.97)	(4.67)	(4.71)	(15.15)
4	Mean	20.11	21.2	16.38	32.29	19.88	52.98	4.62	5.07	45,372.32	30,725.2	35	28	46,611.17	3.03	11.79	88.87
	(Std Dev.)	(2.94)	(2.16)	(2.63)	(6.57)	(3.08)	(8.81)	(1.32)	(7.77)	(22,774.81)	(11,949.02)	(24)	(20)	(8,579.17)	(2.34)	(4.66)	(9.23)
5	Mean	18.95	22.55	18.12	30.13	18.18	57.21	3.79	3.36	45,459.41	36,704.2	28	26	52,845.28	5.18	10.49	88.22
	(Std Dev.)	(2.61)	(2.18)	(3.32)	(6.82)	(2.78)	(9.67)	(1.03)	(4.47)	(14,776.06)	(15,346.83)	(19)	(15)	(9,372.86)	(4.49)	(4.54)	(7.85)
6	Mean	16.71	25.93	17.09	34.28	17	49.3	4.46	3.4	44,402.6	45,405.92	27	24	51,737.35	36.05	22.36	49.61
	(Std Dev.)	(4.48)	(4.44)	(2.39)	(11.28)	(4.87)	(9.29)	(2.48)	(4.48)	(17,210.5)	(22,465.96)	(25)	(10)	(12,610.29)	(23.56)	(6.66)	(21.61)
7	Mean	24.53	19.17	15.82	34.02	17.61	50.77	4.35	8.47	43,609.69	35,855.98	35	26	47,975.76	8.94	17.41	78.52
	(Std Dev.)	(6.77)	(3.69)	(2.13)	(8.49)	(3.12)	(9.96)	(1.19)	(11.08)	(15,388.06)	(14,795.84)	(24)	(14)	(8,810.56)	(8.08)	(5.35)	(13.5)
8	Mean	22.28	21	19.19	26.4	15.39	64.14	3.65	1.27	54,181.79	47,355.76	29	21	54,956.83	5.95	11.39	87.52
	(Std Dev.)	(4.5)	(3.23)	(2.52)	(8.77)	(2)	(9.56)	(1.45)	(1.93)	(33,784.21)	(25,743.06)	(30)	(17)	(11,087.8)	(6.1)	(4.38)	(10.3)

Appendix B: Prevalence and Statistical Summary of Relevant Variables (cont.)

	Factors	Physical Environment			Urban-Rural					
Community Profiles		Average Polluted Days (#)	Limited Access to Healthy Food (%)	Manufacturing (%)	Fast Food Locations per 100,000	Metro (%)	Population (#)	Rural (%)	Severe Housing Cost (%)	Violent Crime Rate (#)
1	Mean	10.01	3.8	1	29.44	99	1,098,334.67	4.19	16.89	384.37
	(Std Dev.)	(1.77)	(2.68)	(9)	(78.7)	(9)	(1,271,736.14)	(5.53)	(4.24)	(220.69)
2	Mean	9.18	5.98	2	103.55	88	157,328.07	36.9	12.21	223.93
	(Std Dev.)	(1.67)	(3.7)	(15)	(416.46)	(32)	(189,940.69)	(25.84)	(3.54)	(130.21)
3	Mean	10.12	9.88	19	245.72	37	94,012.91	54.05	13.69	455.96
	(Std Dev.)	(0.84)	(6.78)	(39)	(1044.72)	(48)	(182,037.16)	(30.1)	(3.29)	(268.23)
4	Mean	10.2	6.19	3	278.14	35	41,603.57	70.46	10.39	203.65
	(Std Dev.)	(1.11)	(5.21)	(16)	(706.43)	(48)	(57,870.86)	(24.99)	(2.31)	(129.7)
5	Mean	10.19	5.79	100	134.76	32	55,047.83	58.82	9.43	204.66
	(Std Dev.)	(1.48)	(4.24)	(0)	(367.79)	(47)	(87,633.49)	(23.69)	(2.03)	(154.31)
6	Mean	7.68	14.74	70	429.28	20	54,305.6	54.55	10.37	272.55
	(Std Dev.)	(1.96)	(12.98)	(25)	(901.81)	(40)	(127,465.57)	(31.58)	(3.59)	(188.91)
7	Mean	8.83	8.56	19	525.17	24	61,152.27	67.25	11.85	255.36
	(Std Dev.)	(1.62)	(8.46)	(39)	(1343.16)	(43)	(113,644.9)	(28.4)	(2.66)	(153.74)
8	Mean	6.88	11.06	1	1322.56	4	18,764.16	76	9.99	166.01
	(Std Dev.)	(1.55)	(10.41)	(11)	(5010.68)	(21)	(26,019.01)	(27.35)	(3.54)	(135.88)

REFERENCES

Diez Roux, Ana V. "Conceptual Approaches to the Study of Health Disparities." *Annual Review of Public Health* 33, no: 1 (2012): 41-58.

Grömping, Ulrike. "Relative Importance for Linear Regression in R: The Packagerelaimpo." *Journal of Statistical Software* 17, no.1 (2006).

Grömping, Ulrike "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition." *The American Statistician* 61, no.2 (2007): 139-147.

Johnson, J.W. and LeBreton, J.M. "History and Use of Relative Importance Indices in Organizational Research." *Organizational Research Methods* 7 (2004): 238-257.

LaVeist, Thomas A. "Disentangling Race and Socioeconomic Status: A Key to Understanding Health Inequalities." *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 82 (2005): iii26-iii34.

Milken Institute. "Community Explorer." 2021. (Community-Explorer (miresearch.github.io))

Smedley, Brian D., Adrienne Y. Stith, and Alan R. Nelson. "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care." *Choice Reviews Online* 40, no. 10 (2003) doi:10.5860/choice.40-5843.

Tibshirani, R., Walther, G., and Hastie, T. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, no. 2 (2001): 411-423. doi:10.1111/1467-9868.00293.

Yong, A. G., and Pearce, S. "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis." *Tutorials in Quantitative Methods for Psychology* 9, no. 2 (2013): 79-94. doi:10.20982/tqmp.09.2.p079.

ACKNOWLEDGMENTS

The authors would like to thank Keith Savard, Jihad Dagher, and the Milken Institute Center of Public Health for insightful discussions about this topic. They also would like to thank participants at the Milken Institute Research Department brown bag meetings for useful discussions.

ABOUT THE AUTHORS

Claude Lopez, PhD, is the head of the Research Department at the Milken Institute. She leads data-driven efforts to influence global policy issues on International Finance, Health Economics, and Regional Economics. She is an active member of the T20 task force on international financial architecture for stability and development and a contributor to W20 (Women 20), two advisory committees to the G20. Lopez has more than 20 years of experience in academic and policy research in the US and abroad. Before joining the Institute, Lopez headed multiple research teams at the Banque de France, the nation's central bank, and was an economics professor at the University of Cincinnati. She has an MS in econometrics from the Toulouse School of Economics and a PhD in economics from the University of Houston.

Brittney Butler is a health economics research analyst within the Research Department at the Milken Institute. Her current work focuses on health equity and disparities within the United States, considering social, demographic, economic, and environmental factors. Butler holds her bachelor's degree in integrative biology from the University of California, Berkeley, and a master of science in global health from the University of California, San Francisco.