

Protocol

The Diabetes Location, Environmental Attributes, and Disparities Network: Protocol for Nested Case Control and Cohort Studies, Rationale, and Baseline Characteristics

Annemarie G Hirsch¹, MPH, PhD; April P Carson², MPH, PhD; Nora L Lee³, PhD; Tara McAlexander³, MPH, PhD; Carla Mercado⁴, PhD; Karen Siegel⁴, MPH, PhD; Nyasha C Black⁵, PhD; Brian Elbel⁶, MPH, PhD; D Leann Long⁷, MS, PhD; Priscilla Lopez⁶, MPH; Leslie A McClure³, MS, PhD; Melissa N Poulsen¹, MPH, PhD; Brian S Schwartz^{1,8}, MS, MD; Lorna E Thorpe⁶, MPH, PhD

¹Department of Population Health Sciences, Geisinger, Danville, PA, United States

²Department of Epidemiology, University of Alabama at Birmingham School of Public Health, Birmingham, AL, United States

³Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health, Philadelphia, PA, United States

⁴Centers for Disease Control and Prevention, Atlanta, PA, United States

⁵Noire Analytics, LLC, Birmingham, AL, United States

⁶Department of Population Health, NYU Langone Health, New York, NY, United States

⁷Department of Biostatistics, University of Alabama at Birmingham School of Public Health, Birmingham, AL, United States

⁸Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:

Annemarie G Hirsch, MPH, PhD

Department of Population Health Sciences

Geisinger

100 N Academy Avenue

Danville, PA, 17822

United States

Phone: 1 267 626 8110

Email: aghirsch@geisinger.edu

Abstract

Background: Diabetes prevalence and incidence vary by neighborhood socioeconomic environment (NSEE) and geographic region in the United States. Identifying modifiable community factors driving type 2 diabetes disparities is essential to inform policy interventions that reduce the risk of type 2 diabetes.

Objective: This paper aims to describe the Diabetes Location, Environmental Attributes, and Disparities (LEAD) Network, a group funded by the Centers for Disease Control and Prevention to apply harmonized epidemiologic approaches across unique and geographically expansive data to identify community factors that contribute to type 2 diabetes risk.

Methods: The Diabetes LEAD Network is a collaboration of 3 study sites and a data coordinating center (Drexel University). The Geisinger and Johns Hopkins University study population includes 578,485 individuals receiving primary care at Geisinger, a health system serving a population representative of 37 counties in Pennsylvania. The New York University School of Medicine study population is a baseline cohort of 6,082,146 veterans who do not have diabetes and are receiving primary care through Veterans Affairs from every US county. The University of Alabama at Birmingham study population includes 11,199 participants who did not have diabetes at baseline from the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study, a cohort study with oversampling of participants from the Stroke Belt region.

Results: The Network has established a shared set of aims: evaluate mediation of the association of the NSEE with type 2 diabetes onset, evaluate effect modification of the association of NSEE with type 2 diabetes onset, assess the differential item functioning of community measures by geographic region and community type, and evaluate the impact of the spatial scale used to measure community factors. The Network has developed standardized approaches for measurement.

Conclusions: The Network will provide insight into the community factors driving geographical disparities in type 2 diabetes risk and disseminate findings to stakeholders, providing guidance on policies to ameliorate geographic disparities in type 2 diabetes in the United States.

International Registered Report Identifier (IRRID): DERR1-10.2196/21377

(*JMIR Res Protoc* 2020;9(10):e21377) doi: [10.2196/21377](https://doi.org/10.2196/21377)

KEYWORDS

type 2 diabetes; built environment; social environment; disparities

Introduction

Background

An estimated 10.5% of the US population has diabetes, and these 34 million individuals are at an increased risk for coronary artery disease, cerebrovascular disease, and other complications [1,2]. Approximately 90% to 95% of people with diabetes have type 2 diabetes (T2D) [1]. Another 88 million individuals have prediabetes, defined as having elevated glucose levels above normal but below the threshold for diabetes, and are at elevated risk of developing T2D [3,4]. Diabetes prevalence and incidence vary substantially by geographic region [5-7]. In 2013, there was a six-fold difference between counties with the lowest and highest diabetes prevalence [6]. A large body of literature links community and environmental factors (hereafter referred to as community factors) to T2D and obesity, one of the risk factors for T2D [8-15]; however, the mechanisms for these links remain poorly understood. Moreover, there are inconsistencies in this body of literature. Identifying the community factors driving T2D disparities and the pathways through which these factors influence T2D is essential to informing geographically targeted policy interventions that reduce the risk of T2D and related outcomes.

Researchers have identified consistent associations of community-level socioeconomic factors (eg, community poverty rate) with T2D prevalence and incidence. However, findings related to aspects of the built (eg, food and physical activity establishment environment, land use environment) and natural environment (eg, greenness) and T2D risk have been less uniform [9-19]. This may be due, in part, to methodologic variations in measuring community factors, including differences in spatial scales, data sources, and measurement approaches [20].

Among the challenges to creating a cohesive body of research is a lack of consistent approaches to conceptualizing and operationalizing the geographic area in which community factors are thought to be relevant to health [19-22]. Furthermore, the size and boundaries of spatial scales most relevant to health may vary according to community type (eg, across the gradient from urban to rural). Community type is also an important consideration in measurement development, as measurement of the same community factors may require different approaches [23]. For example, car ownership may be a basic necessity for individuals living in rural areas but more of a luxury for individuals living in urban areas with good public transportation options. Thus, car ownership may work differently as an indicator of the neighborhood socioeconomic environment

(NSEE) in urban versus rural areas [24]. This differential item functioning may contribute to inconsistencies observed in the literature as the same measure (eg, proportion who own a car) could hold different meanings in different community types.

Community and health research is also vulnerable to structural confounding, which occurs when individual and contextual factors strongly predict residence in a certain community [25,26]. In the presence of structural confounding, certain measures, owing to social sorting, are largely nonoverlapping, resulting in an inability to examine their independent influences on the outcome of interest. For example, in some settings, the distribution of persons across categories of the NSEE and racial residential segregation may reveal a lack of comparable groups across key strata, resulting in analytical challenges of nonpositivity that prohibit causal contrasts across levels of exposure [26]. Finally, capturing the complex interactions among multiple community factors on health can be challenging [27]. As a result, previous research has largely evaluated co-occurring community factors in isolation.

Objectives

The increasing availability of longitudinal, individual-level data from electronic health record (EHR) networks [28] and cohort studies, coupled with advances in geographic information systems (GISs), provides new opportunities to examine the effects of community factors on health. In 2017, the Diabetes LEAD (Location, Environmental Attributes, and Disparities) Network was established to identify the contributions of modifiable community factors on T2D risk. The Network includes researchers from 4 academic institutions who collaborate to address the methodological challenges previously described to investigate a range of community factors across the United States. The Network aims to guide policy decision making to reduce the burden of T2D across the United States. This paper aims to describe the Diabetes LEAD Network, its study populations, and the methodologies used to investigate the community factors that are associated with T2D onset and related outcomes.

Methods

Network Overview

The Diabetes LEAD Network is a research collaboration of 4 academic centers: Drexel University, Geisinger and Johns Hopkins University (G/JHU), New York University School of Medicine (NYU), and the University of Alabama at Birmingham (UAB). The Centers for Disease Control and Prevention (CDC) funded the Network to bring together institutions with diverse

but complementary expertise and a rich array of data assets. Three study sites—G/JHU, NYU, and UAB—use longitudinal data, such as EHRs, administrative claims, and survey data on distinct populations and geographies in the United States (Tables 1-3; Figures 1-3). Drexel, the data coordinating center (DCC), is leading the development of a set of harmonized community factors, health outcomes, and analysis plans (Tables 4 and 5) that will be applied to each study site's cohort and geography.

Each site has its own set of study aims that examine community factors and T2D outcomes, including T2D onset, obesity, and other cardiometabolic conditions. Working collaboratively, the study sites and the CDC also developed a shared set of aims that complement site-specific aims (Textbox 1). We first describe the shared Network aims and then describe the site-specific aims.

Table 1. Diabetes Location, Environmental Attributes, and Disparities Network study site populations for Network-specific aims.

Study site, G/JHU ^a (n=578,485 ^b)	NYU ^c (n=6,082,246)	UAB ^d (n=11,199)
Study design		
Nested case control and cohort	Cohort ^e	Cohort
Source population		
All Geisinger patients (N=1,605,922)	All patients in the Veterans Affairs EHR ^f with at least one primary care visit since 1999 (N=8,346,280)	REGARDS ^g participants at baseline (2003-2007; N=30,239)
Exclusion criteria (sample size excluded)		
Patients with <2 primary care visits ^h (January 1, 2006 - December 31, 2016; n=970,785)	Patients with <2 primary care visits at least 30 days apart during the 5 years (before January 1, 2008; n=4,270,462)	Participants without a second visit (2013-2016; n=14,089)
Patients with a residential address outside of one of the 37 counties in the Geisinger primary service area (n=56,652)	Patients with diabetes ⁱ before or on January 1, 2008 (n=1,049,423)	Participants with diabetes ⁱ at baseline (n=2729) or missing diabetes status at baseline (n=521; combined, n=3250)
N/A ^j	N/A	Participants missing diabetes status at second visit (n=1580)
N/A	N/A	Unable to assign census tract using RECVD ^k data set (n=121)

^aG/JHU: Geisinger and Johns Hopkins University.

^bStudy population will vary based on study design (ie, nested case control or cohort).

^cNYU: New York University.

^dUAB: University of Alabama at Birmingham.

^eThese numbers reflect only those entering the cohort on inception date (January 1, 2008). The dynamic cohort allowed patients to enter the cohort through December 31, 2016 (n=3,113,391). Total cohort population was 6,082,246.

^fEHR: electronic health record.

^gREGARDS: Reasons for Geographic and Racial Differences in Stroke study.

^hIncludes internal medicine, family medicine, pediatrics, and obstetrics or gynecology.

ⁱSee Table 5 for type 2 diabetes definitions.

^jN/A: not applicable.

^kRECVD: Retail Environment and Cardiovascular Disease.

Table 2. Location, Environmental Attributes, and Disparities Network study population characteristics by site.

Characteristics	G/JHU ^a	NYU ^b	UAB ^c
Study population, n	578,485	6,082,246	11,199
Sex, n (%)			
Male	254,218 (43.94)	5,578,056 (91.71)	4946 (44.16)
Female	324,267 (56.05)	504,020 (8.28)	6253 (55.83)
Age (years), n (%)^d			
<18	132,341 (22.87)	0 (0)	0 (0)
18-29	92,458 (15.98)	450,504 (7.40)	0 (0)
30-39	67,185 (11.61)	550,910 (9.05)	0 (0)
40-49	67,996 (11.75)	753,811 (12.39)	649 (5.79)
50-59	74,641 (12.90)	1,168,452 (19.21)	3445 (30.76)
60-69	63,946 (11.05)	1,566,257 (25.75)	4547 (40.60)
≥70	79,918 (13.81)	1,592,179 (26.17)	2558 (22.84)
Race, n (%)			
Black	23,302 (4.02)	920,596 (15.13)	3672 (32.78)
American Indian or Alaska Native	655 (0.11)	56,928 (0.93)	0 (0)
Asian	4616 (0.79)	58,555 (0.96)	0 (0)
Native Hawaiian or other Pacific Islander	2746 (0.47)	60,441 (0.99)	0 (0)
White	542,128 (93.71)	4,411,233 (72.52)	7527 (67.21)
Ethnicity, n (%)			
Hispanic	25,274 (4.36)	331,376 (5.44)	0 (0)
Non-Hispanic	553,211 (95.63)	5,750,870 (94.55)	11,199 (100.00)
Setting of residential address, n (%)			
Higher density urbanized area	48,374 (8.36)	688,488 (11.31)	1809 (16.15)
Lower density urbanized area	76,301 (13.18)	2,139,912 (35.18)	4524 (40.39)
Suburban and small town (UC) ^e	178,548 (30.86)	1,328,278 (21.83)	2644 (23.60)
Rural	275, 272 (47.58)	1,781,743 (29.29)	2222 (19.84)
Diagnosed with type 2 diabetes by the end of the follow-up period, n (%) ^f	64,214 (11.10)	936,627 (15.39)	1408 (12.57)

^aG/JHU: Geisinger and Johns Hopkins University.

^bNYU: New York University.

^cUAB: University of Alabama at Birmingham.

^dG/JHU: age as of date of the data pull (2016); NYU: age calculated by subtracting year of birth from cohort entry year; UAB: age at baseline.

^eUC: urban cluster

^fSee Table 5 for diabetes definitions. G/JHU: 2008 to 2016; NYU: 2008 to 2016; UAB: 2003 to 2016.

Table 3. Location, Environmental Attributes, and Disparities Network individual-level data elements available by study site for Network aims.

Data elements	G/JHU ^{a,b}	NYU ^c	UAB ^{d,e}
Individual-level data			
Demographic data (yes/no)	Yes	Yes	Yes
Residential address data	Most recent only	Longitudinal	Longitudinal
Socioeconomic data	Longitudinal	Longitudinal	Longitudinal
Health-related data			
BMI	Longitudinal	Longitudinal	Longitudinal
Vital signs (eg, blood pressure)	Longitudinal	Longitudinal	Longitudinal
Diagnoses	Longitudinal	Longitudinal	Longitudinal
Treatment	Longitudinal	Longitudinal	Longitudinal ^f
Biomarkers (type)	EHR ^g laboratory data	EHR laboratory data	HbA _{1c} ^h available on subset only ⁱ

^aG/JHU: Geisinger and Johns Hopkins University.

^bIncludes only data in the EHR-based study. See [Multimedia Appendix 1](#) for additional data collected in the primary data collection study.

^cNYU: New York University School of Medicine.

^dUAB: University of Alabama at Birmingham.

^eLongitudinal data are available at 2 time points—baseline (2003-2007) and second in-home exam (2013-2016).

^fAdjudicated (confirmed by review of medical records) coronary heart disease, stroke, end-stage renal disease, and death available throughout follow-up.

^gEHR: electronic health record.

^hHbA_{1c}: glycated hemoglobin.

ⁱn=2694 at baseline and n=2527 at follow-up examination.

Figure 1. Geographic coverage of Pennsylvania in the Geisinger and Johns Hopkins University study population: participants in each site by census tract. G/JHU: Geisinger and Johns Hopkins University.

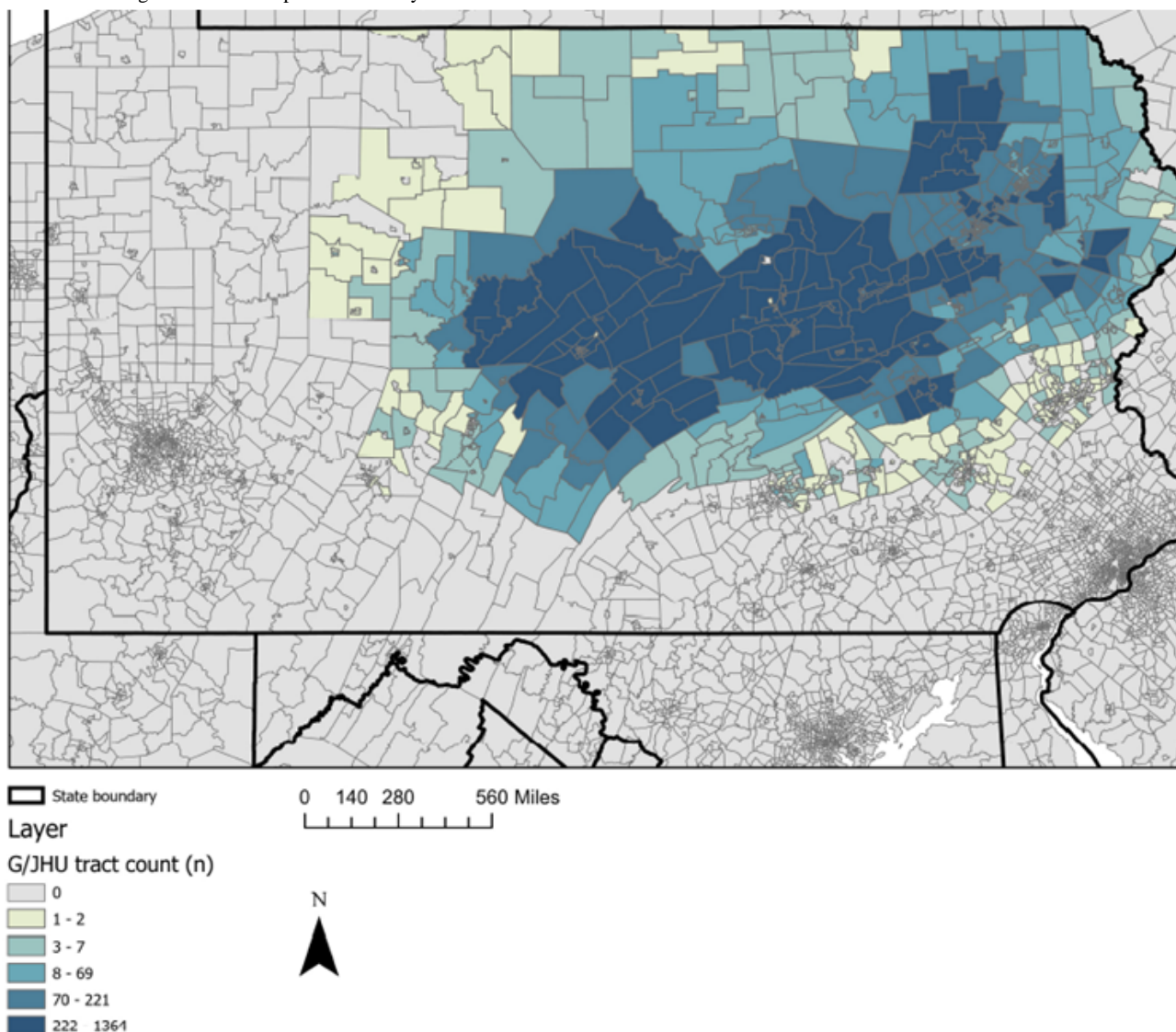


Figure 2. Geographic coverage of the New York University study population: participants in each site by census tract. NYU: New York University.

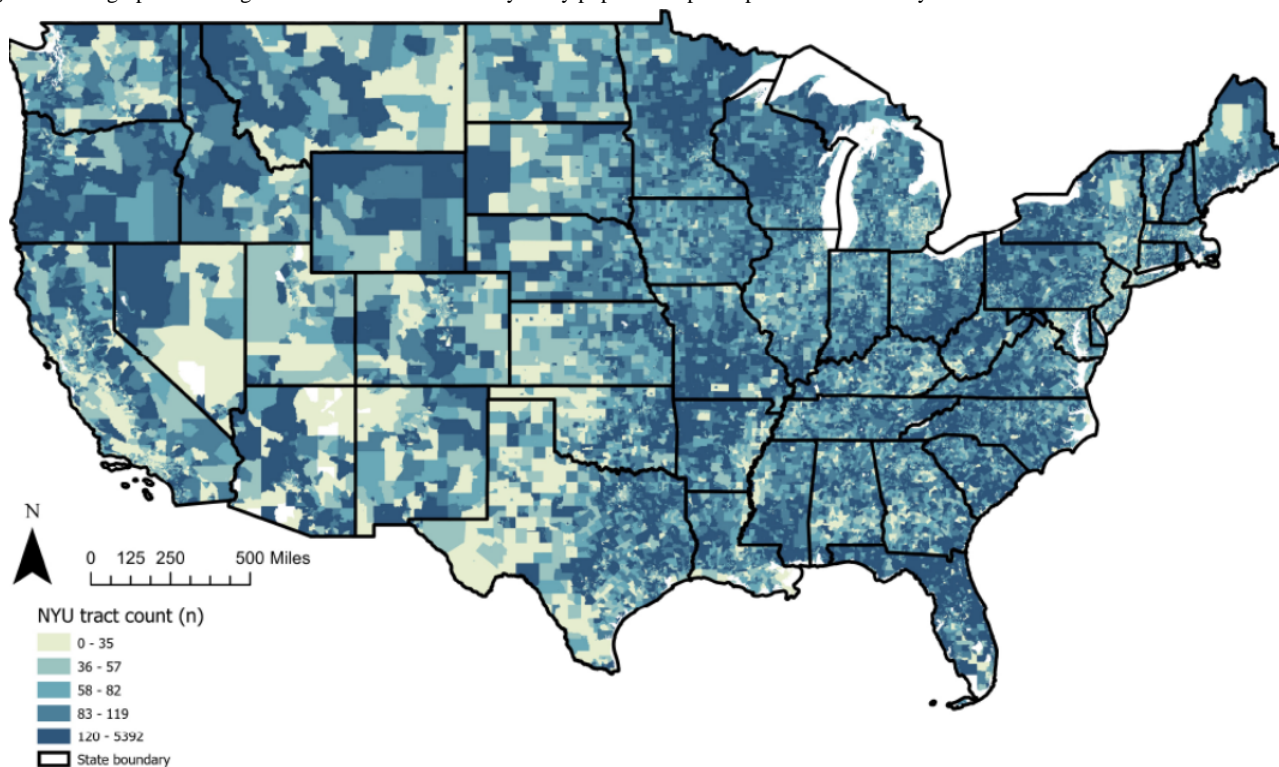


Figure 3. Geographic coverage of the University of Alabama at Birmingham study population in the Stroke Belt region and surrounding states: participants in each site by census tract. Participants are from all 48 contiguous states. The map reflects the Stroke Belt region and surrounding states. UAB: University of Alabama at Birmingham.

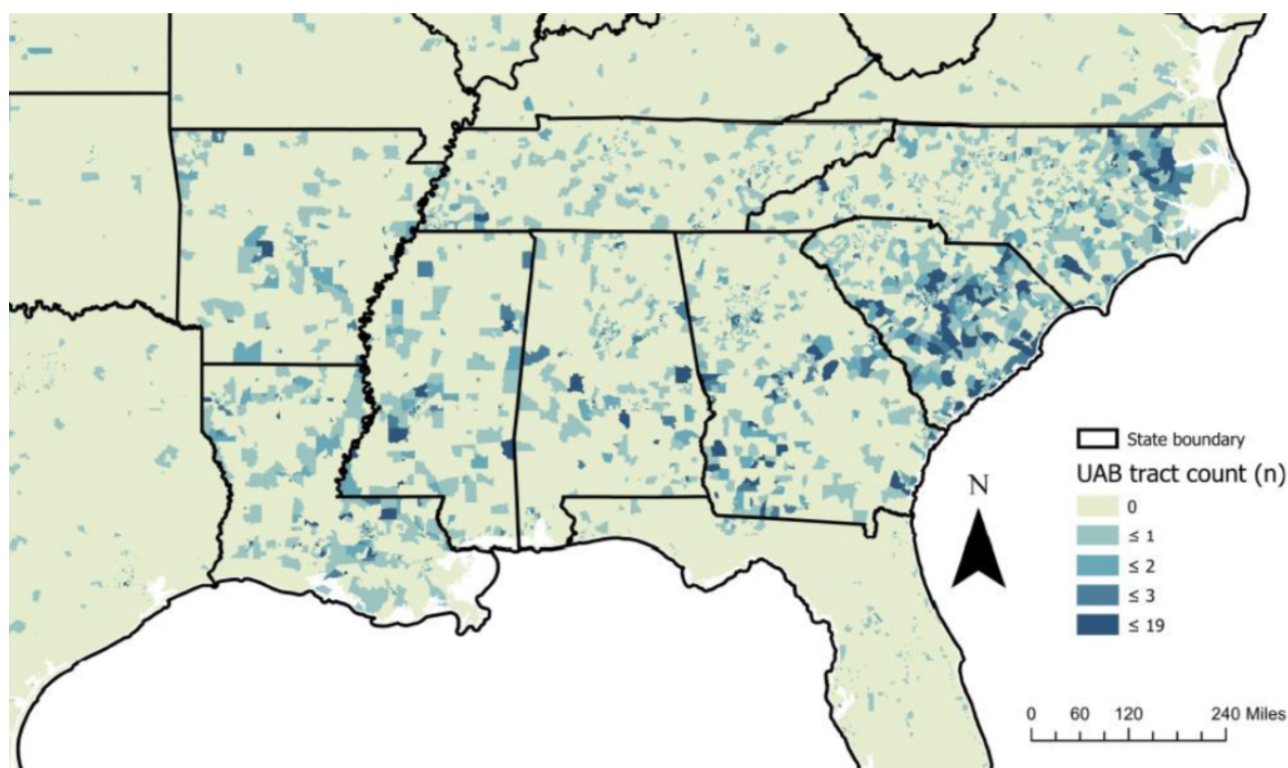


Table 4. Community and environmental domains for Network-wide aims.

Domain	Data source and years	Spatial scale	Description
Neighborhood socioeconomic environment	US Decennial Census and American Community Survey (5-year estimates; 2000, 2010, 2006-2010, and 2008-2012)	Census tract	Area-level index derived from a z-score sum of indicators of the community's social and economic characteristics [29]: percentage of males and females with less than a high school education, percentage of males and females unemployed, percentage of households earning less than US \$30,000 per year, percentage of households in poverty, percentage of households on public assistance, and percentage of households with no cars
Food establishment environment	RECVD ^a Geocoded Business Level Data set, derived from NETS ^b (1997-2014)	Network buffer around the population-weighted centroid of census tracts	Area-level absolute and relative measures: density of supermarkets (including medium-sized grocers) and fast food restaurants (per square kilometer), ratio of supermarkets to all food stores, and ratio of fast food restaurants to all restaurants and eating places
Physical fitness establishment environment	RECVD Geocoded Business Level Data set, derived from NETS ^b (1997-2014)	Network buffer around the population-weighted centroid of census tracts	Area-level density of physical activity venues per square kilometer (eg, gyms, membership sports and recreation clubs, athletic organizations)
Land use environment	RECVD National Land Cover Database, ESRI StreetMap, RECVD NETS, US Decennial Census (2006, 2009, and 2010)	Census tract	Area-level index derived from a z-score sum of 7 indicators: average block length, average block size, intersection density, street connectivity, density of walkable establishments per square mile, percent developed land, and household density per square mile
Leisure-time physical activity environment	CDC's ^c Division of Population Health, National Center for Chronic Disease Prevention and Health Promotion. Derived from the Homeland Security Infrastructure Program Gold 2011 database ^d , Environmental Systems Research Institute Arc Geographic Information System (ESRI ArcGIS) 10.1 Data DVD 2010	Census tract	Spatial access to parks measured by population-weighted distance to the 7 closest parks from the census tract's population-weighted centroid [30,31]

^aDrexel University Urban Health Collaborative. The Retail Environment and Cardiovascular Disease (RECVD) Project.

^bNational Establishment Time-Series (NETS) database.

^cCDC: Centers for Disease Control and Prevention.

^dThe Homeland Infrastructure Foundation-level Data Working Group NAVTEQ from Homeland Security Infrastructure Program Gold 2011 Database.

Table 5. Definition of type 2 diabetes.

Study site	Diagnoses codes	Laboratory measures ^a	Medication orders ^b	Exclusions
Geisinger and Johns Hopkins University	ICD ^c -9, ICD-10, or EDG ^d code for T2D on two separate dates	≥1 elevated HbA _{1c} ^e or glucose measure and ≥1 diagnosis code for T2D	≥1 T2D medication order	≥10 years of T1D ^f diagnoses and <5 years of T2D ^g diagnoses or first diabetes code before 10 years of age or only meet criteria during pregnancy
New York University School of Medicine	ICD-9 or ICD-10 code for T2D on two separate dates	≥2 elevated HbA _{1c} or glucose measure and ≥1 diagnosis code for T2D	≥1 T2D medication order	N/A ^h
University of Alabama at Birmingham ⁱ	N/A	Elevated glucose measure at study visit	Self-report of T2D medication	N/A

^aHbA_{1c}≥6.5%, random glucose≥200 mg/dl, and fasting glucose≥126 mg/dl.

^bExcluding Metformin and Acarbose: Reasons for Geographic and Racial Differences in Stroke project.

^cICD: International Classification of Diseases.

^dEDG: Epic Diagnostic code Groupers.

^eHbA_{1c}: glycated hemoglobin.

^fT1D: type 1 diabetes.

^gT2D: type 2 diabetes.

^hN/A: not applicable.

ⁱFrom the Reasons for Geographic and Racial Differences in Stroke Study.

Textbox 1. Site-specific aims.

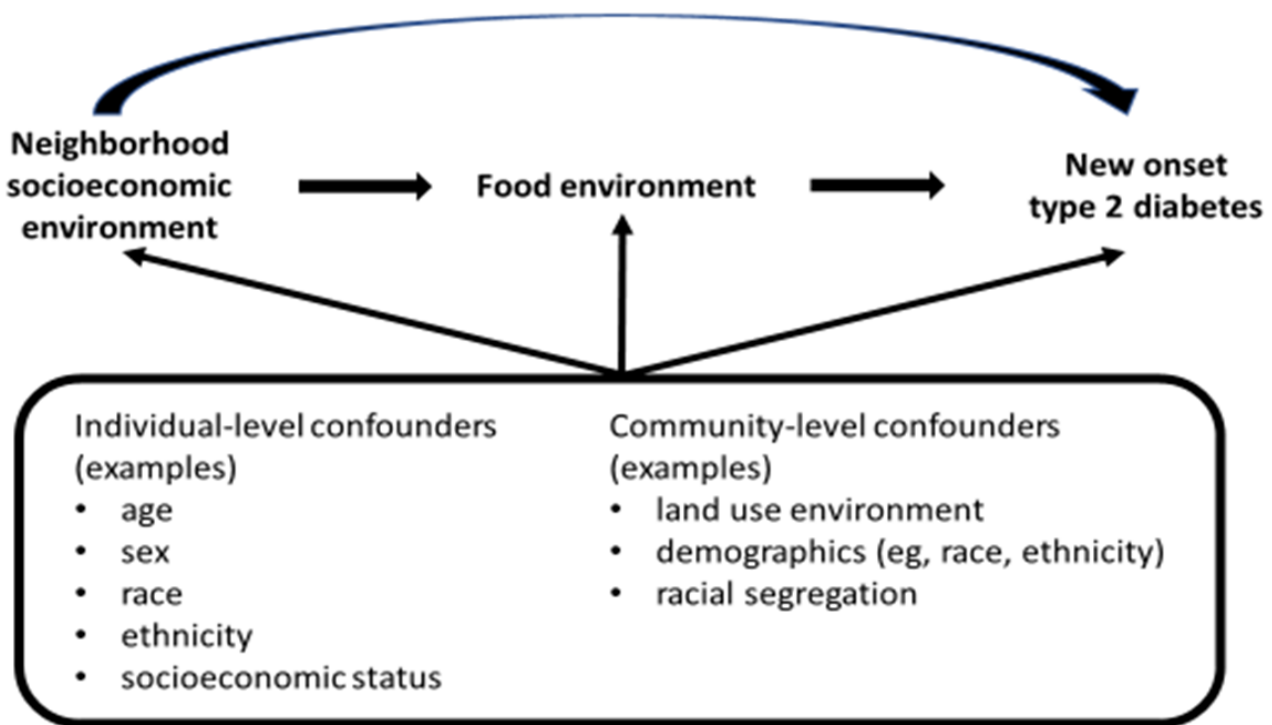
<p>Geisinger and Johns Hopkins University:</p> <ul style="list-style-type: none"> To evaluate associations of chronic environmental contamination [30] (eg, abandoned coal mine lands); the food environment; the physical activity environment; land use environment, the natural environment (eg, greenness); community type (eg, urban/rural); and community socioeconomic deprivation (CSD) with type 2 diabetes (T2D) onset and control and coronary heart disease (CHD) onset within communities. To evaluate mediating pathways (eg, food, physical activity environment) between the neighborhood socioeconomic environment and T2D onset (through LEAD Network Aim 1). To evaluate mediating pathways (eg, stress, health behaviors) between community factors and T2D control among 1000 individuals with T2D living in 40 communities. To evaluate potential effect modification by key individual (eg, age, Medical Assistance) and community factors (eg, CSD) of relations between community factors and T2D and CHD within communities. <p>New York University School of Medicine:</p> <ul style="list-style-type: none"> Using public-use data sources, determine independent and joint association between novel community measures and county-level prevalence of outcomes (diabetes, obesity, and diabetes-obesity prevalence discordance profile), controlling for other county measures (eg, population density, socioeconomic status, and demographic distributions). Measure the impact of modifiable community characteristics such as food and housing environments on (a) risk of a new T2D diagnosis or (b) being obese (BMI≥30 kg/m²) in a large cohort of Veterans Affairs patients, adjusting for community and individual-level covariates in multilevel regression models. Use mediation analysis to examine mediating pathways between modifiable community contexts and T2D. <p>University of Alabama at Birmingham:</p> <ul style="list-style-type: none"> To determine the association of community-level social determinants of health with the prevalence and incidence of T2D and hypertension, separately. To determine if pharmacologic treatment patterns and hospitalization rates vary by community-level social determinants of health for those with T2D and hypertension, separately. To determine if awareness and treatment of T2D and risk of cardiovascular complications varies by community-level and individual-level social determinants of health.

Network Aims

The Network aims to evaluate the association of community factors and T2D outcomes (aims 1 and 2) and to evaluate and address the previously described methodological challenges of community and health research (aims 3 and 4):

1. Evaluate the mediation of the association of NSEE with new-onset T2D. This aim reflects a conceptual framework (Figure 4) that proposes that NSEE influences T2D onset through other community pathways, including the food, physical activity (fitness and leisure) environments, and exposure to fine particulate matter ($\leq 2.5 \mu$, particulate matter_{2.5}).
2. Define and test effect modifiers (eg, age, sex, race) of the association of NSEE with new-onset T2D.
3. Assess the differential item functioning of community measures by geographic region and community type.
4. Evaluate the impact of the spatial scale used to measure community factors (eg, buffer, census tract, county) on associations with new-onset T2D.

Figure 4. Conceptual framework for mediation of the association between neighborhood socioeconomic environment and type 2 diabetes: food environment as an example mediator.



Network Populations and Geographic Coverage

The Diabetes LEAD Network draws from individuals living in all 50 US states (Figures 1-3). The G/JHU participants were selected from among 1.6 million individuals in the Geisinger EHR, spanning 37 counties in central and northeastern Pennsylvania (Table 1). The G/JHU participants range in age from 10 to 97 years, are predominately White (542,128/578,458, 93.71%) and non-Hispanic (553,211/578,458, 95.66%), reflecting the region's population [32]. More than 40% of G/JHU participants reside in areas that the US Census Bureau categorizes as rural. The NYU population of veterans spans all US counties. NYU assembled a baseline cohort of veterans who do not have diabetes (n=6,082,246) and are receiving primary care through Veterans Affairs. Veterans are mostly male (5,578,056/6,082,246, 91.71%) and predominantly White (4,411,233/6,082,246, 72.53%) followed by Black (920,596/6,082,246, 15.14%).

These demographic groups are represented by large sample sizes. The UAB population includes participants from the

Reasons for Geographic and Racial Differences in Stroke (REGARDS) study cohort [33]. At baseline (2003-2007), the REGARDS study enrolled 30,239 non-Hispanic Black and non-Hispanic White adults aged 45 years and older with oversampling of participants from the Stroke Belt region (Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee) [34]. To assess T2D onset, participants without T2D at baseline and who completed the follow-up in-home examination (n=11,199) will be evaluated. Patients were not invited to comment on the cohort development or study design.

Network Data Sources and Measurement

The DCC is leading the development of harmonized, Network-wide approaches to measuring community factors of interest and T2D outcomes. To develop measures of community factors (Table 4), the DCC is using archival data available at the national level, including publicly available data (eg, US Census) and data elements previously created for the Retail Environment and Cardiovascular Disease (RECVD) study. The RECVD study has longitudinal measures of food, fitness, and

social establishments based on the National Establishment Time Series (NETS), a data source that includes information on more than 58 million US business establishments from 1990 to 2014. For each community factor, the DCC has partnered with a study site with relevant expertise to make decisions regarding data sources, spatial scale, exposure assignment, and approach to measurement.

The DCC is applying a range of measurement techniques to define community factors, including data reduction and measurement models. Measurement development is stratified by community type at the census tract level using a modification of the Rural-Urban Commuting Area (RUCA) from the US Department of Agriculture developed by the Network [35]. After collapsing the original 10 RUCA categories into 3, the DCC further divided census tracts within urbanized areas into 2 categories based on land area, resulting in 4 community-type categories that reflect distinct typologies along the rural-urban continuum.

To the extent possible, the Network is harmonizing approaches to measure T2D onset (Table 5) and diabetes-related outcomes. G/JHU and NYU have worked together to develop EHR-based algorithms based on their previous work [36,37] and diagnosis criteria from the American Diabetes Association [38], using a combination of diagnosis codes, medications, and laboratory measures. With coordination from the DCC, the sites are also standardizing approaches to measure potential confounders, mediators, and effect modifiers.

Network Analyses

For each Network-wide aim, the study sites will conduct analyses among their study populations based on a common analytic plan. The DCC is coordinating the development of the analytic plan, harmonizing analytical approaches, including the selection of confounding, mediating, and modifying variables of interest, model building, and model diagnostics. Sites will conduct site-specific sensitivity analyses that include relevant data elements that may not be available Network-wide. This approach allows us to examine consistency in results while leveraging the unique data available at individual sites.

For aims 1 and 2, the Network will employ methods to account for group-level and individual-level data, including multilevel models, Bayesian approaches, and generalized estimating equation models. The Network will conduct causal mediation analysis for aim 1 [39]. For aim 2, we will evaluate effect modification through the inference of interaction terms, creating cross-products between our contextual domains of interest and a predetermined set of individual- and community-level variables, such as age, sex, and race. To guide model development for these aims, sites are developing causal diagrams to formulate and test theoretically based pathways, identify potential confounding influences, and account for potential interaction between measures (Figure 4) [40,41]. To assess spatial residual autocorrelation, the Network will calculate I statistics by Moran (local and global) [42] and use modeling approaches that account for spatial residual autocorrelation, if needed. The Network will conduct sensitivity analyses to evaluate how approaches to measurement of outcomes and community factors impact observed associations.

For aim 3, the Network is exploring strategies to evaluate and address nonpositivity, including propensity scores [43], stratification by community type for all analyses, and latent profile analysis to evaluate community typology [44]. For aim 4, we are evaluating community factors at multiple spatial scales (eg, census tract, network buffer around population-centroid). We will compare the associations of community factors measured at different scales with T2D to assess the influence of the modifiable areal unit problem [21]. The study sites also have site-specific aims, as described below.

Site Descriptions

Geisinger and Johns Hopkins University

The Environmental Health Institute, a joint collaboration between Geisinger, Johns Hopkins Bloomberg School of Public Health, and Johns Hopkins School of Medicine, is evaluating the influence of community factors on T2D onset and control and cardiometabolic outcomes in Pennsylvania (Textbox 1) using a combination of primary and secondary data collection. The team is conducting the study among patients from Geisinger, a health system serving 1.6 million patients in central and northeastern Pennsylvania. To be eligible for study, individuals had to reside in one of 37 counties in Geisinger's service area and have at least two Geisinger primary care visits from 2006 to 2016 (Table 1). The Geisinger primary care patient population represents the age, sex, and racial and ethnic distribution of the general population of the region [32]. The region's population is residentially stable, with an annual out-migration rate of approximately 1% in all but two counties according to US Census Bureau data.

G/JHU is evaluating the main effects of 8 community factors: NSEE, food environment, fitness environment, leisure-time physical activity environment, land use environment, greenness, blue space (aquatic environments such as coasts, lakes, and rivers), and chronic environmental contamination [45]. G/JHU has previously reported associations between these factors and obesity and glycated hemoglobin (HbA_{1c}) [46-51]. G/JHU is using data from publicly available sources (eg, US Census, American Community Survey [ACS], Moderate Resolution Imaging Spectroradiometer from the National Aeronautics and Space Administration's Terra satellite, Pennsylvania Department of Transportation, TeleAtlas) and commercial data to generate measures for these factors. For the Network aims, the team is working with the DCC to guide decisions on the land use and physical fitness environment measures.

For site-specific aims, G/JHU outcomes include T2D onset, T2D control, and cardiometabolic outcomes (Textbox 1). G/JHU is conducting 2 types of studies to evaluate the associations between community factors and T2D outcomes and mediation and moderation of these associations. EHR-based analyses will be used for both Network- and site-specific aims. A primary data collection study will be used for additional site-specific aims, as much of the data collected in this study will be uniquely available at the G/JHU site (Multimedia Appendix 1).

G/JHU is using a mix of nested case-control and retrospective cohort study designs to achieve site-specific aims, using logistic and linear regression as appropriate. To account for correlation

due to both place and space, G/JHU is using generalized estimating equations and multilevel modeling. To examine mediators of the association between NSEE and T2D onset (Network Aim 1), Geisinger will apply a nested case-control design and formal mediation models that include T2D onset cases (n=15,888) matched to controls (n=79,435) on age, sex, and year of encounter.

New York University School of Medicine

Investigators at NYU are examining the relationship between modifiable community factors and risk for T2D and obesity using a retrospective cohort assembled through EHR data from the Veterans Affairs Corporate Data Warehouse, a national repository of clinical and administrative data. The 2 primary exposures of interest are the food and housing environments. The assembled cohort includes more than 6 million veteran patients who were diabetes-free upon entry into the cohort from 2008 to 2016. Entry eligibility includes 2 primary care visits with no indication of diabetes within the 5 years before cohort entry, with at least two follow-up visits at least 30 days apart during the study period (2008-2018). The population has a well-documented high incidence of diabetes [36], providing adequate variation in contexts and outcomes to examine community factors in relation to T2D incidence.

For site-specific analyses, NYU's primary community factors of interest are the food and housing environments. Food environment metrics include 2 absolute measures and 2 relative measures created from the RECDV data (Table 4). The NYU team also has store-level Nielsen Retail Scanner data from 2006 to 2014, which will be used to examine potential mechanistic pathways, including whether risks associated with living in select food environments are partially mediated through per capita sales of sugar-sweetened beverages. The NYU team is guiding Network decisions on the food environment measure development and harmonization, in collaboration with the DCC. They are also engaging in site-specific analyses to examine the influence of housing affordability per ACS and Veterans Affairs data on T2D risk.

NYU study outcomes include diabetes incidence and control as well as obesity prevalence and incidence. Outcome data are extracted from EHRs, capturing demographic, clinical, and utilization data. To ensure participants in the cohort do not have diabetes at cohort entry, individuals with any diabetes (type 1 or type 2) International Classification of Disease version 9 or 10 (ICD-9/10) code or elevated HbA_{1c} at enrollment are excluded. Time-to-event analyses (Cox proportional hazards models with frailty to account for clustering within a community) will be used to examine the main effects of the food environment on T2D risk and its role in mediating the association between NSEE and T2D risk. Person-time is calculated as the date of a censoring event (diabetes diagnosis, death, loss to follow-up, or end of study period) minus the date of cohort entry. The date of death is obtained from the Veterans Affairs Vital Status and Beneficiary Identification Records Locator. Loss-to-follow-up is defined as no Veterans Affairs encounter for more than 2 years but patients can re-enter the cohort if they meet entry criteria again.

University of Alabama at Birmingham

The UAB site is investigating the association of NSEE with a greater burden of T2D and cardiovascular risk, particularly in southeastern United States. To address site-specific and Network-wide research questions, UAB is leveraging resources from the REGARDS study [33]. The REGARDS study is a longitudinal, population-based closed cohort study of 30,239 adults aged 45 years and older at baseline (2003-2007), designed to identify factors associated with higher stroke mortality. The study was designed to oversample non-Hispanic Black adults and residents of the Stroke Belt region, with 56% of the sample selected from the Stroke Belt and the remaining 44% selected from the other 40 contiguous states. Demographics, medical history, and lifestyle factors were assessed at baseline and an in-home physical exam was performed with blood and urine collection. Follow-up is ongoing every 6 months to assess vital status and hospitalizations and obtain medical records for adjudication of possible cardiovascular events. A second in-home physical exam was completed between 2013 and 2016.

For site-specific analyses, the primary exposure includes the NSEE as assessed using principal component analysis for measures of community-level income or wealth, education, housing, health systems or services, employment, social environment, and physical environment. The data to assess these characteristics include both publicly available databases (eg, US Census) and commercial databases (eg, Dun & Bradstreet). The primary outcomes are incident T2D and cardiovascular outcomes. Incident T2D will be assessed among 11,199 REGARDS study participants without prevalent T2D at baseline and who completed the follow-up in-home physical exam during which objective measurements (eg, glucose, use of medications) were collected (Table 2). Cardiovascular outcomes include hypertension (ie, mean blood pressure 140/90 mm Hg or use of hypertension medications) and expert adjudicated clinical events (ie, coronary heart disease, stroke).

Separate from the analysis of the REGARDS study data, UAB will utilize Medicare administrative claims data to investigate the association of NSEE with T2D and hypertension incidence. These data consist of several federal health care insurance programs that cover adults aged 65 years and older and younger individuals who are disabled or have end-stage renal disease. Broadly, Medicare Part A covers hospital services, Medicare Part B covers outpatient and physician services, and Medicare Part D covers prescription drugs. UAB will use the 5% random sample of Medicare claims data available from 1999 to 2015 to investigate community-level determinants of T2D incidence, diabetes hospitalizations, and treatment patterns. An overview of the Medicare sample population and diabetes definitions used for site-specific analyses is provided in [Multimedia Appendix 2](#). Statistical approaches include generalized linear models and spatial generalized linear mixed models.

Drexel Data Coordinating Center

The Drexel DCC provides the study sites with project coordination and statistical support, including advanced methodological and analytic approaches to data analyses driven by the Network aims and heterogeneous data from each study site. The expertise needed for this work is reflected in the

backgrounds of DCC team members, including biostatisticians and epidemiologists from the Dornsife School of Public Health and the Drexel Urban Health Collaborative (UHC), postdoctoral fellows, doctoral-level biostatistics students, data analysts and managers, and GIS experts. Through exploratory analytic work, including principal component analysis, exploratory factor analysis, GIS analysis and mapping, and correlation analysis of contextual indices against individual variables, the DCC supports the Network's collective decision making around defining exposure metrics for addressing Network aims. The relationship with the UHC also allows for access to data from RECDV and other sources; provides support for GIS methods, data distribution, and storage; and provides access to data engineering experts. Furthermore, the UHC has a Policy and Outreach Core, which helps provide guidance on disseminating LEAD Network findings.

Results

The Network has developed metrics for the community factors of interest: NSEE, food establishment, physical fitness establishment, leisure-time physical activity, and land use environments (Table 4). The Network has created these measures using data that are consistently available and contextually applicable to all geographies in the contiguous United States. This underscores the importance of the Network's development of a method for categorizing community types for stratified evaluation of community factors with T2D onset. With harmonized measures, the Network is poised to compare findings across the varying study sites.

The Network has reported findings based on work from the initial years of funding. Preliminary results have been presented at annual meetings of the Society for Epidemiologic Research, the American Diabetes Association, and the American Public Health Association [52,53]. The Network recently published a paper describing county-level determinants of diabetes status in the United States from 2003 to 2012 [54]. The NYU team published a paper describing the impact of changes in the built and social environment on BMI in US counties using data from the Behavioral Risk Factor Surveillance System [13]. Additional manuscripts are in press or in development.

Discussion

Strengths and Limitations

The Diabetes LEAD Network leverages a breadth of expertise and data to advance knowledge regarding modifiable community risk factors for T2D onset and related outcomes. The Network brings strengths to its collective mission to provide scientific evidence for targeted interventions and policies. First, the sites provide the Network with community data sources that collectively ensure widespread geographic coverage and variation of community types across a rural-to-urban spectrum. It was of particular importance to ensure representation by rural communities, since CDC reports that diabetes is 17% more prevalent in rural than urban areas [55]. Furthermore, understanding the association between community and health requires the assessment of a heterogeneous set of communities [12,56]. Each of the study sites contributes unique data sources

to achieve this goal. The NYU cohort spans the nation, the UAB REGARDS study cohort offers a national study with in-depth data from a high-risk region (Stroke Belt), and the G/JHU population is a regionally representative sample with high rural representation and primary data collection.

A second contribution of the Network is the development of measures of 6 community factors (Table 4) to be examined across diverse geographies and community types. These measures are being developed with consideration of community types, examining community factors within strata of different community types defined along a rural-urban spectrum to avoid potential differential item functioning and nonpositivity [23]. In addition, Network investigators are examining individual and joint associations to better understand how these community factors work in concert to contribute to excess T2D risk. With access to individuals' residential and commercial addresses, the Network is evaluating spatial scales of various types and sizes to better understand the impact of scale on the findings.

Third, the range of expertise across institutions allows the Network to address methodological challenges common to community and health research [20]. In addition, the Network is advancing methods for conducting research on the role of community and health using data obtained from EHR systems, including extracting historical residential addresses to allow for time-varying exposure estimation. Finally, the Network is harmonizing community factor definitions and analytical approaches to facilitate comparable analyses and replicating analyses across 3 different study populations. This effort to harmonize approaches across multiple settings and populations will advance both the field of community and health research and generate data needed to guide evidence-based policies for T2D prevention in the United States.

There are some limitations to the Diabetes LEAD Network research portfolio. First, although access to longitudinal data will mitigate issues of temporality, the available data do not allow for investigation of early life exposures (eg, childhood) that may influence T2D risk [57]. Second, while the diversity of populations and geographies across the study sites is advantageous to expand the generalizability of findings and include previously underrepresented settings (ie, rural), it complicates comparison of results across sites. However, by harmonizing measurement and analytical approaches the Network will be well positioned to pinpoint reasons for any potential conflicting results that arise. Finally, despite employing advanced analytic approaches, the studies are all observational in design; thus, they are potentially constrained with respect to causal inference due to the risk of residual confounding and neighborhood self-selection [58,59]. The Network will consider methodological approaches such as propensity scores to address this limitation [43].

Conclusions

T2D is a leading cause of morbidity in the United States, with select populations, often defined by geography, affected by a disproportionate burden of disease. The Diabetes LEAD Network identifies modifiable community factors that influence geographic disparities in T2D risk across diverse communities and identifies policy levers to ameliorate these disparities.

Acknowledgments

This research was conducted by the Diabetes LEAD Network, funded by the CDC cooperative agreements U01DP006293 (Drexel University), U01DP006296 (Geisinger-Johns Hopkins University), U01DP006299 (New York University School of Medicine), and U01DP006302 (University of Alabama at Birmingham), along with collaboration with the US CDC Division of Diabetes Translation. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of primary data collection study for Geisinger and Johns Hopkins University site-specific analyses.

[\[DOCX File , 16 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Description of Medicare study population for University of Alabama at Birmingham site-specific analyses.

[\[DOCX File , 15 KB-Multimedia Appendix 2\]](#)

References

1. National Diabetes Statistics Report. Centers for Disease Control and Prevention. 2020. URL: <https://www.cdc.gov/diabetes/data/statistics-report/index.html> [accessed 2020-09-25]
2. Bhupathiraju SN, Hu FB. Epidemiology of obesity and diabetes and their cardiovascular complications. *Circ Res* 2016 May 27;118(11):1723-1735 [FREE Full text] [doi: [10.1161/CIRCRESAHA.115.306825](https://doi.org/10.1161/CIRCRESAHA.115.306825)] [Medline: [27230638](https://pubmed.ncbi.nlm.nih.gov/27230638/)]
3. American Diabetes Association. Prevention or delay of type 2 diabetes: standards of medical care in diabetes. *Diabetes Care* 2018 Dec 17;42(Supplement 1):S29-S33. [doi: [10.2337/dc19-s003](https://doi.org/10.2337/dc19-s003)] [Medline: [29222376](https://pubmed.ncbi.nlm.nih.gov/29222376/)]
4. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr* 2010 Oct 22;8:29 [FREE Full text] [doi: [10.1186/1478-7954-8-29](https://doi.org/10.1186/1478-7954-8-29)] [Medline: [20969750](https://pubmed.ncbi.nlm.nih.gov/20969750/)]
5. Dwyer-Lindgren L, Mackenbach JP, van Lenthe FJ, Flaxman AD, Mokdad AH. Diagnosed and undiagnosed diabetes prevalence by county in the US, 1999-2012. *Diabetes Care* 2016 Sep;39(9):1556-1562. [doi: [10.2337/dc16-0678](https://doi.org/10.2337/dc16-0678)] [Medline: [27555622](https://pubmed.ncbi.nlm.nih.gov/27555622/)]
6. Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ. Geographic distribution of diagnosed diabetes in the US: a diabetes belt. *Am J Prev Med* 2011 Apr;40(4):434-439. [doi: [10.1016/j.amepre.2010.12.019](https://doi.org/10.1016/j.amepre.2010.12.019)] [Medline: [21406277](https://pubmed.ncbi.nlm.nih.gov/21406277/)]
7. Cunningham SA, Patel SA, Beckles GL, Geiss LS, Mehta N, Xie H, et al. County-level contextual factors associated with diabetes incidence in the United States. *Ann Epidemiol* 2018 Jan;28(1):20-5.e2 [FREE Full text] [doi: [10.1016/j.annepidem.2017.11.002](https://doi.org/10.1016/j.annepidem.2017.11.002)] [Medline: [29233722](https://pubmed.ncbi.nlm.nih.gov/29233722/)]
8. Fang M. Trends in the prevalence of diabetes among US Adults: 1999-2016. *Am J Prev Med* 2018 Oct;55(4):497-505. [doi: [10.1016/j.amepre.2018.05.018](https://doi.org/10.1016/j.amepre.2018.05.018)] [Medline: [30126668](https://pubmed.ncbi.nlm.nih.gov/30126668/)]
9. Creatore MI, Glazier RH, Moineddin R, Fazli GS, Johns A, Gozdyra P, et al. Association of neighborhood walkability with change in overweight, obesity, and diabetes. *J Am Med Assoc* 2016;315(20):2211-2220. [doi: [10.1001/jama.2016.5898](https://doi.org/10.1001/jama.2016.5898)] [Medline: [27218630](https://pubmed.ncbi.nlm.nih.gov/27218630/)]
10. Mezuk B, Li X, Cederin K, Rice K, Sundquist J, Sundquist K. Beyond access: characteristics of the food environment and risk of diabetes. *Am J Epidemiol* 2016 Jun 15;183(12):1129-1137 [FREE Full text] [doi: [10.1093/aje/kwv318](https://doi.org/10.1093/aje/kwv318)] [Medline: [27240801](https://pubmed.ncbi.nlm.nih.gov/27240801/)]
11. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: a systematic review. *Health Place* 2012 Sep;18(5):1172-1187 [FREE Full text] [doi: [10.1016/j.healthplace.2012.05.006](https://doi.org/10.1016/j.healthplace.2012.05.006)] [Medline: [22717379](https://pubmed.ncbi.nlm.nih.gov/22717379/)]
12. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 2010 Mar;16(2):175-190. [doi: [10.1016/j.healthplace.2009.09.008](https://doi.org/10.1016/j.healthplace.2009.09.008)] [Medline: [19880341](https://pubmed.ncbi.nlm.nih.gov/19880341/)]
13. Rummo PE, Feldman JM, Lopez P, Lee D, Thorpe LE, Elbel B. Impact of changes in the food, built, and socioeconomic environment on BMI in us counties, BRFSS 2003-2012. *Obesity (Silver Spring)* 2020 Jan;28(1):31-39. [doi: [10.1002/oby.22603](https://doi.org/10.1002/oby.22603)] [Medline: [31858733](https://pubmed.ncbi.nlm.nih.gov/31858733/)]
14. Walker BB, Shashank A, Gasevic D, Schuurman N, Poirier P, Teo K, et al. The local food environment and obesity: evidence from three cities. *Obesity (Silver Spring)* 2020 Jan;28(1):40-45 [FREE Full text] [doi: [10.1002/oby.22614](https://doi.org/10.1002/oby.22614)] [Medline: [31774254](https://pubmed.ncbi.nlm.nih.gov/31774254/)]

15. Bilal U, Auchincloss AH, Diez-Roux AV. Neighborhood environments and diabetes risk and control. *Curr Diab Rep* 2018 Jul 11;18(9):62. [doi: [10.1007/s11892-018-1032-2](https://doi.org/10.1007/s11892-018-1032-2)] [Medline: [29995252](https://pubmed.ncbi.nlm.nih.gov/29995252/)]
16. Hipp JA, Chalise N. Spatial analysis and correlates of county-level diabetes prevalence, 2009-2010. *Prev Chronic Dis* 2015 Jan 22;12:E08 [FREE Full text] [doi: [10.5888/pcd12.140404](https://doi.org/10.5888/pcd12.140404)] [Medline: [25611797](https://pubmed.ncbi.nlm.nih.gov/25611797/)]
17. Christine PJ, Auchincloss AH, Bertoni AG, Carnethon MR, Sánchez BN, Moore K, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: the multi-ethnic study of atherosclerosis (MESA). *JAMA Intern Med* 2015 Aug;175(8):1311-1320 [FREE Full text] [doi: [10.1001/jamainternmed.2015.2691](https://doi.org/10.1001/jamainternmed.2015.2691)] [Medline: [26121402](https://pubmed.ncbi.nlm.nih.gov/26121402/)]
18. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the multi-ethnic study of atherosclerosis. *Arch Intern Med* 2009 Oct 12;169(18):1698-1704 [FREE Full text] [doi: [10.1001/archinternmed.2009.302](https://doi.org/10.1001/archinternmed.2009.302)] [Medline: [19822827](https://pubmed.ncbi.nlm.nih.gov/19822827/)]
19. Booth GL, Creatore MI, Luo J, Fazli GS, Johns A, Rosella LC, et al. Neighbourhood walkability and the incidence of diabetes: an inverse probability of treatment weighting analysis. *J Epidemiol Community Health* 2019 Apr;73(4):287-294. [doi: [10.1136/jech-2018-210510](https://doi.org/10.1136/jech-2018-210510)] [Medline: [30696690](https://pubmed.ncbi.nlm.nih.gov/30696690/)]
20. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 2004 May;58(10):1953-1960. [doi: [10.1016/S0277-9536\(03\)00414-3](https://doi.org/10.1016/S0277-9536(03)00414-3)] [Medline: [15020010](https://pubmed.ncbi.nlm.nih.gov/15020010/)]
21. Openshaw S. *The Modifiable Areal Unit Problem*. Norwich, CT: GeoBooks; 1984.
22. Boyle MH, Willms JD. Place effects for areas defined by administrative boundaries. *Am J Epidemiol* 1999 Mar 15;149(6):577-585. [doi: [10.1093/oxfordjournals.aje.a009855](https://doi.org/10.1093/oxfordjournals.aje.a009855)] [Medline: [10084248](https://pubmed.ncbi.nlm.nih.gov/10084248/)]
23. Messer LC, Jagai JS, Rappazzo KM, Lobdell DT. Construction of an environmental quality index for public health research. *Environ Health* 2014 May 22;13(1):39 [FREE Full text] [doi: [10.1186/1476-069X-13-39](https://doi.org/10.1186/1476-069X-13-39)] [Medline: [24886426](https://pubmed.ncbi.nlm.nih.gov/24886426/)]
24. Reading R, Raybould S, Jarvis S. Deprivation, low birth weight, and children's height: a comparison between rural and urban areas. *Br Med J* 1993 Dec 4;307(6917):1458-1462 [FREE Full text] [doi: [10.1136/bmj.307.6917.1458](https://doi.org/10.1136/bmj.307.6917.1458)] [Medline: [8281087](https://pubmed.ncbi.nlm.nih.gov/8281087/)]
25. Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *Am J Epidemiol* 2010 Mar 15;171(6):664-673. [doi: [10.1093/aje/kwp435](https://doi.org/10.1093/aje/kwp435)] [Medline: [20139129](https://pubmed.ncbi.nlm.nih.gov/20139129/)]
26. Oakes JM. Commentary: advancing neighbourhood-effects research--selection, inferential support, and structural confounding. *Int J Epidemiol* 2006 Jun;35(3):643-647. [doi: [10.1093/ije/dyl054](https://doi.org/10.1093/ije/dyl054)] [Medline: [16556642](https://pubmed.ncbi.nlm.nih.gov/16556642/)]
27. Carroll-Scott A, Gilstad-Hayden K, Rosenthal L, Peters SM, McCaslin C, Joyce R, et al. Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: the role of built, socioeconomic, and social environments. *Soc Sci Med* 2013 Oct;95:106-114 [FREE Full text] [doi: [10.1016/j.socscimed.2013.04.003](https://doi.org/10.1016/j.socscimed.2013.04.003)] [Medline: [23642646](https://pubmed.ncbi.nlm.nih.gov/23642646/)]
28. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016;37:61-81 [FREE Full text] [doi: [10.1146/annurev-publhealth-032315-021353](https://doi.org/10.1146/annurev-publhealth-032315-021353)] [Medline: [26667605](https://pubmed.ncbi.nlm.nih.gov/26667605/)]
29. Xiao Q, Berrigan D, Powell-Wiley TM, Matthews CE. Ten-year change in neighborhood socioeconomic deprivation and rates of total, cardiovascular disease, and cancer mortality in older US adults. *Am J Epidemiol* 2018 Dec 01;187(12):2642-2650 [FREE Full text] [doi: [10.1093/aje/kwy181](https://doi.org/10.1093/aje/kwy181)] [Medline: [30137194](https://pubmed.ncbi.nlm.nih.gov/30137194/)]
30. Wen M, Zhang X, Harris CD, Holt JB, Croft JB. Spatial disparities in the distribution of parks and green spaces in the USA. *Ann Behav Med* 2013 Feb;45(Suppl 1):S18-S27 [FREE Full text] [doi: [10.1007/s12160-012-9426-x](https://doi.org/10.1007/s12160-012-9426-x)] [Medline: [23334758](https://pubmed.ncbi.nlm.nih.gov/23334758/)]
31. Zhang X, Lu H, Holt JB. Modeling spatial accessibility to parks: a national study. *Int J Health Geogr* 2011 May 9;10:31 [FREE Full text] [doi: [10.1186/1476-072X-10-31](https://doi.org/10.1186/1476-072X-10-31)] [Medline: [21554690](https://pubmed.ncbi.nlm.nih.gov/21554690/)]
32. Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, et al. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology* 2016 Mar;27(2):163-172 [FREE Full text] [doi: [10.1097/EDE.0000000000000387](https://doi.org/10.1097/EDE.0000000000000387)] [Medline: [26426945](https://pubmed.ncbi.nlm.nih.gov/26426945/)]
33. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* 2005;25(3):135-143. [doi: [10.1159/000086678](https://doi.org/10.1159/000086678)] [Medline: [15990444](https://pubmed.ncbi.nlm.nih.gov/15990444/)]
34. Obisesan TO, Vargas CM, Gillum RF. Geographic variation in stroke risk in the United States. Region, urbanization, and hypertension in the third national health and nutrition examination survey. *Stroke* 2000 Jan;31(1):19-25. [doi: [10.1161/01.str.31.1.19](https://doi.org/10.1161/01.str.31.1.19)] [Medline: [10625710](https://pubmed.ncbi.nlm.nih.gov/10625710/)]
35. Rural-Urban Continuum Codes. United States Department of Agriculture. URL: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/> [accessed 2020-04-30]
36. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the department of veterans affairs based on computerized patient data. *Diabetes Care* 2004 May;27(Suppl 2):B10-B21. [doi: [10.2337/diacare.27.suppl_2.b10](https://doi.org/10.2337/diacare.27.suppl_2.b10)] [Medline: [15113777](https://pubmed.ncbi.nlm.nih.gov/15113777/)]

37. Hirsch AG, Scheck McAlearney A. Measuring diabetes care performance using electronic health record data: the impact of diabetes definitions on performance measure outcomes. *Am J Med Qual* 2014;29(4):292-299. [doi: [10.1177/1062860613500808](https://doi.org/10.1177/1062860613500808)] [Medline: [24006028](https://pubmed.ncbi.nlm.nih.gov/24006028/)]
38. American Diabetes Association. Classification and diagnosis of diabetes. *Diabetes Care* 2017 Dec 8;41(Supplement 1):S13-S27. [doi: [10.2337/dc18-s002](https://doi.org/10.2337/dc18-s002)] [Medline: [29222376](https://pubmed.ncbi.nlm.nih.gov/29222376/)]
39. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17-32. [doi: [10.1146/annurev-publhealth-032315-021402](https://doi.org/10.1146/annurev-publhealth-032315-021402)] [Medline: [26653405](https://pubmed.ncbi.nlm.nih.gov/26653405/)]
40. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008 Oct 30;8:70 [FREE Full text] [doi: [10.1186/1471-2288-8-70](https://doi.org/10.1186/1471-2288-8-70)] [Medline: [18973665](https://pubmed.ncbi.nlm.nih.gov/18973665/)]
41. Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. *Nephrol Dial Transplant* 2015 Sep;30(9):1418-1423. [doi: [10.1093/ndt/gfu325](https://doi.org/10.1093/ndt/gfu325)] [Medline: [25324358](https://pubmed.ncbi.nlm.nih.gov/25324358/)]
42. Moran PA. The interpretation of statistical maps. *J R Stat Soc Series B Stat Methodol* 1948 Dec 5;10(2):243-251. [doi: [10.1111/j.2517-6161.1948.tb00012.x](https://doi.org/10.1111/j.2517-6161.1948.tb00012.x)]
43. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41-55. [doi: [10.1093/biomet/70.1.41](https://doi.org/10.1093/biomet/70.1.41)]
44. Lekkas P, Stankov I, Daniel M, Paquet C. Finite mixture models in neighbourhoods-to-health research: a systematic review. *Health Place* 2019 Sep;59:102140. [doi: [10.1016/j.healthplace.2019.05.018](https://doi.org/10.1016/j.healthplace.2019.05.018)] [Medline: [31374380](https://pubmed.ncbi.nlm.nih.gov/31374380/)]
45. Couch SR, Coles CJ. Community stress, psychosocial hazards, and EPA decision-making in communities impacted by chronic technological disasters. *Am J Public Health* 2011 Dec;101(Suppl 1):S140-S148. [doi: [10.2105/AJPH.2010.300039](https://doi.org/10.2105/AJPH.2010.300039)] [Medline: [21836109](https://pubmed.ncbi.nlm.nih.gov/21836109/)]
46. Hirsch AG, Durden ET, Nordberg C, Berger A, Schwartz BS. Associations of Four Community Factors With Longitudinal Change in Hemoglobin A Levels in Patients With Type 2 Diabetes. *Diabetes Care* 2018 Mar;41(3):461-468 [FREE Full text] [doi: [10.2337/dc17-1200](https://doi.org/10.2337/dc17-1200)] [Medline: [29258994](https://pubmed.ncbi.nlm.nih.gov/29258994/)]
47. Liu AY, Curriero FC, Glass TA, Stewart WF, Schwartz BS. The contextual influence of coal abandoned mine lands in communities and type 2 diabetes in Pennsylvania. *Health Place* 2013 Jul;22:115-122. [doi: [10.1016/j.healthplace.2013.03.012](https://doi.org/10.1016/j.healthplace.2013.03.012)] [Medline: [23689181](https://pubmed.ncbi.nlm.nih.gov/23689181/)]
48. Schwartz BS, Stewart WF, Godby S, Pollak J, Dewalle J, Larson S, et al. Body mass index and the built and social environments in children and adolescents using electronic health records. *Am J Prev Med* 2011 Oct;41(4):e17-e28. [doi: [10.1016/j.amepre.2011.06.038](https://doi.org/10.1016/j.amepre.2011.06.038)] [Medline: [21961475](https://pubmed.ncbi.nlm.nih.gov/21961475/)]
49. Nau C, Ellis H, Huang H, Schwartz BS, Hirsch A, Bailey-Davis L, et al. Exploring the forest instead of the trees: an innovative method for defining obesogenic and obesoprotective environments. *Health Place* 2015 Sep;35:136-146 [FREE Full text] [doi: [10.1016/j.healthplace.2015.08.002](https://doi.org/10.1016/j.healthplace.2015.08.002)] [Medline: [26398219](https://pubmed.ncbi.nlm.nih.gov/26398219/)]
50. Nau C, Schwartz BS, Bandeen-Roche K, Liu A, Pollak J, Hirsch A, et al. Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity (Silver Spring)* 2015 Jan;23(1):207-212 [FREE Full text] [doi: [10.1002/oby.20903](https://doi.org/10.1002/oby.20903)] [Medline: [25324223](https://pubmed.ncbi.nlm.nih.gov/25324223/)]
51. Poulsen MN, Glass TA, Pollak J, Bandeen-Roche K, Hirsch AG, Bailey-Davis L, et al. Associations of multidimensional socioeconomic and built environment factors with body mass index trajectories among youth in geographically heterogeneous communities. *Prev Med Rep* 2019 Sep;15:100939 [FREE Full text] [doi: [10.1016/j.pmedr.2019.100939](https://doi.org/10.1016/j.pmedr.2019.100939)] [Medline: [31360629](https://pubmed.ncbi.nlm.nih.gov/31360629/)]
52. Poulsen M, Hirsch A, Nordberg C. Community Determinants of High Blood Pressure Among Patients With Type 2 Diabetes. In: American Public Health Association. 2019 Presented at: APHA'19; November 2-6, 2019; New York, USA.
53. Lopez PM, Feldman J, Rummo PE, Elbel B, Thorpe L. 1649-P: community predictors of diabetes prevalence and change in burden over time: US counties, 2003-2012. *Diabetes* 2019 Jun;68(Supplement 1):1649-P. [doi: [10.2337/db19-1649-p](https://doi.org/10.2337/db19-1649-p)]
54. Feldman J, Lee D, Lopez P. Assessing county-level determinants of individual-level diabetes status in the United States. - 2020 May 20;2012:102324 (forthcoming). [doi: [10.1016/j.healthplace.2020.102324](https://doi.org/10.1016/j.healthplace.2020.102324)] [Medline: [32217279](https://pubmed.ncbi.nlm.nih.gov/32217279/)]
55. Division of Diabetes Translation At A Glance. Centers for Disease Control and Prevention. 2019. URL: <https://www.cdc.gov/chronicdisease/resources/publications/aag/diabetes.htm> [accessed 2020-01-20]
56. Richardson AS, Boone-Heinonen J, Popkin BM, Gordon-Larsen P. Are neighbourhood food resources distributed inequitably by income and race in the USA? Epidemiological findings across the urban spectrum. *BMJ Open* 2012;2(2):e000698. [doi: [10.1136/bmjopen-2011-000698](https://doi.org/10.1136/bmjopen-2011-000698)] [Medline: [22505308](https://pubmed.ncbi.nlm.nih.gov/22505308/)]
57. Barker DJ. Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition* 1997 Sep;13(9):807-813. [doi: [10.1016/s0889-9007\(97\)00193-7](https://doi.org/10.1016/s0889-9007(97)00193-7)] [Medline: [9290095](https://pubmed.ncbi.nlm.nih.gov/9290095/)]
58. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004 May;58(10):1929-1952. [doi: [10.1016/j.socscimed.2003.08.004](https://doi.org/10.1016/j.socscimed.2003.08.004)] [Medline: [15020009](https://pubmed.ncbi.nlm.nih.gov/15020009/)]
59. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010 Feb;1186:125-145. [doi: [10.1111/j.1749-6632.2009.05333.x](https://doi.org/10.1111/j.1749-6632.2009.05333.x)] [Medline: [20201871](https://pubmed.ncbi.nlm.nih.gov/20201871/)]

Abbreviations

ACS: American Community Survey

CDC: Centers for Disease Control and Prevention
CHD: coronary heart disease
CSD: community socioeconomic deprivation
DCC: data coordinating center
EDG: Epic Diagnostic code Groupers
EHR: electronic health record
GIS: geographic information systems
ICD: International Classification of Diseases.
LEAD: Location, Environmental Attributes, and Disparities
NETS: National Establishment Time Series
NSEE: neighbourhood socioeconomic environment
RECVD: Retail Environment and Cardiovascular Disease
REGARDS: Reasons for Geographic and Racial Differences in Stroke
RUCA: Rural-Urban Commuting Area
T2D: type 2 diabetes
UHC: Urban Health Collaborative

Edited by G Eysenbach; submitted 19.06.20; peer-reviewed by J Casey, T Taveira-Gomes; comments to author 25.08.20; revised version received 03.09.20; accepted 08.09.20; published 19.10.20

Please cite as:

Hirsch AG, Carson AP, Lee NL, McAlexander T, Mercado C, Siegel K, Black NC, Elbel B, Long DL, Lopez P, McClure LA, Poulsen MN, Schwartz BS, Thorpe LE

The Diabetes Location, Environmental Attributes, and Disparities Network: Protocol for Nested Case Control and Cohort Studies, Rationale, and Baseline Characteristics

JMIR Res Protoc 2020;9(10):e21377

URL: <http://www.researchprotocols.org/2020/10/e21377/>

doi: [10.2196/21377](https://doi.org/10.2196/21377)

PMID: [33074163](https://pubmed.ncbi.nlm.nih.gov/33074163/)

©Annemarie G Hirsch, April P Carson, Nora L Lee, Tara McAlexander, Carla Mercado, Karen Siegel, Nyesha C Black, Brian Elbel, D Leann Long, Priscilla Lopez, Leslie A McClure, Melissa N Poulsen, Brian S Schwartz, Lorna E Thorpe. Originally published in JMIR Research Protocols (<http://www.researchprotocols.org>), 19.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.researchprotocols.org>, as well as this copyright and license information must be included.