Original Research Article

ISEE Environmental Epidemiology

OPEN

Community profiles in northeastern and central Pennsylvania characterized by distinct social, natural, food, and physical activity environments and their relation to type 2 diabetes

Katherine A. Moon[a,*], Melissa N. Poulsen[b], Karen Bandeen-Roche[c], Annemarie G. Hirsch[b], Joseph DeWalle[b], Jonathan Pollak[a], Brian S. Schwartz[a,b]

Background: Understanding geographic disparities in type 2 diabetes (T2D) requires approaches that account for communities' multidimensional nature.

Methods: In an electronic health record nested case–control study, we identified 15,884 cases of new-onset T2D from 2008 to 2016, defined using encounter diagnoses, medication orders, and laboratory test results, and frequency-matched controls without T2D (79,400; 65,069 unique persons). We used finite mixture models to construct community profiles from social, natural, physical activity, and food environment measures. We estimated T2D odds ratios (OR) with 95% confidence intervals (CI) using logistic generalized estimating equation models, adjusted for sociodemographic variables. We examined associations with the profiles alone and combined them with either community type based on administrative boundaries or Census-based urban/rural status.

Results: We identified four profiles in 1069 communities in central and northeastern Pennsylvania along a rural-urban gradient: "sparse rural," "developed rural," "inner suburb," and "deprived urban core." Urban areas were densely populated with high physical activity resources and food outlets; however, they also had high socioeconomic deprivation and low greenness. Compared with "developed rural," T2D onset odds were higher in "deprived urban core" (1.24, CI = 1.16–1.33) and "inner suburb" (1.10, CI = 1.04–1.17). These associations with model-based community profiles were weaker than when combined with administrative boundaries or urban/rural status.

Conclusions: Our findings suggest that in urban areas, diabetogenic features overwhelm T2D-protective features. The community profiles support the construct validity of administrative-community type and urban/rural status, previously reported, to evaluate geographic disparities in T2D onset in this geography.

Keywords: Built environment; Environmental factors; Food environment; Natural environment; Social environment; Type 2 diabetes

[a]Department of Environmental Health and Engineering, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; [b]Department of Population Health Sciences, Geisinger, Danville, PA; [c]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

This publication was made possible by Cooperative Agreement Number DP006293 funded by the US Centers for Disease Control and Prevention, Division of Diabetes Translation.

Data are available on reasonable request (i.e., IRB approval and a data use agreement).

K.A.M.: Conceptualization, methodology, formal analysis, data curation, writing—original draft, writing—review & editing, visualization; M.N.P.: conceptualization, writing – review & editing; K.B.-R.: methodology, software, writing—review & editing; A.G.H.: conceptualization, writing—review & editing, supervision, project administration, funding acquisition; J.D.: formal analysis, visualization, writing—review & editing; J.P.: formal analysis, writing – review & editing; B.S.S, conceptualization, writing—review & editing, supervision, project administration, funding acquisition.

SDC *Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.environepidem.com).*

**Corresponding Author. Address: Department of Environmental Health, Boston University School of Public Health, 715 Albany Street Boston, MA 02118. E-mail: katmoon@bu.edu (K. A. Moon).*

Environmental Epidemiology (2024) 8:e328

Received 1 October, 2023; Accepted 15 July, 2024

Published online 20 August 2024

DOI: 10.1097/EE9.0000000000000328

Introduction

In 2019, an estimated 11.3% of persons in the United States (US) had diabetes.[1] Type 2 diabetes (T2D) and its complications—including cardiovascular, renal, and peripheral nervous system diseases—present a significant public health challenge.[1] In the US, geographic disparities exist across neighboring counties in T2D incidence and prevalence and across an urban-rural gradient.[1] Racial and ethnic minority groups and disadvantaged communities bear an inequitable burden of diabetes.[2]

Understanding the relationship between community features that affect T2D and related outcomes could help inform population-level interventions to improve T2D outcomes.[2] Numerous prior studies have identified community features associated with increasing or decreasing T2D risk, including socioeconomic deprivation, a "walkable" utilitarian physical activity environment, resources for leisure-time fitness and recreation,

What this study adds

We constructed a four-category typology of the social, natural, physical activity, and food environments using latent profile modeling in 1413 communities in central and northeastern Pennsylvania. In a case–control study, urban profiles with the highest socioeconomic deprivation had elevated odds of type 2 diabetes (T2D) onset. In urban communities, we found the profiles provided independent information on T2D risk, although the T2D associations were weaker than those we observed by categorizing communities using administrative boundaries and Census urban/rural status.

greenspace, and food environment characteristics.[3] These community features likely influence T2D through multiple pathways related to health behaviors (e.g., diet and physical activity), psychosocial factors (e.g., stress and social interactions), and environmental toxicants (e.g., air pollution).[4]

Most studies of community-level features and health outcomes have examined specific features individually. Statistical approaches that account for communities' multidimensional nature, where harmful and protective features coexist, may provide a more complete understanding of contextual determinants of health.[5] Prior studies in the US and Canada have created neighborhood profiles based on finite mixture models or cluster analysis of social and built environment indicators and linked them to health outcomes including adiposity,[6,7] physical activity,[6] and lung function.[8]

This study's objectives were three-fold: (1) use finite mixture modeling to construct community profiles based on multidimensional community features hypothesized to be related to T2D, (2) describe the community profiles in relation to existing administrative boundaries previously evaluated in relation to T2D in the study area, and (3) evaluate whether the profiles were associated with T2D onset and how this may inform understanding of community features that influence T2D risk.

Methods

This study was conducted as a collaboration between Geisinger and Johns Hopkins Bloomberg School of Public Health, as part of the Centers for Disease Control and Prevention-funded Diabetes LEAD (Location, Environmental Attributes, and Disparities) Network. The Diabetes LEAD Network was created to provide scientific evidence to develop targeted interventions and policies to prevent T2D and related health outcomes across the US.[9] The Geisinger Institutional Review Board approved this study. In this analysis, we created community profiles using a finite mixture model of community-level measures of the social, natural, physical activity, and food environments. We then estimated the association between these community profiles and new onset T2D in a case–control analysis.

Case–control study population

As previously described,[10] we conducted a case–control study of new-onset T2D nested within an open and dynamic cohort of Geisinger patients who receive care from hospitals and outpatient and urgent care centers in central and northeastern Pennsylvania. Briefly, we identified cases of new-onset T2D (n = 15,888) and control encounters (n = 79,435, representing 65,084 unique persons) among persons without T2D, frequency matched on age, sex, and year, using electronic health record (EHR) data. Eligible participants received care from Geisinger between 2008 and 2016 and had a geocoded address within the 37-county study area. Most eligible participants (88.9%) were geocoded to their home address, 2.8% were geocoded to a 9-digit zip code (ZIP+4) centroid, and 8.3% were geocoded to a 5-digit ZIP code centroid. We geocoded participants' addresses at the last contact with the health system using ArcGIS version 10.4 (ESRI Inc., Redlands, CA). We identified persons with T2D using an algorithm that incorporated encounter diagnoses, medication orders, and laboratory test results (Supplemental Table S1; http://links.lww.com/EE/A294).[10] Briefly, cases were required to meet at least one of three criteria: (1) At least two clinical encounters associated with a T2D diagnosis (International Classification of Disease-9, International Classification of Disease-10, or electronic diagnosis group on two separate dates; (2) At least one T2D medication order, other than metformin or acarbose if female; or (3) At least one encounter with T2D diagnosis and an abnormal laboratory value (random glucose ≥200 mg/dl; fasting glucose ≥126 mg/dl; or hemoglobin A1c ≥6.5%). To ensure there was sufficient health care system contact to identify T2D if present, we required at least two encounters on different days with a primary care provider before selection as a case or control. To distinguish new-onset T2D from a prevalent case entering the health system, cases had to have at least one encounter at least 2 years before meeting case status, with no evidence of T2D during those 2 years. EHR algorithms using similar clinical and laboratory data for diabetes have sensitivity, specificity, and positive predictive values that exceed 90%.[11–13]

Individual-level data

We defined other individual-level variables, including age, sex, race, ethnicity, and tobacco use, using EHR data.[14] We used the percent of time using Medical Assistance, Pennsylvania's needs-based health insurance, as a surrogate for household socioeconomic status.[14]

Community-level data

In this analysis, our aim was to classify communities rather than individuals; therefore, all community-level variables were assigned to the township, borough, or city census tract (hereafter "administrative community type"). We assigned persons to their administrative-community type based on their geocoded address within minor civil division boundaries (boroughs and townships) and census tract boundaries (cities). For distance metrics, we measured from a population-weighted centroid calculated in ArcGIS using 2010 Census population within block groups. Townships range from agriculturally focused rural areas to low-density suburbs, boroughs are generally walkable small towns of 5000 to 10,000 persons with a core with a gridded street network, and cities are small to medium-sized urban areas. We used this definition, instead of census tract alone, because rural census tracts are often very large while city minor civil divisions are frequently heterogeneous regarding underlying community characteristics.[10,15] We also classified individuals using the 2010 Census, hereafter referred to as "urban/rural status," which defined areas with at least 50,000 persons as urbanized areas (i.e., major urban) and areas with at least 2500 persons and less than 50,000 persons as urban clusters (i.e., smaller urban).[16] Other geographic areas were considered rural.

Social environment

We measured the socioeconomic environment, an important subdomain of the social environment, with a previously described scale of community socioeconomic deprivation based on a factor analysis of Census indicators.[17] Briefly, we used the sum of six z-transformed indicators (% unemployed, % less than a high school education, % below poverty level, % on public assistance, % not in the workforce, and % without a car) from 2010 to 2014 American Community Survey.[18]

Natural environment

We measured the natural environment using three variables: greenness, forested land cover, and blue space. We measured greenness using the normalized difference vegetation index during peak greenness (July 3–19, 2011) from the Moderate Resolution Imaging Spectroradiometer (version 6, level 3, 250-meter resolution) Aqua satellite.[19] Using the pixel nearest to the population-weighted centroid, we calculated the mean of the central pixel and 24 surrounding pixel values, essentially smoothing across a 1250 × 1250 meter grid.[10,20] We measured the percentage of forested land cover from the 2011 National Land Cover Database derived from Landsat satellite imagery (30-meter resolution).[21] We measured blue space using the distance in miles from a population-weighted centroid to the nearest water polygon (i.e., lake, river, tributary, or large stream) in ArcGIS, as previously described.[22]

Physical activity environment

We hypothesized that both leisure-time physical activity (i.e., for recreation or fitness) and utilitarian physical activity (i.e., for everyday activities) were important for T2D.[3,23] We measured the utilitarian physical activity environment with six variables: households per square mile, percentage of developed land, intersection density, average block size, average block length, and street connectivity.[24] Together, these community design, land use, and street network variables indicate whether a community's land-use environment is conducive to walking. Households per square mile were obtained from the 2010 census.[25] The percentage of developed land (class 22–24) was obtained from the 2011 National Land Cover Database.[21] All other utilitarian physical activity measures were calculated in ArcGIS 10.4 (ESRI Inc., Redlands, CA) using 2010 data from the Pennsylvania Department of Transportation. We defined the leisure-time physical activity environment using three variables: distance to nearest local park, distance to nearest state or national park, and density (count per square mile) of physical activity establishments (e.g., exercise facilities, gyms, parks, outdoor recreational facilities). We obtained 2015 park location data from the Pennsylvania Department of Conservation and Natural Resources. We defined physical activity establishments using standard industrial classification (SIC) codes from 2010 InfoUSA and Dun and Bradstreet data. The underlying SIC codes are no longer available; however, a description of the fitness and recreational businesses, clubs, and organizations, that include public and private businesses, and indoor and outdoor establishments, is available in Supplemental Table S6; http://links.lww.com/EE/A294.

Food environment

We measured the food environment with relative measures that compared three food outlet types to a total count of food establishments: percentage grocery stores, percentage chain fast food restaurants, and percentage convenience, drug, or dollar stores (hereafter, "convenience stores"). We also evaluated a proximity metric for grocery stores, defined as the distance from the population-weighted centroid to the nearest grocery store in miles. We identified food outlets using SIC codes and keyword searches of 2010 InfoUSA and Dun and Bradstreet data. The underlying SIC codes are no longer available, but a brief description of our approach is available in Supplemental Table S7; http://links.lww.com/EE/A294. We were guided by a similar approach developed by the retail environment and cardiovascular disease study team,[26,27] although we were not able to exactly replicate their methodology because we had access only to the less granular 6-digit SIC codes. For grocery stores, we identified national-chain supermarkets, superstores, and wholesale clubs and added smaller grocery stores through a manual review of business names. For chain fast food restaurants, we identified eating places that specialize in low preparation time foods that are eaten cafeteria-style (no waiter service) or take-away. Given the recent shift of food purchasing from traditional grocery stores,[28] we were also interested in examining the nontraditional food retailers that offer a limited selection of predominantly shelf-stable foods, which we grouped together in this analysis as convenience, drug, or dollar stores.

Statistical analysis

Identification of community profiles

Finite mixture models, a generalization of latent class analysis,[29,30] can categorize communities into homogenous subgroups (i.e., typologies) that capture unobserved heterogeneity by assuming that the sample consists of *K* homogeneous subgroups with distinct patterns of measured variables. Finite mixture models are similar to cluster analysis but they accommodate a range of indicator distributions (e.g., count, categorical, continuous) and allow for specific correlation between subgroups if theoretically warranted.[30] We fit finite mixture models on a data set of 1413 communities using Mplus version 8.1 (Muthén and Muthén 2017) through the MplusAutomation R package.[31] We fit models with maximum likelihood estimators with robust standard errors to account for non-normal distributions and nonindependence of communities. Data management and visualization were conducted in Stata 17 (StataCorp 2021; College Station, TX) and R version 4.2.3 (R Core Team 2023; Vienna, Austria).

Before modeling, we examined the candidate variable distributions using histograms (Supplemental Figure S1; http://links.lww.com/EE/A294) and bivariate correlation plots. Several variables were highly correlated: forested land cover and greenness (Spearman rho: 0.92) and the utilitarian physical activity variables except street connectivity (Spearman rho: 0.94–0.98). Before entering the model, highly skewed variables were categorized (Supplemental Table S1; http://links.lww.com/EE/A294). Community socioeconomic deprivation was approximately normally distributed; otherwise, all other variables were categorized. Physical activity establishment density and the food environment variables had a high proportion of zero values (43.5–74.5.%) and so were dichotomized (0% vs. >0%). We also examined sensitivity analyses using three categories (0%, <median, ≥median). All other variables were entered into the model as ordinal categorical variables based on quintiles.

We first compared models with one to six classes to determine the number of latent classes yielding the best fit while maintaining a parsimonious model. For these models, we selected a core set of variables representing each domain: socioeconomic deprivation, greenness, household density, distance to the nearest local park, physical activity establishment density, and the three relative measures of the food environment. We included two leisure-time physical activity variables to represent indoor and outdoor physical activity resources and all three food environment variables to represent the range of food outlets in the region. We then selected the number of classes based on model fit and classification ability: Bayesian information criterion, entropy, and two likelihood-based tests of model fit (Lo-Mendell-Rubin test and a parametric bootstrap test).[30] After selecting the number of latent classes, we identified the final model in an iterative process by evaluating whether the classification ability of the model improved by adding variables or relaxing model assumptions, including allowing continuous variables to have unequal variance by latent class and allowing residual correlation between variables within a latent class (Supplemental Table S2; http://links.lww.com/EE/A294).

The final model included only the eight original variables and allowed residual correlation between the food environment variables. We assigned each community to a community profile based on the most likely latent class membership. The average latent class probability for most likely class membership ranged from 0.86 to 0.93. We used the estimated means of continuous variables and threshold probabilities of categorical variables (Supplemental Figure S2; http://links.lww.com/EE/A294), along with boxplots of the observed variables grouped by community profile (Supplemental Figure S3; http://links.lww.com/EE/A294) to interpret and assign descriptive labels.

Association between community profiles and type 2 diabetes

We estimated odds ratios (OR) and 95% confidence intervals (CI) for three community typology variables in association with T2D

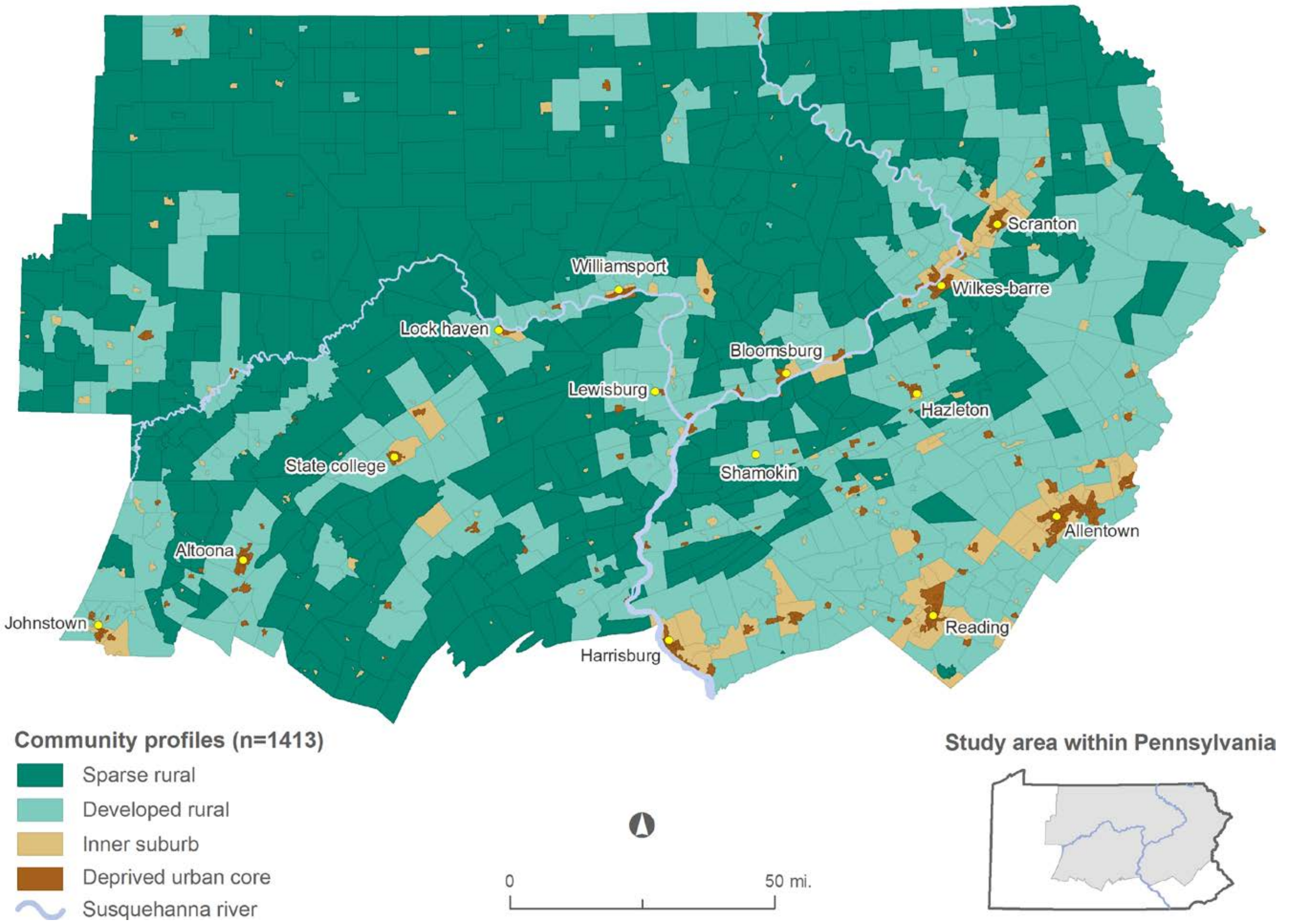

Figure 1. Community profiles characterizing distinct social, natural, physical activity, and food environments in central and northeastern Pennsylvania. Spatial distribution of four community profiles constructed from 1413 communities in the central and northeastern Pennsylvania study area, using community-level measures from the social, natural, physical activity, and food environment.

onset using logistically generalized estimating equation models with robust standard errors. One township did not have an assigned community profile; thus, we excluded four T2D cases and 15 controls (Supplemental Figure S4; http://links.lww.com/EE/A294). In separate models, we evaluated the associations between three community typology variables: (1) the community profiles, (2) community profiles combined with administrative-community type (hereafter, "combined administrative-community profile"), and (3) community profiles combined with urban/rural status (hereafter, "combined Census-community profile"). We created these two combined variables to evaluate community type at a more granular level by cross-tabulating the community profiles with existing community typology variables. We merged some strata of the combined typology variables with adjacent strata when data were sparse: city census tracts were combined with boroughs in "developed rural," townships were combined with boroughs in "deprived urban core," urbanized areas were combined with urban clusters in "sparse rural," and rural was combined with urban cluster in "deprived urban core." We adjusted for the following potential confounders: age (mean-centered age, age^2, and age^3), sex (male vs. female), race (White vs. all other racial groups), ethnicity (Hispanic vs. non-Hispanic), and medical assistance (<0% vs. ≥0% time receiving).

Results

Community profile characteristics

In the overall study area (n = 1413 communities), we constructed four community profiles from a finite mixture model of social, natural, physical activity, and food environment measures (Figure 1). The profiles generally characterized communities along a gradient from rural to urban. In the subset of communities with T2D cases or controls (n = 1069), 30.7% of communities were "sparse rural," 20.0% were "developed rural," 21.0% were "inner suburb," and 19.4% were "deprived urban core" (Table 1). Compared to the overall study area, T2D case–control study communities were more likely to be townships and more likely to have physical activity facilities and chain fast food restaurants, but less likely to have convenience stores (Supplemental Table S3; http://links.lww.com/EE/A294); however, the distinguishing characteristics and proportions of each community profile were similar (Supplemental Table S4; http://links.lww.com/EE/A294).

In the subset of communities with T2D cases or controls (n = 1069), townships were divided evenly between "sparse rural" and "developed rural" while city census tracts were most often classified as "deprived urban core," followed by "inner suburb" (Table 1). Boroughs were mostly split across "inner suburb," "deprived urban core," and "developed rural." "Sparse rural" communities had the lowest population density, and least "walkable" land use, and the highest greenness and forested land cover. "Sparse rural" communities were overwhelmingly townships (93.6%). In comparison to "sparse rural," "developed rural" communities had higher population density and more "walkable" land use, more physical activity facilities and food outlets, and lower greenness and forest cover. "Developed rural" communities were mostly townships. Compared with "developed rural," "inner suburb" communities were more

Table 1.

Characteristics of community profiles in 1069 communities in central and northeastern Pennsylvania with T2D cases and controls

Characteristic	All communities n = 1069 (100%)	"Sparse rural" n = 328 (30.7%)	"Developed rural" n = 310 (20.0%)	"Inner suburb" n = 224 (21.0%)	"Deprived urban core" n = 207 (19.4%)
Community classifications, n (ROW%)					
Administrative-community type, n (%)					
Township	632 (100)	316 (50.0)	278 (44.0)	35 (5.5)	3 (0.5)
Borough	291 (100)	12 (4.1)	29 (10.0)	169 (58.1)	81 (27.8)
City census tract	146 (100)	0 (0.0)	3 (2.1)	20 (13.7)	123 (84.2)
Census urban-rural definition, n (%)					
Rural	596 (100)	317 (53.2)	199 (33.4)	76 (12.8)	4 (0.7)
Urban cluster	151 (100)	10 (6.6)	48 (31.8)	50 (33.1)	43 (28.5)
Urbanized area	322 (100)	1 (0.3)	63 (19.6)	98 (30.4)	160 (49.7)
Community classifications, n (COL%)					
Administrative-community type, n (%)					
Township	632 (59.1)	316 (96.3)	278 (89.7)	35 (15.6)	3 (1.4)
Borough	291 (27.2)	12 (3.7)	29 (9.4)	169 (75.4)	81 (39.1)
City census tract	146 (13.7)	0 (0.0)	3 (1.0)	20 (8.9)	123 (59.4)
Census urban-rural definition, n (%)					
Rural	596 (55.8)	317 (96.6)	199 (64.2)	76 (33.9)	4 (1.9)
Urban cluster	151 (14.1)	10 (3.0)	48 (15.5)	50 (22.3)	43 (20.8)
Urbanized area	322 (30.1)	1 (0.3)	63 (20.3)	98 (43.8)	160 (77.3)
Community features					
Socioeconomic deprivation, unitless, mean (SD)[a]	0.0 (3.1)	−0.9 (3.2)	−0.6 (2.5)	0.1 (2.8)	2.5 (2.8)
Greenness, peak NDVI, median (IQR)[a]	0.8 (0.7, 0.8)	0.8 (0.8, 0.9)	0.8 (0.8, 0.8)	0.7 (0.7, 0.8)	0.6 (0.5, 0.6)
% Forest land cover, median (IQR)	48.8 (16.9, 69.7)	72.7 (58.2, 83.6)	59.0 (45.0, 69.8)	29.4 (12.7, 43.3)	4.0 (0.0, 12.0)
Distance to nearest blue space, miles, median (IQR)	0.7 (0.3, 1.2)	0.8 (0.4, 1.5)	0.8 (0.3, 1.2)	0.4 (0.2, 1.0)	0.6 (0.3, 1.0)
Households per mile^2, median (IQR)[a]	73.6 (21.2, 822.7)	13.8 (8.6, 22.1)	56.4 (35.8, 95.5)	561.6 (358.5, 915.0)	1833.6 (1307.1, 2933.9)
% Developed land cover, median (IQR)	13.6 (6.1, 62.0)	5.1 (3.6, 6.5)	10.6 (7.5, 15.9)	48.1 (34.7, 66.0)	90.6 (77.1, 98.8)
Intersection density per mile^2, median (IQR)	19.3 (7.5, 107.2)	6.0 (4.4, 8.3)	14.1 (9.5, 21.5)	87.0 (52.3, 126.8)	181.0 (138.1, 249.9)
Average block size, feet^2, median (IQR)	4.9e+06 (515235.1, 1.5e+07)	2.0e+07 (1.3e+07, 3.0e+07)	6.8e+06 (3.9e+06, 1.1e+07)	755014.3 (439484.6, 1.2e+06)	248897.7 (164830.5, 375282.2)
Average block length, feet, median (IQR)	972.0 (453.0, 1732.0)	1993.0 (1635.0, 2440.5)	1174.5 (898.0, 1485.0)	490.0 (404.5, 599.5)	359.0 (300.0, 415.0)
Street connectivity, %, median (IQR)	0.9 (0.8, 0.9)	0.8 (0.8, 0.9)	0.8 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.9, 1.0)
Distance to nearest local park, miles, median (IQR)[a]	0.6 (0.2, 2.0)	3.2 (1.6, 4.5)	0.8 (0.4, 1.4)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)
Distance to nearest state or national park, miles, median (IQR)	7.4 (4.7, 11.0)	8.0 (4.8, 12.3)	7.3 (4.4, 10.6)	7.2 (4.1, 10.7)	7.0 (5.3, 9.3)
Physical activity facilities per mile^2, median (IQR)[a]	0.0 (0.0, 0.9)	0.0 (0.0, 0.0)	0.1 (0.0, 0.2)	0.5 (0.0, 1.4)	3.2 (1.6, 5.9)
Distance to nearest grocery store, miles, median (IQR)	2.7 (0.8, 6.1)	6.6 (4.4, 10.1)	1.2 (0.7, 3.7)	2.8 (1.4, 4.9)	0.4 (0.2, 0.7)
% Grocery stores per total food outlets, median (IQR)[a]	0.0 (0.0, 6.1)	0.0 (0.0, 0.0)	0.0 (0.0, 5.0)	0.0 (0.0, 3.9)	2.6 (0.0, 8.3)
% Fast food restaurants per total food outlets, median (IQR)[a]	6.3 (0.0, 14.3)	0.0 (0.0, 0.0)	0.0 (0.0, 14.3)	0.0 (0.0, 16.7)	9.1 (0.0, 15.4)
% Convenience stores per total food outlets, median (IQR)[a]	16.7 (0.0, 25.0)	0.0 (0.0, 0.0)	15.4 (0.0, 28.6)	16.2 (0.0, 26.5)	17.6 (11.1, 23.5)

[a]These eight variables were used to construct the community profiles in a finite mixture model.

COL%, column percentage; IQR, interquartile range (25th percentile–75th percentile); NDVI, normalized difference vegetation index; ROW%, row percentage.

densely populated and more likely to have physical activity facilities but were similar in the proportions of the three food outlet types. "Inner suburb" communities were most often boroughs, followed by townships or city census tracts. "Deprived urban core" communities had the highest socioeconomic deprivation, population density, "walkable" land use, density of physical activity facilities and food outlets, the shortest average distance to local parks, and the lowest greenness. "Deprived urban core" communities were a mixture of city census tracts and boroughs.

Case–control study population characteristics

Most T2D cases resided in "inner suburb" communities (40.0%), followed by 24.9% in "deprived urban core," 18.3% in "developed rural," and 17.8% in "sparse rural" (Table 2 and Supplemental Table S5; http://links.lww.com/EE/A294). T2D cases and controls were predominantly middle-aged, non-Hispanic White adults with a Geisinger primary care provider. Although all communities were predominantly non-Hispanic White, the percentage of White and non-Hispanic participants was lowest in "deprived urban core" communities. More participants used Medical Assistance to pay for health care in "deprived urban core" compared with the other community profiles.

Association between community profiles and type 2 diabetes

Adjusted associations between the three community typology variables (profiles alone, profiles combined with administrative-community type, and profiles combined with Census urban/rural categories) are presented in Table 3 and Figure 2. In adjusted models, the T2D odds were higher in "deprived urban core" communities (OR = 1.24; 95% CI = 1.16, 1.33) and "inner suburb" communities (OR = 1.10; 95% CI = 1.04, 1.17) compared to "developed rural" communities. "Sparse rural" communities had similar T2D odds to "developed rural" communities (OR = 1.04; 95% CI = 0.98, 1.11). For both the combined administrative-community profiles and the combined Census-community profiles models, T2D odds generally increased across the rural-to-urban gradient. T2D odds were highest in "deprived urban core" city census tracts (OR = 1.34; 95% CI = 1.23, 1.46), "deprived urban core" boroughs and townships (OR = 1.14; 95% CI = 1.04, 1.24), and "inner suburb" boroughs (OR = 1.09; 95% CI = 1.02, 1.17). Similarly, T2D odds were highest in "deprived urban core" urban areas (OR = 1.24; 95% CI = 1.12, 1.38), followed by "deprived urban core" urban cluster or rural areas (OR = 1.27; 95% CI = 1.17, 1.39), and "inner suburb" urban clusters (OR = 1.16; 95% CI = 1.06, 1.27).

Discussion

We used a finite mixture model to construct community profiles that captured substantive heterogeneity in the social, natural, physical activity, and food environments in a 37-county region in Pennsylvania. The profiles generally characterized communities along a rural-urban continuum. In the case–control analysis, we found elevated odds of T2D onset in the most densely populated profiles: 24% higher in the most urbanized profile, "deprived urban core," and 10% higher in the second most urbanized profile, "inner suburb," each compared with the least developed profile, "developed rural." Our newly identified community profiles, alone and in combination with two existing community typology variables (administrative and Census urban/rural), were generally associated with increasing T2D odds in an exposure-response gradient from the most sparsely populated to the most densely populated geographies. In models where the community profiles were combined with administrative or Census categories, we found that the community profiles provided additional, independent information about the location of elevated T2D onset odds in more urban communities,

Table 2.

Individual-level and community-level characteristics of participants in case–control study of T2D onset in Geisinger EHR 2008–2016

Characteristic	All communities n = 1069 (100%)	"Sparse rural" n = 328 (30.7%)	"Developed rural" n = 310 (20.0%)	"Inner suburb" n = 224 (21.0%)	"Deprived urban core" n = 207 (19.4%)
T2D cases, n	15884 (100)	2824 (17.8)	2908 (18.3)	6194 (40.0)	3958 (24.9)
Control encounters, n	79400 (100)	14501 (18.3)	14169 (17.8)	34321 (43.2)	16409 (20.7)
Unique controls, n	65069 (100)	11791 (18.1)	11680 (18.0)	27945 (42.9)	13643 (21.0)
Age at diagnosis or control selection, years, median (IQR)	55.4 (45.2, 65.2)	55.8 (46.4, 65.3)	55.2 (44.4, 65.5)	55.8 (45.9, 65.3)	54.1 (43.0, 64.6)
Female, n (%)	46768 (49.1)	8237 (47.5)	8568 (50.2)	19516 (48.2)	10447 (51.3)
White[a], n (%)	93257 (97.9)	17227 (99.4)	16789 (98.3)	39555 (97.6)	19686 (96.7)
Hispanic, n (%)	1463 (1.5)	90 (0.5)	210 (1.2)	570 (1.4)	593 (2.9)
Primary care provider, n (%)	72894 (76.5)	13333 (77.0)	13010 (76.2)	31073 (76.7)	15478 (76.0)
Setting of diagnosis/encounter, n (%)					
Outpatient	86032 (90.3)	15860 (91.5)	15313 (89.7)	36932 (91.2)	17927 (88.0)
Medication order	1631 (1.7)	291 (1.7)	316 (1.9)	663 (1.6)	361 (1.8)
Urgent care	2278 (2.4)	299 (1.7)	417 (2.4)	1087 (2.7)	475 (2.3)
Emergency	3257 (3.4)	524 (3.0)	638 (3.7)	1060 (2.6)	1035 (5.1)
Inpatient	1336 (1.4)	219 (1.3)	248 (1.5)	486 (1.2)	383 (1.9)
Medical assistance, % of time receiving, n (%)					
<50%	91587 (96.1)	16853 (97.3)	16406 (96.1)	39429 (97.3)	18899 (92.8)
≥50%	3697 (3.9)	472 (2.7)	671 (3.9)	1086 (2.7)	1468 (7.2)
Administrative-community type, n (%)					
Township	60553 (63.6)	16778 (96.8)	4374 (25.6)	39220 (96.8)	181 (0.9)
Borough	26377 (27.7)	547 (3.2)	11629 (68.1)	1183 (2.9)	13018 (63.9)
City census tract	8354 (8.8)	0 (0.0)	1074 (6.3)	112 (0.3)	7168 (35.2%)
Census urban-rural type, n (%)					
Rural	41483 (43.5)	16247 (93.8)	3164 (18.5)	22002 (54.3)	70 (0.3)
Urban cluster	25497 (26.8)	758 (4.4)	5645 (33.1)	9661 (23.8)	9433 (46.3)
Urbanized area	28304 (29.7)	320 (1.8)	8268 (48.4)	8852 (21.8)	10864 (53.3)

n (%) are column percentages, unless otherwise noted.

[a]Other racial categories included Black or African American, Asian, American Indian and Alaska Native, Native Hawaiian or other Pacific Islander, and "Other."

Table 3.

Associations between community profiles and new onset of type 2 diabetes in the Geisinger EHR, 2008–2016

Exposure variable	Number of communities	Number of T2D cases	Number of control encounters	Number of unique control persons	Odds ratio (95% confidence interval)[a]	
					Unadjusted	Adjusted
Community profile						
"Sparse rural" communities	328	2824	14501	11791	1.04 (0.98, 1.11)	1.04 (0.98, 1.11)
"Developed rural" communities	310	6194	34321	27945	1.00 (Reference)	1.00 (Reference)
"Inner suburb" communities	224	2908	14169	11680	1.12 (1.05, 1.19)	1.10 (1.04, 1.17)
"Deprived urban core" communities	207	3958	16409	13643	1.36 (1.27, 1.47)	1.24 (1.16, 1.33)
Combined administrative-community profiles[b]						
"Sparse rural" & township	316	2734	14044	11417	1.04 (0.98, 1.11)	1.05 (0.98, 1.11)
"Sparse rural" & borough	12	90	457	374	1.06 (0.91, 1.22)	0.99 (0.70, 1.38)
"Developed rural" & township	278	6004	33216	27047	1.00 (Reference)	1.00 (Reference)
"Developed rural" & (borough or city census tract)	32	190	1105	898	0.96 (0.85, 1.08)	0.95 (0.85, 1.07)
"Inner suburb" & township	35	683	3691	3309	1.06 (0.91, 1.22)	1.08 (0.95, 1.24)
"Inner suburb" & borough	169	2021	9608	7908	1.11 (1.04, 1.19)	1.09 (1.02, 1.17)
"Inner suburb" & city census tract	20	204	870	733	1.24 (1.04, 1.47)	1.17 (0.99, 1.38)
"Deprived urban core" & (borough or township)	84	2371	10828	8994	1.20 (1.09, 1.32)	1.14 (1.04, 1.24)
"Deprived urban core" & city census tract	123	1587	5581	4649	1.52 (1.40, 1.64)	1.34 (1.23, 1.46)
Combined Census-community profiles[c]						
"Sparse rural" & rural	317	2658	13589	11034	1.07 (0.99, 1.14)	1.06 (0.99, 1.14)
"Sparse rural" & (urban cluster or urbanized area)	11	166	912	757	0.97 (0.82, 1.15)	0.96 (0.81, 1.13)
"Developed rural" & rural	199	3326	18676	15154	1.00 (Reference)	1.00 (Reference)
"Developed rural" & urban cluster	63	1304	7548	6160	1.11 (1.00, 1.23)	1.07 (0.97, 1.18)
"Developed rural" & urbanized area	48	1564	8097	6631	1.03 (0.93, 1.13)	1.01 (0.92, 1.10)
"Inner suburb" & rural	76	516	2648	2205	1.09 (0.98, 1.21)	1.07 (0.97, 1.18)
"Inner suburb" & urban cluster	98	1419	6849	5654	1.19 (1.09, 1.31)	1.16 (1.06, 1.27)
"Inner suburb" & urbanized area	50	973	4672	3821	1.12 (1.02, 1.23)	1.09 (0.99, 1.20)
"Deprived urban core" & (urban cluster or rural)	164	2138	8796	7402	1.44 (1.31, 1.57)	1.27 (1.17, 1.39)
"Deprived urban core" & urbanized area	43	1820	7613	6241	1.34 (1.20, 1.49)	1.24 (1.12, 1.38)

Bold values indicate that the Odds ratio (95% confidence interval) does not include 1.0 ($P < 0.05$).

[a]Odds ratios from logistic regression models using generalized estimating equations (GEE) with robust standard errors. Models were adjusted for sex (male vs. female), age (mean-centered age, age^2, age^3), race (White vs. all other racial categories), ethnicity (Hispanic vs. non-Hispanic), and Medical Assistance (<0%, ≥0% time receiving).

[b]Administrative type was defined by residence in a township, borough, or city census tract. Some strata had only sparse data and estimates were likely unstable. We combined adjacent strata when there were fewer than five communities in a stratum. The stratum "Sparse rural" & city census tracts (n = 3) were combined with boroughs (n = 29). "Deprived urban core" & township (n = 3) were combined with urban cluster (n = 81). There were no city census tracts classified as sparse rural.

[c]The 2010 Census defined areas with ≥50,000 persons as urbanized areas and areas with ≥2500 and <50,000 persons as urban clusters. All other geographies were considered rural. Some strata had only sparse data and estimates were likely unstable. We adjacent strata when there were fewer than five communities in a stratum. "Sparse rural" urbanized areas (n = 1) were combined with urban cluster (n = 10). "Deprived urban core" rural (n = 4) were combined with urban cluster (n = 160).

beyond categorizations by administrative or Census boundaries. This additional risk information was most evident from the difference in magnitude of associations within city census tracts depending on the community profile of the tract and the similar magnitude of associations for "deprived urban core" regardless of urban/rural status.

This study should also be interpreted in the context of our prior study, where we examined the independent relation of the administrative (township, borough, and city census tract) and Census boundary typologies (urbanized area, urban cluster, and rural) and odds of T2D using the same case–control study population, with negligible differences in exclusion criteria and the same set of adjustment variables.[10] In the prior study, we used an analogous set of models to those presented here, where we examined each typology alone and in combination. Similar to this study's findings, we found that T2D onset odds increased with higher population density: compared with townships in rural areas, city census tracts in smaller urban areas (i.e., urban clusters) or major urban areas (i.e., urbanized areas) had 41% and 33% higher odds of T2D onset, respectively.[10] In this study, we had hypothesized that community profiles derived from a latent variable model of underlying social, natural, built, and food environment variables would allow us to find greater odds of T2D; however, the magnitude of the T2D associations with community profiles was smaller than with the combined administrative-community type and Census urban/rural status.[10] Our cumulative findings of this and our prior study suggests the model-based community profiles do not improve upon existing administrative and Census-based community variables for identifying geographies of elevated T2D risk.

Broadly, our finding of elevated risk of T2D onset in urban geographies is consistent with prior epidemiologic studies in other geographies in the US, although few prior studies have examined rural-urban disparities in onset, rather than prevalence, of T2D. Analyses of geographic disparities in T2D prevalence that adjusted for demographic and socioeconomic measures have found a lower T2D burden in rural areas.[32]

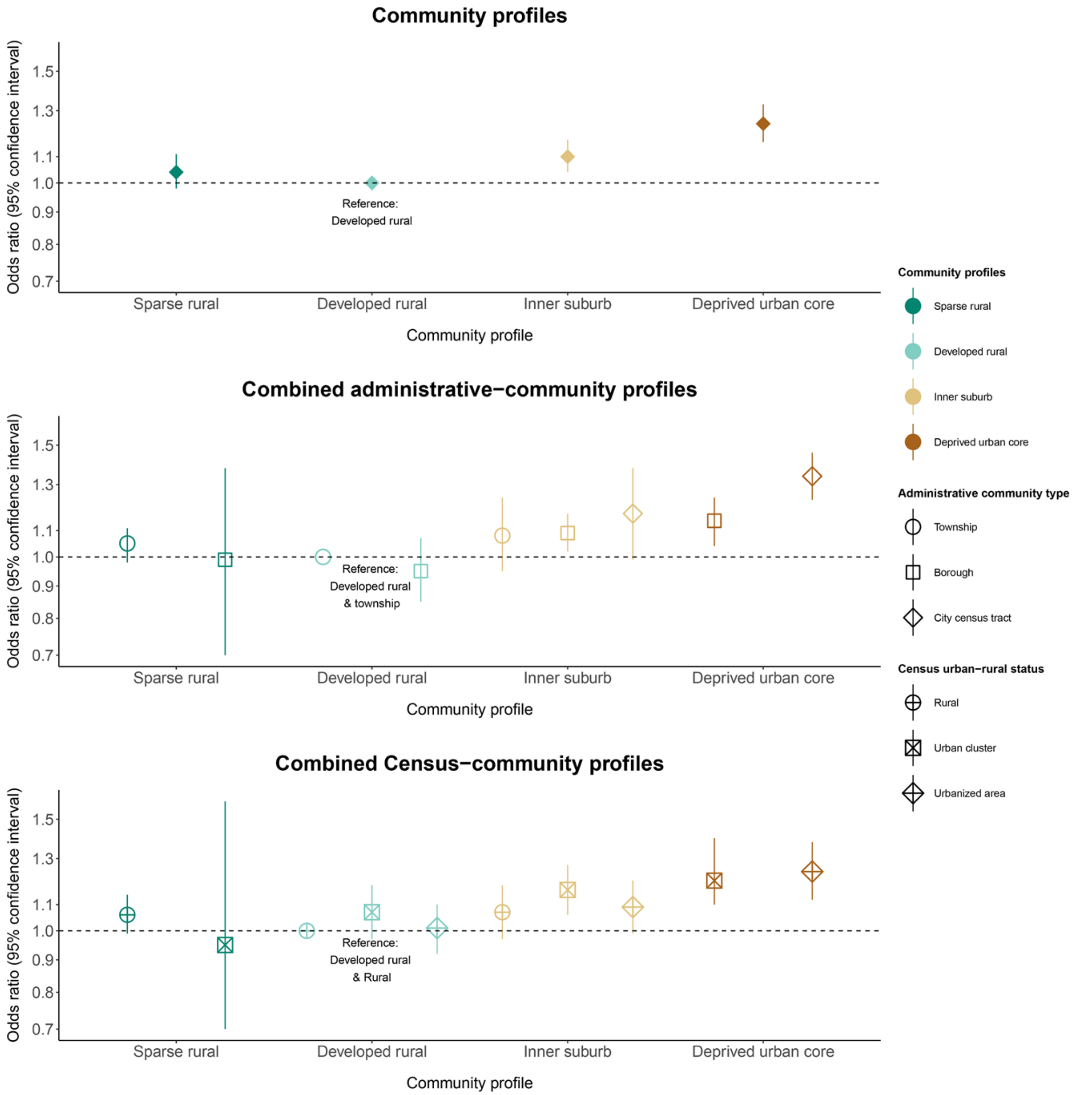

Figure 2. Associations between community profiles and type 2 diabetes onset in a Geisinger EHR case–control study, 2008–2016. Odds ratios from logistic regression models using generalized estimating equations (GEE) with robust standard errors. Models were adjusted for sex (male vs. female), age (mean-centered age, age^2, age^3), race (White vs. all other racial categories), ethnicity (Hispanic vs. non-Hispanic), and medical assistance (<0%, ≥0% time receiving). We combined a stratum with adjacent strata when there were fewer than five communities in a stratum. For the variable combined administrative-community profiles, the stratum "Sparse rural" & city census tracts (n = 3) was combined with boroughs (n = 29). "Deprived urban core" & township (n = 3) were combined with urban cluster (n = 81). There were no city census tracts classified as sparse rural. For the variable combined Census-community profiles, "Sparse rural" urbanized area (n = 1) was combined with urban cluster (n = 10). "Deprived urban core" rural (n = 4) were combined with urban cluster (n = 160).

Regional differences in contextual factors may modify this association; in one study, the prevalence of T2D was higher in rural areas only in the southern US.[33]

In our study, constructing community profiles provided insight into the construct validity of administrative-community type alone and in combination with urban/rural status to evaluate urban-rural geographic disparities in T2D onset.[10] Overall, the community profiles, both independently and combined with the two existing community typology variables, appeared to characterize the urban-rural gradient of community features important for T2D, though not perfectly. For example, the community profiles differentiated between "sparse rural" and "developed rural" latent classes, but both profiles had similar T2D odds. Because the "sparse rural" and "developed rural" classes included essentially all townships, this suggests that for townships, the added differentiation offered by the community profiles did not provide additional benefit beyond administrative-community type categories for capturing geographic disparities in T2D. In contrast, for boroughs and city census tracts, the heterogeneity in how communities were classified by community profile versus administrative-community type or urban/rural status suggests these categorizations capture slightly different constructs.

Both community profiles with elevated T2D onset odds, "deprived urban core" and "inner suburb," were characterized by a mixture of diabetogenic features and T2D-protective features, making it challenging to identify which modifiable community features are responsible for the disparities in T2D onset. However, we hypothesize that elevated T2D onset odds in urban, compared with rural, communities are driven by the higher socioeconomic deprivation observed in most urbanized communities. Socioeconomic deprivation has been consistently associated with poor health outcomes in the study area,[34–36] and with T2D incidence, prevalence, and control in other regions.[3] Socioeconomic deprivation, as an upstream determinant of community resources, can influence the availability, accessibility, and quality of natural, physical activity, and food environments.[37] Thus, higher socioeconomic deprivation in urban communities could overwhelm the coexisting T2D-protective features, such as "walkable" land use and higher access to physical activity facilities.

Limitations

This analysis had limitations. Although we selected community-level variables for inclusion based on a priori hypotheses and data availability, finite mixture models are sensitive to analyst decisions. By defining aggregate community features using an administrative boundary, we likely introduced spatial misclassification (e.g., the modifiable areal unit problem, edge effects);[38] however, in our study area, minor civil divisions are policy-relevant administrative and governmental units. Residual confounding, particularly by individual or household socioeconomic status, could have remained in analyses of T2D onset; however, we adjusted for key sociodemographic potential confounders, including receipt of medical assistance.[14] We were unable to account for residential self-selection, whereby characteristics such as socioeconomic status and race influence residential location choice and individual behaviors, and thus may bias our estimates of the associations between community-level variables with T2D onset. We used the available address data in the EHR, the address at last contact with the health system, to define residential location. Our sample was restricted to adult patients who had at least two encounters with a primary care physician, making it a relatively stable population. Future studies could directly assess the impact of each of the community-level variables on T2D risk using a supervised mixture modeling approach.

Conclusions and future directions

Our study was one of the first to create a typology of the social, natural, physical activity, and food environments in a range of community contexts using latent variable methods, from rural to small to medium-sized urban areas; to evaluate this typology in relation to new onset T2D, and to compare those associations with existing community typology variables. The latent profile analysis-derived community profiles and their associations with T2D onset suggest that the risk of T2D onset increases from rural to small and medium-sized urban communities in central and northeastern Pennsylvania. The community profiles provided granular information on the location of elevated T2D onset risk beyond existing community typology variables, particularly in the most urbanized communities.

Our findings also provide support for the construct validity of using administrative-community type and Census urban-rural status to characterize geographic disparities in T2D onset in this study area, especially in rural communities where we observed substantial overlap in the community profiles and community definitions. Given the slightly weaker magnitude of T2D associations with community profiles compared with those previously observed using administrative-community type and Census urban/rural status,[10] the existing community typology variables appear somewhat superior for identifying geographic disparities in T2D onset in this study geography. They certainly offer a simpler approach without the need for complex latent variable modeling. However, the community profiles provide important clues as to the potentially modifiable community features that most influence disparities in T2D onset.

Our approach to creating community profiles using a finite mixture model could be useful in other geographic contexts to examine geographic disparities in T2D onset, especially those without existing administrative or governmentally relevant boundaries. Only 20 states, including Pennsylvania, have minor civil divisions that function as governmental units, and the meaningfulness of Census designations within counties varies across states.[37] In contrast, the Census urban-rural status definition is often used, de facto, to determine policies and allocation of resources across the US. In addition, community profiles could be used to examine disparities in other cardiometabolic or cardiovascular outcomes that share common causal pathways in the social, natural, physical activity, and food environment domains to T2D. Together with our prior findings in this study population,[10] our analyses of community types and community features in relation to T2D onset in our study region suggest that future research should evaluate how modifiable features of urban communities, primarily socioeconomic deprivation, could be improved to prevent T2D development.

Conflicts of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

References

1. Centers for Disease Control and Prevention. National Diabetes Statistics Report. Available at: https://www.cdc.gov/diabetes/php/data-research/index.html. Accessed 13 June 2023.
2. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social determinants of health and diabetes: a scientific review. *Diabetes Care*. 2021;44:258–279.
3. Bilal U, Auchincloss AH, Diez-Roux AV. Neighborhood environments and diabetes risk and control. *Curr Diab Rep*. 2018;18. doi:10.1007/s11892-018-1032-2.
4. Gullón P, Lovasi GS. Designing healthier built environments. In: Kawachi I, Duncan DT, eds. *Neighborhoods and Health*. 2nd Edition. Oxford University Press; 2018.
5. Lekkas P, Stankov I, Daniel M, Paquet C. Finite mixture models in neighbourhoods-to-health research: a systematic review. *Health Place*. 2019;59:102140.
6. Adams MA, Sallis JF, Kerr J, et al. Neighborhood environment profiles related to physical activity and weight status: a latent profile analysis. *Prev Med*. 2011;52:326–331.
7. Barnett TA, Ghenadenik AE, Van Hulst A, et al. Neighborhood built environment typologies and adiposity in children and adolescents. *Int J Obes (Lond)*. 2022;46:588–596.
8. Humphrey JL, Lindstrom M, Barton KE, et al. Social and environmental neighborhood typologies and lung function in a low-income, urban population. *Int J Environ Res Public Health*. 2019;16:1133–1114.
9. Hirsch AG, Carson AP, Lee NL, et al. 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. doi:10.2196/21377.
10. Schwartz BS, Pollak J, Poulsen MN, et al. Association of community types and features in a case-control analysis of new onset type 2 diabetes across a diverse geography in Pennsylvania. *BMJ Open*. 2021;11. doi:10.1136/bmjopen-2020-043528.
11. Lawrence JM, Black MH, Zhang JL, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *Am J Epidemiol*. 2014;179:27–38.
12. Zhong VW, Obeid JS, Craig JB, et al. An efficient approach for surveillance of childhood diabetes by type derived from electronic health

record data: the SEARCH for diabetes in youth study. *J Am Med Inform Assoc*. 2016;23:1060–1067.
13. Zhong VW, Pfaff ER, Beavers DP, et al; Search for Diabetes in Youth Study Group. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for diabetes in youth study. *Pediatr Diabetes*. 2014;15:573–584.
14. Casey JA, Pollak J, Glymour MM, Mayeda ER, Hirsch AG, Schwartz BS. Measures of SES for electronic health record-based research. *Am J Prev Med*. 2018;54:430–439.
15. Schwartz BS, Stewart WF, Godby 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;41:e17–e28.
16. Urban Area Criteria for the 2010 Census, 76 F.R. 53030 (proposed August 24, 2011). Available at: https://www.federalregister.gov/documents/2011/08/24/2011-21647/urban-area-criteria-for-the-2010-census.
17. Nau C, Schwartz BS, Bandeen-Roche K, et al. Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity (Silver Spring)*. 2015;23:207–212.
18. U.S. Census Bureau. American Community Survey 5-year estimate 2010-2014. Published 2015. Available at: https://www.census.gov/data/developers/data-sets/acs-5year.html. Accessed 1 June 2020.
19. Didan K. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC.* Usgs. Published 2015. Available at: https://cmr.earthdata.nasa.gov/search/concepts/C194001241-LPDAAC_ECS.html. Accessed 10 October 2017.
20. Casey JA, James P, Rudolph KE, Wu CD, Schwartz BS. Greenness and birth outcomes in a range of Pennsylvania communities. *Int J Environ Res Public Health*. 2016;13. doi:10.3390/ijerph13030311.
21. Homer C, Dewitz J, Yang L, et al. Completion of the 2011 national land cover database for the conterminous United States – representing a decade of land cover change information. *Photogramm Eng Remote Sens*. 2015;81:345–354.
22. Poulsen MN, Schwartz BS, DeWalle J, et al. Proximity to freshwater blue space and type 2 diabetes onset: the importance of historical and economic context. *Landsc Urban Plann*. 2021;209:104060.
23. Moon KA, Nordberg CM, Orstad SL, et al. Mediation of an association between neighborhood socioeconomic environment and type 2 diabetes through the leisure-time physical activity environment in an analysis of three independent samples. *BMJ Open Diabetes Res Care*. 2023;11:e003120.
24. Meeker MA, Schwartz BS, Bandeen-Roche K, et al. Assessing measurement invariance of a land use environment construct across levels of urbanicity. *GeoHealth*. 2022;6.
25. U.S. Census Bureau. 2000 Decennial Census, Summary File 1 (SF 1). 2000. Available at: https://api.census.gov. Accessed April 2020.
26. Drexel University Urban Health Collaborative. *Retail Environment and Cardiovascular Disease*. Published (n.d.). https://drexel.edu/uhc/research/projects/retail-environment-cardiovascular-disease/. Accessed 21 June 2024.
27. Kaufman TK, Sheehan DM, Rundle A, et al. Measuring health-relevant businesses over 21 years: refining the National Establishment Time-Series (NETS), a dynamic longitudinal data set. *BMC Res Notes*. 2015;8:1–13.
28. Feng W, Page ET, Cash SB. Dollar stores and food access for rural households in the United States, 2008–2020. *Am J Public Health*. 2023;113:331–336.
29. Collins LM, Stephanie TL. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. John Wiley & Sons, Incorporated; 2009.
30. McLachlan GJ, Peel D. *Finite Mixture Models*. Wiley.; 2000. Available at: https://onlinelibrary.wiley.com/doi/book/10.1002/0471721182
31. Hallquist MN, Wiley JF. MplusAutomation: an R package for facilitating large-scale latent variable analyses in Mplus. *Struct Equ Modeling*. 2018;25:621–638.
32. O'Connor A, Wellenius G. Rural-urban disparities in the prevalence of diabetes and coronary heart disease. *Public Health*. 2012;126:813–820.
33. Redmond ML, Ablah E, Dugan K. A geo-stratified analysis of associations between socioeconomic factors and diabetes risk. *Kans J Med*. 2022;15.
34. Poulsen MN, Glass TA, Pollak J, 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;15. doi:10.1016/j.pmedr.2019.100939.
35. Schwartz BS, Kolak M, Pollak JS, et al. Associations of four indexes of social determinants of health and two community typologies with new onset type 2 diabetes across a diverse geography in Pennsylvania. *PLoS One*. 2022;17:e0274758.
36. Hirsch AG, Nordberg CM, Chang A, et al. Association of community socioeconomic deprivation with evidence of reduced kidney function at time of type 2 diabetes diagnosis. *SSM Popul Health*. 2021;15:100876.
37. U.S. Census Bureau. *County Subdivisions (Chapter 8)*. U.S. Department of Commerce Economics and Statistics Administration; 1994. Available at: https://www2.census.gov/geo/pdfs/reference/GARM/GARMcont.pdf
38. Duncan DT, Regan SD, Chaix B. Operationalizing Neighborhood Definitions in Health Research: Spatial Misclassification and Other Issues. In: *Neighborhoods and Health* (2*nd* Ed).; 2018.