Title

Identifying pre-diabetes hotspots in Northern California using geospatial analysis: opportunities to target diabetes prevention strategies and improve health equity.

Journal

BMJ Open, 14(12)

Authors

Thomas, Tainayah
Duru, Obidiugwu
Yassin, Maher
et al.

Publication Date

2024-12-20

DOI

10.1136/bmjopen-2024-087274

Peer reviewed

BMJ Open

Identifying pre-diabetes 'hotspots' in Northern California using geospatial analysis: opportunities to target diabetes prevention strategies and improve health equity

Tainayah W Thomas ,[1,2] Obidiugwu Kenrik Duru,[3] Maher Yassin,[2] Luis A Rodriguez,[2] Tannaz Moin,[3] Yelba Castellon-Lopez,[4] Julie Schmittdiel[2]

To cite: Thomas TW, Duru OK, Yassin M, *et al*. Identifying pre-diabetes 'hotspots' in Northern California using geospatial analysis: opportunities to target diabetes prevention strategies and improve health equity. *BMJ Open* 2024;**14**:e087274. doi:10.1136/bmjopen-2024-087274

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-087274).

Received 05 April 2024
Accepted 08 November 2024

[1]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, USA
[2]Division of Research, Kaiser Permanente, Pleasanton, California, USA
[3]Department of Medicine, University of California Los Angeles, Los Angeles, California, USA
[4]Cedars-Sinai Comprehensive Cancer Center, Los Angeles, California, USA

Correspondence to
Dr Tainayah W Thomas; tainayah@stanford.edu

ABSTRACT

Objectives The US Preventive Services Task Force recommends screening of adults aged 35–70 with a body mass index ≥25 kg/m^2 for type 2 diabetes and referral of individuals who screen positive for pre-diabetes to evidence-based prevention strategies. The diabetes burden in the USA is predicted to triple by 2060 necessitating strategic diabetes prevention efforts, particularly in areas of highest need. This study aimed to identify pre-diabetes hotspots using geospatial mapping to inform targeted diabetes prevention strategies. A 'hotspot' is defined as a cluster of 3 or more neighbouring census tracts with elevated pre-diabetes prevalence.

Design A cross-sectional study using ArcGIS software to geospatially map pre-diabetes prevalence hotspots. We used health system and census data to identify pre-diabetes hotspots using a systematic five-step geoprocessing approach that made use of incremental spatial autocorrelation and Getis-Ord Gi*.

Setting This study was set in Kaiser Permanente Northern California (KPNC), an integrated health delivery system with over four million members.

Participants KPNC adults ages 35–70 who underwent a haemoglobin A1c (HbA1c) or fasting plasma glucose (FPG) screening test in 2019 were mapped to census tracts in Northern California. People were considered to have pre-diabetes with an HbA1c of 5.7%–6.4% (39–46 mmol/mol) or FPG 100–125 mg/dL.

Primary and secondary outcome measures Individual and census-level characteristics were compared between hotspots and non-hotspots using χ^2 and Wilcoxon rank sum tests, as well as risk differences (RDs) and Hodges-Lehmann (HL) estimates of location shift. Individual-level characteristics were derived from electronic health records and administrative data, while census-level characteristics were derived from the 2019 American Community Survey.

Results A total of 760 044 adults met the study inclusion criteria and 40% had pre-diabetes. Individuals in pre-diabetes hotspots were less likely to be non-Hispanic white (33.6% vs 50.6%, RD: −17.04%, 95% CI −17.26% to −16.81%, p<0.0001) and more likely to have overweight or obesity (72.2% vs 69.2%, RD: 2.95%, 95% CI 2.73% to 3.16%, p<0.0001). Census tracts within hotspots had lower levels of household income (HL estimate: −3651.00, 95% CI −7256.00 to −25.00), per cent of adults with bachelor's degrees or higher (HL estimate: −9.08, 95% CI −10.94 to −7.24) and median home values (HL estimate: −113 200.00, 95% CI −140 600.00 to −85 700.00) and higher rates of household poverty (HL estimate: 0.96, 95% CI 0.55 to 1.37), unemployment (HL estimate: 0.39, 95% CI 0.24 to 0.54), household public assistance (HL estimate: 0.97, 95% CI 0.76 to 1.18) and per cent receiving Medicaid (HL estimate: 4.56, 95% CI 3.40 to 5.76) (p<0.05 for all).

Conclusions We found that individual-level and census tract-level socioeconomic status, obesity prevalence and race and ethnicity categories of patients living in pre-diabetes hotspots differed from those not identified as a hotspot. Policy-makers and care providers can use this information to target diabetes prevention resources and outreach by enacting policies that provide insurance coverage for low-income populations and placing diabetes prevention programmes in communities with highest need.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Employed widely used and accepted spatial methods, including incremental spatial autocorrelation and Getis-Ord Gi.*
⇒ Used a systematic, rigorous and previously vetted five-step approach to derive the distance band that defines neighbouring census tracts while accounting for the large variation in census tract areas.
⇒ Spatial data are subject to numerous sources of bias, including the modifiable areal unit problem, boundary problem and ecologic fallacy.
⇒ Individual and census-level characteristics examined between hotspots and non-hotspots were not adjusted for potential associations with each other.

INTRODUCTION

Over 38% of adults in the USA have pre-diabetes or elevated glucose levels that increase risk for incident type 2 diabetes.[1] Population-based estimates of future diabetes

burden in the USA predict a tripling of diagnosed diabetes by 2060, with a concurrent widening in racial and ethnic disparities in diabetes burden.[1] Research shows that without treatment, approximately 30% of patients with pre-diabetes can develop diabetes over a 3-year period and the lifetime risk of type 2 diabetes for US adults at age 20 is 32.8% with non-Hispanic black and Hispanic adults having markedly higher lifetime risk compared with non-Hispanic whites.[2] To appropriately stem the growing diabetes epidemic and subsequent increase in racial and ethnic diabetes-related disparities, health systems must implement strategies to identify patients with pre-diabetes and promote evidence-based pre-diabetes treatment.

In 2021, the US Preventive Services Task Force (USPSTF) updated its type 2 diabetes and pre-diabetes screening recommendations. The USPSTF now recommends screening of adults aged 35–70 with a body mass index (BMI) ≥25 kg/m^2 for type 2 diabetes and pre-diabetes and referral of individuals who screen positive for pre-diabetes to evidence-based prevention strategies such as the Diabetes Prevention Programme (DPP).[3 4] The DPP demonstrated intensive lifestyle change programmes can reduce the risk of progression from pre-diabetes to type 2 diabetes by 58%.[2] As such, the US public health sector, guided by the US Centers for Disease Control and Prevention (CDC) has played a pivotal role in the development of a nationwide delivery network for DPP and DPP-like programmes via the National Diabetes Prevention Programme (NDPP). The NDPP was established in 2010 to provide infrastructure to support the delivery of evidence-based DPP-like programmes and recognises organisations that implement DPP programmes with fidelity and effectiveness.[5] Unfortunately, recent estimates indicate that only about 0.1% of the US adult population reports participating in a DPP-like programme.[5 6] Furthermore, in California, the most populous US state, only 54 organisations have full CDC recognition as DPP providers[7] which is insufficient for the nearly half of California adults who have pre-diabetes.[8]

Given the limited availability of DPP offerings and the large number of adults in California with pre-diabetes, identifying localities for targeted diabetes prevention is key. However, no studies have been conducted in California to provide this critical health system and public health planning data. Locating geographical areas with patients most at risk for diabetes (ie, those with pre-diabetes) based on clinical data could provide insights into how to increase access to DPP-like programmes in locations where they are most needed. Geospatial mapping has been used to map diabetes prevalence to guide national, city and state-level policy interventions including addressing food environments[9] and directing care resources[9–11] to help improve health outcomes among people already living with diabetes. Geospatial mapping of pre-diabetes may be a useful tool in guiding national, state, city and county-level strategic planning to support policy changes and increased prevention resources to reduce diabetes incidence in pre-diabetes hotspots. A pre-diabetes hotspot is a geographical area with an elevated prevalence of pre-diabetes cases compared with surrounding areas. Understanding the geographical distribution of pre-diabetes prevalence is important in guiding resource allocation for DPP-like programmes and can support the placement of these programmes in geographical areas with large populations of racially and ethnically minoritised patients with pre-diabetes which could potentially alleviate the projected widened diabetes incidence disparities. Our goal in these analyses was to provide information on identifying census tract-level pre-diabetes 'hotspots,' using geospatial mapping methods, to guide DPP and planning that can help reduce rates of, and disparities in, diabetes incidence.[4 12]

RESEARCH DESIGN AND METHODS

Study setting

California is the most populous state in the USA and has the most racial and ethnic diversity in the continental USA.[13] Kaiser Permanente Northern California (KPNC) is a large, integrated healthcare delivery system that provides primary and specialty care, outpatient and inpatient services, and pharmacy and laboratory services to a socioeconomically and racially and ethnically diverse membership of 4.5 million patients, most of which reside in the San Francisco Bay and Greater Bay, Silicon Valley, Sacramento and Central Valley areas in Northern California.[14]

Study cohort

We included KPNC patients who were 35–70 years of age (the USPSTF recommended age range for pre-diabetes and diabetes screening) and had at least one fasting plasma glucose (FPG) or haemoglobin A1c (HbA1c) result available between 1 January 2019 and 31 December 2019. Patients in the cohort were also required to have active health plan enrolment on the day of their lab test and have a home address in the KPNC service area. We defined pre-diabetes as at least one FPG between 100 mg/dL and 125 mg/dL or HbA1c between 5.7% and 6.4% (39–46 mmol/mol) during the study window. We excluded patients with diabetes (defined as having a record in KPNC's diabetes registry before their qualifying lab test) and those without an available home address (figure 1). Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Outcome

Our outcome of interest was pre-diabetes prevalence at the census tract-level to explore more specific geographical characteristics. Patients were first classified within 1 of 2559 census tracts located within the KPNC service area based on their home address on the date of either their first screening or first pre-diabetes indicating HbA1c or FPG drawn in 2019. The number of cohort members with pre-diabetes in each census tract was divided by the total

number of patients who had a pre-diabetes screening test in that census tract, resulting in the prevalence per census tract.

Statistical analysis

We made use of widely accepted spatial methods, including incremental spatial autocorrelation and Getis-Ord Gi* hotspot analysis. Incremental spatial autocorrelation is a common approach to determining the scale of spatial analyses (ie, the distance band), particularly when there is a lack of prior knowledge about the spatial processes that influence clustering for the phenomena being studied (such as pre-diabetes prevalence). It is performed by running Global Moran I tests at successively increasing distances, which results in a z-score for each incremental distance that indicates the intensity of spatial clustering for the phenomena at that distance. The incremental

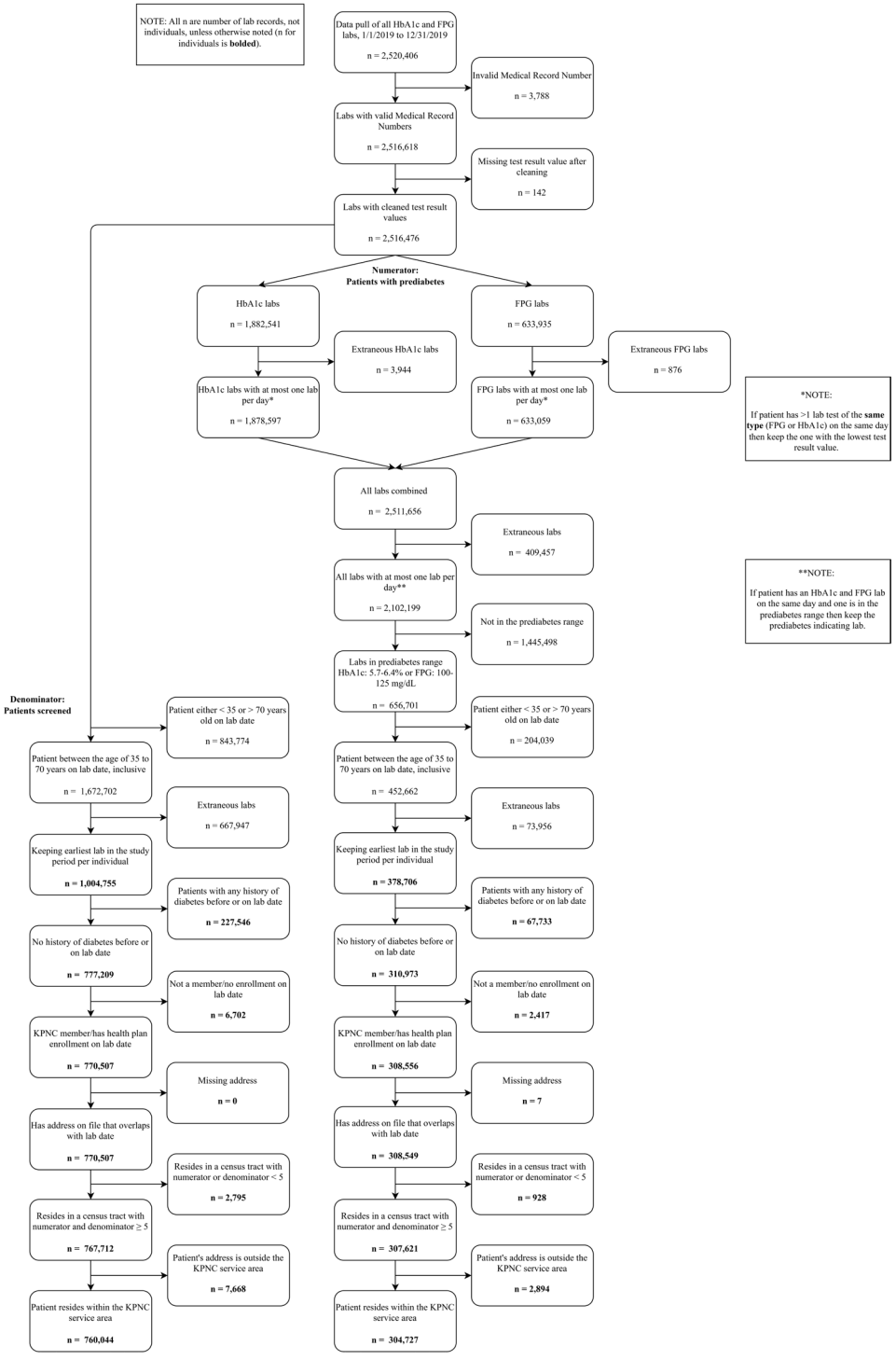

Figure 1 Pre-diabetes hot-spotting cohort identification, 1 January 2019–31 December 2019. FPG, fasting plasma glucose; HbA1c, haemoglobin A1c; KPNC, Kaiser Permanente Northern California.

testing is conducted until the z-score peaks at a particular distance, indicating this is the distance where clustering is the most pronounced. The Global Moran I is a measure of spatial autocorrection, it indicates the extent of overall clustering or dispersion (if any) of phenomena in the area under study. The Moran's I index values range from −1 to +1, with a positive value indicating spatial clustering, a negative value indicating spatial dispersion and a value near 0 indicating spatial randomness. The z-score and p value are used to assess if the null hypothesis that the phenomena is spatially distributed at random can be rejected. A z-score greater than 1.96 and p<0.05 indicate that the null hypothesis can be rejected and indicates clustering that is statistically significant. Conversely, a z-score less than −1.96 and p<0.05 indicate statistically significant dispersion. Getis-Ord Gi* is a statistic used to identify where spatial features (eg, census tracts) with high or low values of interest cluster (referred to as 'hotspots' and 'coldspots', respectively). The statistic is calculated for each feature in the dataset, resulting in a z-score for every feature. Statistically significant positive z-scores (>1.96, p<0.05) denote a hotspot and statistically significant negative z-scores (<−1.96, p<0.05) denote a coldspot, with the magnitude of the z-score reflecting the intensity of the clustering.

For this analysis, 'hotspots' are defined as clusters of neighbouring census tracts (3 or more) with elevated prediabetes prevalence compared with other census tracts in the study area. 'Coldspots' were not a focus of this study and thus were not reported. To create the geographical hotspot maps using geographic information system (GIS) methods, we followed the five steps described by Stopka *et al*.[15 16] This five-step methodology provides a systematic approach to determining key aspects of spatial cluster analyses, including a selection of the distance band that defines which census tracts are to be considered neighbours of each other via the use of incremental spatial autocorrelation. These five steps include:

1. Analysing variation in the size of census tract areas in the KPNC service area. The mean area was 8.17 square miles (SD=47.82), census tracts with a square mile area >1.5 SD above the mean were treated as outliers and temporarily removed from the analysis so they would not have an outsized influence on the determination of the distance band (n=60). For this same reason 'island' tracts, small tracts that were bordering removed outlier tracts and now appear as if they are islands, were also temporarily removed (n=6).
2. Calculating the overall average (1842.24 m) and maximum distance (18 269.81 m) between census tracts and their two nearest neighbours within the study area.
3. Using the Moran I and the aforementioned distances to incrementally test for spatial autocorrelation and determine the scale at which the most intense prediabetes clustering occurs, that is, the distance band. The shortest/beginning distance tested was two-thirds the maximum distance as recommended by Stopka *et al*, with subsequent tests done at incremental increases of half the average distance until the z-score peaked at 87.94 and was statistically significant (p<0.05), at a distance of 19 548.85 m. This distance was used as the distance band to define census tract neighbours.
4. Adding large area outlier and island census tracts back into the analysis and creating a spatial weights matrix

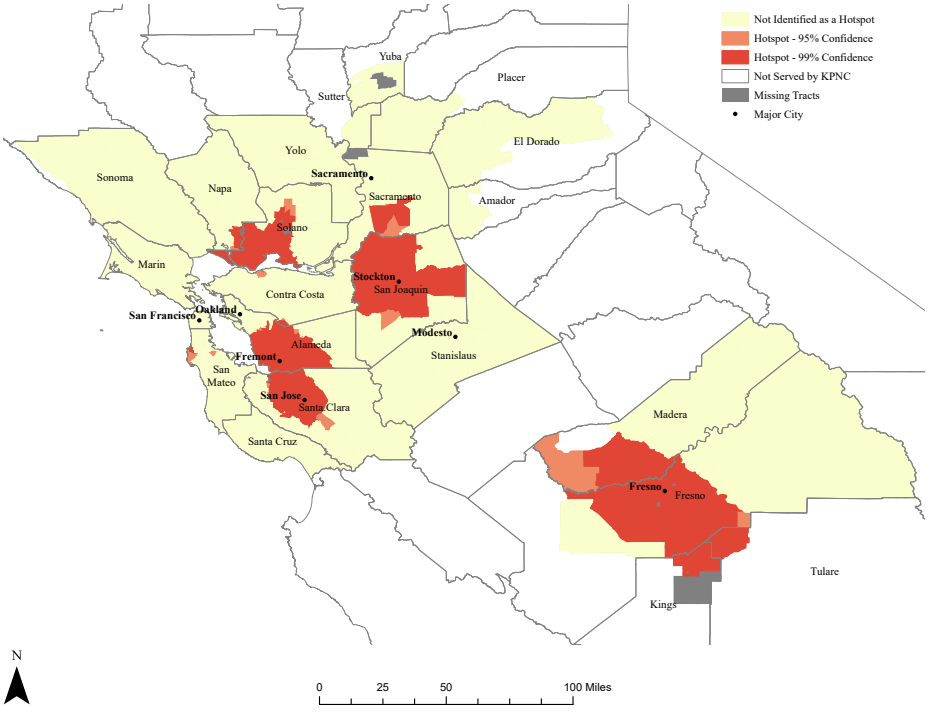

Figure 2 Hotspots of pre-diabetes prevalence, adults ages 35–70, 2019.

to account for census tract size variation. A spatial weights matrix is a file that tracks if each given census tract is a neighbour with each other census tract. The file is formatted as a matrix, where both the number of rows and columns are equal to the total number of census tracts in the study area. The cells of the matrix are populated by either a 1, denoting the given row and column census tract being compared are neighbours, or a 0 if they are not. A census tract was considered a neighbour of another if it was either within the distance band derived during step 3, or for the large area outlier and island census tracts, was one of the two nearest neighbouring tracts regardless of distance from each other. This spatial weight matrix approach allows us to account for differential census tract size and thus guarantee every census tract has at least two neighbours, even if the neighbours fall outside the distance band due to a census tract being too large or an island.

5. Conducting Getis-Ord Gi* hotspot analyses and creating the hotspot maps. To detect any potential within-county clustering patterns, hotspot subanalyses were conducted for the six most populous counties in the KPNC catchment area (Santa Clara, Alameda, Sacramento, Contra Costa, Fresno and San Francisco).

Individual and census tract-level characteristics were compared between hotspots and non-hotspots using χ^2 and Wilcoxon rank sum tests, as well as risk differences (RDs) and Hodges-Lehmann (HL) estimates of location shift. Individual-level characteristics compared included pre-diabetes screening test type, sex, age, race/ethnicity and having a BMI status of overweight/obese. Census tract-level characteristics included screening rate, median household income, median home value, per cent of households below the poverty line, unemployment rate, per cent of Medicaid enrollees, per cent of households on public assistance and per cent with a bachelor's degree or higher. Characteristics of individuals living in identified hotspots versus living in non-hotspots were obtained from electronic health records and administrative data. Data for the census tracts within each hotspot and non-hotspot were obtained from the 2019 American Community Survey.

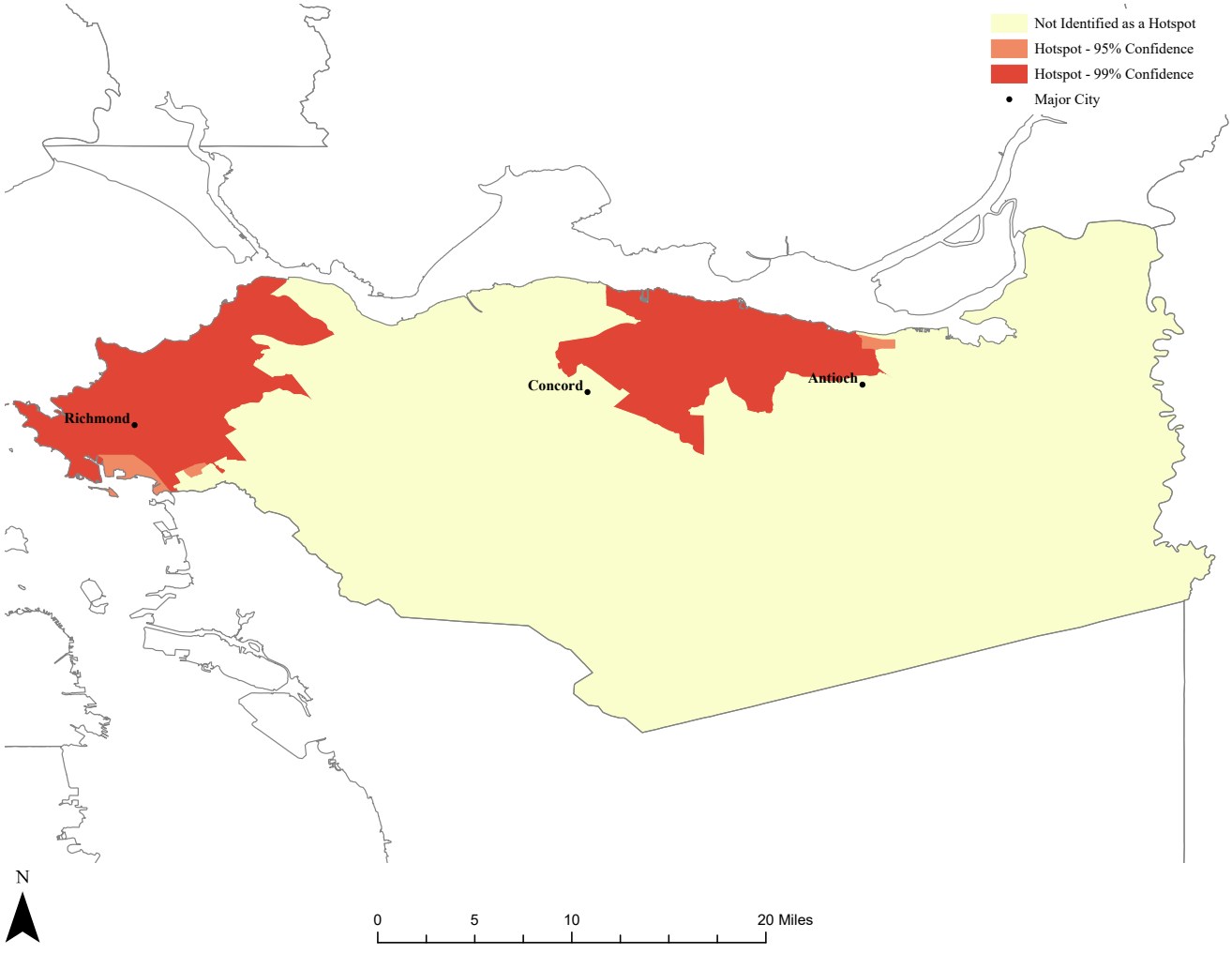

Figure 3 Contra costa county hotspots of pre-diabetes prevalence, adults ages 35–70, 2019.

Table 1 Individual-level characteristics of adults ages 35–70 with pre-diabetes

	Hotspot (95%–99% CI) (n=877)	Not identified as a hotspot (n=1 661)	Total (n=2 538)	Hodges-Lehmann estimate of location shift (95% CI)*	P value
Screening rate, (%)				1.58 (1.20 to 1.96)	<0.0001†
N (Missing)	877 (0)	1 661 (0)	2 538 (0)		
Mean (SD)	36.83 (4.87)	35.37 (4.20)	35.87 (4.49)		
Median	36.96	35.18	35.82		
Range	22.03, 51.54	19.2,3 51.67	19.23, 51.67		
Median household income, ($)				–3 651.00 (–7 256.00 to –25.00)	0.0482†
N (Missing)	877 (0)	1 659 (2)	2 536 (2)		
Mean (SD)	93 670.06 (45 696.67)	98 215.63 (46 598.96)	96 643.68 (46 330.39)		
Median	89 554.00	88 875.00	89 205.00		
Range	16 289.00, 250 001.00	12 340.00, 250 001.00	12 340.00, 250 001.00		
Median home value, ($)				–113 200.00 (–140 600.00 to –85 700.00)	<0.0001†
N (Missing)	872 (5)	1 641 (20)	2 513 (25)		
Mean (SD)	600 206.89 (411 014.89)	730 942.03 (455 480.26)	685 577.51 (444 853.49)		
Median	488 500.00	606 600.00	577 600.00		
Range	47 900.00, 2 000 001.00	41 200.00, 2 000 001.00	41 200.0, 2 000 001.00		
Household poverty, (%)				0.96 (0.55 to 1.37)	<0.0001†
N (Missing)	877 (0)	1 659 (2)	2 536 (2)		
Mean (SD)	9.95 (10.64)	7.62 (7.59)	8.42 (8.83)		
Median	6.16	5.15	5.55		
Range	0.00 60.14	0.00 45.50	0.00 60.14		
Unemployed, (%)				0.39 (0.24 to 0.54)	<0.0001†
N (Missing)	877 (0)	1 660 (1)	2 537 (1)		
Mean (SD)	3.91 (2.30)	3.39 (2.03)	3.57 (2.14)		
Median	3.38	3	3.11		
Range	0.00, 14.29	0.00, 15.87	0.00 15.87		
Medicaid, (%)				4.56 (3.40 to 5.76)	<0.0001†
N (Missing)	877 (0)	1 660 (1)	2 537 (1)		
Mean (SD)	26.60 (17.44)	20.81 (14.39)	22.81 (15.75)		
Median	22.19	17.01	18.64		
Range	0.68, 79.81	0.00, 73.83	0.00 79.81		
Household public assistance, (%)				0.97 (0.76 to 1.18)	<0.0001†
N (Missing)	877 (0)	1 660 (1)	2 537 (1)		
Mean (SD)	4.50 (4.61)	2.79 (3.05)	3.38 (3.75)		
Median	3.03	1.82	2.19		
Range	0.00 27.46	0.00, 27.09	0.00 27.46		
Bachelors or greater, (%)				–9.08 (–10.94 to –7.24)	<0.0001†
N (Missing)	877 (0)	1 660 (1)	2 537 (1)		
Mean (SD)	33.90 (21.52)	42.93 (22.55)	39.81 (22.61)		
Median	30.45	41.17	37.14		
Range	0.50, 90.57	2.22, 93.65	0.50 93.65		

*Hotspot minus non-hotspot.
†Wilcoxon rank sum p value.
BMI, body mass index; FPG, fasting plasma glucose; HbA1c, haemoglobin A1c.

Table 2 Selected characteristics of pre-diabetes hotspots and non-hotspots, by census tracts

	Hotspot (95%–99% CI) (n=262 076)	Not identified as a hotspot (n=497 968)	Total (n=760 044)	Risk difference (95% CI)*	P value
Test type, n (%)					<0.0001†
FPG	66 387 (25.3%)	129 218 (25.9%)	195 605 (25.7%)	−0.62% (−0.82 to −0.41)	
HbA1c	195 689 (74.7%)	368 750 (74.1%)	564 439 (74.3%)	0.62% (0.41 to 0.82)	
Sex, n (%)					0.9266†
Female	144 052 (55.0%)	273 731 (55.0%)	417 783 (55.0%)	0% (-0.24 to 0.23)	
Male	118 012 (45.0%)	224 215 (45.0%)	342 227 (45.0%)	0% (−0.23 to 0.24)	
Other/unknown	12 (0.0%)	22 (0.0%)	34 (0.0%)	0% (0 to 0)	
Age				¶	<0.0001§
Mean (SD)	52.65 (10.13)	53.87 (10.23)	53.45 (10.21)		
Median	52.84	54.56	53.97		
Missing, n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
Race/ethnicity, n (%)					<0.0001†
American Indian or Alaska Native	1018 (0.4%)	1814 (0.4%)	2832 (0.4%)	0.02% (0 to 0.05)	
Asian	84 380 (32.2%)	98 803 (19.8%)	183 183 (24.1%)	12.36% (12.15 to 12.57)	
Black	16 688 (6.4%)	34 043 (6.8%)	50 731 (6.7%)	−0.47% (−0.59 to −0.35)	
Latino	57 812 (22.1%)	85 577 (17.2%)	143 389 (18.9%)	4.87% (4.68 to 5.06)	
Multiracial	1487 (0.6%)	3243 (0.7%)	4730 (0.6%)	−0.08% (−0.12 to −0.05)	
Native Hawaiian or Pacific Islander	2641 (1.0%)	3828 (0.8%)	6469 (0.9%)	0.24% (0.19 to 0.28)	
Unknown	10 013 (3.8%)	18 552 (3.7%)	28 565 (3.8%)	0.10% (0 to 0.19)	
Non-Hispanic white	88 037 (33.6%)	252 108 (50.6%)	340 145 (44.8%)	−17.04% (−17.26 to −16.81)	
Overweight/obese BMI‡, n (%)					<0.0001†
No	51 406 (19.6%)	113 579 (22.8%)	164 985 (21.7%)	−3.19% (−3.39 to −3)	
Yes	189 088 (72.2%)	344 611 (69.2%)	533 699 (70.2%)	2.95% (2.73 to 3.16)	
Missing	21 582 (8.2%)	39 778 (8.0%)	61 360 (8.1%)	0.25% (0.12 to 0.38)	

*Hotspot minus non-hotspot.
†$\chi 2$ p value.
‡BMI≥23 kg/m2 for those of Asian race/ethnicity, BMI≥25 kg/m2 for all other groups.
§Wilcoxon rank sum p value.
¶Could not compute Hodges-Lehmann estimate due to memory limits.
BMI, body mass index; FPG, fasting plasma glucose; HbA1c, haemoglobin A1c.

SAS V.9.4 (SAS Institute) was used for data extraction and cohort creation and ArcGIS Pro V.2.7 (Esri, Redlands, California, USAA) was used to conduct spatial analyses.

Patients and public involvement

None.

RESULTS

A total of 760 044 patients residing in 2538 census tracts were included in the study cohort after excluding census tracts due to the number of patients with pre-diabetes or total patients screened being below 5 (n=21 census tracts containing 161 patients). Of these, 304 727 patients (40%) had pre-diabetes in 2019. We identified five pre-diabetes hotspots concentrated in Alameda, Fresno, Madera, Sacramento, San Joaquin, Santa Clara and Solano counties (figure 2). When the six largest counties in the catchment area were examined, an additional two hotspots were confirmed in Contra Costa County (figure 3). The mean screening rate was 36.8% in hotspots and 35.4% in non-hotspots (HL estimate: 1.58, 95% CI 1.20 to 1.96, p<0.0001). Adults living in pre-diabetes hotspots were more likely to be Latino (22.1% vs 17.2%, RD: 4.87%, 95% CI 4.68% to 5.06%) or Asian (32.2% vs 19.8%, RD: 12.36%, 95% CI 12.15% to 12.57%) compared with those living in non-hotspots (p<0.0001) (table 1). They were also more likely to be overweight or obese (72.2% vs 69.2%, RD: 2.95%, 95% CI 2.73% to 3.16%, p<0.0001). Table 2 compares the census tract-level characteristics of tracts within hotspots versus those in non-hotspots. Census tract hotspots had lower levels of household income (HL estimate: −3651.00, 95% CI −7256.00 to −25.00), per cent of adults with bachelors' degrees or higher (HL estimate: −9.08, 95% CI −10.94

to −7.24), and median home values (HL estimate: −113 200.00, 95% CI −140 600.00 to −85 700.00); and higher rates of household poverty (HL estimate: 0.96, 95% CI 0.55 to 1.37), unemployment (HL estimate: 0.39, 95% CI 0.24 to 0.54), household public assistance (HL estimate: 0.97, 95% CI 0.76 to 1.18) and per cent receiving Medicaid (HL estimate: 4.56, 95% CI 3.40 to 5.76) (p<0.05 for all).

DISCUSSION

Our study found that individuals in pre-diabetes hotspots were less likely to be non-Hispanic white and more likely to be overweight or obese. We also found that census tracts within hotspots had lower household incomes, per cent of adults with bachelors' degrees or higher, and median home values; and higher rates of household poverty, unemployment and per cent receiving Medicaid. This is critical information that can help health systems and communities identify geographical areas at the highest need for diabetes prevention services and at the highest risk for diabetes-related disparities. Increasing preventive health services and relevant community resources (eg, affordable healthful food and physical activity options) in these hotspots could help reduce socioeconomic and race and ethnicity-based disparities in diabetes incidence.

To our knowledge, this study is the first to use geospatial mapping methodology to identify pre-diabetes hotspots in California and to identify pre-diabetes hotspots using health system and administrative data. It is also one of only a handful of studies examining geospatial pre-diabetes hotspots in the USA. In our study, we found that individual-level and census tract-level socioeconomic status, obesity prevalence and race and ethnicity categories of patients living in pre-diabetes hotspots differed from those not identified as a hotspot. Prior geospatial mapping research was limited in that they focused on identifying clusters of patients at high diabetes risk in places outside of the USA,[17–19] identified clusters in a different state with national surveillance data[20 21] and/or were focused on small subpopulations at the city level.[17 18] Our study demonstrates that geospatial mapping techniques, using health system and census data, can be used to discover hotspots of pre-diabetes that can be adapted to other health systems, states and US regions.

Nevertheless, our study does have limitations. It was conducted with patient data from one health system in Northern California which could limit generalisations to other regions. However, the demographic characteristics of Kaiser Permanente members are similar to the general population of their geographic regions suggesting that research within Kaiser Permanente is reflective of the broader, regional population.[14] While this cohort is likely representative of the broader geographical region, one additional limitation is that this study does not include people without health plan enrollment. Spatial data are subject to various sources of bias, including the modifiable areal unit problem,[22] boundary problem and ecologic fallacy. In regard to the modifiable areal unit problem, our data were mapped at the census tract level due to a lack of finer geographical units being available and privacy issues, and thus there is an arbitrary level of aggregation involved that may not reflect the actual underlying spatial process. The boundary problem is also a potential source of bias since census tracts are rather arbitrarily drawn to include a roughly similar number of residents and relatively homogeneous population characteristics. The ecologic fallacy is also at play, as census tract-level characteristics may not reflect the characteristics of all the individuals residing in a given census tract. Additionally, the individual and census-level characteristics compared were crude measures not adjusted for potential associations with each other. Furthermore, these socioeconomic and demographic characteristics could vary significantly within small geographic areas and location may not be the only factor driving pre-diabetes hotspots. Several other systemic factors could contribute to the observed clustering of pre-diabetes cases (eg, education and healthcare access, health system practices). We also did not assess whether participants were from the same household or family in this analysis. We acknowledge that diabetes risk factors may cluster in households and families and that families may cluster in neighbourhoods. We plan to explore this further in future analyses. Finally, due to screening practices in our health system that primarily leverage HbA1c and FPG tests, we did not include oral glucose tolerance tests to assess diabetes screening.

One California study found a relationship between neighbourhood socioeconomic status and HbA1c levels in patients with diabetes, noting that geospatial mapping may be a useful tool for addressing diabetes disparities.[23] Our study found a correlation between census tract-level socioeconomic characteristics and race and ethnicity demographic composition and pre-diabetes hotspots, suggesting that geospatial mapping may also be a useful tool in addressing diabetes prevention disparities and identifying areas that need additional focus and may benefit from additional resources.

To successfully address area-level disparities, it may be critical to leverage community-based resources to support patients in pre-diabetes hotspots. Efforts to integrate geocoded social determinants of health data into EHRs could support health system-level community outreach and engagement efforts. For example, the American Board of Family Medicine's PRIME registry has developed a community-oriented primary care tool, Population Health Assessment Engine (PHATE), that incorporates patient addresses, diagnoses and other quality measures along with area-level social determinants of health indices to leverage geography to improve patient risk assessment.[24] This could be one strategy to support linkages between primary care and community resources. Future studies could use tools like PHATE to incorporate geocoded social determinants of health data into EHRs to identify community characteristics that influence patient health outcomes, identify community partners to link patients to resources and develop visualisations of service

areas so clinicians can integrate community and practice-level data to improve population health.

Successfully reducing diabetes incidence rates by improving population-based screening to identify and then treat pre-diabetes is critical. Research suggests that guideline-concordant diabetes screening rates are low.[25] Identifying pre-diabetes hotspots could be used to better target screening interventions for those most at risk and given that these hotspots also have a higher proportion of non-white adults and household poverty, targeting these geographical areas could serve as a strategy to reduce disparities in diabetes incidence and prevalence by improving screening rates and providing increased access to evidence-based DPPs for underserved populations. This hot-spotting approach can help identify geographical areas that need additional support and can lead to partnerships at the community level to address health disparities and target interventions and resources. Policy-makers and care providers can use this information to target diabetes prevention resources and outreach. This could include national or state policies that provide health insurance coverage for evidence-based DPPs for low-income adults with pre-diabetes to address socioeconomic disparities or health system strategies that place evidence-based DPPs in racially and ethnically diverse communities with the highest need to address racial and ethnic disparities.

Conclusion

Our study found that pre-diabetes cases clustered in geographical hotspots with lower levels of socioeconomic characteristics, higher prevalence of obesity and more individuals from minoritised populations. Policy-makers, state, county, and city health officials, and healthcare systems and providers can use this information to target diabetes prevention resources and outreach to improve health equity and reduce diabetes-related disparities.

X Luis A Rodriguez @Luis_A_Rguez

Contributors TWT: conceptualisation, investigation, writing-original draft, writing-review and editing. OKD: conceptualisation, funding acquisition, supervision, investigation, writing-review and editing. MY: data curation, formal analysis, writing-original draft, writing-review and editing. LAR: conceptualisation, data curation, writing-review and editing. TM: conceptualisation, data curation, investigation, writing- review and editing. YC-L: conceptualisation, investigation, writing-review and editing. JS: conceptualisation, funding acquisition, supervision, writing-original draft, writing-review and editing. Guarantor, TWT, takes full responsibility for the work as a whole, including the study design, access to data and the decision to submit and publish the manuscript.

Funding This work was funded by the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK), R18DK122372-01A1. TWT receives support from a National Institute on Minority Health and Health Disparities Administrative Supplement (3R01MD016738-03S1). LAR receives support from an NIDDK Administrative Supplement (3R18DK122372-03S1). YC-L also receives support from a mentored career development award funded by the NIDDK (1K23DK129828). JS also received support from the NIDDK-funded DREAMS Center for Diabetes Translational Research (P30 DK092924).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the University of California at Los Angeles Institutional Review Board (IRB# 20-000583). Our IRB waived the requirement of patient consent as the research involved no more than minimal risk to subjects, could not be carried out practicably without the waiver and the waiver did not adversely affect the rights and welfare of the subjects.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. The datasets generated during and/or analysed during the current study are not publicly available due to patient data confidentiality.

ORCID iD

Tainayah W Thomas http://orcid.org/0000-0001-9094-8556

REFERENCES

1 Lin J, Thompson TJ, Cheng YJ, *et al*. Projection of the future diabetes burden in the United States through 2060. *Popul Health Metr* 2018;16:9.
2 Knowler WC, Barrett-Connor E, Fowler SE, *et al*. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002;346:393–403.
3 US Preventive Services Task Force, Davidson KW, Barry MJ, *et al*. Screening for Prediabetes and Type 2 Diabetes: US Preventive Services Task Force Recommendation Statement. *JAMA* 2021;326:736–43.
4 Prediabetes and Type 2 Diabetes: Screening. US Preventive Services Task Force, Available: https://uspreventiveservicestaskforce.org/uspstf/recommendation/screening-for-prediabetes-and-type-2-diabetes
5 Ackermann RT, O'Brien MJ. Evidence and Challenges for Translation and Population Impact of the Diabetes Prevention Program. *Curr Diab Rep* 2020;20:9.
6 Ali MK, McKeever Bullard K, Imperatore G, *et al*. Reach and Use of Diabetes Prevention Services in the United States, 2016-2017. *JAMA Netw Open* 2019;2:e193160.
7 Diabetes Prevention Recognition Program Application. Centers for Disease Control and Prevention, Available: https://dprp.cdc.gov/Registry
8 Babey SH, Wolstein J, Diamant AL, *et al*. Prediabetes in California: Nearly Half of California Adults on Path to Diabetes. *Pol Brief UCLA Cent Health Policy Res* 2016;1–8.
9 Lee DC, Gallagher MP, Gopalan A, *et al*. Identifying Geographic Disparities in Diabetes Prevalence Among Adults and Children Using Emergency Claims Data. *J Endocr Soc* 2018;2:460–70.
10 Curtis AB, Kothari C, Paul R, *et al*. Using GIS and secondary data to target diabetes-related public health efforts. *Public Health Rep* 2013;128:212–20.
11 Lee DC, Jiang Q, Tabaei BP, *et al*. Using Indirect Measures to Identify Geographic Hot Spots of Poor Glycemic Control: Cross-sectional Comparisons With an A1C Registry. *Diabetes Care* 2018;41:1438–47.
12 Mikhail N, Wali S, Brown AF. Ethnic Disparities in Diabetes. *Endocrinol Metab Clin North Am* 2021;50:475–90.
13 New World Encyclopedia Contributers. California in New World Encyclopedia. 2023.
14 Gordon NP. *Similarity of adult Kaiser Permanente Members to the adult population in Kaiser Permanente's Northern California service area: comparisons based on the 2017/2018 cycle of the California health interview survey*. Oakland, CA: Kaiser Permanente Division of Research, 2020.

15 Stopka TJ, Krawczyk C, Gradziel P, *et al*. Use of spatial epidemiology and hot spot analysis to target women eligible for prenatal women, infants, and children services. *Am J Public Health* 2014;104 Suppl 1:S183–9.
16 Meyers DJ, Hood ME, Stopka TJ. HIV and hepatitis C mortality in Massachusetts, 2002-2011: spatial cluster and trend analysis of HIV and HCV using multiple cause of death. *PLoS One* 2014;9:e114822.
17 Noble D, Smith D, Mathur R, *et al*. Feasibility study of geospatial mapping of chronic disease risk to inform public health commissioning. *BMJ Open* 2012;2:e000711.
18 Smurthwaite K, Bagheri N. Using Geographical Convergence of Obesity, Cardiovascular Disease, and Type 2 Diabetes at the Neighborhood Level to Inform Policy and Practice. *Prev Chronic Dis* 2017;14:E91.
19 Penney TL, Rainham DGC, Dummer TJB, *et al*. A spatial analysis of community level overweight and obesity. *J Hum Nutr Diet* 2014;27 Suppl 2:65–74.
20 Lord J, Roberson S, Odoi A. Geographic disparities, determinants, and temporal changes in the prevalence of pre-diabetes in Florida. *PeerJ* 2021;9:e10443.
21 Lord J, Roberson S, Odoi A. Investigation of geographic disparities of pre-diabetes and diabetes in Florida. *BMC Public Health* 2020;20:1226.
22 Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environ Plan A* 1991;23:1025–44.
23 Geraghty EM, Balsbaugh T, Nuovo J, *et al*. Using Geographic Information Systems (GIS) to assess outcome disparities in patients with type 2 diabetes and hyperlipidemia. *J Am Board Fam Med* 2010;23:88–96.
24 Bambekova PG, Liaw W, Phillips RL Jr, *et al*. Integrating Community and Clinical Data to Assess Patient Risks with A Population Health Assessment Engine (PHATE). *J Am Board Fam Med* 2020;33:463–7.
25 Shealy KM, Wu J, Waites J, *et al*. Patterns of Diabetes Screening and Prediabetes Treatment during Office Visits in the US. *J Am Board Fam Med* 2019;32:209–17.