


# BMJ Open Demographic and geographic distribution of diabetes and pre-diabetes risk in rural settings: results from a cross-sectional, countywide rural health survey in Sullivan County, New York

David C Lee ,<sup>1,2</sup> Leah Ross,<sup>3</sup> Carolina Quintero Arias,<sup>1</sup> Melissa Rony,<sup>1</sup> Rahi Patel,<sup>4</sup> Erica Jensen,<sup>5</sup> Robert Petcu,<sup>6</sup> Daniel Imas,<sup>1</sup> Brian Elbel,<sup>2,7</sup> Lorna E Thorpe,<sup>2</sup> Rebecca Anthopoulos<sup>2</sup>

**To cite:** Lee DC, Ross L, Quintero Arias C, *et al*. Demographic and geographic distribution of diabetes and pre-diabetes risk in rural settings: results from a cross-sectional, countywide rural health survey in Sullivan County, New York. *BMJ Open* 2024;**14**:e080831. doi:10.1136/bmjopen-2023-080831

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2023-080831>).

Received 12 October 2023  
Accepted 26 July 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr David C Lee;  
David.Lee@nyulangone.org

## ABSTRACT

**Objective** To perform a detailed characterisation of diabetes burden and pre-diabetes risk in a rural county with previously documented poor health outcomes in order to understand the local within-county distribution of diabetes in rural areas of America.

**Design, setting, and participants** In 2021, we prospectively mailed health surveys to all households in Sullivan County, a rural county with the second-worst health outcomes of all counties in New York State. Our survey included questions on demographics, medical history and the American Diabetes Association's Pre-diabetes Risk Test.

**Primary outcome and methods** Our primary outcome was an assessment of diabetes burden within this rural county. To help mitigate non-response bias in our survey, raking adjustments were performed across strata of age, sex, race/ethnicity and health insurance. We analysed diabetes prevalence by demographic characteristics and used geospatial analysis to assess for clustering of diagnosed diabetes cases.

**Results** After applying raking procedures for the 4725 survey responses, our adjusted diagnosed diabetes prevalence for Sullivan County was 12.9% compared with the 2019 Behavioural Risk Factor Surveillance System (BRFSS) estimate of 8.6%. In this rural area, diagnosed diabetes prevalence was notably higher among non-Hispanic Black (21%) and Hispanic (15%) residents compared with non-Hispanic White (12%) residents. 53% of respondents without a known history of pre-diabetes or diabetes scored as high risk for pre-diabetes. Nearest neighbour analyses revealed that hotspots of diagnosed diabetes were primarily located in the more densely populated areas of this rural county.

**Conclusions** Our mailed health survey to all residents in Sullivan County demonstrated higher diabetes prevalence compared with modelled BRFSS estimates that were based on small telephone samples. Our results suggest the need for better diabetes surveillance in rural communities, which may benefit from interventions specifically tailored for improving glycaemic control among rural residents.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Our study performed a survey of all households within a single rural county in order to understand the geographic distribution of rural diabetes at a local level using geographically precise health survey data.
- ⇒ This detailed examination of a rural county with particularly poor health outcomes identified the local areas where rural diabetes burden was geographically concentrated using geospatial analysis.
- ⇒ This approach can identify geographic clustering of diabetes within rural counties with local geographic precision, providing a means of directing key resources such as diabetes screening and prevention to high-need areas.
- ⇒ Our study is limited to a single rural county and our findings may not be generalisable to other rural regions across the USA.

## INTRODUCTION

Though most studies of diabetes have focused on individuals living in large cities, diabetes is not solely an urban phenomenon.<sup>1 2</sup> Several studies have demonstrated that diabetes rates are consistently up to 15% higher in rural regions when compared with urban areas of the USA.<sup>3 4</sup> These high rates of rural diabetes in part reflect differentially higher rates of poverty, obesity and smoking in rural areas of the country.<sup>3 5</sup> Among those with diabetes, rural residents are approximately 20% less likely to obtain the healthcare recommended by most diabetes guidelines.<sup>6–8</sup> The result is a substantially higher rate of diabetes-related mortality in the most sparsely populated rural counties of the USA when compared with the most densely populated urban counties (mortality rate of 26.2 vs 20.7 per 100

000 people in 2016).<sup>9</sup> Evidence also suggests that these geographic disparities are increasing. The relative risk of death from diabetes in rural vs urban areas increased from 1.04 in 1969 to 1.17 in 2009.<sup>10 11</sup> Though rates of diabetes-related mortality have substantially improved in urban areas, high rates of diabetes-related mortality have persisted in many rural regions of the USA.<sup>9</sup>

Evaluating diabetes prevalence and pre-diabetes risk in rural settings can be difficult due to sparse population coverage and limited data. For instance, in Sullivan County, a rural county in New York State, only 40 to 50 residents a year are analysed by national surveys like the Centers for Disease Control and Prevention's (CDC) Behavioural Risk Factor Surveillance System (BRFSS), which may not be a large enough sample to provide an accurate estimate of diabetes prevalence for the county.<sup>12</sup> The CDC has attempted to overcome this small sample size by developing county-level estimates of diabetes prevalence in rural areas that are largely based on mathematical models that estimate disease burden using several demographic factors (eg, age, sex and status as a minority).<sup>13</sup> Importantly, these models do not control for the differences in health outcomes that may vary among different geographic regions and by socioeconomic status. Not controlling for these factors ignores key characteristics of counties like Sullivan County, a rural region where there are few minorities, and many non-Hispanic White residents have low income, poor health and a high risk of developing chronic disease like diabetes.<sup>12</sup>

### Research objectives

To improve our understanding of diabetes prevalence and pre-diabetes risk in rural America, we performed a survey of all households in Sullivan County, New York, a rural county that has the second-poorest county-level health outcomes among all counties in New York State. We performed a geographically distributed, mailed health survey with the research objective to estimate overall diagnosed diabetes prevalence and pre-diabetes risk in the county. We relied on a mailed survey design because it ensured maximum inclusion of all county residents, and because response rates to telephone surveys have declined significantly over the past two decades and tend to exclude certain segments of the population, a pattern exacerbated by the transition to cellphones from landlines. Another objective was to perform a geospatial analysis to identify local clusters of diabetes cases within the county to understand the geographic distribution of rural diabetes.

## METHODS

### Study design

In January 2021, we mailed health surveys to all households in Sullivan County, a rural county with the second-worst health outcomes of all counties in New York State. Our survey included questions with wording matching previously validated national surveys (ie, the American

Community Survey, the Behavioural Risk Factor Surveillance System) on demographics and medical history.<sup>14 15</sup> We also used the American Diabetes Association's Pre-diabetes Risk Test.<sup>16</sup> A score of 5 points or higher on this test has been previously validated as a measure for identifying individuals at high risk of having pre-diabetes (see online supplemental appendix). We then estimated diagnosed diabetes prevalence for the county using raking to adjust for non-response bias and also used geospatial analysis within the county to assess for clustering of diagnosed diabetes cases.

### Mailed health surveys

While the survey was mailed to all households in the county, we analysed the data on only households of permanent residents. To achieve this, perform our cross-sectional survey in Sullivan County, we obtained a comprehensive list of all households from the Marketing Systems Group (Horsham, PA). The list was obtained in October 2020 and contained all non-seasonal and non-vacant residential households including those with a post office box address. The list contained a total of 28 284 households, which compared favourably to the estimated 28 184 households in the 2019 Census estimates for Sullivan County. In order to maximise coverage across the county, all households received one mailed health survey that consisted of questions that first confirmed residence within Sullivan County and then asked a brief selection of health and demographic questions. Survey respondents were offered a \$10 gift card for participation and a pre-stamped return envelope was enclosed in the mailing. Responses were returned to the Sullivan County Public Health Services and then sent to the NYU School of Medicine where responses were uploaded into REDCap for data entry and analysis.

### Main study outcomes

Our main study outcomes were the prevalence of diagnosed diabetes and pre-diabetes risk scores among permanent residents that responded to the countywide mailed health survey. Participants were asked if they had ever been told by a doctor or other health professional that they had any of the following: (1) pre-diabetes or borderline diabetes, (2) gestational or pregnancy-related diabetes or (3) diabetes, in addition to specifying which type of diabetes. To obtain the responses needed to calculate the pre-diabetes risk scores, they were also asked about their age and sex; if they had a mother, father, sister or brother with diabetes; if they had ever been diagnosed with hypertension or high blood pressure; and, if they were physically active. Participants were also asked to provide their height and weight in order to calculate their Body Mass Index. Responses were then scored with a result of 5 points or higher being considered high risk for having pre-diabetes.<sup>16</sup>

### Demographic factors

Survey respondents were also asked to fill out household and demographic questions. These questions included

their age, sex, race/ethnicity, health insurance, marital status and the number of adults and children living in their household. Given the large number of seasonal residents in Sullivan County, we also asked whether survey respondents were full-time or part-time residents in the county. For all analyses, we excluded part-time residents so that our study population would only include full-time residents of the county especially given the known high influx of population that occurs in the summer. In order to compare survey respondents to the adult population in Sullivan County and to account for non-response bias, we obtained Census estimates for these demographic and household characteristics from the 2019 American Community Survey.

### Statistical analysis

We first performed a descriptive analysis on the Census characteristics and health survey respondents in terms of age, sex, race/ethnicity, health insurance and household characteristics. We also compared these characteristics among respondents with a prior history of diagnosed

diabetes and respondents with elevated pre-diabetes risk score but no prior known history of diabetes or pre-diabetes. Testing for statistically significant differences from Census estimates was performed by calculating 95% CI for each proportion among survey respondents, those with a history of diagnosed diabetes and those high risk for pre-diabetes with a prior diagnosis of diabetes. For our analysis of the main study outcomes, we performed raking adjustments using age, sex, race/ethnicity and health insurance strata to help mitigate non-response bias.<sup>17</sup> The Census data for the county was used as a reference for the target population. Approximately 2% of survey respondents had missing data for one of the raking variables. Given this small proportion, we restricted this analysis to complete cases. The race/ethnicity categories for Asian and Other were combined in order to avoid small cell sizes. We compared the raking-based estimates to the crude estimates, which used the results of survey responses without any adjustments.

**Table 1** Study population characteristics

Population characteristics	Census estimates	Health survey respondents	History of diabetes	High risk for pre-diabetes
Total adults	59 174	4725	740	1838
Age				
18 to 39	32%	12%*	4%*	2%*
40 to 59	35%	29%*	23%*	21%*
60 to 79	28%	49%*	60%*	64%*
80 or older	5%	10%*	13%*	13%*
Sex				
Male	51%	43%*	52%	56%*
Female	49%	57%*	48%	44%*
Race/ethnicity				
White	70%	85%*	80%*	87%*
Black	8%	4%*	7%	4%*
Hispanic	16%	8%*	10%*	6%*
Asian	2%	1%*	2%	1%*
Other	5%	2%*	1%*	2%*
Health insurance				
Private	49%	38%*	25%*	30%*
Medicare	20%	41%*	49%*	53%*
Medicaid	19%	16%*	20%	13%*
Self-pay	13%	5%*	5%*	4%*
Households				
Single, no kids	45%	42%*	46%	45%
Single, with kids	10%	7%*	6%*	4%*
Married, no kids	28%	37%*	41%*	43%*
Married, with kids	17%	14%*	7%*	8%*

\*Statistically significant differences from Census estimates with a p value of 0.05.

## Geographic analysis

We also performed geospatial analysis within the county to identify significant clusters of cases of diagnosed diabetes. Mailing addresses were geocoded to pinpoint the exact location of residence. Survey respondents with a post office box address (5% of the analytic sample) were excluded from this analysis given the lack of a precise location. The more prevalent clusters of diabetes cases were identified using Getis-Ord  $G_i^*$  hotspot analysis. We used the K-nearest neighbours to model spatial proximity as there was significant variation in distance between nearest neighbourhoods within the rural county. The number of nearest neighbours was specified as 100 neighbours, with sensitivity analyses at 50 and 200 neighbours. A false discovery rate correction was applied to account for multiple testing and spatial dependency. The resultant z-scores and p values were used to assess whether a given point and its K-nearest neighbours represented a hot or cold spot of diagnosed diabetes prevalence.

Statistical analyses were performed using Stata 16.1 (Statacorp; College Station, TX, 2019) and R 4.1.2 (2021-11-01). Geographic analysis and mapping were performed using ArcGIS Pro 2.8.3 (ESRI; Redlands, CA, 2021).

## Patient and public involvement

The Sullivan County Public Health Services, the local health department in Sullivan County, was involved in the study design and was a key collaborator given their health expertise in this rural area of New York State.

## RESULTS

### Mailed survey responses

Of the 28 284 mailed health surveys, 2706 (10%) were returned to sender. The five most common reasons for these returns were vacant address (n=777), unable to forward (n=660), no mail receptacle (n=373), insufficient

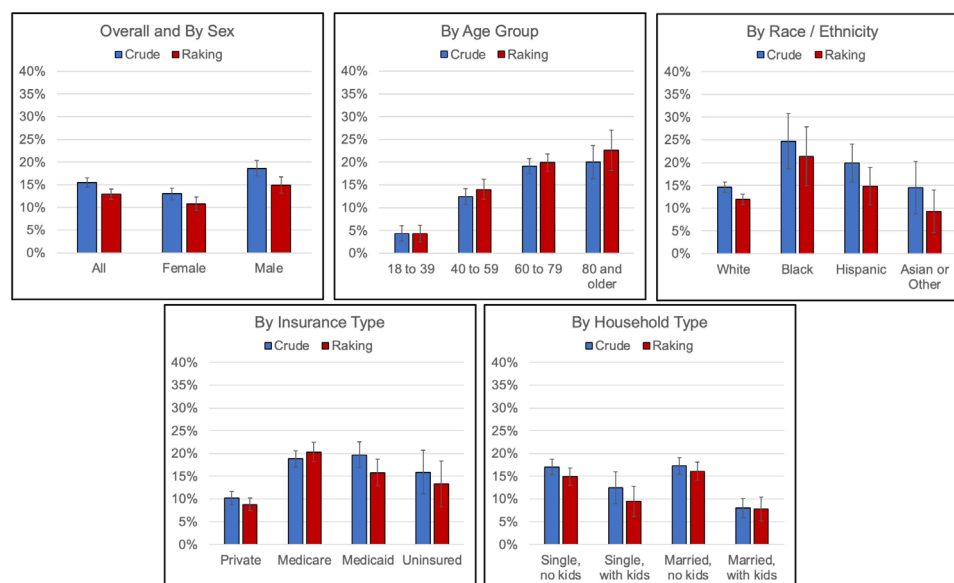
address (n=325) and addressee unknown (n=234). Of the remaining 25 578 surveys, we received 5230 responses (20%), which was similar to our expectation, based on our prior experience performing a one-time mailed health survey in this rural region, which previously outperformed telephone-based approaches.<sup>18</sup> 28 of the returned surveys were partially completed and could not be used for the analysis as they were missing key demographic data. In addition, we excluded an additional 468 survey respondents who reported being part-time residents of the county and nine participants that lived outside of the borders of Sullivan County. Our final analytic sample included a total of 4725 full-time residents of Sullivan County who completed the health survey, which accounts for 8% of the estimated 59 174 adults that live in Sullivan County.

### Population characteristics

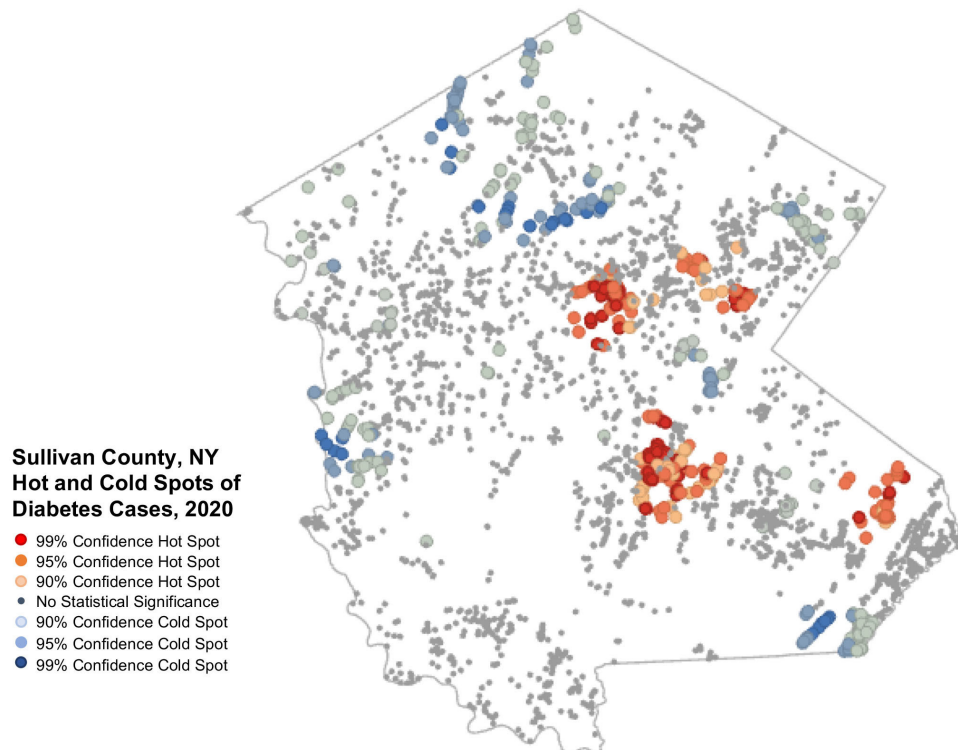
Compared with Census estimates for the adult population in Sullivan County, survey respondents were more frequently older, female, White, insured by Medicare or married without children. Survey respondents with a prior history of diabetes were proportionally older, male, non-White and more often insured by Medicare or Medicaid compared with the overall sample of survey respondents (table 1).

### Diagnosed diabetes prevalence estimates

Of the 4725 survey responses analysed, the unadjusted crude diagnosed diabetes prevalence in our sample was 15.5% with a 95% CI of 14.4% to 16.5%. After performing raking (adjusting for age, sex, race/ethnicity and health insurance) in order to obtain an estimate of diabetes prevalence that would be representative for Sullivan County as a whole, we obtained an adjusted diagnosed diabetes prevalence of 12.9% with a 95% CI of 11.8% to 14.1%. Our estimate was 1.5 times higher than



**Figure 1** Diabetes prevalence by demographic group in Sullivan County, NY.



**Figure 2** Geographic clustering of cases of diagnosed diabetes in Sullivan County, NY.

BRFSS's estimate for Sullivan County in 2019, which used a Bayesian mathematical model to arrive at an estimate of 8.6%. In comparing crude and raking-adjusted results for specific demographic groups, raking-adjusted results were lower for all groups except for older subgroups and those insured by Medicare, which were over-represented in the survey sample.

Analysing the raking-adjusted results, diagnosed diabetes prevalence was higher among men (14.9%) compared with women (10.8%), progressively higher by age and significantly higher among Black (21.4%) and Hispanic (14.8%) residents compared with White (11.9%) and Asian or other race/ethnicity (9.3%) residents. While diabetes prevalence was highest among residents with Medicare (20.3%), likely due to their older age, diabetes prevalence was also respectively 1.8 and 1.5 times higher among residents with Medicaid (15.8%) and uninsured or self-pay residents (13.3%) when compared with residents with private (8.8%) health insurance. As for household composition, diabetes prevalence was higher among single or married households without children compared with those with children (figure 1).

#### Elevated pre-diabetes risk

Excluding the 740 (15.7%) respondents with a history of diagnosed diabetes and the 482 (10.2%) respondents with a known history of pre-diabetes, we also performed an analysis for the subgroup of 3503 study respondents who did not have a known history of diabetes or pre-diabetes in order to assess risk of pre-diabetes within Sullivan County. Of these respondents without known diabetes or pre-diabetes, 1826 (52.5%) had a high pre-diabetes risk

score of 5 points or higher. In general, the demographic characteristics of these respondents with an elevated pre-diabetes risk score were similar to the demographic characteristics of respondents with a known history of diabetes. However, it should be noted that this analysis did not account for bias through raking, as it was not possible to know the underlying Census characteristics of residents without a known history of diabetes or pre-diabetes. Therefore, this high rate of pre-diabetes risk scores may only be representative of survey respondents rather than the rural county as a whole.

#### Geospatial clustering of diabetes cases

In our geospatial analysis, we assessed whether there was any statistically significant clustering of diagnosed diabetes cases using the closest 100 neighbours for each geocoded residential location. The map of Sullivan County in figure 2 demonstrated the hot and cold spots with increasing levels of confidence. The two largest hotspots of diabetes were concentrated in the two most densely populated towns, which are known as urban clusters based on Census definitions. Our sensitivity analysis using the closest 50 and 200 neighbours demonstrated a similar clustering of these hot and cold spots but only differed slightly in the size of areas identified.

#### DISCUSSION

Using a cross-sectional household survey, our study assessed diabetes burden in Sullivan County, a rural country known to have particularly poor health, as it was ranked second worst for health outcomes among all



counties in New York State. Our study estimated overall diagnosed diabetes prevalence in Sullivan County based on a countywide survey and raking adjustments. We found rates of diabetes were 1.5 times higher than CDC estimates, which are modelled based on BRFSS data from 2019.<sup>12</sup> In addition, we also found that more than half of our participants surveyed without a history of diabetes or pre-diabetes were at an elevated risk for pre-diabetes.

As our county-level estimates of diagnosed diabetes prevalence in Sullivan County were significantly higher than modelled CDC estimates, our findings suggest that current surveillance methods might underestimate disease prevalence especially in rural counties where surveillance coverage is low. In addition, modelled estimates might underestimate diabetes prevalence due to the large portion of residents in Sullivan County that have low socioeconomic status though they are predominately White. Therefore, using models that only consider age, sex and race/ethnicity might very well underestimate disease risk in a place like Sullivan County, as the average health for White individuals in other parts of the country may not be applicable to those White residents in Sullivan County who face significant poverty and unemployment.<sup>13</sup> Mathematical models used to estimate disease prevalence may need to incorporate socioeconomic factors with more robustness to produce better estimates of disease burden.<sup>19</sup>

Our study's response rates align with expectations from our previous papers, suggesting that mail-in surveys may be a reasonable approach for reaching remote rural populations, especially, given the steep decline in responses to telephone surveys, which are now estimated to be in the single digits.<sup>18</sup> Given the preference of different demographic subgroups for specific survey modalities (eg, mail, phone, email, vs other online sources), there may be a need to increase use of multimodal approaches.<sup>20</sup> Alternatively, surveillance efforts may benefit from considering an entirely different approach, such as using other data streams like electronic health records, in order to estimate disease prevalence with data already collected at a large scale.<sup>21</sup> In any case, there needs to be modernisation of surveillance in such a way that is able to accurately estimate disease burden especially in areas where it may be hard to reach residents by traditional survey approaches and in regions where certain socioeconomic disadvantages may make the population in question be more prone to disease than would otherwise be expected by typical demographic factors.

Evaluating our findings based on demographic subgroups, several of our findings align with the known literature, including men having a higher prevalence of diabetes compared with women and diabetes prevalence increasing progressively with age.<sup>22</sup> In addition, we found the prevalence of diabetes to be disproportionately higher among Black (23%) and Hispanic (16%) respondents, which is concerning even though they only account for only 8% and 16% respectively of the adult population in Sullivan County according to Census estimates.<sup>23</sup> We also

noted disparities in diagnosed diabetes prevalence based on health insurance type. While diabetes prevalence was highest among those with Medicare as would be expected given their older age, we also found that diabetes prevalence was higher among those who were uninsured or self-pay and was even higher among those who had Medicaid when compared with those with private insurance.<sup>24</sup> In terms of household type, our study found that diabetes prevalence was higher among households without children compared with those with children. This finding is likely due to the age distributions of these households within the county. However, it should be noted that other studies have found that raising children might actually increase a parent's risk of diabetes due to socioeconomic and lifestyle factors.<sup>25</sup> Overall, these disparities in diagnosed diabetes prevalence are likely driven by differences in physical activity and diet among other key factors that increase the risk of diabetes.

We also performed geospatial analysis to understand the geographic distribution of diabetes across rural Sullivan County. The clustering of diabetes we found suggests that diabetes is more prevalent in the more densely populated areas of this rural county, which may be due to clustering of socioeconomic status or other factors that confer a higher risk of diabetes. For instance, rural residents with lower socioeconomic status may live closer to town centres because food, housing, transportation and social programmes are more accessible without the expense of a car which otherwise would be necessary to live in a rural community.<sup>26</sup> The geographic associations between diagnosed diabetes prevalence and predisposing risk factors may differ in rural areas when compared with urban areas meaning that factors like the food environment need to be quantitatively measured through a different approach.<sup>27</sup> We should also note that though we found clustering of diagnosed diabetes prevalence in the more urban areas of this rural county, our geospatial analysis does not mean that residents in the most remote rural areas are not at risk for diabetes as well.

There are limitations to our survey. Survey respondents were skewed towards certain demographic groups, for example, older, White or female residents, and those insured by Medicare. As certain subgroups were disproportionately more likely to respond to our survey, there is a concern for selection bias. In addition, we performed a mailed survey, which may have introduced other biases in the responses received. However, to help adjust for these biases, we performed raking to adjust estimates to be reflective of the overall population sample for Sullivan County with respect to key demographic factors and health insurance status. Another limitation is that our survey was exclusive to full-time residents of Sullivan County. Therefore, our findings may not be generalisable to other rural populations of interest. Additionally, our measure of diabetes prevalence was based on if a healthcare provider had ever told respondents they had a health condition. However, if respondents did not regularly see a healthcare provider, there was a chance of misreporting

their true disease status, which would result in an underestimation of diabetes prevalence in our sample. Prior studies have suggested approximately one in four adults with diabetes in the USA are undiagnosed; therefore, the true diabetes prevalence may be about 33% higher than these diagnosed diabetes prevalence estimates.<sup>28</sup>

## CONCLUSION

Prior studies have demonstrated that rates of diabetes are higher in rural compared with urban areas of the USA.<sup>29,30</sup> This higher burden of diabetes may be due to specific challenges these populations face as individuals in rural areas, including being less physically active, having a higher prevalence of obesity and higher rates of other cardiometabolic conditions compared with residents of urban areas.<sup>31</sup> Our study finds a higher prevalence of diabetes using a countywide mailed survey in comparison to modelled surveillance estimates, which suggests the need to improve more geographically detailed surveillance in rural areas of the country. Rural parts of the USA may suffer from a lack of robust disease surveillance and underestimated disease burden. Additional research is needed to consider if similar trends are occurring in other rural counties, and whether other strategies are needed to improve rural health surveillance. With diabetes affecting rural and urban communities in different ways, public health policy needs to address these health inequities in different geographic areas and provide interventions that fit the geographic locale in question.<sup>32</sup>

## Author affiliations

<sup>1</sup>Emergency Medicine, New York University School of Medicine, New York, New York, USA

<sup>2</sup>Population Health, New York University School of Medicine, New York, New York, USA

<sup>3</sup>Graduate School of Arts and Sciences, NYU, New York, New York, USA

<sup>4</sup>Northeast Ohio Medical University, Rootstown, Ohio, USA

<sup>5</sup>Touro College of Osteopathic Medicine Harlem Campus, New York, New York, USA

<sup>6</sup>Medical College of Georgia, Augusta, Georgia, USA

<sup>7</sup>Wagner Graduate School of Public Service, NYU, New York, New York, USA

**Contributors** DCL, BE, LET and RA contributed to the study concept and design. CQA, MR, RP, EJ, RP and DI contributed to data acquisition, analysis or interpretation. DCL and LR drafted the manuscript. CQA, MR, RP, EJ, RP and DI critically revised the manuscript for important intellectual content. DCL and RA conducted the statistical analyses. DCL obtained funding. DCL is the guarantor of this work and, as such, had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Funding** This study was funded by grant R01-DK124400 from the National Institute of Diabetes and Digestive and Kidney Diseases, which is focused on understanding the environmental and geographic risk factors for diabetes in rural areas of the country.

**Map disclaimer** The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

**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 Institutional Review Board at the NYU School of Medicine (study number: s19-01920). The Institutional Review Board provided a waiver of authorisation and consent for this part of the study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Deidentified study data will be shared with researchers who provide a methodologically sound proposal upon reasonable request and agree to the requirements of a data use agreement with NYU Langone Health. Requests may be directed to: David.Lee@nyulangone.org.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

David C Lee <http://orcid.org/0000-0002-7202-1893>

## REFERENCES

- 1 Strom JL, Lynch CP, Egede LE. Rural/urban variations in diabetes self-care and quality of care in a national sample of US adults with diabetes. *Diabetes Educ* 2011;37:254–62.
- 2 Ablah E, Dong F, Cupertino AP, et al. Prevalence of diabetes and pre-diabetes in Kansas. *Ethn Dis* 2013;23:415–20.
- 3 O'Connor A, Wellenius G. Rural-urban disparities in the prevalence of diabetes and coronary heart disease. *Public Health (Fairfax)* 2012;126:813–20.
- 4 Krishna S, Gillespie KN, McBride TM. Diabetes burden and access to preventive care in the rural United States. *J Rural Health* 2010;26:3–11.
- 5 Long AS, Hanlon AL, Pellegrin KL. Socioeconomic variables explain rural disparities in US mortality rates: Implications for rural health research and policy. *SSM Popul Health* 2018;6:72–4.
- 6 Lutfiyya MN, McCullough JE, Mitchell L, et al. Adequacy of diabetes care for older U.S. rural adults: a cross-sectional population based study using 2009 BRFSS data. *BMC Public Health* 2011;11:940.
- 7 Andrus MR, Kelley KW, Murphey LM, et al. A comparison of diabetes care in rural and urban medical clinics in Alabama. *J Community Health* 2004;29:29–44.
- 8 Brown-Guion SY, Youngerman SM, Hernandez-Tejada MA, et al. Racial/ethnic, regional, and rural/urban differences in receipt of diabetes education. *Diabetes Educ* 2013;39:327–34.
- 9 Callaghan T, Ferdinand AO, Akinlotan MA, et al. The Changing Landscape of Diabetes Mortality in the United States Across Region and Rurality, 1999–2016. *J Rural Health* 2020;36:410–5.
- 10 Singh GK, Siahpush M. Widening rural-urban disparities in life expectancy, U.S., 1969–2009. *Am J Prev Med* 2014;46:e19–29.
- 11 Ferdinand AO, Akinlotan MA, Callaghan T, et al. Diabetes-related hospital mortality in the U.S.: A pooled cross-sectional study of the National Inpatient Sample. *J Diabetes Complicat* 2019;33:350–5.
- 12 Estimated county-level prevalence of diabetes and obesity - United States, 2007. *MMWR Morb Mortal Wkly Rep* 2009;58:1259–63.
- 13 Barker LE, Thompson TJ, Kirtland KA, et al. Bayesian small area estimates of diabetes incidence by United States county, 2009. *J Data Sci* 2013;11:269–80.
- 14 Spielman SE, Folch D, Nagle N. Patterns and causes of uncertainty in the American community survey. *Appl Geogr* 2014;46:147–57.
- 15 Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol* 2013;13:49.



- 16 Bang H, Edwards AM, Bombback AS, *et al.* Development and validation of a patient self-assessment score for diabetes risk. *Ann Intern Med* 2009;151:775–83.
- 17 Battaglia MP, Hoaglin DC, Frankel MR. Practical considerations in raking survey data. *Surv Pract* 2009;2:1–10.
- 18 Lee DC, McGraw NA, Doran KM, *et al.* Comparing methods of performing geographically targeted rural health surveillance. *Emerg Themes Epidemiol* 2020;17:3.
- 19 Dunson DB. Commentary: practical advantages of Bayesian analysis of epidemiologic data. *Am J Epidemiol* 2001;153:1222–6.
- 20 Burkhart Q, Haviland AM, Kallaur PN, *et al.* How much do additional mailings and telephone calls contribute to response rates in a survey of medicare beneficiaries? *Field methods* 2015;27:409–25.
- 21 Lee DC, Feldman JM, Osorio M, *et al.* Improving the geographical precision of rural chronic disease surveillance by using emergency claims data: a cross-sectional comparison of survey versus claims data in Sullivan County, New York. *BMJ Open* 2019;9:e033373.
- 22 Dugani SB, Wood-Wentz CM, Mielke MM, *et al.* Assessment of disparities in diabetes mortality in adults in US rural vs nonrural counties, 1999–2018. *JAMA Netw Open* 2022;5:e2232318.
- 23 Mainous AG, King DE, Garr DR, *et al.* Race, rural residence, and control of diabetes and hypertension. *Ann Fam Med* 2004;2:563–8.
- 24 Pinchevsky Y, Butkow N, Raal FJ, *et al.* Demographic and clinical factors associated with development of type 2 diabetes: a review of the literature. *Int J Gen Med* 2020;13:121–9.
- 25 Peters SAE, Yang L, Guo Y, *et al.* Parenthood and the risk of diabetes in men and women: a 7 year prospective study of 0.5 million individuals. *Diabetologia* 2016;59:1675–82.
- 26 Arcury TA, Preisser JS, Gesler WM, *et al.* Access to transportation and health care utilization in a rural region. *J Rural Health* 2005;21:31–8.
- 27 McAlexander TP, Malla G, Uddin J, *et al.* Urban and rural differences in new onset type 2 diabetes: Comparisons across national and regional samples in the diabetes LEAD network. *SSM Popul Health* 2022;19:101161.
- 28 Menke A, Casagrande S, Geiss L, *et al.* Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA* 2015;314:1021–9.
- 29 Chin MH, Cook S, Jin L, *et al.* Barriers to providing diabetes care in community health centers. *Diabetes Care* 2001;24:268–74.
- 30 Dugani SB, Mielke MM, Vella A. Burden and management of type 2 diabetes in rural United States. *Diabetes Metab Res Rev* 2021;37:e3410.
- 31 Patterson PD, Moore CG, Probst JC, *et al.* Obesity and physical inactivity in rural America. *J Rural Health* 2004;20:151–9.
- 32 Massey CN, Appel SJ, Buchanan KL, *et al.* Improving diabetes care in rural communities: an overview of current initiatives and a call for renewed efforts. *Clin Diabetes* 2010;28:20–7.