Identifying priority and bright spot areas for improving diabetes care: a geospatial approach

Michael Topmiller,1 Peter J Mallow,2 Kyle Shaak,3 Autumn M Kieber-Emmons3,4

ABSTRACT

The objective of this study was to describe a novel geospatial methodology for identifying poor-performing (priority) and well-performing (bright spot) communities with respect to diabetes management at the ZIP Code Tabulation Area (ZCTA) level. This research was the first phase of a mixed-methods approach known as the focused rapid assessment process (fRAP). Using data from the Lehigh Valley Health Network in eastern Pennsylvania, geographical information systems mapping and spatial analyses were performed to identify diabetes prevalence and A1c control spatial clusters and outliers. We used a spatial empirical Bayes approach to adjust diabetes-related measures, mapped outliers and used the Local Moran’s I to identify spatial clusters and outliers. Patients with diabetes were identified from the Lehigh Valley Practice and Community-Based Research Network (LVPBRN), which comprised primary care practices that included a hospital-owned practice, a regional practice association, independent small groups, clinics, solo practitioners and federally qualified health centres. Using this novel approach, we identified five priority ZCTAs and three bright spot ZCTAs in LVPBRN. Three of the priority ZCTAs were located in the urban core of Lehigh Valley and have large Hispanic populations. The other two bright spot ZCTAs have fewer patients and were located in rural areas. As the first phase of fRAP, this method of identifying high-performing and low-performing areas offers potential to mitigate health disparities related to diabetes through targeted exploration of local factors contributing to diabetes management. This novel approach to identification of populations with diabetes performing well or poor at the local community level may allow practitioners to target focused qualitative assessments where the most can be learnt to improve diabetic management of the community.

INTRODUCTION

Uncontrolled diabetes results in complications, avoidable hospitalisations and growing healthcare costs in the USA and disproportionately affects rural and high poverty areas.1 Primary care physicians play an important role in managing diabetes in a number of ways, including improving access to annual retinal examinations, haemoglobin A1c (A1c) tests and foot examinations, in these areas.2 Practice redesign over the last decade has focused on how primary care providers within patient-centred medical homes can improve diabetes care.3 4 However, beyond a practice’s walls, several factors related to community environments affect diabetes management of patients. Rural and poverty-stricken patients with less access to physical activity resources and healthy foods often have worse diabetes-related outcomes, including death.5 6 In urban areas, the lack of safe walkable communities and food deserts have been associated with worse outcomes.7 Policy differences in insurance access across settings have also shown to be associated with disparities in diabetes outcomes.8 More research is needed to understand these community factors. Identifying and locating geographical areas with patients most in need of intervention based on clinical data may provide insights into how to improve disparities in diabetes management and outcomes.

Over the past decade there has been a growing interest among primary care researchers using geographical information systems (GIS) to better understand the social determinants of health of patients.8–11 Berke8 discusses the potential of GIS for better understanding the context or place of patients to improve patient care. Bazemore et al12 integrated clinical and community data to create service area and penetration maps for a network of community health centres, revealing gaps in access to care. Dulin et al15 used GIS as part of the Multiple Attribute Primary Care Targeting Strategy (MAPCATS), which integrated various sources of data (both clinical and community) to identify areas that would benefit most from increased access to primary care.10 The MAPCATS was later used to help to develop interventions to improve access to care for underserved populations.11

Several studies have used GIS mapping to target diabetes interventions using public, community survey and clinical data.12–15 Kruger et al16 mapped data from a community...
survey to identify high-risk areas for diabetes and explore haemoglobin A1c testing rates in these areas; focus groups and needs assessment were then conducted in these high-risk areas to explore potential social and environmental factors. Curtis et al.14 used publicly available county-level data to explore high-risk areas based on diabetes prevalence and low-resource areas based on the number of physicians and diabetes self-management education programmes in Michigan. Using registry data from the University of California Davis electronic health record (EHR) system, researchers mapped out more than 8000 patients with diabetes and linked them with census tract-level socioeconomic and demographic data. Findings included an association between census tract-level socio-economic status (SES) and A1c control, where low-income areas had higher rates of poor A1c control.14 A mixed-methods study in Boston identified census tracts with large numbers of patients with diabetes in poor control (priority areas) in order to target in-depth qualitative research (photovoice) to better understand local factors that might contribute to poor diabetes outcomes.15 Other studies have used various geospatial approaches to identify target areas for a variety of health outcomes.16-19 MacQuillan et al.20 focused on mapping birth outcomes in Kalamazoo, Michigan using kernel density estimation methods to identify target areas and changes over time. Hardt et al.21 created ‘hot spot’ density maps to highlight priority areas based on social and health measures from various sources; these maps were shared widely, resulting in service providers targeting high-needs areas, the relocation and expansion of health services, and plans for a new location of local health department. One study used outlier maps and hot spot mapping to identify priority areas for increased treatment of opioid use disorder, while research focused on the Appalachian region used a conditional mapping approach to identify priority and bright spot counties for targeted diabetes interventions.18

While several of the referenced studies have integrated multiple data sources and used GIS approaches to target in-depth research to better understand management of diabetes,12 15 the research on using EHR data for use by a health system to identify target populations for in-depth research to better understand factors contributing to uncontrolled diabetes is lacking. Linking targeted geospatial mapping with qualitative enquiry is a novel mixed-methods approach developed by Lehigh Valley Health Network (LVHN) researchers and national collaborators entitled focused rapid assessment process (fRAP)20 This paper describes the first phase of an fRAP study focused on diabetes management.

The objective of this study was to describe an exploratory geospatial methodology for identifying priority and bright spot areas (defined as ZIP Code Tabulation Areas, ZCTAs) in the LVHN region of service. The results of this research are being used by researchers at LVHN to identify and target underserved patients for in-depth, qualitative research to better understand social and physical environment factors that may be influencing self-management of diabetes.

**GEOGRAPHICAL BACKGROUND**

LVHN is the largest healthcare provider in the Lehigh Valley region which is located in eastern Pennsylvania about 60 miles north of Philadelphia and 100 miles west of New York City. LVHN includes health centres caring for communities in six counties, numerous primary and specialty care physician practices throughout the region, and pharmacy, imaging and laboratory services, and preferred provider services. The Lehigh Valley Practice and Community-Based Research Network (LVPRBN), supported by the Department of Family Medicine, comprised approximately 250 primary care physicians in over 70 practices. LVPRBN, formerly Eastern Pennsylvania Inquiry Collaborative Network or EPICnet, is an Agency for Healthcare Research and Quality (AHRQ)-recognised practice and practice-based research network whose membership spans the spectrum of practice models, including a hospital-owned practice, a regional practice association, independent small groups, clinics, solo practitioners and federally qualified health centres. Practices serve semi-urban, suburban and rural areas.

The primary care practices of LVHN included in this research serve a diverse patient population. The Hispanic population has grown quickly in the past decade, comprising almost 13% of the population. Nearly 5% of the population speaks Spanish as their primary language. Approximately, 12% of the population lives in poverty. The insurance mix of the population includes 27% Medicare, 12% Medicaid and 4% uninsured.

**STEPS FOR PERFORMING GEOSPATIAL MAPPING**

**Step 1: ethical considerations**

All data necessary were retrospective in nature and non-interventional. We received approval from LVHN to use all data in this study.

**Step 2: data collection**

LVHN’s EHR was queried for data on its primary care (family medicine and internal medicine) patients visiting between August 2016 and August 2018. A patient was included if they met the following criteria: an in-office encounter at a primary care practice in the past 24 months including A1c test, a primary care practice (PCP) belonging to a primary care practice, at least 18 years old and not deceased. This query identified over 180 000 patients from 38 practices spread across 98 ZCTAs (ranging from 55 to 12 630 patients per ZCTA). The roughly 180 000 patients make up about 14% of the population across the 98 ZCTAs. Of these 180 000 patients, the electronic management record’s (EMR) diabetes registry was used to identify 21 664 primary care patients with type II diabetes; patients were not included in the diabetic analysis if they had type I or gestational diabetes. Based
on estimates of diabetes prevalence across these ZCTAs, the 21 664 patients with diabetes make up about 13% of the population with diabetes. Patient data were aggregated at the ZCTA level for diabetes prevalence, the rates of controlled and uncontrolled diabetes, and the percentages of Hispanic and non-Hispanic patients. We identified patient diabetes control based on their most recent A1c test. The EHR data were rounded to the nearest whole number. Controlled diabetes was defined as an A1c result less than 8% and uncontrolled if the result was greater than 9%.2 4 21 We chose less than 8 and greater than 9 as the uncontrolled threshold given the conflicting literature around appropriate A1c levels for different age groups, helping to ensure that the thresholds for this study would be consistent with most standards. All ZCTAs with less than 10 patients with diabetes were removed.

Community-level estimates of the population as well as urban versus rural status of the ZCTAs were obtained from the American Community Survey 2013–201722 and the Robert Graham Center estimates of patients with diabetes, which used data from the Behavioral Risk Factor Surveillance System 2017.23 The use of the urban and rural census designations was chosen to support the second phase of the fRAP approach.

**Step 3: identification of priority and bright spot communities**

Using the open-source GIS software GeoDa,24 geospatial analyses were performed to identify priority and bright spot areas (defined as ZCTAs) based on diabetes prevalence and rates of A1c control (or lack of control). We created spatial empirical Bayes (SEB) adjusted measures for diabetes prevalence, controlled A1c (<8) and uncontrolled A1c (>9). Due to the wide variation in the number of total patients per ZCTA (ranging from 55 to 12 630) and evidence of spatial clustering of diabetes prevalence and A1c control, we used the SEB approach. The SEB approach adjusts rates (eg, diabetes prevalence) towards a local mean (based on a ZCTA and its neighbours) based on the number of patients (ZCTAs with more patients are adjusted less, ZCTAs with fewer patients are adjusted more).20 SEB approaches are useful when there is instability in population variance (wide ranges in population) and wide geographical variations; using a standard empirical Bayes (EB) approach that adjusts rates based on the overall population mean may cause us to overlook smaller population regions as the approach smooths out low population regions. Previous literature on using SEB approaches in healthcare research is limited, although one study used an SEB approach to identify priority areas for community health centres.25

**Step 4: outlier maps**

Using the SEB adjusted measures, we created outlier maps to identify controlled and uncontrolled A1c outliers. Outlier maps are a GIS method for highlighting both positive and negative outliers, and can be done by using box maps, SD maps or percentile maps.26 We used SD outlier maps and focused on outliers that were 2 SD or above the mean.

**Step 5: determine significant spatial clusters**

The next step was to explore and identify statistically significant spatial clusters and outliers for uncontrolled and controlled A1c. Specifically, we conducted a Local Moran’s I to identify uncontrolled A1c hot spots and controlled A1c spatial outliers. The Local Moran’s I tool allows users to identify statistically significant geographical clusters and outliers by comparing observed spatial patterns against permuted random patterns.27 It is important to note that we conducted the spatial analysis on the original raw rates, and not the SEB adjusted rates, as that would have introduced spatial bias into the analysis. Spatial clusters (or priority areas) of uncontrolled A1c ZCTAs were identified as ZCTAs with high rates of uncontrolled diabetes surrounded by ZCTAs with high rates of uncontrolled diabetes. Spatial outlier ZCTAs were defined as ZCTAs with high rates of controlled diabetes surrounded by ZCTAs with low rates of controlled diabetes. Significance was determined at the 0.05 level after running 99 999 permutations, and queen contiguity was used to define neighbours.27

**Step 6: establish criteria for defining priority areas and bright spots**

Figure 1 displays the criteria and location of priority and bright spot ZCTAs, which were defined by the following criteria: priority ZCTAs have SEB adjusted diabetes prevalence in the top 25th percentile (from step 5) and uncontrolled A1c rates 2 SD or greater than the mean (from step 5); priority ZCTAs are also identified as uncontrolled diabetes hot spots, which are ZCTAs with high rates of uncontrolled diabetes surrounded by ZCTAs with high rates of uncontrolled diabetes (from step 5). These criteria allowed for the identification of those ZCTAs most in need (priority) or performing the best (bright spots). Bright spot ZCTAs have SEB adjusted diabetes prevalence in the top 25th percentile (from step 3) and controlled A1c rates 2 SD or greater above the mean (from step 4); bright spot ZCTAs are also identified as controlled diabetes spatial outliers, which are defined as ZCTAs with high rates of controlled diabetes surrounded by ZCTAs with low rates of controlled diabetes (from step 5). Including both outliers (SD) and geographical clusters as priority areas was identified by the LVHN research team as an important component to capture both the ZCTAs with the highest rates but also potential ZCTAs that were part of geographical clusters that would allow for targeting ZCTAs and their surrounding areas in later phases of the fRAP process.

**LVHN EXAMPLE OF PRIORITY AREAS AND BRIGHT SPOTS**

Table 1 shows the characteristics of the almost 175 000 patients across 98 ZCTAs that were included in the study. There were five priority (three in the urban core) and two...
bright spot ZCTAs identified based on the criteria chosen in step 6 (figure 1). The data were separated by urban and rural priority ZCTAs based on input from the LVHN qualitative research team to differentiate the populations. Overall, the majority of LVHN patients were female (57.3%) and white (82.4%), with large percentages of Hispanic (12.7%), Medicare (27.3%) and patients with diabetes (12.0%). The patient sample is representative of the overall population, where the majority are white (75.4%) with similar shares of Hispanic (14.9%) and diabetic (13.2%) populations as evident from the community characteristics. The community characteristics provide further context of ZCTAs allowing a more holistic view of the patient population, although it must be noted these data were from estimates created by the American Community Survey22 and Robert Graham Center23 rather than direct measurements of the population.

The majority of the almost 15 000 patients living in priority ZCTAs are in the three urban ZCTAs (18101, 18102, 18109), which were the most populated ZCTAs in the service area. Patients in these ZCTAs were also much more likely to be Hispanic, with Spanish as primary language and on Medicaid. These ZCTAs have high rates of poverty relative to the LVHN service area. The urban priority ZCTAs contain one hospital and several community clinics. However, they are somewhat isolated geographically from the majority of LVHN’s practices and outpatient services. The two other priority ZCTAs (19507, 17901) were located on the rural edge of the LVHN service area, contained small numbers of patients with diabetes and have limited access to any LVHN facilities (based on distance). In addition to having much lower poverty rates, majority of patients in these ZCTAs were white and have much lower rates of Medicaid compared with urban core priority ZCTAs.

Initially, three bright spot ZCTAs (18056, 18105, 19503) were identified. Bright spot ZCTA 18105 was removed as it is primarily a PO Box, and few patients live there. As shown in figure 1, the remaining two bright spots were near each other in the northeast corner of Berks County (south of the urban core of Lehigh Valley) and have relatively small numbers of patients, with only 43 patients with diabetes residing in the two ZCTAs. Similar to the overall population, the large majority of patients living in these

---

**Figure 1** The bright spots (yellow) and priority areas (red) identified in the Lehigh Valley Health Network. ZCTAs, ZIP Code Tabulation Areas.
ZCTAs were white (~98%), with relatively higher percentages of Medicare patients. The two bright spot ZCTAs include a small town with some amenities, good public schools and large areas of open farmland surrounding the town centre. ZCTA 19503 has multiple churches and community programmes, sidewalks for ease of exercise and a community park. ZCTA 18056 is a smaller town centre and more rural, with abundant open farm and forestland.

CONCLUSION
As the first phase of fRAP, this research used a geospatial approach to identify diabetes management priority and bright spot areas and builds on LVHN efforts to use population health mapping for targeting diabetes care and addressing health disparities. LVHN researchers are now conducting a series of focus groups with patients from the priority and bright spot ZCTAs to explore local social, environmental and clinical factors that may improve disparities in diabetes care. Patients are being recruited to participate based on ethnicity, diabetes control and use of medication to ensure that information from focus groups is representative of a wide range of local experiences. Additionally, patient responses during focus groups are informing a retrospective chart review to explore local differences in clinical diabetes indicators. Identifying local disparities, barriers and facilitators may lead to tailored interventions designed to improve diabetes care and adherence to guidelines.

This research builds on the growing interest in GIS mapping approaches to help identify target areas for improving diabetes care and highlight areas for future in-depth research. While this research could be used a model for future researchers, variations of this method can also be used to identify target areas. For example, the decision to adjust our measures based on local averages rather than the overall mean (ie, for all of the LVHN service area) was made by the research team to ensure that the more rural areas with lower numbers of patients were not overlooked. In preliminary analyses using a

<table>
<thead>
<tr>
<th>Measure</th>
<th>All</th>
<th>Urban priority</th>
<th>Rural priority</th>
<th>Bright spots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ZCTAs</td>
<td>98</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Patient characteristics</td>
<td>DM</td>
<td>All</td>
<td>DM</td>
<td>All</td>
</tr>
<tr>
<td>% female</td>
<td>49.0</td>
<td>57.3</td>
<td>57.8</td>
<td>61.2</td>
</tr>
<tr>
<td>% white</td>
<td>79.5</td>
<td>82.4</td>
<td>42.7</td>
<td>45.2</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>16.1</td>
<td>12.7</td>
<td>60.9</td>
<td>56.7</td>
</tr>
<tr>
<td>% Spanish language</td>
<td>9.4</td>
<td>4.9</td>
<td>44.2</td>
<td>30.1</td>
</tr>
<tr>
<td>% Medicaid</td>
<td>10.3</td>
<td>12.0</td>
<td>27.5</td>
<td>37.2</td>
</tr>
<tr>
<td>% Medicare</td>
<td>54.4</td>
<td>27.3</td>
<td>49.2</td>
<td>24.2</td>
</tr>
<tr>
<td>% with diabetes</td>
<td>n/a</td>
<td>12.0</td>
<td>n/a</td>
<td>17.0</td>
</tr>
<tr>
<td>% A1c control</td>
<td>73.5</td>
<td>n/a</td>
<td>66.8</td>
<td>n/a</td>
</tr>
<tr>
<td>% A1c uncontrolled</td>
<td>13.4</td>
<td>n/a</td>
<td>19.8</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Community characteristics

<table>
<thead>
<tr>
<th>Measure</th>
<th>All</th>
<th>Urban priority</th>
<th>Rural priority</th>
<th>Bright spots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>1 228 754</td>
<td>n/a</td>
<td>71 154</td>
<td>n/a</td>
</tr>
<tr>
<td>% non-Hispanic white</td>
<td>75.4</td>
<td>n/a</td>
<td>27.2</td>
<td>n/a</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>14.9</td>
<td>n/a</td>
<td>57.3</td>
<td>n/a</td>
</tr>
<tr>
<td>% limited English proficiency</td>
<td>3.2</td>
<td>n/a</td>
<td>14.1</td>
<td>n/a</td>
</tr>
<tr>
<td>% age 65 and older</td>
<td>16.7</td>
<td>n/a</td>
<td>10.0</td>
<td>n/a</td>
</tr>
<tr>
<td>% poverty</td>
<td>11.5</td>
<td>n/a</td>
<td>33.0</td>
<td>n/a</td>
</tr>
<tr>
<td>% diabetes</td>
<td>13.2</td>
<td>n/a</td>
<td>10.5</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Urban priority includes ZCTAs 18101, 18102 and 18109 (identified as hot spots).
Rural priority includes ZCTAs 17901 and 19507 (identified as SD outliers).
Bright spots include ZCTAs 18056 and 19503 (identified as spatial outliers).
Patient characteristics in table 1 include primary care (family medicine and internal medicine) patients visiting between August 2016 and August 2018. A patient was included if they met the following criteria: an in-office encounter at a primary care practice in the past 24 months, a PCP belonging to a primary care practice, at least 18 years old and not deceased. The denominator for the patient characteristics was the total number of LVHN patients within the ZCTA(s).
Community characteristics in table 1 include data from the American Community Survey 2013–2017 and estimates (for diabetes) created by the Robert Graham Center using data from the Behavioral Risk Factor Surveillance System 2017.
DM, type II diabetes mellitus; LVHN, Lehigh Valley Health Network; n/a, not available; ZCTAs, ZIP Code Tabulation Areas.
standard EB approach, which adjusts based on the overall population, the only ZCTAs that were identified were the ones with the largest number of patients. This demonstrates the exploratory nature of this research and the flexibility of the methods presented in this paper, where various methods are available and priority and bright spot areas can be defined in many ways. This is important as primary care researchers attempting to replicate these methods may not have access to GIS experts or software and may need to explore more accessible methods and tools that allow non-GIS experts to map and explore service area based on publicly available data or aggregate clinical data.26–30

While having flexibility in methods is important, it also highlights one of the main limitations of geospatial approaches for targeting priority areas. Most of the decisions that define the criteria for identifying priority and bright spot areas were arbitrary. We could have used different cut-offs for the outlier maps and statistical significance for identifying hot spots and spatial outliers (eg, p=0.01 rather than p=0.05). Modifying any of these criteria may have changed the results. An additional limitation involves the geographical unit of analysis, ZCTAs. It is likely that using a more refined geography (such as census tracts) or clinic-level data may have changed the results, a common issue in geographical research known as themodifiable areal unit problem.31 Finally, the approach described is intended to identify communities for further research and should not be construed as a means to target interventions to individual patients.

Despite the limitations described above, the results of this research are allowing LVHN to target specific areas for better understanding of community factors that impact diabetes management and outcomes. By using fRAP to identify priority and bright spot areas, LVHN has been able to tailor qualitative enquiry to uncover barriers and facilitators that are contributing to disparities in diabetes outcomes for our patient population. Future research will explore different criteria and methods for identifying priority and bright spot areas and clinics, including methods that adjust for population factors such as poverty, race or ethnicity. This will help ensure that we identify priority and bright spot ZCTAs that have similar characteristics so lessons learnt from patients in bright spots can be applied directly to patients in priority areas.

This research presented an innovative geospatial method for identifying areas for targeted interventions to improve diabetes care. We identified several priority and bright spot areas for targeted in-depth research to better understand how local community factors impact diabetes management.

Contributors MT, AMK-E and KS conceived the study. KS and MT completed the literature review and wrote the background. KS extracted and cleaned the data. MT and PJM performed the analytics. MT created the map. AMK-E took the lead on the conclusion section. MT is responsible for the content of the article, although all authors contributed and approved the final manuscript.

Funding This research was supported by the Dorothy Rider Pool Health Care Trust (1573-020).

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Institutional Review Board (IRB) at Lehigh Valley Health Network.

Provenance and peer review Not commissioned; externally peer reviewed.

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
Peter J Mallow http://orcid.org/0000-0001-9642-2977

REFERENCES


