# Using Geographic Information Systems (GIS) to Assess Outcome Disparities in Patients with Type 2 Diabetes and Hyperlipidemia

*Estella M. Geraghty, MD, MS, MPH/CPH, FACP, Thomas Balsbaugh, MD, Jim Nuovo, MD, and Sanjeev Tandon, MBBS, MD*

*Objectives:* Geographic information systems (GIS) tools can help expand our understanding of disparities in health outcomes within a community. The purpose of this project was (1) to demonstrate the methods to link a disease management registry with a GIS mapping and analysis program, (2) to address the challenges that occur when performing this link, and (3) to analyze the outcome disparities resulting from this assessment tool in a population of patients with type 2 diabetes mellitus.

*Methods:* We used registry data derived from the University of California Davis Health System's electronic medical record system to identify patients with diabetes mellitus from a network of 13 primary care clinics in the greater Sacramento area. This information was converted to a database file for use in the GIS software. Geocoding was performed and after excluding those who had unknown home addresses we matched 8528 unique patient records with their respective home addresses.

Socioeconomic and demographic data were obtained from the Geolytics, Inc. (East Brunswick, NJ), a provider of US Census Bureau data, with 2008 estimates and projections. Patient, socioeconomic, and demographic data were then joined to a single database. We conducted regression analysis assessing A1c level based on each patient's demographic and laboratory characteristics and their neighborhood characteristics (socioeconomic status [SES] quintile). Similar analysis was done for low-density lipoprotein cholesterol.

*Results:* After excluding ineligible patients, the data from 7288 patients were analyzed. The most notable findings were as follows: There was, there was found an association between neighborhood SES and A1c. SES was not associated with low-density lipoprotein control.

*Conclusion:* GIS methodology can assist primary care physicians and provide guidance for disease management programs. It can also help health systems in their mission to improve the health of a community. Our analysis found that neighborhood SES was a barrier to optimal glucose control but not to lipid control. This research provides an example of a useful application of GIS analyses applied to large data sets now available in electronic medical records. (J Am Board Fam Med 2010;23:88–96.)

Type 2 diabetes mellitus (DM) is one of the most common chronic medical conditions in the United States. The Centers for Disease Control and Prevention estimates that there are approximately 17 million people with DM.[1] End-organ complica-

tions from DM are a substantial source of morbidity and mortality. Disparities for adverse outcomes associated with DM are well documented.[2] Factors that are associated with these disparities include language barriers, inadequate access to care, low socioeconomic status, and suboptimal self-care behaviors.[3] Significant disparities exist among racial/ethnic minorities in DM health outcomes and quality of care, including poorer measures of glycemic control and higher rates of end-organ complications.[4]

Efforts to improve outcomes of patients with chronic conditions has included implemention of the Chronic Care Model (CCM).[5] The CCM was developed to identify the essential elements of a health care system that encourage high-quality chronic disease care. These elements include the

community, the health system, self-management support, delivery system design, decision support, and clinical information systems.[5] Clinical information systems may serve as reminders to help teams comply with practice guidelines, as report cards to providers about measures of care, and as registries for planning individual patient care and conducting population-based care.[6] When conducting population-based care, geographic information systems (GIS) tools expand our understanding of disparities in health outcomes within a community.[7]

The purpose of this project was 3-fold. We wanted to demonstrate the methods to link a disease management registry with a GIS mapping and analysis program; to address the challenges that occur in performing this link; and to describe the contributing factors associated with differences in outcomes in a population of patients with DM, specifically the factors associated with optimal glucose control and treatment of hyperlipidemia.

## Methods

The Chronic Disease Management Program at the University of California Davis Health System (UCDHS) is a system-wide effort to improve the quality of care for patients with chronic conditions. The CCM served as the paradigm by which the program was structured.[5] The Chronic Disease Management Program maintains a registry of all patients with DM who receive care in the UCDHS.

### University of California Davis Medical Center and Primary Care Network

The patient registry data from this study included all patients with DM who were seen within the clinics at the Sacramento-based medical center and our regional primary care network (13 clinics represented in this analysis). Clinic addresses were geocoded with 100% matching.

### Diabetes Registry Data

Patient data were obtained by querying the UCDHS's electronic medical record (Epic, Verona, WI). We included all patients with a diagnosis of DM who had a clinic visit with a UCDHS family physician or internist during the previous 1-year period (April 18, 2008, to April 19, 2009). Confirmation of DM was based on International Classification of Diseases version 9 diagnosis codes

250.00 through 250.99. Patient variables requested included patient address, age, sex, primary care provider name and office location, last glycohemoglobin (A1c), last low-density lipoprotein (LDL) cholesterol, last urine microalbumin/creatinine, insurance type, race/ethnicity, and primary language. The resulting table contained 9722 unique patient records.

### Application of Registry Data to GIS Mapping Tool

The diabetes registry data were incorporated into a spreadsheet format (Excel 2003, Microsoft, Bellevue, WA). These data were converted to a database file structure for use in the GIS software (ArcInfo version 9.3, ESRI, Redlands, CA). Geocoding was performed to obtain latitude and longitude coordinates for each patient's home address. Initially, 8051 (83%) records were matched and 1671 (17%) were left unmatched. Numerous errors were discovered during the rematching process, including abbreviated street names, lack of appropriate spacing between words and/or street numbers, and misspellings of street names. All unmatched records were reviewed and corrected where possible, leading to another 477 matched addresses (total of 8528 [88%]). Most of the remaining unmatched addresses were post office boxes and therefore could not be geocoded. Individuals younger than the age of 25 were excluded (n = 53) so family physicians could be directly compared with internists. In addition, this age cutoff correlates with the census bureau's cutoff for education attainment reporting. Another 115 patients did not have an A1c on file and were excluded. Resulting point features were then spatially joined to year 2000 census tracts (www.census.gov). A spatial join is a process that allows the GIS user to append the attributes of one data layer (patient address points) to the attributes of another layer (census data) based on a common location (census tracts). For cases in which a census tract had fewer than 10 patients, those tracts and patients (763) were excluded from the analysis to avoid the rate instability associated with small sample sizes, leaving 7288 patient records for analysis (Figure 1).

These data were used to construct 2 example maps as demonstrations of the capabilities of this GIS tool. The first (Figure 2) addressed the diabetic population of the primary care network; the second (Figure 3) addressed the comparison of

**Figure 1.  Study design showing initial patient database and progressive exclusion criteria. Shaded boxes indicate numbers of patients who were excluded.**
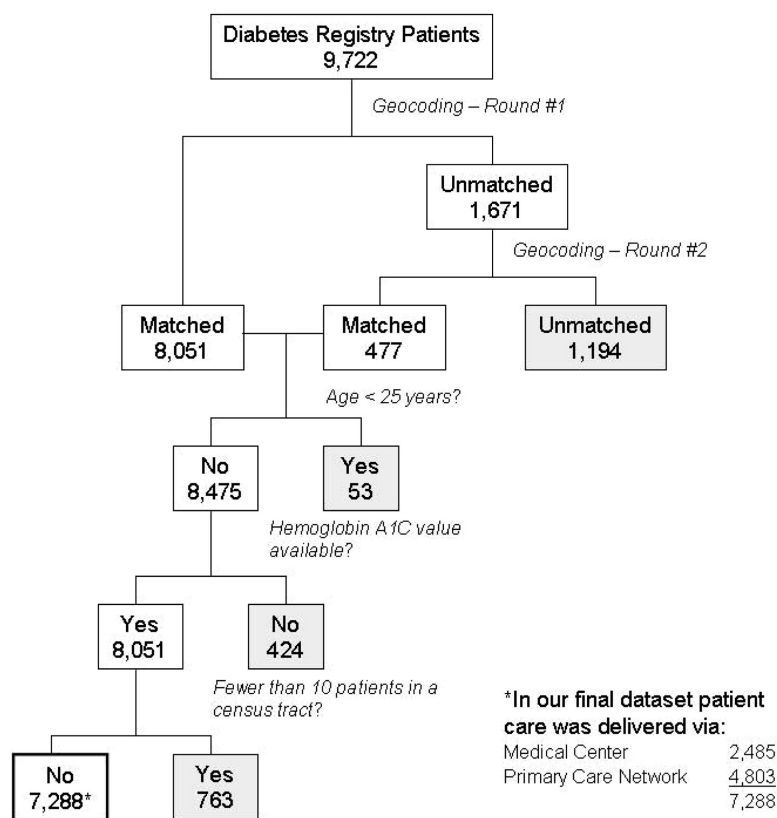


driving distance to each patient's primary care provider.

### Socioeconomic and Demographic Data

Socioeconomic and demographic data were obtained from GeoLytics, Inc. (East Brunswick, NJ). We collected 2008 data at the census tract level. GeoLytics bases their estimates on US Census Bureau reports and limited population estimates, then expands on those to provide multiple population-based variables. Variables included were median income, education attainment, unemployment, and white and black race.
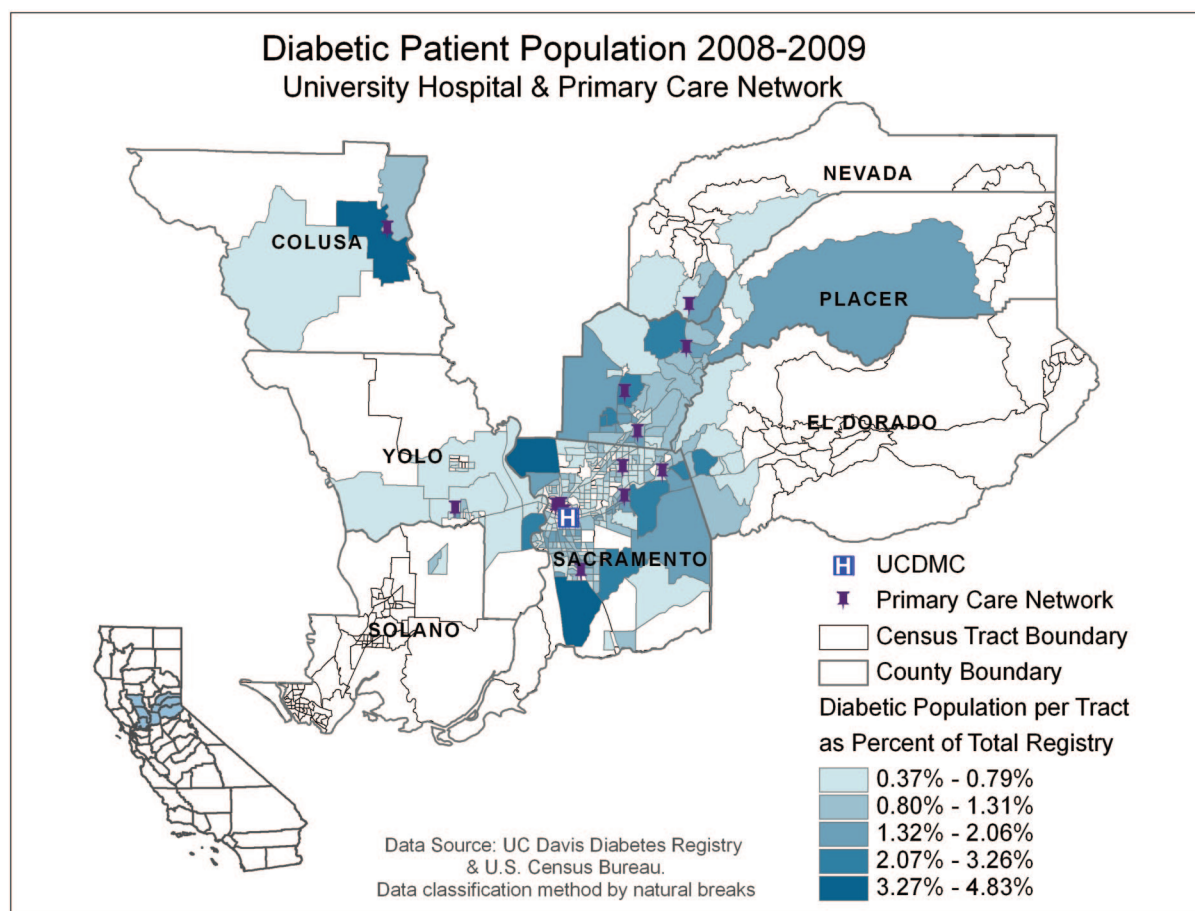
### Data Analysis

To address study purpose—describing the contributing factors associated with differences in outcomes associated with optimal glucose and lipid control—-we performed the following data analysis. We obtained patient data and demographic information from our disease management registry. This aforementioned registry is derived from our electronic medical records. Information about so-

cioeconomic status (SES) was obtained from the US Census Bureau's 2008 report. Patient, socioeconomic, and demographic data were joined to a single database using ArcInfo (version 9.3, ESRI). Then, Euclidean distance was calculated from each patient's home to their primary care clinic. The resulting file was loaded into STATA MP 11 (Stata Corp., College Station, TX). Sociodemographic variables were combined using factor analysis. We followed the technique described by Diez Roux et al[8] to create a summary measure of census tract level socioeconomic status. Variables in the factor analysis included median income in 2008, proportion of population with a below-average education, proporation of the population that was unemployed, and the proportion of the population that was black or white. The resulting factor (labeled as SES) was categorized by quintiles.

We sought to determine whether there was an association between optimal glucose and lipid control with these demographic and socioeconomic variables. In our first model we conducted a random-intercept and random-slope regression analy-

**Figure 2. Diabetic population as a percent of the total diabetic patient registry population by census tract. Medical center and primary care network locations are also noted.**



Diabetic Patient Population 2008-2009
University Hospital & Primary Care Network

Data Source: UC Davis Diabetes Registry
& U.S. Census Bureau.
Data classification method by natural breaks

sis assessing A1c level based on each patient's personal and laboratory characteristics (age, sex, insurance status, race/ethnicity, primary language, LDL cholesterol level, urinary microalbumin/creatine ratio, and distance from practice clinic); the practice characteristics of their primary care physician (PCP) (medical center vs satellite primary care clinic); clinic specialty (family medicine, internal medicine, or combined) as fixed effects. The census tract in which the patient lives and their neighborhood's SES quintile were considered random effects. Within the model age was treated as a categorical variable. A second model regressed LDL cholesterol level on the same patient, practice, and neighborhood characteristics but also included whether a statin medication had been prescribed. In this case a mixed-effects logistic regression was used, given that consensus statements suggested that LDL cholesterol should be at least <100 in all patients with DM.[9] Model assessment included the

intracluster correlation coefficient (ICC or rho) of the random effects, and variance explained by the model ($R^2$). The ICC is a measure of the relatedness of clustered data. Similarity among patients within groups or clusters reduced the variability of responses within a cluster. As rho approaches 1, all responses within a cluster are identical.

## Results

A total of 7288 patients were analyzed. Baseline characteristics for the study population are shown in Table 1 and the spatial distribution of our patients with DM within the UCDHS is shown in Figure 2. The mean A1c of the patients in our registry was 7.26% (range, 4.0% to 18.4%). Mean LDL cholesterol among our patients with DM was 96.0 mg/dL (range, 6.0 to 313.0 mg/dL).

Our regression model provided adjusted estimates of the effects of individual and provider based on A1C

**Figure 3. Diabetes patients in our health system drive farther, on average, to see their primary care physicians at the medical center. If our health services were to be expanded, the area of greatest need for our diabetic patients is in Nevada and Placer counties.**
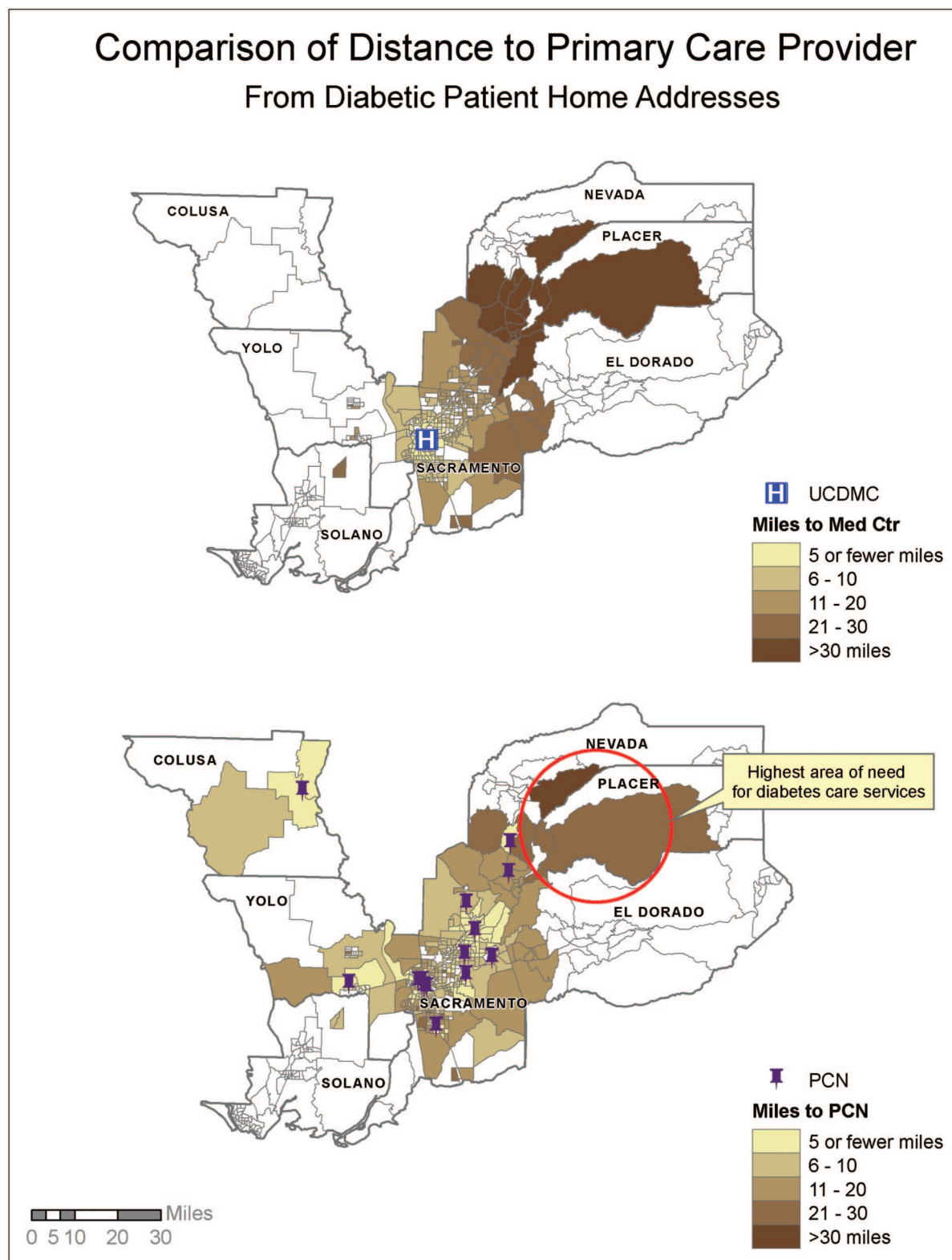
**Table 1. Baseline Study Population Characteristics of 7288 Patients**

| Characteristic | Mean (SD) |
|---|---|
| Female (%) | 47.39 |
| Age (years) | 62.00 (14.08) |
| Hemoglobin A1c | 7.26 (1.66) |
| LDL cholesterol (mg/dL) | 96.00 (35.15) |
| Urine microalbumin/creatinine | 56.45 (391.39) |
| Race/Ethnicity (%) | |
| White | 44.67 |
| Black | 10.50 |
| Hispanic | 9.17 |
| Asian | 4.82 |
| Other | 30.84 |
| Insurance type (%) | |
| Fee for service | 1.49 |
| HMO | 41.67 |
| MediCal | 3.99 |
| MediCare | 41.47 |
| PPO | 9.56 |
| Self pay | 0.72 |
| Workman's compensation | 0.25 |
| Other | 0.86 |
| Primary language (%) | |
| English | 94.73 |
| Other | 5.27 |
| Driving distance to PCP (miles) | 5.56 (5.05) |
| Driving distance to medical center (miles) | 9.25 (10.10) |

LDL, low-density lipoprotein; HMO, health management organization; PPO, preferred provider organization; PCP, primary care physician.

(Table 2). The ICC of the SES neighborhood was 0.09 and the model $R^2$ was 8%. In general, women had a higher A1c than men in our study ($P < .001$). Age was not a significant predictor of A1c level; however, being black was predictive of a higher A1c ($P = .002$). Most of the insurance categories we examined were not related to A1c control except for self-pay patients and those being seen under workman's compensation programs. Both kinds of insurance were associated with having a lower A1c. LDL cholesterol and urine microalbumin/creatinine were both directly correlated with glucose control. There was no significant difference between family practice and internal medicine specialties in predicting A1c levels, but we did find that patients being seen in our primary care network were more likely to have a lower A1c than those seen at the medical center ($P < .001$). The distance from a patient's home address to their primary care clinic was not related to their A1c level.

Table 3 shows the resulting odds, 95% CIs, and $P$ from our mixed effects logistic regression of LDL cholesterol on the same set of variables and whether a statin medication was prescribed. In this model, an odds ratio <1 predicted a greater probability of LDL cholesterol being <100. Variables that predicted having an LDL <100 included female sex, younger age, statin prescription, and being seen in an internal medicine clinic. Our model also indicated that patients were less likely to have a controlled LDL cholesterol level if they had a higher A1c level. An individual's race and the distance that a patient lived from their PCP's office did not significantly affect LDL control. There was no clear association between type of insurance and LDL control.

## Discussion

This study explored the potential of applying GIS methodology to analyze data from a chronic disease registry in a large health system. We had 2 main purposes with this project: (1) to investigate the relationship of SES and disease control within our health system, and (2) to explore the future opportunities and limitations of using GIS analysis with our diabetes registry.

Neighborhood SES was associated with A1c levels, with lower income neighborhoods having higher A1c, which indicating less controlled DM. This was true even though individual SES was not associated with A1c levels. This suggests that neighborhoods could effect diabetes control despite the SES of the individual. The association between SES and glucose control was small, with a contribution to the total variance in A1c on the order of 9%. Glucose control was also significantly associated with LDL control. However, there was not a significant correlation between SES and LDL control. The data suggest that SES was a barrier to better glucose control but not a barrier to lipid control. Although there have been numerous studies that show disparities related to ethnicity,[3,10–13] there is less evidence for an association between SES and diabetes control. The Translating Research Into Action for Diabetes study was a multicenter analysis of quality of care for adults with DM who were enrolled in managed-care health plans. The outcomes of the Translating Research Into Action for Diabetes study showed no significant differences in glucose or lipid control across socioeconomic vari-

**Table 2. Regression Model 1: Adjusted Relationships between Individual and Practice Characteristics and Glycated Hemoglobin\***

| | Parameter Estimate | 95% CI |
|---|---|---|
| Patient characteristics | | |
| Female | 0.24 | 0.13 to 0.36 |
| Age (years)† | — | — |
| 25–40 | | |
| 41–55 | −0.02 | −0.71 to 0.67 |
| 56–64 | −0.04 | −0.4173 to 0.65 |
| 65–74 | −0.25 | −0.95 to 0.45 |
| >74 | −0.47 | −1.17 to 0.23 |
| Race/ethnicity | | |
| White† | — | — |
| Black | 0.33 | 0.12 to 0.54 |
| Hispanic | 0.18 | −0.04 to 0.39 |
| Asian | 0.14 | −0.15 to 0.44 |
| Other | 0.16 | 0.02 to 0.30 |
| Primary language | | |
| English† | — | — |
| Not English | 0.04 | −0.24 to 0.33 |
| LDL cholesterol | 0.007 | 0.005 to 0.008 |
| Urine microalbumin/creatinine | 0.0002 | 0.0001 to 0.0003 |
| Insurance type | | |
| Fee for service† | — | — |
| HMO | −0.29 | −0.77 to 0.20 |
| MediCal | −0.11 | −0.67 to 0.46 |
| Medicare | −0.34 | −0.84 to 0.15 |
| Other | −0.95 | −1.74 to −0.17 |
| PPO | −0.26 | −0.77 to 0.26 |
| Self pay | −0.75 | −1.58 to 0.84 |
| Workman's compensation | −2.37 | −3.62 to −1.12 |
| Distance to clinic | −0.005 | −0.015 to 0.005 |
| Practice-related characteristics | | |
| PCN (vs medical center) | −0.37 | −0.54 to −0.20 |
| Specialty | | |
| Family practice† | — | — |
| Internal medicine | −0.10 | −0.25 to 0.06 |
| Family practice/internal medicine Combined | 0.16 | −0.02 to 0.33 |

\*Census tract number and socioeconomic status were included in the model as random effects.
†Reference.
LDL, low-density lipoprotein; HMO, health management organization; PPO, preferred provider organization; PCN, primary care network.

ables.[14] Harris et al[15] analyzed the National Health and Nutrition Examination Survey III data and found that race had a significant relationship with glycemic control but SES did not. Our data suggest that some communities may have socioeconomic barriers to DM control not seen in larger studies. Many health systems, including our own, are putting additional resources into chronic disease management. Socioeconomic barriers that affect quality of care will need to be factored into strategies for disease management. Like most health systems, SES is not available as part of patient demographics in our electronic medical records and disease registry. Unless special efforts are taken to include this information in the electronic medical records, health systems will need to use GIS methodologies to obtain this information about their practice population.

**Table 3. Regression Model 2***

| | Odds Ratio (95% CI) |
|---|---|
| Patient characteristics | |
|   Female | 0.58 (0.52–0.67) |
|   Age (years)[†] | |
|     25–40 | — |
|     41–55 | 2.45 (1.15–5.19) |
|     56–64 | 1.96 (0.92–4.16) |
|     65–74 | 1.64 (0.76–3.51) |
|     >74 | 1.34 (0.62–2.88) |
|   Race/ethnicity | |
|     White[†] | — |
|     Black | 1.18 (0.94–1.47) |
|     Hispanic | 1.22 (0.96–1.54) |
|     Asian | 1.13 (0.81–1.56) |
|     Other | 1.05 (0.90–1.22) |
|   Primary language | |
|     English[†] | — |
|     Not English | 0.88 (0.65–1.21) |
|   Hemoglobin A1c | 1.07 (1.04–1.11) |
|   Urine microalbumin/creatinine | 1.00 (1.00–1.00) |
|   Distance to clinic | 0.10 (0.99–1.01) |
|   Prescribed a statin | 0.52 (0.45–0.60) |
|   Insurance type | |
|     Fee for service[†] | — |
|     HMO | 0.90 (0.54–1.51) |
|     MediCal | 0.52 (0.29–0.96) |
|     Medicare | 0.70 (0.41–1.19) |
|     Other | 0.84 (0.36–1.98) |
|     PPO | 0.81 (0.47–1.40) |
|     Self pay | 1.03 (0.42–2.50) |
|     Workman's compensation | 0.09 (0.01–0.80) |
| Practice-related characteristics | |
|   PCN (vs medical center) | 1.18 (0.99–1.41) |
|   Specialty | |
|     Family practice[†] | — |
|     Internal medicine | 0.68 (0.57–0.80) |
|     Family practice/internal medicine combined | 0.79 (0.66–0.94) |

*Low-density lipoprotein (LDL) cholesterol level was logistically regressed on individual patient characteristics, including whether they have been prescribed a statin and their primary care physician's practice characteristics. Census tract number and socioeconomic status were included as random effects in the model. The odds ratio represents the probability of LDL cholesterol being >100.

[†]Reference.

HMO, health management organization; PPO, preferred provider organization; PCN, primary care network.

Our study also explored the association between a patient's home address and his or her PCP's office. Previous work indicated that patients are less likely to be taking insulin as the distance to their PCP increases.[16] We also analyzed 2 types of offices in this study: academic practice (family medicine and internal medicine) versus our nonteaching primary care offices. Patients seen in the academic practices had both higher A1c and LDL cholesterol. We also found that patients who came to the academic practice traveled significantly greater distances to see their PCP. We considered the possibility that this may be because of a preference for specialized care available at the university or that patients who had Medicaid insurance may be more likely to be treated in the academic practice. We did not find a significant effect of insurance type on the distance driven for care. Nevertheless, the distance patients live from their PCP may still be an important factor for health systems to consider. This information may help them plan for the most appropriate location for educational services and opportunities for practice growth (Figure 3).

This study has a number of limitations. Of 9722 patients identified in the registry, 2434 patients were excluded from the analysis, in large part because of a lack of an exact match for the geocoding database. We used Euclidian distance ("as the crow flies") instead of driving distance or driving time analysis. Driving time analysis is probably a more important measure for most primary care applications of GIS. Our factor analysis for SES did not include all races. Our analysis only explains a small amount of the variance in DM control. This is not necessarily surprising. Our study had a limited array of variables to examine. We also know that DM has significant intrinsic biologic variability. PCPs frequently encounter medication failure in this group as well as difficulty in managing the behavioral aspects of glucose control.

This study also provides important lessons learned for incorporating GIS into primary care research:

1. *Better systems are needed for gathering patient addresses and achieving an "exact match."* A significant number of patients in our study were excluded because of the difficulty in establishing a precise geocode for their address, ie, address abbreviations, post office boxes, misspellings, etc. Registration systems and electronic medical records could be enabled to create an exact match with existing GIS/address databases. Establishing this directly from the patient would improve data reliability.

2. *The enumeration unit limits the types of conclusions you can make.* Our study used the census

tract because there were too few numbers in some census blocks to make comparisons. The census block has less heterogeneity for data like SES and may therefore provide a better proxy for the patient's actual SES.

3. *Availability of better population data are needed.* We used 2008 sociodemographic data obtained from GeoLytics, Inc. These data are created using information from the US Census Bureau with additional estimates and projections. Although not as accurate as the Census, it provides a viable option for obtaining current census-tract level data that is so important to health research. Although the American Community Survey, a product of the US Census Bureau, provides annual population data, it is not currently available for the smaller enumeration units needed in health research.

4. *Incorporate measurements that make the data more useful.* Using Euclidian distance is a crude method for estimating driving distance. If one is using GIS methods to create a strategic plan for disease management, driving time analysis would be more meaningful to leadership. Other analyses could include data that would be helpful to public health officials, like the distance to grocery stores or parks.

## Conclusion

GIS methodology is an important tool for primary care and can provide guidance for disease management programs. GIS can also help health systems in their mission to improve the health of a community. In our analysis, socioeconomic factors made a difference in glucose control for patients who received care in our health system. This research method should be applied to the large volumes of data now available from different health systems' electronic medical records.

## References

1. Centers for Disease Control and Prevention. Diabetes Data & Trends. Number (in millions) of civilian/noninstitutionalized persons with diagnosed diabetes, United States, 1980–2006. Available at http://www.cdc.gov/diabetes/statistics/prev/national/figpersons.htm. Accessed 3 February 2009.

2. Miech RA, Kim J, McConnell C, Hamman RF. A growing disparity in diabetes-related mortality. U.S. trends, 1989–2005. Am J Prev Med 2009;36:126–32.

3. Kirk JK, Passmore LV, Bell RA, et al. Disparities in A1c levels between Hispanic and non-Hispanic white adults with diabetes: a meta-analysis. Diabetes Care 2008;31:240–6.

4. Peek ME, Cargill A, Huang ES. Diabetes health disparities: a systematic review of health care interventions. Med Care Res Rev 2007;64(5 Suppl):101S–56S.

5. Improving Chronic Illness Care. The Chronic Care Model. Available at http://www.improvingchroniccare.org/index.php?p=Model_Elements&s=18. Accessed 6 February 2009.

6. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002;288:1775–9.

7. Choi M, Afzal B, Sattler B. Geographic information systems: a new tool for environmental health assessments. Public Health Nurs 2006;23:381–91.

8. Diez Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med 2001;345:99–106.

9. American Diabetes Association. Standards of medical care in diabetes–2009. Diabetes Care 2009;32(1 Suppl):S13–61.

10. Wendel CS, Shah JH, Duckworth WC, Hoffman RM, Mohler MJ, Murata GH. Racial and ethnic disparities in the control of cardiovascular disease risk factors in Southwest American veterans with type 2 diabetes: the Diabetes Outcomes in Veterans Study. BMC Health Serv Res 2006;6:58–65.

11. Sequist TD, Adams A, Zhang F, Ross-Degnan D, Ayanian JZ. Effect of quality improvement on racial disparities in diabetes care. Arch Intern Med 2006;166:675–81.

12. Misra R, Lager J. Ethnic and gender differences in psychosocial factors, glycemic control, and quality of life among adult type 2 diabetic patients. J Diabetes Complications 2009;23:54–64.

13. Heisler M, Smith DM, Hayward RA, Krein DL, Kerr EA. Racial disparities in diabetes care processes, outcomes, and treatment intensity. Medical Care 2003;41:1221–32.

14. Brown A, Gregg EW, Stevens MR, et al. Race, ethnicity, socioeconomic positions, and quality of care for adults with diabetes enrolled in managed care: the Translating Research Into Action for Diabetes (TRIAD) study. Diabetes Care 2005;28:2864–70.

15. Harris M, Eastman RC, Cowie CC, Flegal KM, Eberhardt MS. Racial and ethnic differences in glycemic control of adults with type 2 diabetes. Diabetes Care 1999;22:403–8.

16. Littenberg B, Strauss K, MacLean CD, Troy AR. The use of insulin declines as patients live farther from their source of care: results of a survey of adults with type 2 diabetes. BMC Public Health 2006;6:198.