

Determining Chronic Disease Prevalence in Local Populations Using Emergency Department Surveillance

David C. Lee, MD, MS, Judith A. Long, MD, Stephen P. Wall, MD, MS, Brendan G. Carr, MD, MS, Samantha N. Satchell, BA, R. Scott Braithwaite, MD, MS, and Brian Elbel, PhD, MPH

In its 2012 report on measures for population health, the Institute of Medicine prioritized understanding local population health to improve health care for populations with the highest need.¹ Generally, health care providers have used the term “population health” when referring to patients linked to a specific health care provider or insurance group.² However, the discipline of public health more broadly defines population health as the health of all individuals living in specific geographic regions.³

To estimate disease burden, traditional methods include performing population-based telephone health surveys.⁴ Unless large numbers of individuals are surveyed, it is difficult to determine prevalence in small geographic areas such as census tracts, and yearly estimates have significant noise because of small sample sizes.⁵ Low response rates can lead to errors in estimating disease prevalence, and larger surveys can be costly and difficult to perform.⁶

With increasing use of big data in the form of large administrative data sets with clinical data,⁷ there is an opportunity to create more precise measures of population health by reducing the variance associated with small sample sizes.⁸⁻¹⁰ These methods may be biased as they only track individuals who register a medical claim, which makes for a type of convenience sample. Nevertheless, a significant proportion of all individuals, regardless of insurance type, interact with the health care system, especially through emergency services. Nearly 1 in 5 individuals report having gone to an emergency department (ED) in the past year.¹¹ Previous studies have demonstrated the promise of using emergency claims data for tracking acute illnesses; however, there is potential to extend these methods to the surveillance of chronic disease.^{12,13} One of the advantages of using administrative claims data is the achievement of large sample sizes without the need to conduct large surveys.^{14,15}

Objectives. We sought to improve public health surveillance by using a geographic analysis of emergency department (ED) visits to determine local chronic disease prevalence.

Methods. Using an all-payer administrative database, we determined the proportion of unique ED patients with diabetes, hypertension, or asthma. We compared these rates to those determined by the New York City Community Health Survey. For diabetes prevalence, we also analyzed the fidelity of longitudinal estimates using logistic regression and determined disease burden within census tracts using geocoded addresses.

Results. We identified 4.4 million unique New York City adults visiting an ED between 2009 and 2012. When we compared our emergency sample to survey data, rates of neighborhood diabetes, hypertension, and asthma prevalence were similar (correlation coefficient = 0.86, 0.88, and 0.77, respectively). In addition, our method demonstrated less year-to-year scatter and identified significant variation of disease burden within neighborhoods among census tracts.

Conclusions. Our method for determining chronic disease prevalence correlates with a validated health survey and may have higher reliability over time and greater granularity at a local level. Our findings can improve public health surveillance by identifying local variation of disease prevalence. (*Am J Public Health*. 2015;105:e67–e74. doi:10.2105/AJPH.2015.302679)

In this study, we have introduced a novel geographic method of public health surveillance and determined whether we could use ED administrative claims to estimate chronic disease prevalence at a local level over time. As the ED is generally a place where all individuals can access care regardless of socioeconomic or insurance status, it offers an ideal environment for public health surveillance among all types of individuals within a heterogeneous population.¹⁶

METHODS

To determine whether we could use ED data for chronic disease surveillance, we identified the proportion of unique ED patients who had a diagnosis of diabetes, hypertension, or asthma coded within an administrative claims database in New York City from 2009 to 2012. We then compared these rates of chronic disease prevalence with estimates from the New York City Community Health Survey

(CHS), which is used by the city’s Department of Health and Mental Hygiene (DOHMH) to estimate citywide and neighborhood-level disease prevalence.¹⁷ For diabetes prevalence, we also compared the performance of our surveillance method over time and determined whether it could be used to identify prevalence at the census tract level.

Data Sources

New York City Community Health Survey. The New York City CHS has been conducted annually by the city’s DOHMH since 2002. The CHS uses a telephone-based stratified random sample to track chronic diseases and health behaviors.¹⁸ Previous studies have shown that the rates of self-reported disease prevalence from the CHS compare within reason with estimates obtained by methods also employing physical examinations and laboratory testing; however, self-report can underestimate prevalence because of undiagnosed disease burden.^{19,20} To estimate

diabetes, hypertension, and asthma prevalence, the survey asks whether respondents have ever been told by a doctor, nurse, or other health professional that they had these conditions. The survey samples New York City residents from each of the United Hospital Fund's (UHF) 34 neighborhoods, which are groupings of zip codes that match local communities with a distinct identity and mirror New York City community planning districts.²¹ For the New York City CHS, we used both multiyear and single-year survey sampling weights provided by the city's DOHMH to determine chronic disease prevalence by UHF neighborhood for the overall study period and for longitudinal estimates.

New York State SPARCS database. Within the New York State Department of Health, the Statewide Planning and Research Cooperative System (SPARCS) was established as a comprehensive all-payer data reporting system to collect data on patients' characteristics, diagnoses, treatment and services for all hospital discharges, ambulatory surgery procedures, and ED visits in New York State.²² Hospitals are required to report 95% of all data within 60 days of a patient's discharge or visit, or they can be subjected to reimbursement penalties. SPARCS ensures that nearly all hospital facilities meet these standards before releasing data to researchers. The database also contains encrypted unique identifiers for tracking specific individuals and patient-level address information, which can be used to locate the exact residence of an individual.

American Community Survey. To compare study populations from SPARCS emergency department data and the New York City CHS, we used census data from the American Community Survey (ACS). We used a mean of annual data over the study period to determine the distribution of age, gender, race, ethnicity, and insurance status of city residents.

Participants

The CHS includes noninstitutionalized adults aged 18 years and older living in New York City households reachable by a landline telephone; since 2009, it has also included households with only cell phone access.¹⁸ The survey excludes the approximately 3% of city residents living in nonresidential group quarters (i.e., nursing homes, college dormitories, and correctional facilities).¹⁸ To match these

criteria, we analyzed data from ED patients aged 18 years and older, excluding those whose claims were paid by a correctional facility and those transferred from a nursing home or other health care facility.

Main Outcome

The primary outcome of our study was the correlation of neighborhood, age-adjusted chronic disease prevalence between our method, which used ED surveillance, and the New York City CHS. We used the unique identifiers within SPARCS to identify visits by the same individual throughout the study period. We then determined what proportion of these unique ED patients had ever received a diagnosis of diabetes, hypertension, or asthma during any ED visit, including ED admissions.¹⁵ We identified these diagnoses as a primary or any secondary *International Classification of Diseases, Ninth Revision (ICD-9)*²³ code starting with 250, 401 through 405, or 493. We then divided the number of unique ED patients who carried each of these diagnoses by the total number of unique ED patients.

Previous studies have suggested that, to specify cases of chronic disease when outpatient administrative data are used, a repeat diagnosis at a separate visit is necessary²⁴; however, other studies have suggested that this requirement is unnecessary.¹⁵ We chose not to include this limitation since requiring a repeat diagnosis in 2 separate ED visits would effectively exclude individuals who had only 1 ED visit during the study period. We also chose to include individuals with a single identification of chronic disease within administrative data since the New York City CHS had asked respondents whether they had ever been told that they had these conditions. To match the CHS design, we also determined chronic disease prevalence by UHF neighborhood using zip codes and age-adjusted crude rates of chronic disease, using the direct method previously described by the Centers for Disease Control and Prevention.²⁵

Statistical Analyses

To determine similarity of prevalence estimates between the 2 methods, we analyzed the correlation between age-adjusted diabetes, hypertension, and asthma rates by UHF neighborhoods. We calculated the correlation

coefficient between estimates using CHS and ED surveillance for each disease.

For diabetes, we also analyzed year-to-year trends among UHF neighborhoods based on the CHS and ED surveillance. Using previously described methods, we analyzed data at the patient-level using logistic regression to determine whether there were significant increases in the age-adjusted prevalence of diabetes over time.²⁶ In this logistic regression, we regressed diabetes status against the year of ED visit or CHS after adjusting the frequency of observations by age strata (logit [diabetes status] = $\beta_0 + \beta_1 \times \text{year}$). We performed this regression separately for each UHF neighborhood. To account for multiple comparisons among 34 UHF neighborhoods, we report only *P* values of .0015 or less as statistically significant.

As SPARCS has patient-level address information, we also studied diabetes prevalence at the census tract level. To do this, we geocoded patient addresses to determine the census tract in which patients resided. We excluded census tracts with a total population count of zero (e.g., nonresidential areas). We mapped overall diabetes prevalence for the 4-year study period and then analyzed year-to-year trends in prevalence again using logistic regression separately for each census tract.

We performed statistical analyses with Stata version 12.1 (StataCorp LP, College Station, TX). We performed geographic analysis using ArcGIS Desktop 10.1 (ESRI, Redlands, CA). We geocoded addresses using the ArcGIS 10 style North American Address Locator with a default spelling sensitivity of 80% and minimum match score of 75% (TomTom, Amsterdam, Netherlands).

RESULTS

We identified nearly 4.4 million unique adults who visited an ED at least once in New York State between 2009 and 2012 and who had an address located in New York City. This was a substantial proportion of the estimated 6.5 million adults in New York City. Overall, the population characteristics of both the survey-weighted New York City CHS data and the unweighted SPARCS emergency department data were generally similar to characteristics found in the ACS data (Table 1).

However, the SPARCS emergency department population had fewer individuals with their race coded as White (25.6% in SPARCS, vs 35.6% in CHS and 33.6% in ACS) and a higher proportion of individuals with race coded as other (15.0% in SPARCS, vs 2.1% in CHS and 2.5% in ACS). In addition, the SPARCS emergency department population had a higher proportion of individuals who were uninsured (25.0% in SPARCS, vs 18.8% in CHS and 16.4% in ACS) or had public health insurance (44.0% in SPARCS, vs 32.2% in CHS and 33.5% in ACS).

Main Results

The overall, citywide, age-adjusted prevalences of diabetes, hypertension, and asthma were 10.0% (95% confidence interval [CI]=9.6%, 10.5%), 27.9% (95% CI=27.3%, 28.5%), and 11.7% (95% CI=11.2%, 12.2%), respectively, based on the CHS and 11.0% (95% CI=11.0%, 11.1%), 23.5% (95% CI=23.5%, 23.5%), and 7.6% (95%

CI=7.5%, 7.6%) based on SPARCS emergency department data. Rates based on ED data were highly correlated by neighborhood to those estimated using the CHS, with correlation coefficients of 0.86 for diabetes, 0.88 for hypertension, and 0.77 asthma (Figure 1). When we mapped the results of the 2 methods, age-adjusted diabetes rates were similarly distributed among New York City neighborhoods (Figure A, available as a supplement to the online version of this article at <http://www.ajph.org>).

Longitudinal Comparisons

Comparing year-to-year diabetes prevalence, we found large swings in estimates based on the CHS. Annual prevalence by UHF neighborhood changed on average by 3.5 percentage points and up to 10.1 percentage points a year (Figure 2a). Ninety-five-percent confidence intervals around annual neighborhood diabetes prevalence based on the CHS ranged from 2.5% to 21.4% in width

(mean = 9.3%). By contrast, we found that the year-to-year estimates of diabetes prevalence based on SPARCS emergency department data were more stable and exhibited less variation (Figure 2b). Ninety-five-percent confidence intervals around annual neighborhood diabetes prevalence based on SPARCS emergency department data ranged from 0.3% to 1.0% in width (mean = 0.6%). Using SPARCS emergency department data, we also observed statistically significant increases in diabetes prevalence in 27 of the 34 UHF neighborhoods, which ranged from an absolute increase of 0.2 to 1.1 percentage points a year (Table A, available as a supplement to the online version of this article at <http://www.ajph.org>).

Geographic Specificity

At the census tract level, we found marked variability within UHF neighborhoods (Figures 3 and 4). Each UHF neighborhood contained anywhere from 31 to 145 census tracts. In some cases, only a few census tracts were responsible for very high diabetes prevalence within a neighborhood. On an annual basis, there was a mean of 585 unique ED patients per census tract, and only 1% of estimates were based on fewer than 100 observations. From the logistic regression analysis, we found that annual diabetes prevalence increased by a mean of 0.5 percentage points a year among the census tracts. Ninety-five percent of changes in prevalence ranged from -0.6 to 1.7 percentage points a year. Twenty-two percent of census tracts demonstrated significant changes in prevalence, which was a higher percentage than would be expected by chance assuming a *P* value of .05.

DISCUSSION

Based on surveillance of administrative emergency claims data, our novel geographic method for determining chronic disease prevalence correlated well to an established standard for measuring population health. By using ED data, we analyzed the segment of the health care system that, for the most part, can be accessed by all individuals regardless of socioeconomic or health insurance status.¹⁶ In this way, the ED offers a means for surveying the health of heterogeneous populations with varied types of health insurance and differential

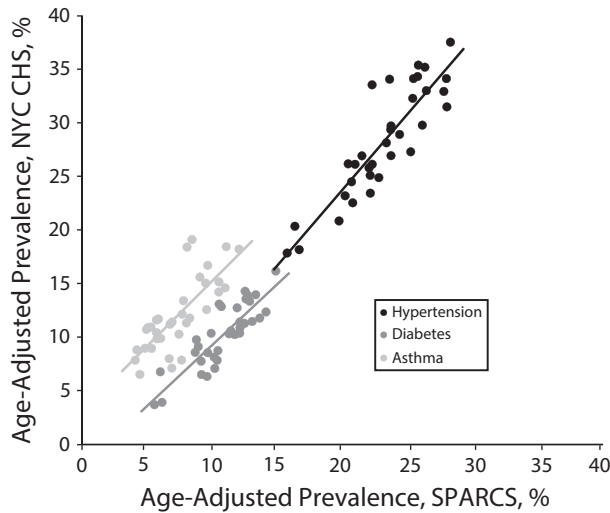
TABLE 1—Characteristics of Study Populations: New York City, 2009–2012

Population Characteristics	ACS Population Estimates ^a (n = 6 485 488), %	Community Health Survey ^b (n = 36 188), %	SPARCS Emergency Data (n = 4 392 152), %
Age, y			
18–24	12.9	13.0	16.5
25–44	40.0	39.6	39.6
45–64	31.4	30.7	27.6
≥ 65	15.7	16.7	16.3
Gender			
Male	46.7	46.3	44.4
Female	53.3	53.7	55.6
Race/ethnicity			
White	33.6	35.6	25.6
Black	22.8	22.0	29.6
Hispanic	28.5	26.8	23.4
Asian	12.6	13.5	6.4
Other	2.5	2.1	15.0
Insurance			
Private	50.1	49.0	31.0
Public	33.5	32.2	44.0
Uninsured	16.4	18.8	25.0

Source. American Community Survey (ACS), New York City Community Health Survey, and New York Statewide Planning and Research Cooperative System (SPARCS).

^aRace/ethnicity distributions for the ACS were calculated on entire population as full breakdown for adults only was not available.

^bWe weighted the study population from the New York City Community Health Survey using survey weights supplied by the New York City Department of Health and Mental Hygiene.



Note. CHS = Community Health Survey; NYC = New York City; SPARCS = Statewide Planning and Research Cooperative System.

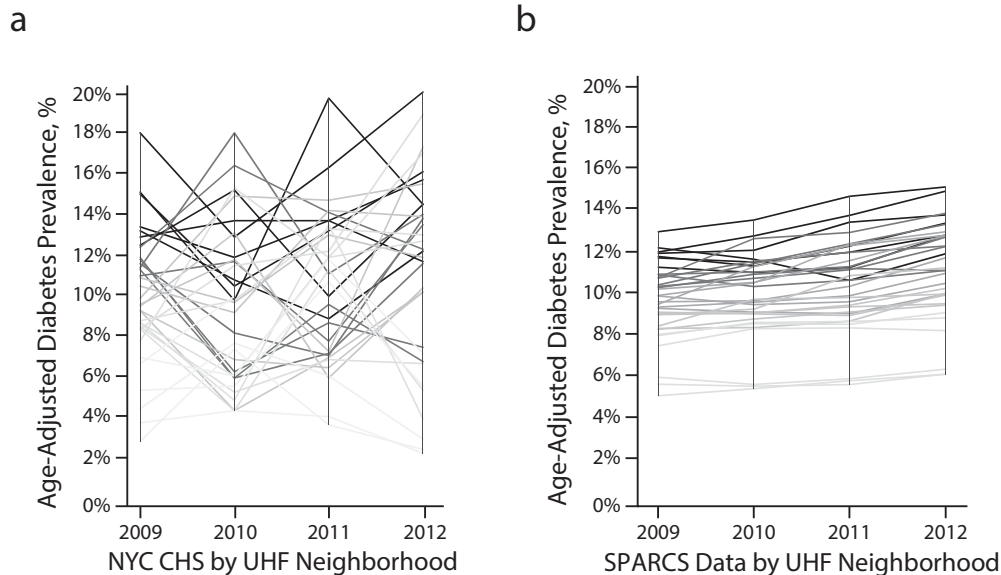
FIGURE 1—Correlation of age-adjusted chronic disease prevalence, by neighborhood: New York City, 2009–2012.

access to care.^{27,28} Comparing our method with the New York City CHS, we found that rates of diabetes, hypertension, and asthma were generally similar when analyzed by New York City neighborhood and citywide.

Previous studies have demonstrated the validity of using outpatient and inpatient administrative claims data to identify chronic disease for specific individuals within a given insurance group (Medicare patients¹⁴) or in

single-payer countries (Canada²⁴). However, in regions where there are marked differences in health care accessibility or limitations in data availability, using ED surveillance may better match disease prevalence at a population level. In addition to disease prevalence, ED data may also be useful in analyzing severity of disease (i.e., frequency of high-acuity emergency admissions), which may help further characterize disease burden in specific geographic regions. Although we only analyzed the data of individuals who registered an ED claim, nearly 1 in 5 individuals report visiting an ED in a given year.¹¹ In our 4-year study period, we found that our analysis captured a significant proportion of all estimated adults living in New York City.

Examining the demographics of our study population, we found that data for our ED population was surprisingly close to available census data in terms of age and gender. In general, ED patients are thought of as critically ill patients who are older and have a higher prevalence of chronic disease; however, these individuals may represent the minority of patients who visit the ED more frequently. Our study actually adjusted for frequent ED use



Note. SPARCS = Statewide Planning and Research Cooperative System; UHF = United Hospital Fund. The figure compares the variation in year-to-year estimates of diabetes prevalence for each of the 34 UHF neighborhoods. The shading for each neighborhood trend line is based on the quintile of diabetes prevalence in the starting year. Darker lines represent neighborhoods with higher initial diabetes prevalence, and lighter lines represent neighborhoods with lower initial diabetes prevalence.

FIGURE 2—Comparison of age-adjusted diabetes prevalence over time and by neighborhood based on (a) the New York City Community Health Survey (NYC CHS) and (b) SPARCS emergency department (ED) surveillance: New York City, 2009–2012.

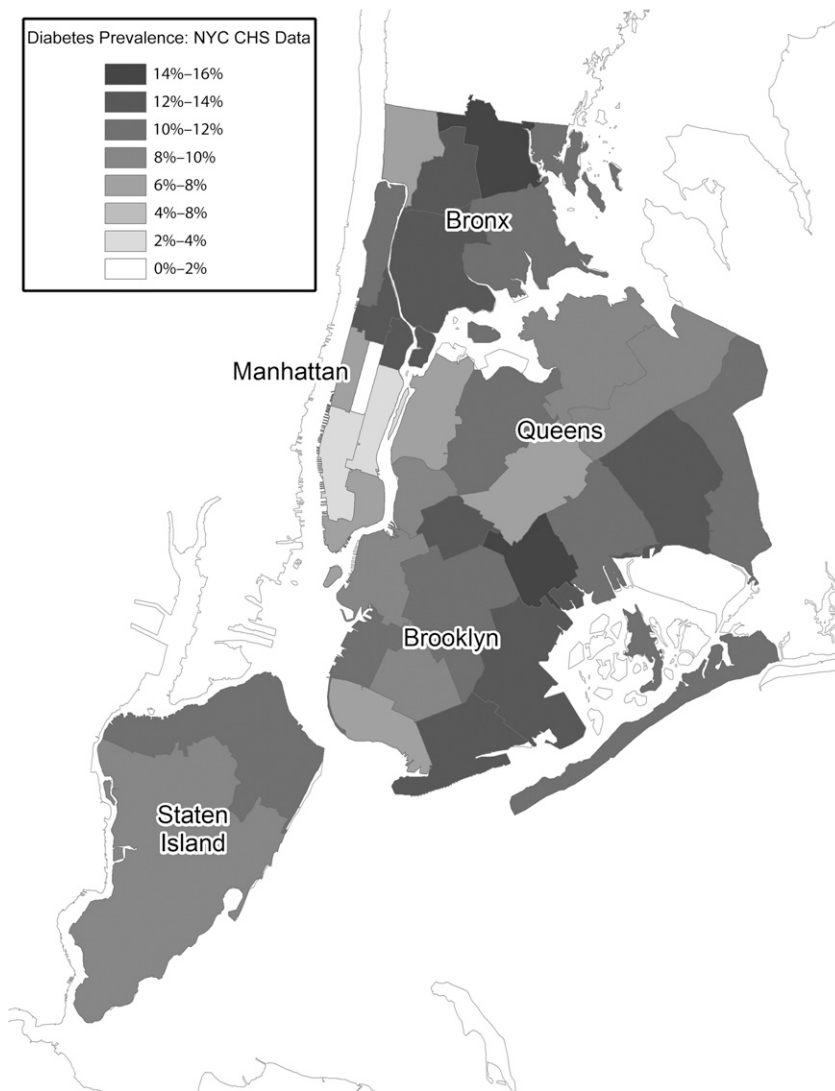


FIGURE 3—Age-adjusted diabetes prevalence based on the New York City Community Health Survey (NYC CHS) by United Hospital Fund neighborhood: New York City 2009–2012.

by evaluating only unique ED users, which may be why our sample population was closer to census data than expected. In addition, we also used secondary diagnoses to identify patients with diabetes, hypertension, or asthma; thus, a patient could have reported having 1 of these chronic conditions during an unrelated low-acuity visit, which are also common among the ED population.

However, a known limitation of SPARCS emergency department data is systematic error in racial classifications.²⁹ Some hospitals are not consistent in coding variables, and their

rate of coding race as “other” is erroneously high.²⁹ Well-validated health surveys, such as the New York City CHS, have advantages over administrative claims surveillance because they can be more reliable at coding these demographic variables. They can also collect more specific information on socioeconomic and health information, especially behavioral data, which would not be available in large administrative claims databases.⁵

Because we used an existing large administrative database of ED claims, we determined chronic disease prevalence based on a larger

sample of individuals than traditional telephone-based health surveys.⁵ Thus, our method may create more stable longitudinal estimates of chronic disease burden compared with traditional methods, which demonstrated significant year-to-year variation in our analysis of diabetes prevalence.³⁰ We confirmed these findings with longitudinal analyses of hypertension and asthma rates as well (data not shown). Thus, our method may represent a better estimate of the true underlying disease prevalence in local populations and provide a more reliable indicator of changes over time.

In addition, our method allows for analyses of disease prevalence in small geographic units over time, which previously was virtually impossible because of the sample size needed.^{5,31} As an example, we have analyzed diabetes prevalence at the census tract level, demonstrating that at a local level, there was a sufficient number of observations to reliably estimate diabetes prevalence in very small geographic regions.^{31–33} We also found that diabetes prevalence can differ markedly within large neighborhoods and can be localized to specific areas within them. Identifying disease prevalence in smaller areas would also allow public health researchers and institutions to use more advanced geospatial models with methods involving spatial autocorrelation that take advantage of neighboring estimates to improve local hot-spotting of disease burden.

With this patient-level address data and analysis, we have greater granularity to identify hyperlocal clusters of disease prevalence,³¹ which can improve efforts to target and tailor care for specific communities with poor health.³³ Although not all large administrative claims databases carry this level of information, even zip code analyses would be closer to local populations than counties, large neighborhoods, and traditionally studied hospital referral regions and hospital service areas.^{34,35} For instance, in New York City, all of Manhattan is considered its own hospital referral region and hospital service area. Our analysis characterizes geographic variation in population health at a much lower level.

Limitations

Certain limitations to our methods exist. First, administrative claims data can contain coding errors by the institutions providing

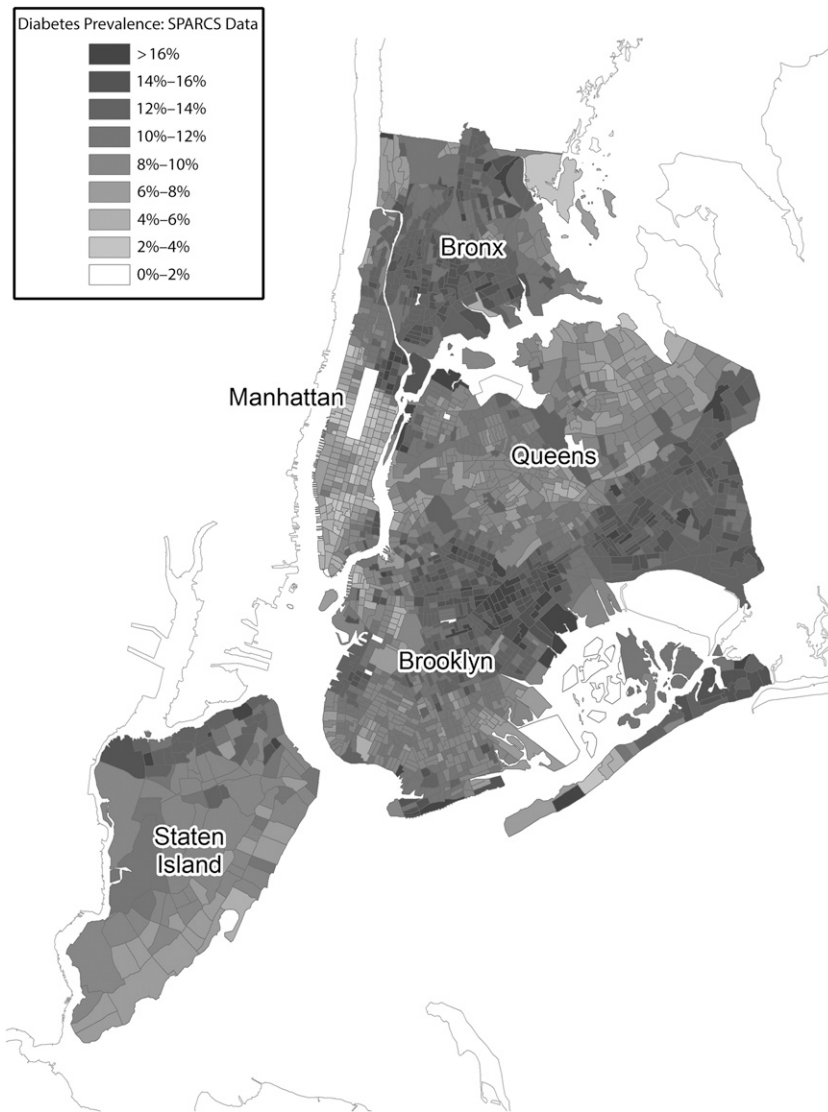


FIGURE 4—Age-adjusted diabetes prevalence based on Statewide Planning and Research Cooperative System (SPARCS) emergency department (ED) surveillance by census tract: New York City, 2009–2012.

data.²⁹ These errors can lead to double counting of individuals, counting nonresidents as residents, and difficulty with mapping of individuals by address. In fact, 8% of ED visits were not geocodable to a specific census tract, largely because of inaccuracies in address information. If patients with diabetes were more or less likely to be geocoded than patients without diabetes, these missing observations may have introduced bias to the analysis of prevalence by census tract. In addition, identification of

individuals with chronic disease was based on diagnosis codes, the accuracy of which depends on the fidelity of recording these diagnoses.

Second, because we studied individuals who registered an ED claim, there was sampling bias that skewed observations toward patients more likely to come to an ED for care, especially those with public insurance.³⁶ Because our study was observational, there may have been unmeasured trends that accounted for increases in diabetes prevalence found among ED patients. Changes

may have been attributable to shifts in the mix of patients presenting to EDs, rather than actual changes in prevalence among the population.

Third, the standard by which we validated our method for determining chronic disease prevalence is a population-based health survey, which relies on self-report and may not represent the actual disease burden in a given population. Both methods of surveillance may not adequately address the silent burden of undiagnosed chronic disease, as nearly a quarter of New York City residents may not be aware that they have chronic diseases such as diabetes or hypertension.

Fourth, our study was also limited to New York City, a unique and dense urban environment. Access to ED care in New York City is not greatly limited by distance compared with more suburban and rural regions. As previous studies have shown, farther distances can mean markedly lower rates of ED use.³⁷ Thus, we caution that our method may be limited to analyses of urban regions, and adjustments may be needed when considering this methodology for use in nonurban regions.³⁸

Conclusions

We present a novel geographic method for determining the prevalence of chronic disease in local populations based on ED surveillance. Further research should be done to validate these methods for regions outside of New York City as other states and countries can provide access to patient-level address data. Our approach could be especially useful in locales where population-based health surveillance is not regularly or systematically performed. Additional studies should expand these methods to other chronic diseases to further quantify population health.⁵ Expanding the range of indications that could be tracked longitudinally could be used to measure the impact of specific health interventions with higher resolution and greater frequency. It would also allow for monitoring of other chronic diseases not usually included in health surveys. Also, poststratification weighting may improve estimates of disease prevalence by matching the distribution of ED patients to the general population using available census data.³⁹ In our own attempts to poststratify this ED sample, we found some evidence of improved correlation with the New York City CHS (Table B, available

as a supplement to the online version of this article at <http://www.ajph.org>.

More importantly, now that variation in chronic disease prevalence can be characterized in small geographic areas,⁴⁰ further research should be done to determine what factors drive high versus low rates of disease in different communities.⁴¹ For example, we must identify what is unique about those regions that would be expected to have high diabetes prevalence based on socioeconomic factors,⁴² but that actually have lower-than-expected rates of diabetes.^{43,44} Doing so may help us identify protective and modifiable factors that can be used to perform targeted interventions in areas where chronic disease burden is overwhelming.^{45,46} ■

About the Authors

David C. Lee and Stephen P. Wall are with the Ronald O. Perleman Department of Emergency Medicine and R. Scott Braithwaite and Brian Elbel are with the Department of Population Health, New York University School of Medicine, New York, NY. Judith A. Long is with the Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA. Brendan G. Carr is with the Department of Emergency Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA. Samantha N. Satchell is with the Milken Institute School of Public Health, George Washington University, Washington, DC.

Correspondence should be sent to David C. Lee, MD, MS, Departments of Emergency Medicine and Population Health, 462 First Ave, Room A345, New York, NY 10016 (e-mail: david.lee@nyumc.org). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

This article was accepted March 8, 2015.

Contributors

D. C. Lee collected the data and performed the statistical analysis and takes full responsibility for their accuracy. D. C. Lee, J. A. Long, and B. Elbel conceptualized and designed the study and drafted the article. All of the authors analyzed and interpreted the results of the study and critically revised the article for important intellectual content.

Acknowledgments

This study was funded by 2 intramural grants from the University of Pennsylvania: a health services research grant from the Leonard Davis Institute of Health Economics and a population health research grant from the Robert Wood Johnson Foundation Health and Society Scholars Program. D. C. Lee was also supported by the fellowship and faculty mentorship at the Robert Wood Johnson Foundation Clinical Scholars Program.

Note. The study funders had no role in the design or conduct of the study; the collection, analysis, or interpretation of the data; or the preparation, review, or approval of the article; nor in the decision to submit the article for publication. B. G. Carr spends a portion of

his time as the director of the Emergency Care Coordination Center, Office of the Assistant Secretary for Preparedness & Response, US Department of Health and Human Services. The views represented here do not necessarily represent those of the US Government.

Human Participant Protection

This study was approved by the New York State Department of Health Data Protection Review Board and institutional review boards at the University of Pennsylvania and New York University Schools of Medicine.

References

1. Institute of Medicine. *Toward Quality Measures for Population Health and the Leading Health Indicators*. Washington, DC: National Academies Press; 2012.
2. Eggleston EM, Finkelstein JA. Finding the role of health care in population health. *JAMA*. 2014;311(8):797–798.
3. Jacobson DM, Teutsch S. An environmental scan of integrated approaches for defining and measuring total population health by the clinical care system, the government public health system, and stakeholder organizations. National Quality Forum, 2012. Available at: <http://www.improvingpopulationhealth.org/PopHealthPhaseICommissionedPaper.pdf>. Accessed April 13, 2015.
4. Centers for Disease Control and Prevention. Estimated county-level prevalence of diabetes and obesity—United States, 2007. *MMWR Morb Mortal Wkly Rep*. 2009;58(45):1259–1263.
5. Schlundt DG, Hargreaves MK, McClellan L. Geographic clustering of obesity, diabetes, and hypertension in Nashville, Tennessee. *J Ambul Care Manage*. 2006;29(2):125–132.
6. Boland M, Sweeney MR, Scallan E, Harrington M, Staines A. Emerging advantages and drawbacks of telephone surveying in public health research in Ireland and the UK. *BMC Public Health*. 2006;6:208.
7. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Aff (Millwood)*. 2014;33(7):1163–1170.
8. Buck MD, Anane S, Taverna J, Amirfar S, Stubbs-Dame R, Singer J. The Hub Population Health System: distributed ad hoc queries and alerts. *J Am Med Inform Assoc*. 2012;19(e1):e46–e50.
9. Kho ANJ, Hota BN, Sims SA, Malin BA, Galanter WL. The Chicago Health Atlas: a public resource to visualize health conditions and resources in Chicago. Presented at the American Medical Informatics Association Annual Symposium; November 3–7, 2012; Chicago, IL.
10. Lenert L, Sundwall DN. Public health surveillance and meaningful use regulations: a crisis of opportunity. *Am J Public Health*. 2012;102(3):e1–e7.
11. Centers for Disease Control and Prevention. *Health, United States, 2013: With Special Feature on Prescription Drugs*. Hyattsville, MD: National Center for Health Statistics; 2014.
12. Hope KG, Merritt TD, Durrheim DN, et al. Evaluating the utility of emergency department syndromic surveillance for a regional public health service. *Commun Dis Intell Q Rep*. 2010;34(3):310–318.

13. Fowlkes A, Dasgupta S, Chao E, et al. Estimating influenza incidence and rates of influenza-like illness in the outpatient setting. *Influenza Other Respir Viruses*. 2013;7(5):694–700.
14. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM. Identifying persons with diabetes using Medicare claims data. *Am J Med Qual*. 1999;14(6):270–277.
15. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care*. 1997;35(9):932–947.
16. Curtis AJ, Lee WA. Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. *Int J Health Geogr*. 2010;9:43.
17. McVeigh KH, Newton-Dame R, Perlman S, et al. *Developing an Electronic Health Record-Based Population Health Surveillance System*. New York, NY: New York City Dept of Health and Mental Hygiene; 2013.
18. Norton JM, Sanderson M, Gupta L, et al. Methodology Updates to the New York City Community Health Survey. Epi Research Report, New York City Department of Health and Mental Hygiene, September 2012. Available at: <http://www.nyc.gov/html/doh/downloads/pdf/epi/epiresearch-chsmethods.pdf>. Accessed April 6, 2015.
19. Yi SS, Johns M, Lim S. Use of regional data to validate and recalibrate self-reported hypertension: highlighting differences in immigrant groups in New York City. *J Immigr Minor Health*. 2015;Epub ahead of print.
20. Gwynn RC, Garg RK, Kerker BD, Frieden TR, Thorpe LE. Contributions of a local health examination survey to the surveillance of chronic and infectious diseases in New York City. *Am J Public Health*. 2009;99(1):152–159.
21. Buchholz N, Resnick S, Konty K. *The New York City Community Health Survey Atlas, 2010*. New York, NY: New York City Dept of Health and Mental Hygiene; 2012.
22. Quan JM. SPARCS: the New York State health care data system. *J Clin Comput*. 1980;8(6):255–263.
23. *International Classification of Diseases, Ninth Revision*. Geneva, Switzerland: World Health Organization; 1980.
24. Quan H, Khan N, Hemmelgarn BR, et al. Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54(6):1423–1428.
25. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected US population. *Healthy People 2010 Stat Notes*. 2001;(20):1–10.
26. Koo BK, Kim EK, Choi H, Park KS, Moon MK. Decreasing trends of the prevalence of diabetes and obesity in Korean women aged 30–59 years over the past decade: results from the Korean National Health and Nutrition Examination Survey, 2001–2010. *Diabetes Care*. 2013;36(7):e95–e96.
27. Hirshon JM. The rationale for developing public health surveillance systems based on emergency department data. *Acad Emerg Med*. 2000;7(12):1428–1432.
28. Gordon JA, Goldfrank LR, Andrulis DP, D'Alessandri RM, Kellermann AL. Emergency department initiatives to improve the public health. *Acad Emerg Med*. 1998;5(9):935–937.
29. Kim M, Berger D, Matte T. *Diabetes in New York City: Public Health Burden and Disparities*. New York, NY: New York City Dept of Health and Mental Hygiene; 2006.

30. Bilheimer LT. Evaluating metrics to improve population health. *Prev Chronic Dis*. 2010;7(4):A69.
31. Drewnowski A, Rehm CD, Moudon AV, Arterburn D. The geography of diabetes by census tract in a large sample of insured adults in King County, Washington, 2005–2006. *Prev Chronic Dis*. 2014;11:E125.
32. Liese AD, Lawson A, Song HR, et al. Evaluating geographic variation in type 1 and type 2 diabetes mellitus incidence in youth in four US regions. *Health Place*. 2010;16(3):547–556.
33. Tompkins JW, Luginaah IN, Booth GL, Harris SB. The geography of diabetes in London, Canada: the need for local level policy for prevention and management. *Int J Environ Res Public Health*. 2010;7(5):2407–2422.
34. Institute of Medicine. *Variation in Health Care Spending: Target Decision Making, Not Geography*. Washington, DC: National Academies Press; 2013.
35. Wrobel JS, Mayfield JA, Reiber GE. Geographic variation of lower-extremity major amputation in individuals with and without diabetes in the Medicare population. *Diabetes Care*. 2001;24(5):860–864.
36. Morganti KG, Bauhoff S, Blanchard JC, et al. *The Evolving Role of Emergency Departments in the United States*. Santa Monica, CA: RAND Corporation; 2013.
37. Lee JE, Sung JH, Ward WB, Fos PJ, Lee WJ, Kim JC. Utilization of the emergency room: impact of geographic distance. *Geospat Health*. 2007;1(2):243–253.
38. Andrus MR, Kelley KW, Murphey LM, Herndon KC. A comparison of diabetes care in rural and urban medical clinics in Alabama. *J Community Health*. 2004;29(1):29–44.
39. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. *Int J Forecast*. 2014;Epub ahead of print.
40. Staines A, Bodansky HJ, McKinney PA, et al. Small area variation in the incidence of childhood insulin-dependent diabetes mellitus in Yorkshire, UK: links with overcrowding and population density. *Int J Epidemiol*. 1997;26(6):1307–1313.
41. Leese GP, Feng Z, Leese RM, Dibben C, Emslie-Smith A. Impact of health-care accessibility and social deprivation on diabetes related foot disease. *Diabet Med*. 2013;30(4):484–490.
42. LaVeist TA, Thorpe RJ Jr, Galarraga JE, Bower KM, Gary-Webb TL. Environmental and socio-economic factors as contributors to racial disparities in diabetes prevalence. *J Gen Intern Med*. 2009;24(10):1144–1148.
43. Naumova EN. A cautionary note for population health: disproportionate emphasis on personal responsibility for health and wellbeing. *J Public Health Policy*. 2014;35(3):397–400.
44. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, Bilo H, Denig P. Review: relation between quality-of-care indicators for diabetes and patient outcomes: a systematic literature review. *Med Care Res Rev*. 2011;68(3):263–289.
45. Arday DR, Fleming BB, Keller DK, et al. Variation in diabetes care among states: do patient characteristics matter? *Diabetes Care*. 2002;25(12):2230–2237.
46. Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ. Geographic distribution of diagnosed diabetes in the US: a diabetes belt. *Am J Prev Med*. 2011;40(4):434–439.