POPULATION HEALTH MANAGEMENT
Volume 20, Number 6, 2017

DOI: 10.1089/pop.2016.0168

Original Articles

Assessing the Reliability of Performing Citywide Chronic Disease Surveillance Using Emergency Department Data from Sentinel Hospitals

David C. Lee, MD, MS,[1,2] Jordan L. Swartz, MD, MA,[1] Christian A. Koziatek, MD,[1] Andrew J. Vinson, MS,[1] Jessica K. Athens, PhD, MS,[2] and Stella S. Yi, PhD, MPH[2]

Abstract

Given the inequalities in the distribution of disease burden, geographically detailed methods of disease surveillance are needed to identify local hot spots of chronic disease. However, few data sources include the patient-level addresses needed to perform these studies. Given that individual hospitals would have access to this geographically granular data, this study assessed the reliability of estimating chronic disease prevalence using emergency department surveillance at specific hospitals. Neighborhood-level diabetes, hypertension, and asthma prevalence were estimated using emergency claims data from each individual hospital in New York City from 2009–2012. Estimates were compared to prevalence obtained from a traditional health survey. A multivariable analysis also was performed to identify which individual hospitals were more accurate at estimating citywide disease prevalence. Among 52 hospitals, variation was found in the accuracy of disease prevalence estimates using emergency department surveillance. Estimates at some hospitals, such as NYU Langone Medical Center, had strong correlations for all diseases studied (diabetes: 0.81, hypertension: 0.84, and asthma: 0.84). Hospitals with patient populations geographically distributed throughout New York City had better accuracy in estimating citywide disease prevalence. For diabetes and hypertension, hospitals with racial/ethnic patient distributions similar to Census estimates and higher fidelity of diagnosis coding also had more accurate prevalence estimates. This study demonstrated how citywide chronic disease surveillance can be performed using emergency data from specific sentinel hospitals. The findings may provide an alternative means of mapping chronic disease burden by using existing data, which may be critical in regions without resources for geographically detailed health surveillance.

Keywords: chronic disease prevalence, emergency department surveillance, geographic information systems, administrative claims data

Introduction

The concept of using individual hospitals for population health surveillance, also known as sentinel hospital surveillance, dates back to the 1990s with the HIV epidemic.[1] The Centers for Disease Control and Prevention selected specific hospitals in major US cities to track emerging infectious diseases in the general population.[2] Now that chronic diseases such as diabetes are reaching epidemic levels, more geographically detailed health surveillance is needed to identify the communities that are local hot spots of chronic disease. However, traditional health surveys such as the Behavioral Risk Factor Surveillance System and the National Health and Nutrition Examination Survey require intensive data collection, and the sample sizes yielded are unable to identify neighborhood-level geographic variation in health.[3,4]

It is critical to identify the exact geographic areas with a high burden of chronic disease and target these local neighborhoods with interventions tailored to the specific needs of these communities.[5] The study team recently demonstrated that emergency department (ED) claims data could be used to identify the local geographic distribution of chronic diseases such as diabetes, hypertension, and asthma.[6] Nearly 1 in 5 Americans visits an ED in a given year, and the team's method of

[1]Ronald O. Perelman Department of Emergency Medicine, New York University School of Medicine, New York, New York.
[2]Department of Population Health, New York University School of Medicine, New York, New York.

emergency department surveillance was effective at identifying neighborhood-level chronic disease prevalence when compared to a traditional health survey.[7] Using patient-level addresses and unique patient identifiers to account for repeat ED use, the study team identified local hot spots of chronic disease with extremely high geographic resolution.

Unfortunately, patient-level address data is not available in most federal and statewide claims databases.[8] In addition, regional health information exchanges rarely include this geographically detailed data when consolidating records across different health care institutions.[9] Though now electronically available, health records remain organized in a manner that does not facilitate data analyses of multiple health care institutions in an easy or efficient manner.[10] The present study's main objective was to assess whether ED data from individual hospitals could be used to track chronic disease prevalence at a neighborhood level in New York City (NYC). The secondary objective was to assess what factors would make for an ideal sentinel hospital to perform this citywide surveillance.

Methods

Study design and setting

To assess the reliability of using administrative claims data from specific hospitals to perform chronic disease surveillance, a recently validated method was used that analyzed ED data to identify neighborhood-level chronic disease prevalence in NYC. The method of ED surveillance demonstrated strong correlations with the NYC Community Health Survey (CHS), a traditional, telephone-based health survey. The present study assessed how accurately neighborhood-level chronic disease prevalence could be estimated throughout NYC when using retrospective ED claims data from each hospital in NYC.

Data sources

NYC CHS. Since 2002, the NYC Department of Health and Mental Hygiene has conducted the NYC CHS for approximately 9000 adults each year.[11] The survey tracks chronic disease and health behaviors using a telephone-based stratified random sample of the NYC population among areas identified by the United Hospital Fund's (UHF) neighborhoods.[12] These geographic units are collections of zip codes that roughly correspond to the NYC community planning districts, which are used by city agencies to address needs in specific NYC neighborhoods. Data are weighted to be representative of the adult noninstitutionalized NYC population. To estimate chronic disease prevalence, the survey assessed whether respondents have ever been told by a doctor, nurse or other health professional that they had conditions such as diabetes, hypertension and asthma. Multiyear survey weights were used to determine chronic disease period prevalence estimates by UHF neighborhood from 2009 to 2012.

New York State (NYS) SPARCS database. The Statewide Planning and Research Cooperative System (SPARCS) was established by the NYS Department of Health as a comprehensive data reporting system that collects data on patient characteristics, diagnoses, treatment, and hospital services for all inpatient discharges, ED visits, and other select data from NYS hospitals.[13] Encrypted unique identifiers within the database use a combination of social security numbers, sex, names, and date of birth to track specific individuals who make visits at different health care institutions. SPARCS is one of the few statewide claims databases that offer patient-level address data, which can be geocoded to identify an exact location of residence.

The American Community Survey (ACS). To determine the underlying age, sex, race/ethnicity, insurance, and geographic distribution of the NYC population, Census data from the ACS from 2009 to 2012 were used.[14] These estimates were used to determine how the population of unique ED patients at each hospital compared to the overall distribution of adult residents in NYC.

Participants

To match inclusion criteria of the ED surveillance method to survey respondents in the NYC CHS, all noninstitutionalized NYC adults were included. The study team analyzed patients aged 18 years and older who visited a NYS ED between 2009 and 2012, excluding those whose health care claims were paid by correctional facilities and any patient transferred from a nursing home or other health care facility.[12] Also excluded were any EDs not located at a general acute care hospital or that care for specific patient populations (eg, surgical subspecialty, oncology, Veterans Affairs hospitals).

Main outcome

The primary outcome was the correlation of age-adjusted chronic disease prevalence by neighborhood using the study team's method of ED surveillance versus NYC CHS estimates. Using the unique identifiers from the SPARCS database, the team accounted for repeat ED visits by the same individuals and only analyzed prevalence among unique ED patients. This process allowed accounting for patients who had multiple ED visits across several hospitals. The study team calculated what proportion of these unique individuals had ever received a primary or secondary diagnosis of diabetes, hypertension, or asthma during any ED visit between 2009 and 2012.[6] Diagnosis codes were based on an *International Classification of Diseases, Ninth Revision* code starting with the prefix of 250 for diabetes, 401–405 for hypertension, and 493 for asthma. To obtain the chronic disease prevalence among the ED population, the team divided the number of unique ED patients with each of these diagnoses by the overall number of unique ED patients by UHF neighborhood. Age adjustment was performed using the direct method and 4 age groups. Correlations were then performed by neighborhood to compare prevalence estimates made using ED surveillance versus the NYC CHS.

Statistical analysis

The hypothesis was that the accuracy of using data from a single hospital would depend on 2 factors: (1) how closely the patient population at each hospital matched the overall NYC population, and (2) how accurately each hospital identified individuals with chronic disease based on diagnosis codes. To determine how closely the patient population at each hospital matched the overall NYC population,

the study team analyzed the proportion of patients by age subgroups, sex, race/ethnicity, health insurance status, and geography based on county. For each of these categories, the average absolute difference was calculated across subgroups between the hospital's patient population and ACS Census estimates. To determine the accuracy of each hospital's ability to identify individuals with chronic disease based on diagnosis codes, the diagnosis code capture rate was ascertained. This study defined this metric as how frequently each hospital identified a given patient as having diabetes, hypertension, or asthma when compared to using diagnosis code data available from all NYS hospitals in the entire SPARCS database.

To compare the accuracy among hospitals in identifying chronic disease prevalence using data from each hospital alone, the study team analyzed the correlation between age-adjusted diabetes, hypertension and asthma prevalence using the 34 UHF Neighborhoods. Coefficients of correlation were calculated between chronic disease prevalence estimates from the NYC CHS and ED surveillance using data from all of the NYS hospitals, only the NYC hospitals, and for each of the NYC hospitals individually. For each pairwise correlation, the associated *P* value was calculated to determine the statistical significance of correlations identified.

The multivariable regression analysis included the measures of the similarity of each hospital's patient population in age, sex, race/ethnicity, insurance status, and distribution by county to the overall NYC population based on Census data. The analysis also included the diagnosis code capture rate that calculated how accurately each hospital was able to identify individuals with chronic disease based on administrative claims diagnosis codes versus using claims data from all NYS hospitals. These factors were assessed for evidence of multicollinearity before inclusion in the final model. Post-regression diagnostics included an analysis of the normality of residuals and for influential observations to test the robustness of the study results.

Statistical analyses were performed using Stata 12.1 (StataCorp LP, College Station, TX). Geographic analysis was performed using ArcGIS Desktop 10.2 (ESRI, Redlands, CA).

Results

Study participants

Among the 52 EDs located at general acute care hospitals in NYC, a total of 4.25 million unique adult NYC residents were identified using ED surveillance using data from 2009 to 2012. This population sample accounts for two thirds of the estimated 6.42 million NYC adults based on ACS Census estimates. On average, the population distribution of age, sex, and geography (ie, by county) were not substantially different between the Census data and averages of the ED populations among NYC hospitals. However, there were proportionally more black or African American adults and publicly insured and uninsured adults in the patient population of an average ED. Notably, the average ED patient population at a single NYC hospital had nearly 3 times the number of unique individuals as the number sampled by the NYC CHS. When analyzed individually, each hospital had substantial differences in population characteristics based on their own hospital location and patient mix. Table 1 presents data for ACS Census estimates and ED patient population data averaged among the 52 NYC hospitals and also for 3 specific hospitals affiliated with the study institution.

Primary outcome

The correlation of chronic disease prevalence estimates were essentially the same or slightly better when using all NYC hospitals versus all NYS hospitals (Table 2). When using ED data at individual NYC hospitals, the average diagnosis code capture rate was 81% for diabetes, 80% for hypertension, and 70% for asthma when compared to data from all NYS hospitals. However, there was substantial variation in the accuracy of coding among the 52 NYC hospitals as values ranged from 51% and 92% (see online Supplementary Fig. S1; Supplementary Data are available in the online article at www.liebertpub.com/pop).

Significant variation also was identified in the strength of correlation between chronic disease estimates from the NYC CHS and ED surveillance when using data only from a single NYC hospital (Table 2). On average, the correlation was 0.40 for diabetes (range of 0.01 to 0.88), 0.39 for hypertension (range of 0.03 to 0.88), and 0.46 for asthma (range of 0.12 to 0.84). The range of correlation outcomes among all 52 NYC hospitals is shown by histograms in online Supplementary Figure S2. In this analysis, the only NYC hospital that had a correlation of 0.80 or higher for all 3 chronic diseases was NYU Langone Medical Center. The strength of correlation for prevalence estimates with the NYC CHS was much lower for the 2 other hospitals affiliated with the study institution–Bellevue Hospital Center and NYU Lutheran Medical Center (graphically depicted in Fig. 1). Given the strong correlations identified using ED surveillance with data only from NYU Langone Medical Center, the study team mapped the neighborhood-level prevalence of diabetes, hypertension, and asthma throughout NYC using ED data only from NYU Langone Medical Center in Figure 2.

Multivariable analysis

The multivariable analysis identified factors that were associated with strong correlations between chronic disease prevalence estimates from the NYC CHS versus ED surveillance using data from a single hospital. Factors included comparing the average absolute difference across subgroups in the distributions of age, sex, race/ethnicity, insurance, and by county between the ED populations at each hospital versus Census estimates. The diagnosis code capture rate also was included for each of the chronic diseases for the regressions performed for diabetes, hypertension, and asthma, respectively (Table 3). For diabetes and hypertension, hospitals with an ED population with a racial/ethnic distribution more similar to NYC Census estimates were associated with a higher correlation of prevalence (*P* values 0.041 for diabetes, and 0.017 for hypertension). Higher rates of diagnosis code capture among hospitals also predicted strong correlations for diabetes and hypertension estimates (*P* values 0.023 and 0.003, respectively). However, the only factor that was statistically significant for all 3 chronic diseases was the similarity of geographic distribution between the ED population at a given hospital and Census

Table 1. Study Population Demographic, Insurance, and Geographic Characteristics for 2009 to 2012

Study population	*American Community Survey*	*NYC Community Health Survey*	*NYC hospital average*	*NYU Langone Medical Center*	*Bellevue Hospital Center*	*NYU Lutheran Medical Center*
Unique adults	6,424,312	36,188	106,668	82,608	163,556	119,064
Age subgroups, %						
18–24	13	13	16	13	15	16
25–44	40	39	39	40	45	44
45–64	31	31	28	25	31	24
65 and older	16	17	17	22	8	16
Sex, %						
Male	47	46	44	45	56	45
Female	53	54	56	55	44	55
Race and ethnicity, %						
White and other	36	38	37	77	38	49
Black or African American	23	22	34	10	24	7
Hispanic or Latino	28	27	24	8	31	34
Asian	13	13	5	5	7	10
Insurance, %						
Private	50	49	30	64	10	29
Public	34	32	48	25	36	49
Uninsured	16	19	22	11	54	22
County, %						
Bronx	16	16	21	5	9	1
Brooklyn	30	30	31	23	23	91
Manhattan	21	21	19	56	46	2
Queens	27	27	23	14	21	3
Staten Island	6	6	6	2	1	3

NYC, New York City.

estimates of NYC residents by county (*P* values 0.000, 0.000, and 0.003 for diabetes, hypertension, and asthma).

Discussion

Traditional health survey methods rely on intensive data collection, which limits the sample size that can be obtained.[15] Recently, novel methods such as ED surveillance have been developed to use existing large-scale data sources to identify health patterns at a local geographic level.[10] In NYS, the SPARCS database is one of the few data sources that provides access to patient-level address data that can be used to identify the exact residence of patients and local geographic patterns of health.[16]

The usefulness of geographically detailed health surveillance is underscored by the fact that most health care providers are generally unaware of where their patients live, and therefore are unaware of the contextual factors that may make individuals more susceptible to higher disease burden in specific geographic areas. On the other hand, community members may be more aware of where general problem areas are located, such as neighborhoods with high poverty or high rates of crime. But they would not be focused on specific health issues or be attuned to how the geographic

Table 2. Rates of Diagnosis Code Capture and Correlation of Neighborhood-Level Chronic Disease Prevalence Estimates

	Diabetes			*Hypertension*			*Asthma*		
Hospitals analyzed	*Diag. Code Capt., %*	*Corr. Coeff.*	P	*Diag. Code Capt., %*	*Corr. Coeff.*	P	*Diag. Code Capt., %*	*Corr. Coeff.*	P
All New York State Hospitals	100	0.86	0.000	100	0.88	0.000	100	0.77	0.000
All NYC hospitals	99	0.86	0.000	99	0.89	0.000	99	0.78	0.000
NYC hospital average	81	0.40	N/A[a]	80	0.39	N/A[a]	70	0.46	N/A[a]
NYU Langone Medical Center	85	0.81	0.000	87	0.84	0.000	76	0.84	0.000
Bellevue Hospital Center	65	0.49	0.003	60	0.30	0.084	55	0.59	0.000
NYU Lutheran Medical Center	91	0.54	0.001	92	0.36	0.034	86	0.29	0.098

[a]NYC hospital averages do not have *P* values associated with the correlation coefficient because it is reported as the average value among the 52 hospitals studied.

Corr. Coeff., correlation coefficient; Diag. Code Capt., diagnosis code capture rate; NYC, New York City.

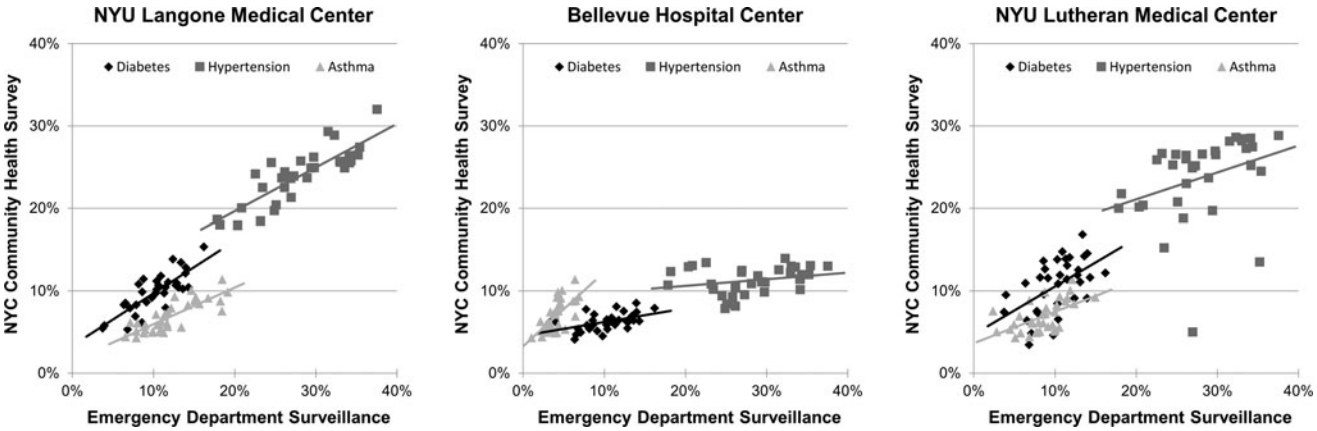

FIG. 1. Neighborhood-level correlation of chronic disease prevalence estimates based on data from individual hospitals. Correlation of diabetes, hypertension, and asthma prevalence estimates based on the NYC Community Health Survey versus emergency department surveillance using claims data from a single hospital. Examples shown here include 3 hospitals associated with the study academic institution. NYC, New York City.

distribution of one disease may differ from another. Furthermore, local hot spots of chronic disease also may be found in areas that do not directly overlap with these types of socioeconomic factors. Thus, a data-driven approach is needed to delineate the exact areas that experience higher disease burden.

However, geographically detailed data of this nature is not readily available in most regions.[17] Patient-level address data is generally absent in the federal- and state-level sources that have already collected data across the fragmented health care system.[18] Several other efforts initiated by regional health information organizations have been focused on consolidating more detailed data from disparate health care institutions, but few include the address data needed to perform detailed local health surveillance using geographic analysis.[19]

Use of a sentinel hospital also would allow for health surveillance that only requires data from a single source, thus obviating the need for complex data integration across institutions.[20] This single hospital could act as a sentinel hospital that provides citywide health surveillance.[21] This study demonstrates what attributes would make for ideal sentinel hospitals: those that have a broad geographic catchment, a patient population more reflective of the population at large, and a higher accuracy of coding chronic diseases. Because it may be difficult to generalize data obtained from a single hospital for a large geographic area, this type of health surveillance also might be improved by

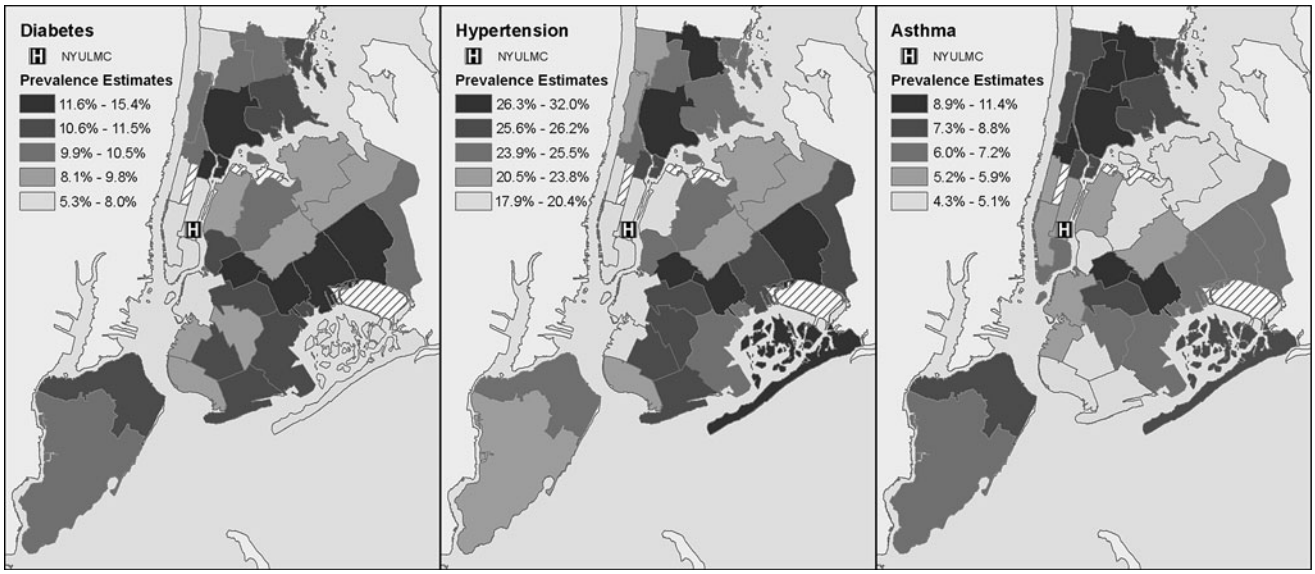

FIG. 2. Mapping neighborhood-level chronic disease prevalence using data from NYULMC. Maps of chronic disease prevalence for diabetes, hypertension, and asthma among NYC neighborhoods using emergency department claims data from a single, sentinel hospital—NYU Langone Medical Center—which demonstrated strong correlation of prevalence estimates compared with survey data. NYULMC, NYU Langone Medical Center.

Table 3. Multivariable Regression Analysis of the Strength of Correlation for Chronic Disease Prevalence Estimates

	Among NYC hospitals		*Diabetes*		*Hypertension*		*Asthma*	
Multivariable factors	*Hospital average, %*	*Hospital range, %*	*Beta Coeff.*	P	*Beta Coeff.*	P	*Beta Coeff.*	P
Difference for age group distribution	3.7	1.7–8.4	+0.06	0.570	+0.03	0.720	–0.02	0.852
Difference for male/female distribution	4.6	0.0–12.9	+0.02	0.844	–0.01	0.914	–0.22	0.082
Difference for race/ethnicity distribution	14.4	4.8–31.7	**–0.24**	**0.041**	**–0.24**	**0.017**	–0.19	0.146
Difference for insurance distribution	15.4	1.5–29.1	–0.22	0.065	–0.12	0.244	–0.13	0.317
Difference for geographic distribution	24.8	10.5–36.2	**–0.46**	**0.000**	**–0.63**	**0.000**	**–0.40**	**0.003**
Diagnosis code capture rate	See Table 1[a]		**+0.27**	**0.023**	**+0.31**	**0.003**	–0.04	0.725

Bold values statistically significant with a P value < 0.05.
[a]Diagnosis code capture rates vary based on the chronic disease studied and are already summarized in Table 1.
Beta Coeff., beta coefficient; NYC, New York City.

selecting a small group of hospitals with these characteristics that would provide a larger population sample that is also more representative of a given region.

The multivariable analysis helped to identify factors associated with stronger correlation of chronic disease prevalence estimated from claims data versus a traditional health survey. Hospitals with stronger correlation of prevalence estimates had a catchment that more similarly reflected the geographic distribution of the NYC population. At NYU Langone Medical Center, the average absolute difference in the proportion of patients over the 5 NYC counties was 14.2% when compared to Census estimates, which was on the low end of the range among NYC hospitals. For diabetes and hypertension, a higher rate of diagnosis code capture among hospitals also was associated with stronger correlation of chronic disease prevalence estimates. This variation in coding practices may be related to the resources available at each hospital. Hospital-based health surveillance ideally would take place at the sentinel hospital with high accuracy of coding individuals with chronic disease.

This study found that Bellevue Hospital Center, a public hospital, had a geographic distribution that most closely matched the overall population of NYC. However, a low diagnosis code capture rate, in addition to racial/ethnic population differences, led to more moderate correlations in chronic disease prevalence estimates. On the other end of the spectrum, NYU Lutheran Medical Center had one of the highest rates of diagnosis code capture rates, but correlation strength was limited by the geographic catchment of the hospital, which was overwhelmingly limited to Brooklyn. For NYU Lutheran Medical Center, substantial scatter in some correlations was caused by low observation counts in certain neighborhoods, leading to substantial error in some prevalence estimates.

Given these findings, the next steps are to explore the validity of performing citywide health surveillance for other conditions using data from specific sentinel hospitals. Sources such as the all-payer SPARCS database are limited to administrative claims data, which does not include important clinical data such as medications used or laboratory values.[22,23] Instead of performing the difficult task of aggregating these data across multiple institutions, data from a single sentinel hospital or a few sentinel hospitals could be used instead to represent geographically detailed health patterns throughout the city.[24] Aside from being easier to perform and less costly, these health surveillance methods would help capture data for a large proportion of the population with address-level data, which would enable identification of the exact geographic areas where disease burden is highest.

In addition, these methods may be extended across the country–especially to areas where traditional population health surveillance is infrequently or inadequately performed.[25] For instance, estimates of chronic disease prevalence in rural regions are limited to the county level.[26] Data from a single, nearby ED or small group of EDs may be helpful in identifying local health patterns in these regions.[20] Specific examples of how this approach can improve health surveillance include identifying local environmental pollutants that lead to higher rates of asthma in specific areas, performing local dietary interventions for large clusters of patients with diabetes who live near food swamps, or identifying specific neighborhoods with a high prevalence of behavioral disorders among children for school-based interventions. By performing more geographically detailed studies of health, it may be possible to identify local communities with a higher burden of chronic disease, which can be critical for targeting interventions to areas where health outcomes are particularly poor.[27]

Limitations

Administrative claims data can contain errors coded by the institutions providing data and lead to the incorrect categorization of patients into demographic categories.[28] Also, the variable analyzing the fidelity of diagnosis coding does not account for false positives in which patients are tagged with a diagnosis code during an ED visit but do not actually have the disease. However, prior studies of identifying ED patients with diagnosis codes have shown the accuracy of this approach. For instance, a diagnosis code of diabetes during an ED visit has been shown to be 95% sensitive and 99% specific.[29] In addition, the current best practice for determining neighborhood-level chronic disease prevalence in NYC is the CHS, which is a population-based health survey that relies on self-report and can miss the substantial proportion of the population who have undiagnosed chronic disease.[30] Finally, this study focused on NYC, a unique urban environment. Though specific results may differ in other populations or geographic regions, these methods may be readily adopted elsewhere.

Conclusions

This study found that that it is possible to perform city-wide surveillance using data from specific sentinel hospitals in NYC with accuracy similar to using data from all hospitals at the state or city level. Important factors that make a given hospital more likely to have strong correlations of prevalence estimates to traditional health surveillance methods include a racial/ethnic and geographic distribution of patients that more closely parallels the underlying city-wide population based on Census estimates. In addition, the accuracy of coding the diagnosis codes also is associated with strong correlation in prevalence estimates with traditional surveillance methods. One of the hospitals affiliated with the study institution, NYU Langone Medical Center, matched many of these characteristics, and thus had one of the strongest correlations of prevalence estimates for diabetes, hypertension, and asthma across NYC.

Author Disclosure Statement

Drs. Lee, Swartz, Koziatek, Athens, and Yi, and Mr. Vinson declared no conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors received the following financial support: This study was funded in part by a grant K23DK110316 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and grant 16-04083 from the New York State Health Foundation to study the geographic distribution of chronic disease among adults.

References

1. St Louis ME, Rauch KJ, Petersen LR, Anderson JE, Schable CA, Dondero TJ. Seroprevalence rates of human immunodeficiency virus infection at sentinel hospitals in the United States. The Sentinel Hospital Surveillance Group. N Engl J Med 1990;323:213–218.
2. Henry K, Campbell S. The potential efficiency of routine HIV testing of hospital patients—data from a CDC sentinel hospital. Public Health Rep 1992;107:138–141.
3. Brown DW, Balluz LS, Giles WH, et al. Diabetes mellitus and health-related quality of life among older adults. Findings from the behavioral risk factor surveillance system (BRFSS). Diabetes Res Clin Pract 2004;65:105–115.
4. Valdez R, Yoon PW, Liu T, Khoury MJ. Family history and prevalence of diabetes in the US population: 6-year results from the National Health and Nutrition Examination Survey (1999–2004). Diabetes 2007;30:2517–2522.
5. Kho AN, Cashy J, Hota BN, Sims SA, Malin BA, Galanter WL. The Chicago Health Atlas: a public resource to visualize health conditions and resources in Chicago. Chicago: American Medical Informatics Association, 2012.
6. Lee DC, Long JA, Wall SP, et al. Determining chronic disease prevalence in local populations using emergency department surveillance. Am J Public Health 2015;105:e67–e74.
7. Centers for Disease Control and Prevention. Health, United States, 2013: with special feature on prescription drugs. Hyattsville, MD: National Center for Health Statistics, 2014.
8. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. Health Aff (Millwood) 2009;28:454–466.
9. Downing NL, Lane S, Eisenberg M, Sharp C, Palma J, Longhurst C. An exponential increase in regional health information exchange with collaborative policies and technologies. Stud Health Technol Inform 2015;216: 931.
10. Galster G. Contextualisation of clinical information from fragmented health records. Stud Health Technol Inform 2013;194:41–47.
11. Buchholz N, Resnick S, Konty K. The New York City Community Health Survey Atlas, 2010. New York: The New York City Department of Health and Mental Hygiene, 2012.
12. Norton JM, Sanderson M, Gupta L, et al. Methodology updates to the New York City Community Health Survey. New York: New York City Department of Health and Mental Hygiene, 2012.
13. Quan JM. SPARCS: the New York State health care data system. J Clin Comput 1980;8:255–263.
14. Census Bureau. American Community Survey Multiyear Accuracy of the Data. 2014. www2.census.gov/programs-surveys/acs/tech_docs/accuracy/MultiyearACSAccuracyofData2013.pdf Accessed January 20, 2017.
15. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001;50:1–35.
16. Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL. Evaluation of the positional difference between two common geocoding methods. Geospat Health 2011; 5:265–273.
17. Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. Public Health Rep 2011;126(Suppl 3):54–61.
18. Phillips AB, Wilson RV, Kaushal R, Merrill JA; HITEC Investigators. Implementing health information exchange for public health reporting: a comparison of decision and risk management of three regional health information organizations in New York state. J Am Med Inform Assoc 2014;21:e173–e177.
19. Goedert J. Are HIEs the answer? Alternatives to regional health information exchanges are emerging. Health Data Manag 2009;17:24–28.
20. Curtis AJ, Lee WA. Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. Int J Health Geogr 2010;9:43.
21. Centers for Disease Control and Prevention. Estimating meningitis hospitalization rates for sentinel hospitals conducting invasive bacterial vaccine-preventable diseases surveillance. MMWR Morb Mortal Wkly Rep 2013;62: 810–812.
22. Chamany S, Silver LD, Bassett MT, et al. Tracking diabetes: New York City's A1C Registry. Milbank Q 2009;87: 547–570.
23. Newton KM, Wagner EH, Ramsey SD, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol 1999;52: 199–207.
24. Karami M, Zahraei SM, Gouya MM. Perspective of Iranian vaccine-preventable diseases surveillance system in the era of newly introduced pentavalent vaccine: possibility of implementing a sentinel hospital-based surveillance system. Iran J Public Health 2015;44:1712–1713.

25. Fan Z, Lackland DT, Lipsitz SR, et al. Geographical patterns of end-stage renal disease incidence and risk factors in rural and urban areas of South Carolina. Health Place 2007; 13:179–187.
26. Centers for Disease Control and Prevention. Estimated county-level prevalence of diabetes and obesity—United States, 2007. MMWR Morbid Mortal Wkly Rep 2009;58: 1259–1263.
27. Stevens CD, Schriger DL, Raffetto B, Davis AC, Zingmond D, Roby DH. Geographic clustering of diabetic lower-extremity amputations in low-income regions of California. Health Aff (Millwood) 2014;33: 1383–1390.
28. Kim M, Berger D, Matte T. Diabetes in New York City: public health burden and disparities. New York: New York City Department of Health and Mental Hygiene, 2006.
29. Zgibor JC, Orchard TJ, Saul M, et al. Developing and validating a diabetes database in a large health system. Diabetes Res Clin Pract 2007;75:313–319.
30. 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;99:152–159.

Address correspondence to:
David C. Lee, MD, MS
Ronald O. Perelman Department of Emergency Medicine
New York University School of Medicine
462 First Avenue
Room A345
New York, NY 10016

E-mail: david.lee@nyumc.org