Innovations in Population Health Surveillance: Using Electronic Health Records for Chronic Disease Surveillance

With 87% of providers using electronic health records (EHRs) in the United States, EHRs have the potential to contribute to population health surveillance efforts. However, little is known about using EHR data outside syndromic surveillance and quality improvement. We created an EHR-based population health surveillance system called the New York City (NYC) Macroscope and assessed the validity of diabetes, hyperlipidemia, hypertension, smoking, obesity, depression, and influenza vaccination indicators. The NYC Macroscope uses aggregate data from a network of outpatient practices. We compared 2013 NYC Macroscope prevalence estimates with those from a population-based, in-person examination survey, the 2013–2014 NYC Health and Nutrition Examination Survey.

NYC Macroscope diabetes, hypertension, smoking, and obesity prevalence indicators performed well, but depression and influenza vaccination estimates were substantially lower than were survey estimates. Ongoing validation will be important to monitor changes in validity over time as EHR networks mature and to assess new indicators.

We discuss NYC's experience and how this project fits into the national context. Sharing lessons learned can help achieve the full potential of EHRs for population health surveillance. (*Am J Public Health*. 2017;107:853–857. doi: 10.2105/AJPH.2017.303813)

Sharon E. Perlman, MPH, Katharine H. McVeigh, PhD, Lorna E. Thorpe, PhD, Laura Jacobson, MSPH, Carolyn M. Greene, MD, and R. Charon Gwynn, PhD

With better sanitation, antibiotics, and improved critical care medicine, average life expectancy has greatly increased in the past century in most parts of the world. In the United States and in many countries, chronic diseases have replaced infectious diseases as leading causes of death, yet most local public health agencies have limited information that can be used to inform prevention and control efforts. One emerging source for surveillance of chronic disease risk factors and conditions is data from electronic health records (EHRs).

Although data from EHRs are collected primarily to manage individuals' health care, they have the potential to be used for monitoring the health of populations. EHR use in the United States has increased from 21% of office-based physicians in 2004 to 87% in 2014.[1] This increased use is partly because of federal and state regulations. Federal meaningful use[2] incentives have encouraged adoption of EHRs and standardization of how data are entered. In New York State, electronic prescriptions are now legally mandated[3] and other states are following suit. Beyond individual patient management, researchers and clinicians have mostly used EHR data for clinical quality improvement initiatives, comparative effectiveness research, and monitoring outbreaks of infectious disease.[4,5] However, in the past few years, there has been increasing interest in using EHRs for population health surveillance, that is, for monitoring the disease burden in a defined geographic area.

Several jurisdictions have made progress in accessing and integrating electronic health data from various sources. One example is the Chicago Health Atlas, which combines hospital EHR data with public health and social service data[6] for chronic diseases (e.g., diabetes, asthma, breast cancer), infectious diseases (e.g., tuberculosis, chlamydia), environmental exposures such as lead, and birth and deaths aggregated at the zip code level. Another example is the Colorado Health Observation Regional Data Service,[7] which collects and integrates data from 10 health care systems to produce a registry for obesity, hypertension, and hyperlipidemia. In Massachusetts, MDPHnet allows health departments to obtain aggregate data from practices' EHRs for key health outcomes.[8] Elsewhere, Regional Health Information Organizations are facilitating the collection and analysis of EHR data from multiple sites.

Using aggregated data to generate population prevalence estimates of chronic diseases is still in its early stages.[9] There is poor understanding of the validity and reliability of EHR data for this purpose and whether EHR data represent the population in a particular geographic area. We briefly describe one recently developed municipal EHR surveillance system and summarize validation study findings published to date.

NEW YORK CITY MACROSCOPE DEVELOPMENT

In 2012, a cross-disciplinary team including epidemiologists, health information technology experts, and clinicians from the NYC Department of Health and Mental Hygiene (DOHMH) and the City University of New York School of Public Health (now at NYU), with the support of several foundations, developed an EHR surveillance system

ABOUT THE AUTHORS

Sharon E. Perlman, Katharine H. McVeigh, and R. Charon Gwynn are, and at the time of this study Carolyn M. Greene and Laura Jacobson were, with the New York City Department of Health and Mental Hygiene, Queens, NY. Lorna E. Thorpe is with the New York University School of Medicine Department of Population Health, New York, NY.

Correspondence should be sent to Sharon E Perlman, Division of Epidemiology, NYC Department of Health, 42-09 28th Street, CN#6, New York, NY 11101 (e-mail: sperlma1@health.nyc.gov). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted February 26, 2017.

doi: 10.2105/AJPH.2017.303813

known as the NYC Macroscope.[10] The NYC Macroscope uses data from a large, distributed EHR network in NYC. This process included defining health indicators, generating prevalence estimates, and validating the estimates.

NYC Macroscope uses EHR data from the DOHMH's Primary Care Information Project (PCIP). PCIP assists ambulatory practices to adopt and use EHRs to increase delivery of clinical preventive services, reduce chronic disease risk factors, and improve disease management.[11] As part of this program, a subset of 700 practices using eClinicalWorks agreed to share aggregate data with the NYC DOHMH.[12] This virtual network, created in collaboration with eClinicalWorks, allows the NYC DOHMH to monitor the health of nearly 1.5 million patients, about one in six New Yorkers. The NYC DOHMH uses these data to create reports for providers that highlight opportunities for delivering clinical preventive services. The reports also allow providers to compare their aggregate patient data with similar providers in the city,[13] enabling them to quickly identify what is working well in their practices and areas for improvement. At the same time, the NYC DOHMH can use aggregate data to assess population health and to guide policy and programmatic initiatives. In many ways, PCIP serves as a bridge between clinical care and public health.

In developing the NYC Macroscope, we selected and defined indicators (Table 1; Appendix, available as a supplement to the online version of this article at http://www.ajph.org as) important to public health surveillance. These indicators included prevalence of diabetes, hyperlipidemia, hypertension, smoking, obesity, depression, and influenza vaccination as well as treatment and control of diagnosed diabetes, hyperlipidemia, and hypertension (Table 1). We developed queries to extract information from PCIP practices aggregated by provider and used information captured by these queries to calculate the NYC Macroscope estimates.

TABLE 1—Selected New York City Macroscope 2013 Indicators Important to Public Health Surveillance

Indicator and Type	Definition
Smoking prevalence	Current smoker recorded in structured fields (checkboxes)[a]
Obesity prevalence	Body mass index ≥ 30 kg/m^2 from most recent height and weight[a] in vitals
Depression prevalence	*ICD* diagnosis of depression or Patient Health Questionnaire-9 score of 10–27
Influenza vaccination prevalence	CVX, CPT, or *ICD* code indicating receipt of influenza vaccination[a]
Hypertension	
Prevalence	*ICD* diagnosis of hypertension
Augmented prevalence	Most recent blood pressure[a] systolic ≥ 140 mm Hg or diastolic ≥ 90 mm Hg or *ICD* diagnosis of hypertension or medication prescribed for hypertension[a]
Treatment	Medication prescribed[a] among those with *ICD* diagnosis of hypertension
Control	Most recent blood pressure < 140/90[a] mm Hg among those with *ICD* diagnosis of hypertension
High cholesterol	
Prevalence	*ICD* diagnosis of high cholesterol
Augmented prevalence	Most recent total cholesterol ≥ 240[b] mg/dL or *ICD* diagnosis of high cholesterol or medication prescribed for high cholesterol[a]
Treatment	Medication prescribed[a] among those with *ICD* diagnosis of high cholesterol
Control	Most recent total cholesterol < 240[b] mg/dL among those with *ICD* diagnosis of high cholesterol
Diabetes	
Prevalence	*ICD* diagnosis of diabetes
Augmented prevalence	Most recent hemoglobin A1c ≥ 6.5[b] or *ICD* diagnosis of diabetes or medication prescribed for diabetes[a]
Treatment	Medication prescribed[a] among those with *ICD* diagnosis of diabetes
Poor control	Most recent A1c > 9[b] among those with *ICD* diagnosis of diabetes

Note. CPT = current procedural terminology; CVX = vaccine administered; *ICD* = *International Classification of Diseases International Classification of Diseases, Ninth Revision, Clinical Modification* (Hyattsville, MD: National Center for Health Statistics; 1980) DHHS publication PHS 80-1260.

[a]In calendar year.

[b]In calendar year or previous calendar year.

We designed the NYC Macroscope to be generalizable to the NYC adult population who sought primary care in the past year because of known health differences between this "in-care" population and those who were not in care in the past year.[14] An important advantage of PCIP's aggregate data model is the absence of personal identifying information, which ensures patient privacy. However, this inability to identify individuals prevented us from eliminating duplicate records across practices. To reduce duplicate records and minimize bias from specialists seeing sicker patients, we limited the NYC Macroscope to primary care providers. We also limited the surveillance system to a group of providers that met minimum criteria for EHR data completeness and quality.

We used estimates from two well-established reference surveys to validate the NYC Macroscope estimates: the gold standard NYC Health and Nutrition Examination Survey (NYC HANES)[15] and the NYC Community Health Survey.[16] NYC HANES is a population-based, in-person examination survey modeled on the national HANES, and the Community Health Survey is an annual random digit dial telephone survey in NYC modeled on the national Behavioral Risk Factor Surveillance System.[17] For comparability, we limited all data sources to those who reported seeking primary care in the past year. To determine if estimates could be used for population health surveillance, we established a priori goodness-of-fit criteria to compare estimates from the NYC Macroscope with those from the two surveys. Criteria included prevalence difference (within

five points), prevalence ratio (0.85–1.15), a Spearman correlation of ≥ 0.80, *t* test, and the two one-sided test of equivalence.[18]

In addition to validating the overall estimates from the NYC Macroscope, we conducted a chart review on a subset of NYC HANES participants who had sought primary care in the past year to assess the validity of EHR measures at the individual level (61% consented to chart review, and 27% of those signed HIPAA [the Health Insurance Portability and Accountability Act of 1996] release forms and met eligibility criteria). We abstracted information from the EHRs from 48 NYC HANES participants whose providers contributed data to the NYC Macroscope and 142 participants whose providers did not (the latter included charts from 22 different EHR vendors). For each study participant, we compared the indicator estimates from primary care EHR data to those from NYC HANES and computed sensitivity and specificity. We also stratified EHR-based indicator estimates by whether the participant's provider was part of the NYC Macroscope to examine generalizability to other EHR surveillance systems.

NYC MACROSCOPE VALIDATION RESULTS

Detailed results of the NYC Macroscope validation have been described previously.[19,20] In brief, NYC Macroscope indicators for diabetes, hypertension, smoking, and obesity prevalence performed well according to a priori criteria for agreement (Table 2), and each of these demonstrated high sensitivity and specificity (0.89–1.00) in the chart review study of records from providers contributing to the NYC Macroscope. Although hyperlipidemia prevalence met many of the a priori criteria, sensitivity and specificity were not as good (0.72 and 0.58, respectively).

Indicators for treatment and control measures for diabetes, hypertension, and hyperlipidemia were less concordant,[20] which might be explained by bias attributable to PCIP's interactions with physicians, including newsletters, summary dashboards for providers, and EHR-based reminders to perform clinical preventive services and for hyperlipdemia, a change in clinical guidelines in 2013.[21] Small NYC HANES sample sizes could also have contributed to poor precision and lower agreement. The NYC Macroscope prevalence estimates for depression and influenza vaccination were substantially lower than were survey estimates and had poor sensitivity.[19] Many providers are not routinely using a standardized screening tool to assess depression (eClinicalWorks software uses the Patient Health Questionnaire), and influenza vaccines given at pharmacies or workplaces may not be documented in primary care settings.

Results of the broader chart review study involving multiple EHR vendors are forthcoming, but preliminary results suggest that obesity and diabetes prevalence estimates translate well across primary care EHR software systems. Smoking and hypertension prevalence also performed well and can likely be used for population health surveillance by other jurisdictions, although it may be helpful to conduct local validation against an established data source if possible. For these indicators, generalizability across EHR software systems may improve with strategic restrictions, such as including only providers who have attested to Meaningful Use standards.

Future validation efforts will focus on evaluating accuracy of neighborhood-level estimates as well as estimates stratified by race/ethnicity to explore how the NYC Macroscope can be used to monitor health inequities. We have also initiated an evaluation of the system's capacity to monitor trends from 2012 to 2015 and have begun to explore additional indicators, including obesity in young children. Ongoing validation will be important to monitor changes in validity as the EHR network matures and to assess new indicators.

The NYC Macroscope complements existing chronic disease monitoring systems and can be used to help the NYC DOHMH and its local partners establish priorities and inform programming. In NYC and other

TABLE 2—Results of Indicator Validation by Comparing NYC Macroscope 2013 (n = 648 816) with NYC HANES 2013–2014 (n = 1135) and the 2013 Community Health Survey (n = 6166) for NYC Adults in Care in the Past Year

	Indicator						
	Hypertension	Diabetes	Smoking	Obesity	Hypercholesterolemia	Depression	Influenza Vaccination
NYC Macroscope, % (95% CI)	32.3 (32.2, 32.4)	13.9 (13.8, 14.0)	15.2 (15.1, 15.3)	27.9 (27.7, 27.9)	49.3 (49.1, 49.5)	8.2 (8.1, 8.2)	20.9 (20.8, 21.0)
NYC HANES, % (95% CI)	32.5 (29.4, 35.7)	12.6 (10.6, 14.8)	17.7 (15.1, 20.8)	31.3 (28.5, 34.5)	46.9 (42.6, 51.3)	19.0 (16.4, 21.9)	47.6 (44.0, 51.3)
Community Health Survey, % (95% CI)	31.6 (30.2, 33.0)	12.5 (11.5, 13.6)	14.9 (13.6, 16.3)	24.7 (23.2, 26.3)	47.9 (45.7, 50.1)	NA	47.3 (45.5, 49.0)
NYC Macroscope vs NYC HANES							
Absolute difference < 5	✓ (0.15)	✓ (1.36)	✓ (2.55)	✓ (3.46)	✓ (2.36)	X (10.80)	X (26.71)
Prevalence ratio of 0.85–1.15	✓(1.00)	✓ (1.11)	✓ (0.86)	✓ (0.89)	✓ (1.05)	X (0.43)	X (0.44)
Test of difference (*t* test), $P \geq .05$	✓($P = .93$)	✓ ($P = .19$)	✓ ($P = .08$)	X ($P = .02$)	✓ ($P = .29$)	X ($P < .01$)	X ($P < .01$)
Test of equivalence, $P < .05$	✓ ($P < .01$)	✓ ($P < .001$)	✓ ($P = .04$)	X ($P = .14$)	X ($P = .12$)	X ($P = .99$)	X ($P = .99$)
Recommendation	Ready for use	Ready for use	Ready for use	Ready for use	Use with caution[a]	Not ready for use	Not ready for use

Note. ✓ = criterion met; X = criterion not met; CI = confidence interval; HANES = Health and Nutrition Examination Survey; NA = not available in this data set; NYC = New York City.

[a]Caution for hypercholesterolemia prevalence is on the basis of the tests shown here and lower sensitivity and specify in the chart review study (data not shown).

jurisdictions, survey response rates are declining, and results from traditional surveillance systems are often unavailable for smaller cities, towns, or neighborhoods because these areas do not have sample sizes large enough to produce reliable estimates. EHR-based surveillance systems can be cost effective and timely and can provide prevalence estimates for local communities and smaller subpopulations. Especially when used in conjunction with other data sources, they can provide a comprehensive and accurate picture of the health of a defined population.

ADVANCING POPULATION HEALTH NATIONALLY

Validation findings from the NYC Macroscope and early results from similar emerging systems[22,23] suggest that prevalence of diabetes, smoking, hypertension, and obesity are good indicators for EHR-based surveillance in jurisdictions with functioning EHR networks (Table 2). Methodological decisions made in developing the NYC Macroscope may be helpful to other jurisdictions establishing similar EHR-based surveillance systems; for example, they can help them develop methods to weight results for generalizability to a larger population, establish inclusion criteria, define indicators with attention to how diagnostic criteria may differ from epidemiological definitions, and select statistical methods for comparing estimates across data sources.[10]

To share ideas and further advance the field of population health surveillance using EHRs, the NYC DOHMH has fostered collaborative meetings with other early adopter agencies and select organizations to advance EHR surveillance methodology. The DOHMH hosted a small conference in October 2014 that was attended by representatives from health departments, federal agencies, clinical data research networks, and health care delivery networks. Common issues discussed included governance models, the need for high level buy-in, support from partnering institutions, legal and technical challenges in establishing a network and permitting data sharing, and difficulties of sustainability and ensuring data quality.

In January 2017, AcademyHealth and the DOHMH, with foundation support, sponsored a workshop to explore methodologies for small area estimation using EHR networks. We also partnered with AcademyHealth to develop a Population Health Community of Practice[24] for sharing ideas and challenges and for encouraging collaboration among those using EHRs for population health.[24] As EHR networks mature, the active sharing of lessons learned can stimulate and improve data-driven initiatives to bridge primary care and population health[25] and help achieve the full potential of EHRs for population health surveillance. AJPH

CONTRIBUTORS

S. E. Perlman, L. E. Thorpe, and C. M. Greene were involved in the conceptualization of the NYC Macroscope. S. E. Perlman, K. H. McVeigh, L. E. Thorpe, and L. Jacobson led the analysis and interpretation of the data. All authors contributed to drafting and editing the article.

ACKNOWLEDGMENTS

This work has been made possible by the financial support of the de Beaumont Foundation (grant 2012009); the Robert Wood Johnson Foundation (grant 69975), including its National Coordinating Center for Public Health Services and Systems Research; the Robin Hood Foundation (grant 2014103); the New York State Health Foundation (grant 11-00765); the Doris Duke Charitable Foundation (grant 2014070); and the US Centers for Disease Control and Prevention (CDC; grant U28EH000939). The efforts of L. E. Thorpe were supported in part by the CDC (grant U48DP001904).

The authors would like to thank the NYC Macroscope team as well as Thomas Farley, Sarah Shih, Remle Newton-Dame, Lauren Schreibstein, Elisabeth Snell, Elizabeth Lurie, and Matthew Romo for their contributions to this work.

Note. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funders.

HUMAN PARTICIPANT PROTECTION

The institutional review boards of the New York City (NYC) Department of Health and Mental Hygiene (DOHMH) and the City University of New York School of Public Health approved the NYC Health and Nutrition Examination Survey. The NYC DOHMH institutional review board approved the NYC Macroscope chart review study.

REFERENCES

1. Office of the National Coordinator for Health Information Technology. Office-based physician electronic health record adoption. Available at: https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php. Accessed April 3, 2017.

2. Office of the National Coordinator for Health Information Technology. Meaningful use definitions and objectives. 2015. Available at: http://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives. Accessed January 23, 2017.

3. New York State Department of Health. Electronic prescribing. Available at: http://www.health.ny.gov/professionals/narcotic/electronic_prescribing. Accessed April 3, 2017.

4. Romano MJ, Stafford RS. Electronic health records and clinical decision support systems: impact on national ambulatory care quality. *Arch Intern Med*. 2011; 171(10):897–903.

5. Chester KG. BioSense 2.0. *Online J Public Health Inform*. 2013;5(1):e200.

6. Chicago Health Atlas. Available at: http://www.chicagohealthatlas.org. Accessed March 31, 2017.

7. Denver Public Health. Colorado health observation regional data service (CHORDS). Available at: http://www.denverhealth.org/public-health-and-wellness/public-health/health-information-and-reports/data-and-informatics-tools/colorado-health-observation-regional-data-service-chords. Accessed March 31, 2017.

8. Vogel J, Brown JS, Land T, Platt R, Klompas M. MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *Am J Public Health*. 2014;104(12):2265–2270.

9. Paul MM, Greene CM, Newton-Dame R, et al. The state of population health surveillance using electronic health records: a narrative review. *Popul Health Manag*. 2015;18(3):209–216.

10. Newton-Dame R, McVeigh KH, Schreibstein L, et al. Design of the New York City Macroscope: innovations in population health surveillance using electronic health records. *EGEMS (Wash DC)*. 2016;4(1):1265.

11. Mostashari F, Tripathi M, Kendall M. A tale of two large community electronic health record extension projects. *Health Aff (Millwood)*. 2009;28(2):345–356.

12. 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.

13. Sebek KM, Virkud A, Singer J, Pulgarin CP, Schreibstein L, Wang JJ. Preliminary evaluation of a comprehensive provider feedback report. *J Med Pract Manage*. 2014;29(6):397–405.

14. Romo ML, Chan PY, Lurie-Moroni E, et al. Characterizing adults receiving primary medical care in New York City: implications for using electronic health records for chronic disease surveillance. *Prev Chronic Dis*. 2016;13:E56.

15. Thorpe LE, Greene C, Freeman A, et al. Rationale, design and respondent characteristics of the 2013–2014 New York City Health and Nutrition Examination Survey (NYC HANES 2013–2014). *Prev Med Rep*. 2015;2: 580–585.

16. New York City Department of Health and Mental Hygiene. Community Health Survey methodology. 2016. Available at: http://www1.nyc.gov/site/doh/data/data-sets/community-health-survey-methodology.page. Accessed March 31, 2017.

17. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System. Available at: https://www.cdc.gov/brfss. Accessed January 6, 2017.

18. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2011; 26(2):192–196.

19. McVeigh KH, Newton-Dame R, Chan PY, et al. Can electronic health records be used for population health surveillance? Validating population health

metrics against established survey data. *EGEMS (Wash DC)*. 2016;4(1):27.

20. Thorpe LE, McVeigh KH, Perlman S, et al. Monitoring prevalence, treatment, and control of metabolic conditions in New York City adults using 2013 primary care electronic health records: a surveillance validation study. *EGEMS (Wash DC)*. 2016;4(1):1226.

21. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 pt B):2889–2934. [Erratum in *J Am Coll Cardiol*. 2015;66 (24):2812; *J Am Coll Cardiol*. 2014;63(25 pt B):3024–3025]

22. Szatkowski L, Lewis S, McNeill A, Huang Y, Coleman T. Can data from primary care medical records be used to monitor national smoking prevalence? *J Epidemiol Community Health*. 2012;66(9): 791–795.

23. Aguilar-Palacio I, Carrera-Lasfuentes P, Poblador-Plou B, Prados-Torres A, Rabanaque-Hernández MJ; Grupo de Investigación en Servicios Sanitarios de Aragón. [Morbidity and drug consumption. Comparison of results between the National Health Survey and electronic medical records]. *Gac Sanit*. 2014;28(1): 41–47.

24. EDMForum. Population health community of practice. Available at: http://www.edm-forum.org/collaborate/cop/popcop. Accessed April 3, 2017.

25. Institute of Medicine. *Primary Care and Public Health: Exploring Integration to Improve Population Health*. Washington, DC: National Academies Press; 2012.