Generalizability of Indicators from the New York City Macroscope Electronic Health Record Surveillance System to Systems Based on Other EHR Platforms

Katharine H. McVeigh, PhD, MPH;[i] Elizabeth Lurie-Moroni, MPH;[i] Pui Ying Chan, MPH;[i] Remle Newton-Dame, MPH;[i] Lauren Schreibstein, MA;[i] Kathleen S. Tatem, MPH;[i] Matthew L. Romo, PharmD, MPH;[i] Lorna E. Thorpe, PhD;[ii] Sharon E. Perlman, MPH[i]

ABSTRACT

Introduction: The New York City (NYC) Macroscope is an electronic health record (EHR) surveillance system based on a distributed network of primary care records from the Hub Population Health System. In a previous 3-part series published in *eGEMS*, we reported the validity of health indicators from the NYC Macroscope; however, questions remained regarding their generalizability to other EHR surveillance systems.

Methods: We abstracted primary care chart data from more than 20 EHR software systems for 142 participants of the 2013-14 NYC Health and Nutrition Examination Survey who did not contribute data to the NYC Macroscope. We then computed the sensitivity and specificity for indicators, comparing data abstracted from EHRs with survey data.

Results: Obesity and diabetes indicators had moderate to high sensitivity (0.81-0.96) and high specificity (0.94- 0.98). Smoking status and hypertension indicators had moderate sensitivity (0.78-0.90) and moderate to high specificity (0.88-0.98); sensitivity improved when the sample was restricted to records from providers who attested to Stage 1 Meaningful Use. Hyperlipidemia indicators had moderate sensitivity (≥0.72) and low specificity (≤0.59), with minimal changes when restricting to Stage 1 Meaningful Use.

Discussion: Indicators for obesity and diabetes used in the NYC Macroscope can be adapted to other EHR surveillance systems with minimal validation. However, additional validation of smoking status and hypertension indicators is recommended and further development of hyperlipidemia indicators is needed.

[i]New York City Department of Health and Mental Hygiene, [ii]New York University School of Medicine, Department of Population Health

Conclusion: Our findings suggest that many of the EHR-based surveillance indicators developed and validated for the NYC Macroscope are generalizable for use in other EHR surveillance systems.

Introduction

Aggregated data from electronic health records (EHRs) are increasingly being used for population health surveillance; however, best practices for using this data source remain to be established.[1] Before EHR data can be used for surveillance, it is necessary to establish that the health indicators reflect, with acceptable accuracy, the information contained within the EHR and the individual's true health status. A number of studies have evaluated the feasibility of using EHRs for surveillance and the validity of indicators, both in the United States and around the world.[2-13] The majority of these studies have compared EHR-based prevalence estimates with population survey estimates[2-5,8-10] or measured agreement of indicators between EHR data and the entire medical record[6-9,11,12] or patient self-report.[13] The results from these studies, however, are largely self-contained and their generalizability to other EHR surveillance systems remains to be established.

Since 2007, the New York City (NYC) Department of Health and Mental Hygiene (DOHMH) has been assisting hundreds of ambulatory primary care practices with the uptake and use of EHRs and engaging in clinical and data quality improvement activities through the Primary Care Information Project (PCIP).[14] A key component of PCIP is the Hub Population Health System (the Hub), a network of practices using eClinicalWorks software. Using the Hub infrastructure, the NYC Macroscope was

designed, in collaboration with the City University of New York School of Public Health, to estimate chronic disease and risk factor prevalence for the NYC adult population in care (defined as having seen a health care provider for primary care in the past year).[15] Practices included in the NYC Macroscope cohort met data documentation inclusion and exclusion criteria described elsewhere.[16]

The development of the NYC Macroscope EHR surveillance system and the validity of indicators relative to the NYC Health and Nutrition Examination Survey (HANES), a gold standard population health examination survey, has been documented in a 3-part series previously published in *eGEMS (Generating Evidence & Methods to improve patient outcomes)*.[4,5,16] These validation studies also compared individual-level NYC HANES data with EHR data from 48 NYC HANES participants whose data were included in the 2013 NYC Macroscope.[4,5] Given the unique relationship the NYC DOHMH has with Hub Population Health System providers and the reliance of the NYC Macroscope on a single EHR software system, it remains unclear if NYC Macroscope indicators could be generalized to EHR surveillance systems using data from other providers or recorded on other EHR platforms. To determine the generalizability of NYC Macroscope indicators beyond that singular EHR platform, we examined the validity of the NYC Macroscope indicators for non-NYC Macroscope records from participants included in NYC HANES.

Methods

Data sources and sample

Data sources for this study were the 2013-14 NYC HANES, a population-based examination survey of NYC residents 20 years of age and older,[17] and EHR data from a sub-sample of NYC HANES participants. Printed copies of EHRs were requested from providers via in-person visits, telephone, fax, and mail. Only records from providers practicing general internal medicine, family medicine, pediatrics, geriatrics, and adolescent medicine, based on the taxonomy associated with the National Provider Identifier,[18] were included.

Of the 1,527 NYC HANES participants, 1,089 met eligibility criteria because they had reported visiting a health care provider for a "check-up, advice about a health problem, or basic care" within the previous year (i.e., "in care"), and did not have a proxy interview (Figure 1). Of these participants, 45 percent (491 individuals) signed a consent form and completed a Health Insurance Portability and Accountability Act waiver, granting us access to their medical records. We obtained printed copies of EHRs for 277 participants, of which 190 contained primary care data recorded within a year prior to the participant's NYC HANES interview.

The 190 valid records consisted of 48 NYC Macroscope records and 142 non-NYC Macroscope records. The 142 non-NYC Macroscope records were received from 133 providers working in 89 medical practices throughout NYC, using more than 20 different EHR software systems. Of these non-NYC Macroscope records, 86 were received from 79 providers working in 49 medical practices that had attested to Stage 1 Meaningful Use, using more than 15 EHR software systems.

Chart review procedures

A team of medical and nursing students was trained to abstract relevant data from the participants' medical records for the period starting from their NYC HANES interview date back to January 1, 2011. The abstracted data were stored on a secure and confidential internal server at the NYC DOHMH. Data were abstracted from the following structured fields: International Classification of Diseases Ninth Revision (ICD-9) diagnosis codes, body mass index (BMI), smoking, blood pressure, and laboratory values including total cholesterol and glycated hemoglobin (A1C). Data were also abstracted from unstructured fields, including provider notes and scanned documents. Scanned documents included external laboratory results not incorporated into the EHR interface, external medical records, consultation notes or records, and other relevant documents.

Data quality control

To ensure data quality, more than half (56 percent [107/190]) of the records were abstracted by multiple reviewers. Duplicate entries were compared, and discrepancies were investigated and corrected. Of the remaining 83 records reviewed by a single reviewer, 59 (71 percent) were abstracted by one of the 3 most accurate reviewers (inter-rater reliability [Kappa coefficient] among these 3 reviewers was 0.96). The remaining 24 charts were entered one time and reviewed by a supervisor to confirm data entry accuracy.

IRB approval

The NYC DOHMH and City University of New York Institutional Review Boards approved collection of HIPAA waivers from NYC HANES study participants, and the NYC DOHMH Institutional Review Board approved this study.

Measures

Self-reported data from NYC HANES were collected through an in-person interview, and objective clinical measures were obtained through physical examination and laboratory testing of a blood specimen. NYC Macroscope indicators were based on data from structured EHR fields that captured diagnoses, medications, vital signs, laboratory values, and smoking status.

Indicators

Appendix table A1 provides definitions of the NYC Macroscope and NYC HANES indicators for prevalence measures of obesity, smoking, hypertension, diabetes, and hyperlipidemia. Obesity was defined as a BMI ≥30, using the most recent BMI recorded in the EHR; BMI in NYC HANES was calculated based on measured height and weight. Smoking was based on the most recently recorded smoking status in the EHR and was self-reported in NYC HANES. The diagnosis indicators for hypertension, diabetes, and hyperlipidemia were based on the presence of an ICD-9 diagnosis code in the EHR, and were contrasted to the NYC HANES measures of self-reported diagnosis. Augmented indicators for hypertension, diabetes, and hyperlipidemia were defined as an ICD-9 diagnosis code, *or* the most recent diagnostic blood pressure/laboratory values, *or* medications electronically prescribed to treat the condition, and were contrasted to the NYC HANES measures based on self-reported diagnosis and/or examination or laboratory findings.[19]

Statistical analysis

Weighted demographic characteristics and health outcomes were compared between the NYC HANES chart review subsample (N=190) and the other NYC HANES in-care participants (N=945). Similar comparisons using unweighted data were made within the chart review sample, stratified by whether the participants' records contributed to the NYC Macroscope (N=48) or not (N=142). Additionally, we described differences in sample distributions between NYC Macroscope (N=716,076) and the chart review sample (N=190). All comparisons were made using χ^2 tests with a 2-sided α level set at 0.05.

NYC Macroscope indicator definitions were applied to the abstracted EHR data and computed outcome classifications were linked with NYC HANES outcome classifications for each individual. Agreement between the NYC Macroscope indicator classifications and the NYC HANES classifications was quantified using Cohen's Kappa, sensitivity, and specificity. Kappa coefficients were interpreted using the criteria established by Landis and Koch that characterize agreement as slight (Kappa: 0.0-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61- 0.80), and near perfect (0.81-1.0).[20] Sensitivity was characterized as high (0.90-1.00), moderate (0.70-0.89), and low (<0.70), and specificity was characterized as high (0.90-1.00), moderate (0.80-0.89), and low (<0.80). To aid interpretation, 95 percent confidence intervals were computed for sensitivity and specificity. All analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC) and SAS-callable SUDAAN® software, version 11.0 (Research Triangle Institute, Research Triangle Park, NC).

Sensitivity analyses

The first sensitivity analysis restricted data to the subset of non-NYC Macroscope records from patients seen by providers who had attested to the Centers for Medicaid & Medicare Services EHR Stage 1 Meaningful use incentive program.[21] Stage 1 Meaningful Use aims to optimize data capture and sharing through documentation and communication specifications (e.g., recording smoking status for patients 13 years of age and older; providing

patients with timely access to their electronic health information). Attestation status was based on Centers for Medicaid & Medicare Services data through December 31, 2013, which coincided most closely with the attestation status achieved during the chart review look-back period.[22,23] At that time, Stage 1 Meaningful Use was the highest level of attestation possible. This analysis was important because NYC Macroscope uses documentation quality criteria aligned with Stage 1 Meaningful Use as an inclusion criterion.[16] The second sensitivity analysis assessed the impact of incorporating unstructured data on sample size and sensitivity for obesity and smoking status indicators, and on sensitivity for diagnosis and augmented indicators of hypertension, diabetes, and hyperlipidemia, using non-NYC Macroscope records.

Results

Sample characteristics

Our final sample consisted of 190 NYC HANES participants in the chart review; 945 participants were excluded for various reasons (Figure 1).
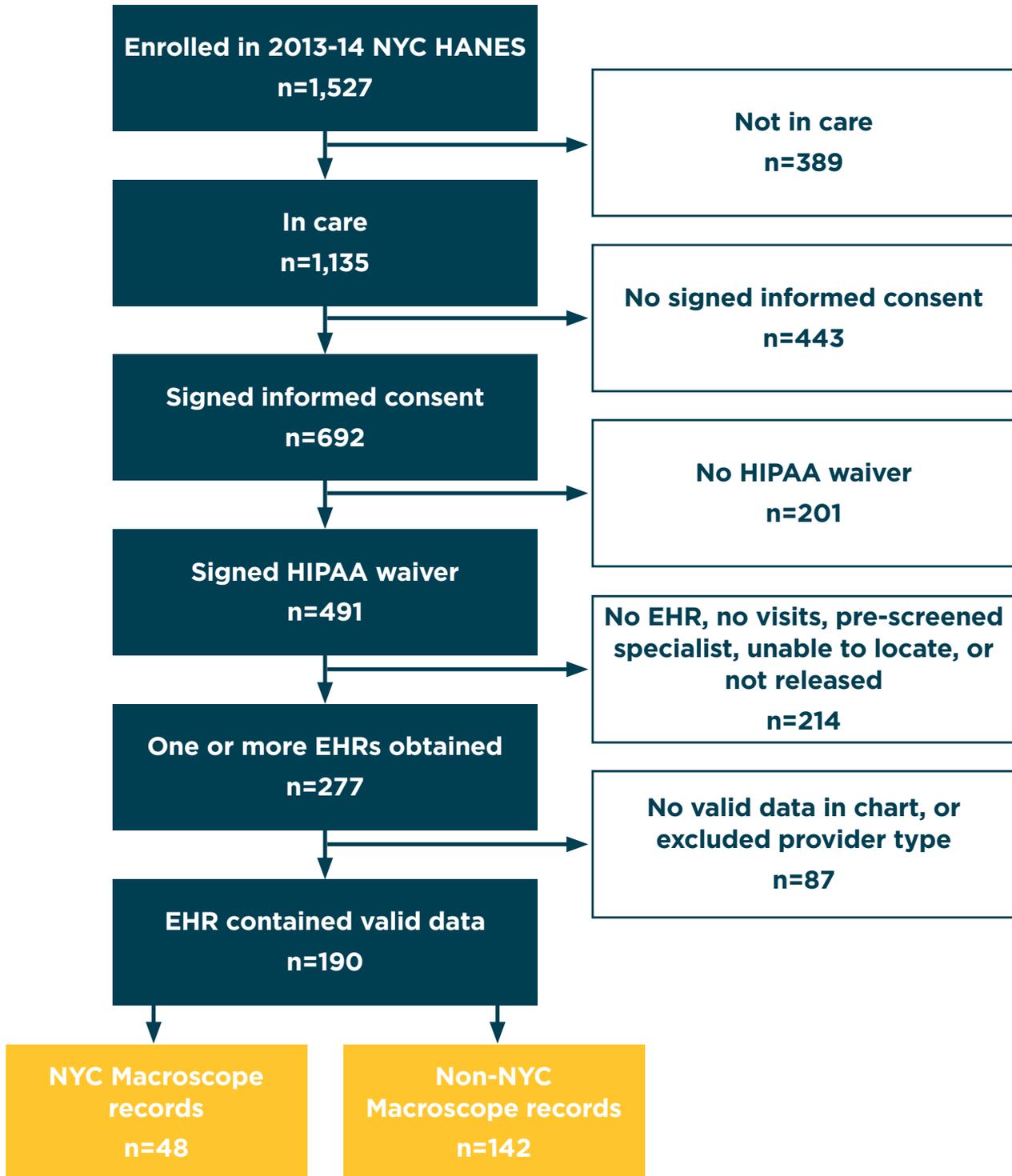
When comparing demographics and health outcomes between people who were included or excluded from the study (Table 1), there were no significant differences with the exception of higher prevalence of hyperlipidemia (diagnosis) in the chart review sample (57.3 percent vs. 45.6 percent; *P* = 0.04). Within the chart review sample, there were no significant differences between participants with NYC Macroscope records and non-NYC Macroscope records. Overall, most participants in the chart review sample were ≥40 years of age (66.8 percent) and female (61.6 percent). In terms of risk factors and chronic diseases, 35.8 percent of the sample were obese, 14.3 percent were smokers, and based on diagnosis/augmented definitions, 15.2 percent/17.8

percent had diabetes, 37.1 percent/43.2 percent had hypertension, and 57.3 percent/62.2 percent had hyperlipidemia. Compared to the unweighted NYC Macroscope sample, the unweighted chart review sample was less likely to be in the 20-39 year age group (-3.1 percentage points), more likely to be female (+5.6 percentage points), and more likely to be from the city's wealthiest neighborhoods (+9.3 percentage points). The weighted chart review sample also had higher prevalence of obesity (+8 percentage points), hypertension (+4.8 percentage points) and hyperlipidemia (+8 percentage points), as measured by NYC HANES, than the weighted NYC Macroscope sample.[4,5,16]

Agreement and validity of indicators

In the total chart review sample combining both NYC Macroscope and non-NYC Macroscope records, indicator agreement between the medical record and NYC HANES ranged from fair (Kappa=0.30 for hyperlipidemia [diagnosis]) to near perfect (Kappa=0.86 for obesity) (Table 2). Sensitivity ranged from moderate (0.71 for hyperlipidemia [diagnosis]) to high (0.94 for diabetes [augmented]) and specificity ranged from low (0.59 for hyperlipidemia [diagnosis]) to high (0.98 for both smoking and diabetes [augmented]). Indicator agreement in the non-NYC Macroscope sample ranged from fair (Kappa=0.30 for hyperlipidemia [diagnosis]) to almost perfect (Kappa=0.89 for diabetes [augmented]). Sensitivity ranged from moderate (0.72 for hyperlipidemia [diagnosis]) to high (0.91 for both obesity and diabetes [augmented]) and specificity was high (>0.90) for all indicators other than hyperlipidemia (0.58 [diagnosis], 0.59 [augmented]). Overall, agreement, sensitivity, and specificity were similar between NYC Macroscope[4,5] and non-NYC Macroscope records.

Figure 1. Participant Inclusion and Exclusion in the Chart Review Study

```
Enrolled in 2013-14 NYC HANES          ──────▶   Not in care
n=1,527                                           n=389

In care                                 ──────▶   No signed informed consent
n=1,135                                           n=443

Signed informed consent                 ──────▶   No HIPAA waiver
n=692                                             n=201

Signed HIPAA waiver                     ──────▶   No EHR, no visits, pre-screened
n=491                                             specialist, unable to locate, or
                                                  not released
                                                  n=214

One or more EHRs obtained               ──────▶   No valid data in chart, or
n=277                                             excluded provider type
                                                  n=87

EHR contained valid data
n=190
    │                    │
    ▼                    ▼
NYC Macroscope       Non-NYC
records              Macroscope records
n=48                 n=142
```

Note: Overall, 190 participants from 2013-14 NYC HANES contributed chart data for review in the present study. Of these participants, 48 had contributed data to the NYC Macroscope EHR surveillance system and 142 did not. EHR, electronic health record; HIPAA, Health Insurance Portability and Accountability Act; NYC HANES, New York City Health and Nutrition Examination Survey.

Table 1. Demographic Characteristics and Health Outcomes of the NYC Macroscope Sample and of the 2013-14 NYC HANES In-Care Participants Stratified by Participation in the Chart Review Study

	NYC MACROSCOPE SAMPLE[a] (N = 716,076)	2013-14 NYC HANES IN-CARE PARTICIPANTS			
		NOT IN CHART REVIEW SAMPLE (N = 945)	OVERALL CHART REVIEW SAMPLE (N = 190)	CHART REVIEW SAMPLE	
				NYC MACROSCOPE (N = 48)	NON-NYC MACROSCOPE (N = 142)
	%	WEIGHTED %	WEIGHTED %	UNWEIGHTED %	UNWEIGHTED %
AGE GROUP					
20-39 years	38.9[b]	36.9	33.2	35.4	35.9
40-59 years	37.1[b]	36.3	38.1	43.8	36.6
≥60 years	24.1[b]	26.8	28.7	18.8	27.5
SEX					
Female	59.1[b]	56.2	61.6	64.6	64.8
Male	40.9[b]	43.8	38.4	35.4	35.2
NEIGHBORHOOD POVERTY LEVEL[c]					
<10% (low)	14.4[b]	23.8	24.7	27.1	22.5
10%-19%	33.7[b]	36.1	34.4	33.3	35.9
20%-29%	28.5[b]	22.9	24.6	20.8	25.4
≥30% (very high)	23.4[b]	17.3	16.3	18.8	16.2
HEALTH OUTCOMES					
Obesity	27.8[d]	30.9	35.8	27.7	37.3
Current Smoking	15.2[d]	18.0	14.3	14.6	15.5
Diabetes (diagnosis)	13.9[d]	13.0	15.2	18.8	14.8
Diabetes (augmented)	15.3[d]	19.1	17.8	23.9	17.3
Hypertension (diagnosis)	32.3[d]	33.7	37.1	31.3	36.6
Hypertension (augmented)	39.2[d]	41.9	43.2	37.5	43.0
Hyperlipidemia (diagnosis)[e]	49.3[d]	45.6	57.3	50.0	54.8
Hyperlipidemia (augmented)	54.5[d]	56.3	62.2	58.3	60.0

[a]Selected estimates are previously published.[4,5,16]
[b]Unweighted estimate.
[c]Proportion of households in one's ZIP code living below the federal poverty threshold per the 2008-12 American Community Survey.
[d]Estimate weighted to NYC HANES in-care population.
[e]The chart review sample is significantly different from non-chart review sample using chi-square test ($P = 0.04$). All other comparisons resulted in a P-value ≥ 0.10.

Table 2. Validity of NYC Macroscope Indicators by Practice[a]

INDICATOR	ALL PRACTICES	NYC MACROSCOPE PRACTICES[b]	NON-NYC MACROSCOPE PRACTICES	
			OVERALL	RESTRICTED TO STAGE 1 MEANINGFUL USE[c]
Obesity	N=159	N=44	N=115	N=72
Kappa	0.86	0.89	0.85	0.94
Sensitivity (95% CI)	0.91 (0.80, 0.97)	0.92 (0.64, 1.00)	0.91 (0.78, 0.97)	0.96 (0.80, 1.00)
Specificity (95% CI)	0.95 (0.89, 0.98)	0.97 (0.83, 1.00)	0.94 (0.86, 0.98)	0.98 (0.88, 1.00)
Smoking Status	N=151	N=43	N=108	N=66
Kappa	0.85	1.00	0.79	0.83
Sensitivity (95% CI)	0.83 (0.63, 0.95)	1.00 (0.54, 1.00)	0.78 (0.52, 0.94)	0.90 (0.56, 1.00)
Specificity (95% CI)	0.98 (0.94, 1.00)	1.00 (0.91, 1.00)	0.98 (0.92, 1.00)	0.96 (0.88, 1.00)
Diabetes (diagnosis)	N=190	N=48	N=142	N=86
Kappa	0.82	0.87	0.80	0.81
Sensitivity (95% CI)	0.87 (0.69, 0.96)	1.00 (0.66, 1.00)	0.81 (0.58, 0.95)	0.83 (0.52, 0.98)
Specificity (95% CI)	0.97 (0.93, 0.99)	0.95 (0.83, 0.99)	0.98 (0.93, 0.99)	0.97 (0.91, 1.00)
Diabetes (augmented)[d]	N=178	N=45	N=133	N=82
Kappa	0.79	0.94	0.89	0.87
Sensitivity (95% CI)	0.94 (0.80, 0.99)	1.00 (0.69, 1.00)	0.91 (0.72, 0.99)	0.86 (0.57, 0.98)
Specificity (95% CI)	0.98 (0.94, 1.00)	0.97 (0.92, 1.00)	0.98 (0.94, 1.00)	0.99 (0.92, 1.00)

CI, confidence interval.
[a]Sample size varied by indicators within each practice type due to (1) missing data in charts (applicable to obesity and smoking only); and (2) non-response or missing data for height and weight, blood pressure, or laboratory values (for obesity and augmented indicators) in NYC HANES.
[b]Previously published.[4,5]
[c]Restricted to records from providers or practices that had attested for Stage 1 Meaningful Use as of December 31, 2013.
[d]The NYC Macroscope data were restricted to providers using an electronic lab interface for at least 10 patients, whereas data in the non-NYC Macroscope records were not.
[e]Restricted to records from participants aged ≥40 years for men or ≥45 years for women at the time of NYC HANES interview.

Table 2. Validity of NYC Macroscope Indicators by Practice[a] (Cont'd)

INDICATOR	ALL PRACTICES	NYC MACROSCOPE PRACTICES[b]	NON-NYC MACROSCOPE PRACTICES	
			OVERALL	RESTRICTED TO STAGE 1 MEANINGFUL USE[c]
Hypertension (diagnosis)	N=190	N=48	N=142	N=86
Kappa	0.79	1.00	0.72	0.79
Sensitivity (95% CI)	0.84 (0.73, 0.92)	1.00 (0.78, 1.00)	0.79 (0.65, 0.89)	0.89 (0.72, 0.98)
Specificity (95% CI)	0.94 (0.89, 0.98)	1.00 (0.89, 1.00)	0.92 (0.85, 0.97)	0.91 (0.81, 0.97)
Hypertension (augmented)	N=190	N=48	N=142	N=86
Kappa	0.74	0.78	0.72	0.71
Sensitivity (95% CI)	0.84 (0.71, 0.89)	0.83 (0.59, 0.96)	0.80 (0.68, 0.89)	0.82 (0.65, 0.93)
Specificity (95% CI)	0.92 (0.85, 0.96)	0.93 (0.78, 0.99)	0.91 (0.83, 0.96)	0.88 (0.77, 0.96)
Hyperlipidemia (diagnosis)[e]	N=110	N=26	N=84	N=52
Kappa	0.30	0.31	0.30	0.23
Sensitivity (95% CI)	0.71 (0.58, 0.82)	0.69 (0.39, 0.91)	0.72 (0.57, 0.84)	0.70 (0.50, 0.86)
Specificity (95% CI)	0.59 (0.44, 0.72)	0.62 (0.32, 0.86)	0.58 (0.41, 0.74)	0.52 (0.31, 0.72)
Hyperlipidemia (augmented)[d,e]	N=103	N=23	N=80	N=50
Kappa	0.41	0.37	0.41	0.30
Sensitivity (95% CI)	0.80 (0.68, 0.89)	0.77 (0.46, 0.95)	0.81 (0.67, 0.91)	0.78 (0.58, 0.91)
Specificity (95% CI)	0.60 (0.43, 0.74)	0.60 (0.26, 0.88)	0.59 (0.41, 0.76)	0.52 (0.31, 0.73)

CI, confidence interval.

[a]Sample size varied by indicators within each practice type due to (1) missing data in charts (applicable to obesity and smoking only); and (2) non-response or missing data for height and weight, blood pressure, or laboratory values (for obesity and augmented indicators) in NYC HANES.

[b]Previously published.[4,5]

[c]Restricted to records from providers or practices that had attested for Stage 1 Meaningful Use as of December 31, 2013.

[d]The NYC Macroscope data were restricted to providers using an electronic lab interface for at least 10 patients, whereas data in the non-NYC Macroscope records were not.

[e]Restricted to records from participants aged ≥40 years for men or ≥45 years for women at the time of NYC HANES interview.

Sensitivity analyses

Restriction to practices attesting Stage 1 Meaningful Use

When the non-NYC Macroscope sample was restricted to providers who attested to Stage 1 Meaningful Use (Table 2), sensitivity of the smoking indicator increased by 0.12 (from 0.78 to 0.90) and the sensitivity of the hypertension (diagnosis) indicator increased by 0.10 (from 0.79 to 0.89). Differences in other indicators were smaller and went in both directions. Differences in specificity were negligible.

Unstructured data

When unstructured data were incorporated into the NYC Macroscope indicators for obesity and smoking status, the sample sizes increased by 1 and 3 records, respectively, and sensitivity changed minimally by 0.02. The incorporation of unstructured data into metabolic indicators had no or little impact on sensitivity of diabetes diagnosis, hypertension diagnosis, and all augmented indicators, but increased the sensitivity of the hyperlipidemia diagnosis indicator by 0.11. Across all metabolic conditions, all of the unstructured diagnoses (9 cases) were obtained from free text fields instead of scanned documents. (Full data not shown for this sensitivity analysis).

Discussion

Overall, the indicators developed for the NYC Macroscope had generally good sensitivity and specificity in a sample of EHR records from outside of the NYC Macroscope when using NYC HANES as a reference. Relative to previously published data from a sample of 48 NYC Macroscope records,[4,5] Kappa coefficients in the non-NYC Macroscope sample ranged from 0.21 points lower to 0.05 points higher than in the NYC Macroscope sample. However,

the impact of these quantitative differences on the characterization of agreement was minimal. We also found that restricting non-NYC Macroscope records to providers who attested to Stage 1 Meaningful Use had mixed influence on agreement across indicators, including no meaningful impact on agreement for diabetes indicators and augmented hypertension, improved agreement for obesity, smoking status, and hypertension diagnosis indicators, and decreased agreement for both hyperlipidemia indicators. Contrary to findings from a study of behavioral health records,[24] we found that incorporating unstructured data did not improve indicator sensitivity. Our results suggest that several of the indicators developed for the NYC Macroscope EHR surveillance system are generally robust for application to populations outside of the Hub Population Health System. Jurisdictions outside of NYC can learn from these findings to accelerate the development of their own EHR surveillance systems and validation of indicators.

Among indicators, estimates of obesity prevalence had the best agreement between the NYC Macroscope and NYC HANES, as shown by the high Kappa coefficients, sensitivity, and specificity across all subsamples. Based on these data and findings from other studies,[3,10,25,26] we expect that, in most situations, EHR-based adult obesity prevalence estimates will accurately reflect the underlying prevalence in the population sampled. Observed differences at the population level will most likely reflect differences in sample composition rather than measurement error, as measurement error associated with height and weight in the primary care setting appears to be nonsystematic and is not differential by socioeconomic status, for example.[27] Other jurisdictions may want to use the obesity indicator to carry out preliminary evaluations of the representativeness of their EHR data source before carrying out a full-scale multi-indicator validation study.

The smoking status indicator had good validity and agreement overall, with improved sensitivity when restricting the sample to records from providers who had attested to Stage 1 Meaningful Use. Smoking status sensitivity in the Stage 1 Meaningful Use sample, although lower than in NYC Macroscope records, was slightly higher than reported in two earlier studies.[11,28] One of these studies (Chen et al.) reported improvements in smoking indicator sensitivity over time.[28] Previous studies reported that roughly one-third of primary care patients lacked data on smoking status in structured fields;[5,29,30] however, we anticipate that further uptake of Meaningful Use provisions will lead to less missing data on smoking status and in turn improve sensitivity. As a cautionary note, misclassification may be a problem among those with documented smoking status.[31,32] For example, a former smoker might be misclassified as a current smoker because the provider did not update the patient's smoking status in the EHR. Therefore, further efforts are warranted to improve documentation quality.

There was good agreement between the two data sources for diabetes indicators in both the non-NYC Macroscope sample overall and the restricted non-NYC Macroscope sample, with the augmented indicator having slightly better agreement than the diagnosis indicator. The observation that sensitivity was at least 0.14 points lower in non-NYC Macroscope records than in the NYC Macroscope sample supports our hypothesis that the unique relationship between the NYC DOHMH and Hub Population Health System providers contributes to high validity of the NYC Macroscope.[4,16,33] However, diabetes indicators in both the full and restricted non-NYC Macroscope samples had high sensitivity, suggesting that these indicators may be acceptable for use in most other EHR systems, with or without Stage 1 Meaningful Use provider restriction, and may need only limited local validation. Our finding of

the robustness of diabetes indicators is consistent with two studies from Britain; one compared self-reported diabetes diagnoses obtained in a survey of patients with EHR data[13] and the other compared the diabetes status classifications assigned by an algorithm with the complete medical record.[34] Both studies found strong agreement between the two data sources. In addition, a study using data from the Women's Health Initiative found a high level of agreement between self-reported diabetes and paper medical records.[35]

The external validity of hypertension indicators was weaker relative to obesity, smoking status, and diabetes indicators. Results for the augmented indicator, overall, were similar to those observed in the NYC Macroscope sample, when restricting non-NYC Macroscope records to Stage 1 Meaningful Use attestation, and similar to results for hypertension indicators used in other studies.[7,10] The diagnosis indicator, on the other hand, was not as sensitive in the non-NYC Macroscope sample, especially when the records were not restricted to Stage 1 Meaningful Use. The higher sensitivity among the NYC Macroscope records might have been the result of NYC DOHMH's data driven quality improvement efforts in improving blood pressure documentation by providers who contribute data to the NYC Macroscope relative to data from other providers. For this reason, we encourage jurisdictions interested in using EHR data for hypertension prevalence to devote resources to local validation research.

The hyperlipidemia indicators had low specificity in non-NYC Macroscope and NYC Macroscope samples alike. In both samples, sensitivity was borderline low for the diagnosis indicator and on the low side of moderate for the augmented indicator. There are many possible explanations for the poor validity of the hyperlipidemia indicators, including diagnoses in the EHR that are not captured in NYC HANES, which seems to be a more pervasive

issue for hyperlipidemia than for diabetes and hypertension;[36-39] the study's 2-year laboratory look-back period, which is shorter than the approximate 5-year recommended cholesterol testing interval;[40] and factors guiding diagnostic decisions that are not captured in these indicators such as low-density lipoprotein levels and 10-year cardiovascular disease risk.[41] If, contrary to our designation of NYC HANES as the reference standard, we accept the diagnoses in the EHR as accurate, the true hyperlipidemia prevalence could be substantially higher than the NYC HANES estimates suggest. Continued work is needed to develop a strong hyperlipidemia surveillance indicator.

An important limitation of this study was the small size of our sample and low statistical power. Even after combining NYC Macroscope and non-NYC Macroscope records, confidence intervals for sensitivity spanned more than 18 percentage points on average, making comparisons across indicators speculative at best. Another limitation was the year of data collection relative to secular changes in health information technology advances. The NYC Macroscope was designed in 2012 and data were captured in 2013 when Stage 1 Meaningful Use was the highest level of attestation. Documentation quality is likely to have improved considerably as the field moves towards the third and final stage of Meaningful Use attestation.[23] It may be that the sensitivity of some of the measures, for example, smoking status, hypertension, and hyperlipidemia, is now higher than our findings suggest. We should also point out that our study was restricted to New York City residents. There may be factors that we have not assessed that would limit the generalizability of our findings to EHR systems in other geographic regions.

This study had several strengths. The chart review sample was small, but not significantly different from the entire NYC HANES cohort with regard to sociodemographic or health characteristics other than having higher prevalence of hyperlipidemia. Our results were consistent across provider subgroups. We evaluated 8 different indicators, 4 of which were based on direct physical or laboratory measurements (BMI and the 3 augmented measures). We also evaluated our indicators with and without provider inclusion criteria pertaining to documentation quality. Another strength is the study's innovative design. The majority of EHR-based surveillance system validation studies are limited to comparisons of population-based prevalence estimates,[2,3,8,10,11,30,42-44] inclusive of both measured and self-reported data. Chart review validation studies using individual-level data have only been carried out for a handful of EHR-based surveillance systems.[6,7,9,11-13] Unlike these self-contained studies that compared indicators using EHR data with review of the complete medical record or patient self-report, our study compared indicators using EHR data with individual-level data from a gold standard population-based survey that employed a physical examination and laboratory testing. Thus, our findings offer a bridge between examination surveys and EHR-based surveillance. Furthermore, our findings are not specific to any single EHR or health care delivery system, as we demonstrated the external validity of NYC Macroscope indicators in a sample of 133 providers and 89 practices that use more than 20 different EHR software systems.

Conclusion

Our findings suggest that many of the EHR-based surveillance indicators developed and validated for the NYC Macroscope are generalizable for use in other EHR surveillance systems; that provider inclusion/exclusion criteria based on attesting to Stage 1 Meaningful use has variable impact on indicator validity; and, that incorporation of unstructured data into EHR-based surveillance indicators may not be cost effective, given the little

added value seen in this study and the expense and difficulty of incorporating natural language processing. Other studies, however, have found that unstructured data may contain speculative and diagnostic data that can be missed by algorithms.[9,24,45] Thus, other jurisdictions should determine at a local level if unstructured data have a relevant impact on indicator validity. The indicators assessed in this study may be transferable to other settings, but we would recommend that other jurisdictions developing their own EHR surveillance systems start with a small chart review of records from contributing practices to help inform indicator development. The results from this study can serve as a guide as they develop and validate indicators specific to their own EHR surveillance system.

Acknowledgements

References

1. Castrucci BC, Rhoades EK, Leider JP, Hearne S. What gets measured gets done: an assessment of local data uses and needs in large urban health departments. *J Public Health Manag Pract*. 2015;21 Suppl 1:S38-48.
2. Tomasallo CD, Hanrahan LP, Tandias A, et al. Estimating Wisconsin asthma prevalence using clinical electronic health records and public health data. *Am J Public Health*. 2014;104(1):e65-e73.
3. Bailey LC, Milov DE, Kelleher K, et al. Multi-institutional sharing of electronic health record data to assess childhood obesity. *PLoS One*. 2013;8(6):e66192.
4. Thorpe LE, McVeigh KH, Perlman SE, 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*. 2016;4(1):28.
5. 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*. 2016;4(1):27.
6. Bailey SR, Heintzman JD, Marino M, et al. Measuring preventive care delivery: Comparing rates across three data sources. *Am J Prev Med*. 2016;51(5):752-761.
7. Williamson T, Green ME, Birtwhistle R, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med*. 2014;12(4):367-372.
8. Booth H, Prevost A, Gulliford M. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf*. 2013;22(12):1357-1361.
9. Coleman N, Halas G, Peeler W, et al. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract*. 2015;16:11.
10. Zellweger U, Bopp M, Holzer BM, et al. Prevalence of chronic medical conditions in Switzerland: exploring estimates validity by comparing complementary data sources. *BMC Public Health*. 2014;14:1157.
11. McGinnis KA, Brandt CA, Skanderson M, et al. Validating smoking data from the Veteran's Affairs Health Factors dataset, an electronic data source. *Nicotine Tob Res*. 2011;13(12):1233-1239.
12. Butt DA, Tu K, Young J, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. *Neuroepidemiology*. 2014;43(1):28-37.
13. Barber J, Muller S, Whitehurst T, et al. Measuring morbidity: self-report or health care records? *Fam Prac*. 2010;27(1):25-30.
14. Buck MD, Anane S, Taverna J, et al. The Hub Population Health System: distributed ad hoc queries and alerts. *J Am Med Inform Assoc*. 2012;19(e1):e46-e50.

15. 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.

16. 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*. 2016;4(1):26.

17. 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.

18. Centers for Medicare & Medicaid Services. National Plan and Provider Enumeration System (NPPES). https://npiregistry.cms.hhs.gov/. Accessed September 18, 2016.

19. City University of New York School of Public Health. New York City Health and Nutrition Examination Survey (NYC HANES). http://nychanes.org/. Acessed September 18, 2016.

20. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33(2):363-374.

21. Centers for Medicare & Medicaid Services. EHR Incentive Programs: Requirements for Previous Years. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/RequirementsforPreviousYears.html. Updated August 24, 2016. Accessed September 18, 2016.

22. New York State Department of Health. Medicaid Electronic Health Records Incentive Program Provider Payments: Beginning 2011. https://health.data.ny.gov/Health/Medicaid-Electronic-Health-Records-Incentive-Progr/6ky4-2v6j. Accessed September 18, 2016.

23. Centers for Medicare & Medicaid Services. Medicare Electronic Health Record (EHR) Incentive Program. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentivePrograms. Updated August 16, 2016, Accessed September 18, 2016.

24. Wu CY, Chang CK, Robson D, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. *PLoS One*. 2013;8(9):e74262.

25. Flood TL, Zhao YK, Tomayko EJ, et al. Electronic health records and community health surveillance of childhood obesity. *Am J Prev Med*. 2015;48(2):234-240.

26. Sidebottom AC, Johnson PJ, VanWormer JJ, et al. Exploring electronic health records as a population health surveillance tool of cardiovascular disease risk factors. *Popul Health Manag*. 2015;18(2):79-85.

27. Lyratzopoulos G, Heller RF, Hanily M, Lewis PS. Risk factor measurement quality in primary care routine data was variable but nondifferential between individuals. *J Clin Epidemiol*. 2008;61(3):261-267.

28. Chen LH, Quinn V, Xu L, et al. The accuracy and trends of smoking history documentation in electronic medical records in a large managed care organization. *Subst Use Misuse*. 2013;48(9):731-742.

29. Greiver M, Aliarzadeh B, Meaney C, et al. Are we asking patients if they smoke? Missing information on tobacco use in Canadian electronic medical records. *Am J Prev Med*. 2015;49(2):264-268.

30. Linder JA, Rigotti NA, Brawarsky P, et al. Use of practice-based research network data to measure neighborhood smoking prevalence. *Prev Chronic Dis*. 2013;10:E84.

31. Murray RL, Coleman T, Antoniak M, et al. The potential to improve ascertainment and intervention to reduce smoking in primary care: a cross sectional survey. *BMC Health Serv Res*. 2008;8:6.

32. Garies S, Jackson D, Aliarzadeh B, et al. Sentinel eye: improving usability of smoking data in EMR systems. Can Fam Physician. 2013;59(1):108, e60-e61.

33. McVeigh KH, Newton-Dame R, Perlman SE, et al. *Developing an electronic health record-based population health surveillance system*. New York: New York City Departmnt of Health and Mental Hygiene, July 2013. Available from: https://www1.nyc.gov/assets/doh/downloads/pdf/data/nyc-macro-report.pdf. Accessed September 18, 2016.

34. Makam AN, Nguyen OK, Moore B, et al. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Med Inform Decis Mak*. 2013;13:81.

35. Jackson JM, DeFor TA, Crain AL, et al. Validity of diabetes self-reports in the Women's Health Initiative. *Menopause*. 2014;21(8):861-868.

36. Martin LM, Leff M, Calonge N, et al. Validation of self-reported chronic conditions and health services in a managed care population. *Am J Prev Med*. 2000;18(3):215-218.

37. Chun H, Kim IH, Min KD. Accuracy of self-reported hypertension, diabetes, and hypercholesterolemia: Analysis of a representative sample of Korean older adults. *Osong Public Health Res Perspect*. 2016;7(2):108-115.

38. Huerta JM, Tormo MJ, Egea-Caparrós JM, et al. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipemia en población adulta española. Resultados del estudio DINO. *Rev Esp Cardiol*. 2009;62(2):143-152.

39. Van Eenwyk J, Bensley L, Ossiander EM, Krueger K. Comparison of examination-based and self-reported risk factors for cardiovascular disease, Washington State, 2006-2007. *Prev Chronic Dis*. 2012;9:E117.

40. US Preventive Services Task Force. Clinical Summary: Lipid Disorders in Adults (Cholesterol, Dyslipemia): Screening, June 2008. http://www.uspreventiveservicestaskforce.org/Page/Document/ClinicalSummaryFinal/lipid-disorders-in-adults-cholesterol-dyslipidemia-screening. Published June 2008. Accessed September 18, 2016.

41. 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.

42. Esteban-Vasallo MD, Dominguez-Berjon MF, Astray-Mochales J, et al. Epidemiological usefulness of population-based electronic clinical records in primary care: estimation of the prevalence of chronic diseases. *Fam Pract*. 2009;26(6):445-454.

43. Szatkowski L, Lewis S, McNeill A, et al. Can data from primary care medical records be used to monitor national smoking prevalence? *J Epidemiol Community Health*. 2012;66(9):791-795.

44. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health*. 2013;13:251.

45. Ford E, Nicholson A, Koeling R, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol*. 2013;13:105.

Appendix

Table A1. Indicators in 2013 NYC Macroscope and 2013-14 NYC HANES

INDICATOR	NYC MACROSCOPE[a]	NYC HANES
Obesity	Most recent BMI ≥30 recorded in structured fields within look-back period.	BMI ≥30 calculated from height and weight measured at interview.
Smoking Status	Indication of current smoking in the most recently recorded structured smoking field within the look-back period.	Smoked at least 100 cigarettes in lifetime and currently smoke every day or some days.
Diabetes (diagnosis)	Any ICD-9 diagnosis of diabetes ever recorded in the problem list or assessment section of the EHR. Patients without the diagnosis were coded as not having the condition.	Ever been told to have diabetes by a health care provider. Participants with a "don't know" response were coded as not having the condition.
Diabetes (augmented)	Last A1C measured in look-back period ≥6.5; or any ICD-9 diagnosis of diabetes; or medication prescribed in look-back period.	A1C measured at interview ≥6.5 or ever been told to have diabetes by a health care provider.
Hypertension (diagnosis)	Any ICD-9 diagnosis of hypertension ever recorded in the problem list or assessment section of the EHR. Patients without the diagnosis were coded as not having the condition.	Ever been told to have hypertension by a health care provider. Participants with a "don't know" response were coded as not having the condition.
Hypertension (augmented)	Last blood pressure measured in look-back period with systolic reading ≥ 140 mmHg or diastolic reading ≥ 90 mmHg; or any ICD-9 diagnosis of hypertension; or, medication prescribed in look-back period.	Blood pressure measured at interview with systolic reading ≥ 140 mmHg or diastolic reading ≥ 90 mmHg; or ever been told to have hypertension by a health care provider.
Hyperlipidemia (diagnosis)[b]	Any ICD-9 diagnosis of Hyperlipidemia ever recorded in the problem list or assessment section of the EHR. Patients without the diagnosis were coded as not having the condition.	Ever been told to have hyperlipidemia by a health care provider. Participants with a "don't know" response were coded as not having the condition.
Hyperlipidemia (augmented)[b]	Last total cholesterol measured in look-back period ≥240 mg/dL (6.20 mmol/L); or any ICD-9 diagnosis of hyperlipidemia; or medication prescribed in look-back period.	Total cholesterol measured at interview ≥240 mg/dL (6.20 mmol/L) or ever been told to have hyperlipidemia by a health care provider.

BMI, body mass index; EHR, electronic health record; ICD-9, International Classification of Diseases Ninth Revision.
[a]All NYC Macroscope indicators are derived from data stored in structured fields that have been recorded within the NYC Macroscope indicator-specific look-back periods. The look-back period extended from the date of the NYC HANES interview back 1 year for BMI, smoking status, blood pressure, and prescribed medications, and 2 years for A1C and total cholesterol.
[b]Restricted to men ≥40 years of age and women ≥45 years of age.