

VALIDATING THE USE OF ELECTRONIC HEALTH RECORD DATA FOR POPULATION HEALTH SURVEILLANCE

Katharine H. McVeigh, PhD, MPH¹; Remle Newton-Dame, MPH²; Lorna E. Thorpe, PhD³; Sharon E. Perlman, MPH¹; Lauren Schreiberstein, MA²; Elisabeth Snell, MPH¹; Claudia Chernov, MPH¹

1) Division of Epidemiology, 2) Primary Care Information Project, New York City Department of Health & Mental Hygiene; and 3) City University of New York School of Public Health

Introduction:

Networked electronic health records (EHRs) could play an important role in population health surveillance by providing data about :

- disease prevalence, management and control
- risk factor prevalence
- uptake of preventive services

The NYC Macroscopic project is evaluating the validity of EHR-derived prevalence estimates for 13 outcomes:

- prevalence, management and control of hypertension, high cholesterol, and diabetes
- the prevalence of obesity, smoking and depression
- the use of preventive services (flu shots)

Research Objective:

To evaluate the construct validity of an EHR-based indicator of smoking prevalence relative to an established survey measure

Data Sets and Sources:

NYC Macroscopic estimates of smoking prevalence for 2013 are generated from EHR data aggregated across 467,983 patients at 384 practices throughout New York City. Reference survey estimates are obtained from the:

- 2013-14 New York City Health and Nutrition Examination Survey (NYC HANES; N = 1,543; 1,135 in care)
- 2013 NYC Community Health Survey (CHS; N = 8,698; 6,531 in care)

Study Design:

This analysis compares NYC Macroscopic prevalence estimates, weighted to the age group, sex and neighborhood poverty distribution of the NYC adult population ages 20 and older that has seen a doctor in the past year (population in care), with estimates from the NYC HANES and CHS in-care populations

Analysis:

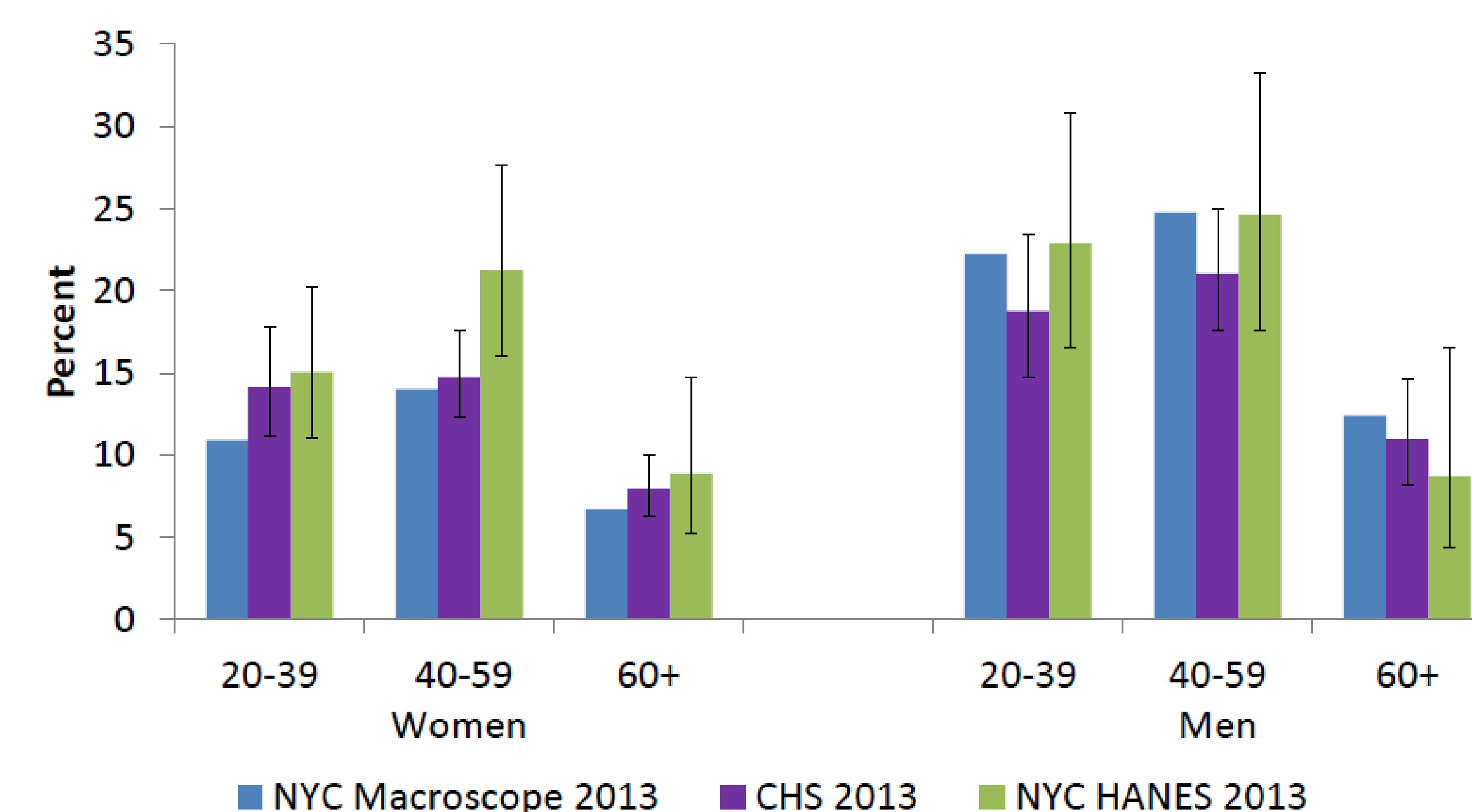
NYC Macroscopic estimates have been stratified into 6 groups defined by age group (3 levels) and sex (2 levels).

For each stratum, we compared:

- 1) The NYC Macroscopic estimate against those of reference survey estimate
- 2) The prevalence ratio and the absolute and standardized differences between the two estimates

Summary measures of goodness of fit include the proportion of strata with significantly different estimates, the average differences between estimates across all 6 strata, and the Spearman correlation coefficient

Smoking Prevalence among Adult New Yorkers who Visited a Doctor in the Past Year



Indicator/Criteria	Total Estimate Comparison					Summary of Comparisons Across 6 Strata				
	CHS Estimate % (95% CI)	HANES Estimate % (95% CI)	Yes	0.85-1.15	+/- 5 points	5/6 (83%)	0.85-1.15	5 points	1.96 SE	r _s < 0.8
CHS vs HANES	CHS Estimate % (95% CI)	HANES Estimate % (95% CI)	CHS within HANES 95% CI?	Prevalence Ratio CHS/HANES	Difference (CHS - HANES)	CHS within HANES 95% CI?	Prevalence Ratio Mean (range)	Absolute Difference Mean (range)	Standardized Difference Mean (range)	Spearman Correlation
CHS and NYC HANES: smoked 100 cigarettes in lifetime and currently smoke daily or some days	14.91 (13.58-16.33)	17.36 (14.73-20.35)	Yes	0.86	-2.45	5/6 (83%)	0.91 (0.69 - 1.26)	2.81 (0.75 - 6.51)	0.97 (0.39-2.20)	r _s = 0.89
Macroscopic vs. HANES	Macroscopic Estimate %	HANES Estimate % (95% CI)	Macroscopic within HANES 95% CI?	Prevalence Ratio Macro/HANES	Difference (Macro - HANES)	Macroscopic within HANES 95% CI?	Prevalence Ratio Mean (range)	Absolute Difference Mean (range)	Standardized Difference Mean (range)	Spearman Correlation
Macroscopic: Current smoker. NYC HANES: smoked 100 cigarettes in lifetime and currently smoke daily or some days.	15.33	17.36 (14.73-20.35)	Yes	0.88	-2.03	4/6 (67%)	0.89 (0.71 - 1.21)	2.90 (0.31 - 7.48)	1.10 (0.04-2.45)	r _s = 0.76
Macroscopic vs. CHS	Macroscopic Estimate %	CHS Estimate % (95% CI)	Macroscopic within CHS 95% CI?	Prevalence Ratio Macro/CHS	Difference (Macro - CHS)	Macroscopic within CHS 95% CI?	Prevalence Ratio Mean (range)	Absolute Difference Mean (range)	Standardized Difference Mean (range)	Spearman Correlation
Macroscopic: Current smoker. CHS: smoked 100 cig lifetime and currently smoke daily or some days.	15.33	14.91 (13.58-16.33)	Yes	1.03	0.42	5/6 (83%)	1.01 (0.78 - 1.17)	2.19 (0.47 - 3.58)	1.32 (0.35 - 1.88)	r _s = 0.94

Principal Findings:

Benchmarking comparisons for the same self-reported measure of smoking behavior between CHS and HANES showed excellent alignment.

- Compared to NYC HANES, Macroscopic estimates were
- Not significantly different from NYC HANES estimates overall
 - Were consistently lower than reference estimates in women but not men, with some significant differences

The fit between NYC Macroscopic and CHS was extremely good and all fit criteria were met

Implications for Public Health Practice and Policy

- Estimates of smoking prevalence derived from EHR data are comparable to those from telephone and examination surveys, but may be somewhat lower than survey estimates for women
- These results are promising, but future studies assessing the sensitivity and specificity of the EHR-based indicator, both within the Macroscopic context and across EHR vendors, are needed before validity can be established
- If we are able to demonstrate similar results across a variety of NYC Macroscopic indicators, we have the potential to revolutionize local health surveillance

