Missing Data in Electronic Health Records: Implications for Population Monitoring

Remle Newton-Dame, MPH[1]; Lauren Schreibstein, MA[1]; Katharine H. McVeigh, PhD, MPH[2]; Sharon E. Perlman, MPH[2]; Lorna E. Thorpe, PhD[3]
1) Primary Care Information Project and 2) Division of Epidemiology, New York City Department of Health & Mental Hygiene; 3) City University of New York School of Public Health

Building the Macroscope Electronic Health Record Surveillance System

Define Priority Indicators for Population Health from Available Clinical Data

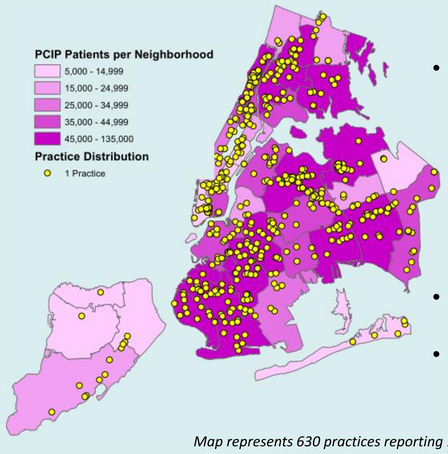
- Electronic Health Record (EHR) indicators were developed to reflect significant sources of morbidity and mortality in NYC: hypertension, cholesterol, diabetes, smoking, depression, influenza and obesity

Indicator of Population Health	NYC HANES 2013	NYC Macroscope
Prevalence/Diagnosis		
Hypertension	Measured BP ≥ 140/90 **OR** prescribed antihypertensive medication in past year	Last BP ≥ 140/90 in past year **OR** prescribed antihypertensive medication in past year
Hypertension expanded prevalence	Measured BP ≥ 140/90 **OR** prescribed antihypertensive medication in past year **OR** ever told BP high	Last BP ≥ 140/90 in past year **OR** prescribed antihypertensive medication in past year **OR** Dx of hypertension ever entered in assessments
Hypertension history/diagnosis	Ever told BP high	Dx of hypertension ever entered in assessments

- All indicators will be compared to similar metrics from the NYC Health And Nutrition Examination Survey (NYC HANES 2013)

Convert to vendor-specific electronic health record queries

Push SQL Queries to Medical Practices Through the Hub Population Health System (the Hub)

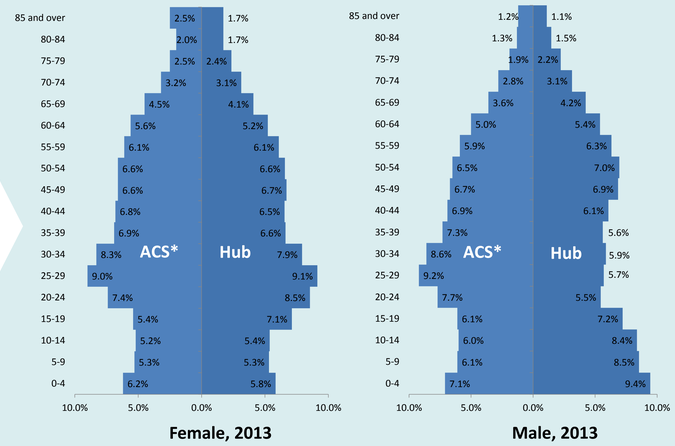


- 700 independent ambulatory practices participating in the Primary Care Information Project (PCIP) and using eClinicalWorks EHR
- 4000+ providers
- 1.9 million patients

Map represents 630 practices reporting 1,628,136 patients with an office visit in 2012 (October 22, 2013)

Receive aggregate patient counts by practice

Compare to Gold Standard at the Patient and Pop Levels



**American Community Survey estimates of the NYC population for 2013*

METHODS: MISSING DATA SUB-STUDY

Research Objective
To understand the potential biases in EHR data introduced by missing data between indicators, patient sub-populations and provider type.

Study Design
For NYC Hub patients aged 20 to 100 years, we calculated percent missing blood pressure, smoking status and total cholesterol laboratory results in 2013, among all patients and those with hypertension/hyperlipidemia. To reflect national recommendations, cholesterol included men aged >=35 and women aged >=45.

Analysis
Chi squared tests and t-tests were used to identify significant differences in level of missing across indicator, provider type and patient group.

Inclusion Criteria
The NYC Macroscope surveillance system uses provider-level inclusion criteria to maximize data quality. Hub providers that contribute data to the NYC Macroscope are:

1. Primary Care (practicing internal medicine, family medicine, pediatrics, geriatrics)
2. Proficient in EHR documentation, aka "Supercohort," modeled on Meaningful Use Stage 1 criteria and literature review:
   - >10 patients seen in 2013
   - Vitals (blood pressure or body mass index) recorded in >=50% of patients
   - Diagnosis code recorded for >=80% of patients
   - Medication recorded for >=20% of patients

The NYC Macroscope queries were returned by 660 practices and 2,229 providers. After applying inclusion criteria, 853 providers and 386 practices remained in the sample. These providers saw 605,118 New Yorkers aged 20-100 in 2013.

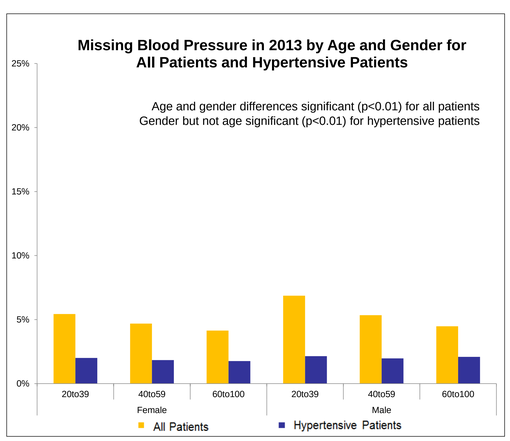
FINDINGS

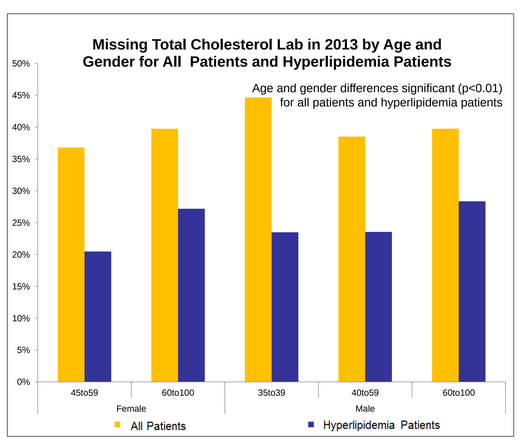
Missing Data Varies Significantly by Provider Specialty and Documentation Ability

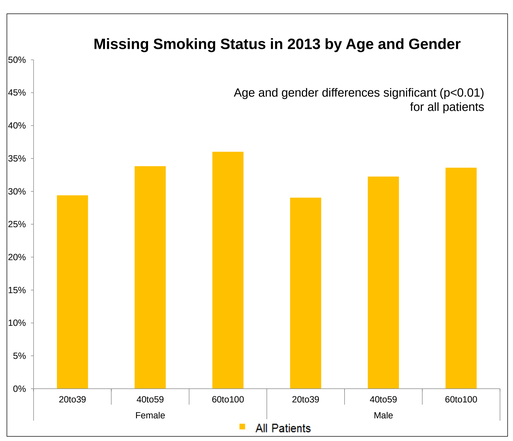
2013 % Missing	Blood Pressure	Cholesterol	Smoking Status
All providers	11%	52%	33%
Primary Care	6%	40%	33%
Specialist	18%[t]	69%[t]	33%
Supercohort	6%	50%	32%
Non-Supercohort	51%[t]	77%[t]	47%[t]
Primary Care & Supercohort	5%	39%	32%

[t] significant at p<0.01

Missing Data Varies Significantly by Age, Gender and Disease Status: Results from the Supercohort Primary Care Providers







CONCLUSIONS

Missing data is a challenge for population health monitoring

The impact of missing data may be mitigated by focusing on:

1. Areas of the EHR that are well populated (vitals, diagnosis)
2. Higher risk populations
   - Patients with relevant diagnosis
   - Younger patients for smoking, older patients for blood pressure
3. Primary care providers
   - Specialists significantly less likely to document chronic disease
4. Appropriately bounded indicators
   - Cholesterol may perform poorly due to a one year look-back period - screening is recommended every five years by USPSTF

For more information on the NYC Macroscope, please contact us at nycmacroscope@health.nyc.gov
This work is supported by the Robert Wood Johnson Foundation, the de Beaumont Foundation, the New York State Health Foundation, the Robin Hood Foundation, the Doris Duke Foundation and the Centers for Disease Control and Prevention