Public Health and Population Health: Leveraging Electronic Health Record Data for Local Population Health Surveillance

by

Emily V. McCormick

B.S., Georgia State University, 2006

M.P.H, Emory University, 2008

A dissertation submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Clinical Sciences

2018
This dissertation for the Doctor of Philosophy degree by

Emily V. McCormick

has been approved for the

Clinical Sciences Program

by

Heather Haugen, Chair

Arthur Davidson, Advisor

John Douglas

Catherine Battaglia

Michael Kahn

Date: May 18, 2018
McCormick, Emily V. (Ph.D., Clinical Sciences)

Public Health and Population Health: Leveraging Electronic Health Record Data for Local Population Health Monitoring

Dissertation directed by Associate Professor Arthur J. Davidson

ABSTRACT

Background: While measurement is a population health priority, traditional data sources provide limited local health insights. Local public health agencies (LPHAs) are evaluating the feasibility of using electronic health record (EHR) data, which could provide more detailed health data in a timelier manner, but must be further explored and validated for surveillance uses.

Purpose: A mixed methods study was conducted to assess LPHAs’ capacity for EHR-based surveillance, conduct a comparison of EHR-based estimates, and explore more granular, EHR-based, population health insights.

Methods: Twenty five LPHA interviews were conducted to understand EHR-based surveillance practices and attitudes. Quantitative methods were then used to validate EHR-based chronic disease prevalence estimates (i.e., diabetes and hypertension) through comparison with survey-based estimates, apply population weights to generate adjusted prevalence estimates, and pilot county and census tract estimation of chronic disease control. EHR data from three healthcare providers for Denver adults (2011-2015) were extracted, geocoded, transformed and analyzed at the county and census tract level.
Results: LPHAs believe EHRs hold meaningful local health data, but less than 20% are using EHR-based surveillance. EHR data suggested that adult diabetes prevalence significantly increased from 10% in 2011 to 11% in 2015 (p≤0.0001) and was higher than survey’s estimate (7% in 2015). EHR-based hypertension prevalence estimates declined from 20% in 2011 to 19% in 2015 (p≤0.0001) and were below survey based estimates (27% in 2015). Census tract hypertension prevalence estimates spanned 4.3% to 41.7%. EHR-based diabetes control decreased from 80% in 2011 to 78% in 2015 (p≤0.0001) and hypertension control decreased from 74% in 2011 to 71% in 2015 (p≤0.0001).

Conclusions: LPHAs face many challenges in EHR-based surveillance and need staff with advanced skills to utilize EHR data to measure new and existing indicators at a granular geographic level. EHR-based prevalence estimates were not statistically equivalent to population surveys. EHR-based prevalence and control estimates were sensitive to year over year changes and socio-demographic disparities and produced novel census tract-level information.

The form and content of this abstract are approved. I recommend its publication.

Approved: Arthur Davidson
DEDICATION

This dissertation would not have been possible without tireless support from a many extraordinary human beings. This work is dedicated to all of them.

First and foremost, my fiancé Matt (probably my husband by now), who flung himself on the sword many times to read draft after draft of this document and remind me that completing it was not only possible but inevitable. Thank you for the lunches and dinners, for picking up the dry cleaning, going to the grocery store, and watering the plants. This man was crazy enough to go on a first date with someone employed full time while pursuing a PhD in the early mornings, late nights, and on the weekends. Thank you for believing in me. You will never know how much your support has meant.

My close friends and family, who followed my progress and allowed me to process the PhD experience during our holiday meals and gatherings. I’m certain that many dinner guests could have survived without a deep discussion concerning electronic health records and the trials and tribulations of several years in academia. Thank you for caring about something so unrelated to most people’s everyday life. Thank you for checking in and asking about classes, tests, papers, proposals, preliminary exams, comps, and this very long document.

I must thank my colleagues at Denver Public Health, who provided me endless amounts of support and lent me their courage and confidence when I headed off to class. I appreciate their discretion when they found me asleep or unraveled on or beneath my
desk. Their persistent and warm inquiries about my progress and sanity kept me moving forward at many difficult points.

My cohort, Rose, Melanie, and Roman were sisters in arms throughout each of our very different journeys. Thank you for reading my drafts, sharing your struggles, and providing tips and tricks to survive this journey.

At the beginning when I was only considering this journey, my mentor Art Davidson pledged to always be in my corner. I find myself unable to properly describe my gratitude for his council and advice amidst many tears and moments of intense frustration. Art sacrificed many Friday afternoons to ponder and discuss possible dissertation topics. He read this document many times. He reminded me that I could become a better writer and told me that my ideas were worthwhile at the many points when I had lost hope. I could not have begun, let alone have finished this tumultuous journey without him.
TABLE OF CONTENTS

CHAPTER

I. INTRODUCTION ........................................................................................................... 1
   Purpose of the Study ................................................................................................. 5
   Specific Aims and Hypothesis .................................................................................. 6
   Significance of the Study .......................................................................................... 8
   Summary .................................................................................................................... 10

II. REVIEW OF RELATED LITERATURE .................................................................. 11
   Literature Search Methods ...................................................................................... 12
   Literature Search Results ......................................................................................... 14
   Summary of Gaps ....................................................................................................... 41

III. METHODS .......................................................................................................... 46
   Aim 1 ....................................................................................................................... 46
   Aims 2 and 3 .......................................................................................................... 50

IV. RESULTS ............................................................................................................ 66
   Aim 1 ....................................................................................................................... 66
   Aim 2 ....................................................................................................................... 91
   Aim 3 ...................................................................................................................... 101

V. DISCUSSION ........................................................................................................ 126
Implications for Public Health Practice ................................................................. 127

Implications for Usability of EHR Data for Surveillance ......................................... 130

Recommendations .................................................................................................. 133

Future Research ...................................................................................................... 136

Challenges ............................................................................................................... 137

Limitations ............................................................................................................... 140

Summary of Contributions and Implications ........................................................... 141

REFERENCES ............................................................................................................ 143

APPENDIX

A. Interview Guide and Questions for Key Informants Guide ................................. 160

B. Diabetes and Hypertension ICD-9 and ICD-10 Diagnosis Codes .......................... 163

C. Quotes Describing Approach to Population Health Surveillance .......................... 177

D. Quotes Describing Organizations Structure and Investment in Population Health Surveillance .................................................................................................................................. 180

E. Quotes Describing Evolution in Population Health Surveillance .......................... 181

F. Quotes Describing Perspectives on EHR data for Population Health Surveillance .... 182

G. Diabetes Phenotype Comparison by Age, Gender, and Race Ethnicity .................. 185

H. Hypertension Phenotype Comparisons by Age, Gender, and Race Ethnicity, 2013 .. 187

I. COMIRB Protocol Describing Aim 1 Analysis ..................................................... 189
LIST OF TABLES

Table 1: Recommended Population Health Measures by Domain........................................ 44
Table 2: Diabetes and Hypertension Case Definitions...................................................... 64
Table 3: Overview of Control Measure Case Definition .................................................. 65
Table 4: Overview of Case and Prevalence Phenotypes.................................................... 65
Table 19: Persons with Diabetes Seen with Hemoglobin A1C Lab Test Status, Denver, Colorado, 2011-2015 ................................................................. 121
Table 20: Denver EHR-Based Diabetes Control Trends by Demographic Group, 2011-2015 ........................................................................................................... 121
Table 21: Persons with Hypertension Seen with Blood Pressure Recorded, Denver, Colorado, 2011-2015 .................................................................................... 122
Table 22: Denver EHR-Based Hypertension Control Trends by Demographic Group, 2011-2015 ........................................................................................................... 123
Table 5: Summary of LPHA Key Informants ................................................................ 105
Table 6: Categories of LPHA Approaches to Population Health Surveillance................... 106
Table 7: Traditional and Non-Traditional Surveillance Data Sources ............................... 107
Table 8: Level of EHR-based Surveillance Activity ......................................................... 108
Table 9: Level of EHR-based Surveillance Activity by Jurisdiction Size ......................... 108
Table 10: Counts and Population Coverage of Care Population by Phenotype, Denver .... 109
Table 11: Demographic Distribution of Care Population by Phenotype, Denver ............... 110
Table 12: Diabetes and Hypertension Prevalence Comparison, Denver, 2013 ............... 112
Table 13: Demographic Description of Care Population, Denver, Colorado, 2011-2015.... 113
Table 14: Care Population Coverage by Demographic Group, Denver, Colorado, 2015 .... 114
Table 15: Trends in EHR-Based Diabetes Prevalence by Demographic Group, 2011-2015 . 116
Table 16: Comparison of EHR-Based and Survey-Based Estimates of Annual Diabetes Prevalence, Denver, Colorado, 2011-2015 .......................................................... 117

Table 17: Trends in Denver EHR-Based Hypertension Prevalence by Demographic Group, 2011-2015 ............................................................................................................. 119

Table 18: Comparison of EHR-Based and Survey-Based Estimates of Annual Hypertension Prevalence, Denver, Colorado, 2011-2015 ......................................................... 119
LIST OF FIGURES

Figure 1: Conceptual Framework for Public Health Surveillance ................................................. 44
Figure 2: The Measures Management System Blueprint .............................................................. 45
Figure 3: Overview of Sequential Exploratory Research Design ................................................ 64
Figure 11: Diabetes Control Map by Census Tract, Denver, Colorado, 2014-2015 .............. 122
Figure 12: Hypertension Control Map by Census Tract, Denver, Colorado, 2014-2015 ..... 124
Figure 13: Comparison of Census Tract Hypertension and Diabetes Prevalence and Control, Denver, Colorado, 2014-2015 ................................................................. 125
Figure 4: Care Population Waterfall, Denver, 2013 ................................................................. 111
Figure 5: Population Coverage Map by Census Tract, Denver, Colorado, 2014-2015 ...... 114
Figure 6: Comparison of Care Population and Coverage by Census Tract, Denver, Colorado, 2014-2015 ........................................................................................................ 115
Figure 7: Comparison of Denver Diabetes Prevalence by Data Source, 2011-2015 ........... 117
Figure 8: Diabetes Prevalence Map by Census Tract, Denver, Colorado, 2014-2015 ........ 118
Figure 9: Comparison of Denver Hypertension Prevalence by Data Source, 2011-2015 ..... 120
Figure 10: Hypertension Prevalence Map by Census Tract, Denver, Colorado, 2014-2015 120
## LIST OF ABBREVIATIONS

ACA: Affordable Care Act  
ACO: Accountable Care Organization  
ACS: American Community Survey  
BP: Blood Pressure  
BRFSS: Behavioral Risk Factor Surveillance System  
CHORDS: Colorado Health Observation Regional Data Service  
COMIRB: Colorado Multiple Institutional Review Board  
CMS: Centers for Medicare & Medicaid Services  
DBP: Diastolic Blood Pressure  
DH: Denver Health and Hospital Authority  
DRN: Distributed Research Network  
EHR: Electronic Health Record  
HGBA1C: Hemoglobin A1C or glycosylated hemoglobin  
HHS: United States Department of Health and Human Services  
HITECH: The Health Information Technology for Economic and Clinical Health Act of 2009  
HIPAA: Health Insurance Portability and Accountability Act  
HSR: Health Statistics Region  
ICD-9: International Classification of Disease Version 9  
ICD-10: International Classification of Disease Version 10  
IHI: Institute for Healthcare Improvement  
IOM: Institute of Medicine  
KPCO: Kaiser Permanente Colorado  
LHS: Learning Health System  
LPHA: Local Public Health Agency  
LPHAs: Local Public Health Agencies  
MU: Meaningful Use  
MDPHnet: Massachusetts Department of Public Health Network  
NACCHO: National Association of City and County Health Officials  
NHANES: National Health and Nutrition Examination Survey  
NYC-HANES: New York City National Health and Nutrition Examination Survey  
NIH: National Institute of Health  
NQF: National Quality Forum  
ONC: Office of the National Coordinator for Health Information Technology  
PHEkB: Phenotype Knowledge Base  
UCH: University of Colorado Health
CHAPTER I
INTRODUCTION

In 2008, Berwick’s Triple Aim Framework reintroduced the importance of “improving the health of populations,” elevating the need for effective population health measurement\(^1\). Population health is a concept defined by Kindig as the health outcomes of a group of individuals, including the distribution of such outcomes within the group\(^2\). Population health pairs a traditional clinical focus on individual determinants of health (e.g., tobacco use) with a public health perspective that considers community-level, place-based factors like community rates of poverty\(^3\)\(^-\)\(^7\). This broader view of health was designed to catalyze change in how health care is provided and place a greater focus on local collaboration between clinical, public health and community stakeholders to identify and address each community’s greatest health needs\(^2\)\(^,\)\(^3\)\(^,\)\(^6\)\(^-\)\(^10\). Despite differing opinions on the exact meaning and target group characteristic(s) for “population health”\(^11\)\(^-\)\(^16\), stakeholders agree, an analytic and local measurement-driven approach is essential\(^17\)\(^,\)\(^18\). Using local data to identify communities with the greatest disease burden can improve the ability of service providers to connect target populations with tailored resources and interventions. Chronic diseases such as diabetes and hypertension are priority issues for which accurate, local, timely information is needed\(^8\)\(^,\)\(^19\)\(^-\)\(^24\).

Population health frameworks suggest that local surveillance activities should explore disparities between socio-demographic groups and geographic areas\(^2\)\(^,\)\(^3\)\(^,\)\(^9\)\(^,\)\(^17\)\(^,\)\(^18\). Measures should venture beyond more general indicators like life expectancy to include
actionable indicators of disease burden\textsuperscript{17,18,24}. Given the need for more detailed geographic analysis, measuring population health trends should be a responsibility of local public health agencies (LPHAs)\textsuperscript{21,23,25}. Many LPHAs have built population health analytic capacity by leading community health assessments, conducting health impact assessments, and supporting non-profit hospital community health needs assessments\textsuperscript{17,26-28}. However, key information describing how LPHAs have approached population health surveillance is missing from the literature. \textbf{How do perspectives on and opportunities for population health surveillance vary across LPHAs and what lessons learned can be shared?} What best practices are guiding LPHAs in population health measurement activities? Who are the stakeholders and how do they engage with local health status information?

Improving capacity for population health measurement relies on robust sources of local data to identify population-specific health priorities and to measure the impact of targeted interventions\textsuperscript{17,18,24}. Population-based surveys are well-established tools for measuring health status, but are expensive to administer; generally limited to results no more frequent than annually, they frequently generate limited community-level (e.g., sub-county) information due to inadequate sample size\textsuperscript{25,29-37}. The 2016 Public Health 3.0 framework suggested that new and innovative data sources are needed and recommended that “timely, reliable, granular (i.e., sub-county), and actionable data should be made accessible to communities throughout the country”\textsuperscript{21}.

Widespread adoption of electronic health records (EHRs) has created volumes of detailed health data suitable for local population health surveillance because data may be:
1) accessed in near real time, 2) geo-located, and 3) used for cross-sectional and longitudinal analyses\textsuperscript{38-41}. The HITECH Act identified surveillance as a valuable secondary use of EHRs\textsuperscript{42} and many publications have described a great opportunity to advance population health using EHR-based assessment strategies\textsuperscript{25,41,43,44}. By monitoring local disease trends, measuring disparities within a community, and providing timely feedback on targeted interventions or policies, EHR information could be an essential population health monitoring resource for local stakeholders\textsuperscript{25,41,43}.

How LPHAs will make secondary use of EHR data remains a major challenge. While EHR data availability has recently improved, little guidance exists for how an end-user might leverage these information resources\textsuperscript{25,45}. How should LPHAs adapt to and harness a progressively growing body of digital health data to inform essential public health service delivery? For most LPHAs, technology, governance, and operations needed to acquire and use EHR data for prevalence estimates are relatively unfamiliar. A national public health informatics steering committee advocated for the importance of resources and further development of the evidence base to support LPHAs, “its [LPHA] systems must be modernized; its workforce must be properly trained; and its vision updated”\textsuperscript{46}.

Potential limitations of EHR data and differences between survey and EHR-based health estimates remain a concern in the literature\textsuperscript{25,43,44,47-50}. LPHAs have decades of knowledge and experience addressing well-understood survey limitations of random error, sample size, and recall bias. LPHA surveillance methods have been designed and modified to accommodate common survey limitations identified over the decades such as: post-
stratification weighting, expanded telephone sampling to include land lines and cell phones, and small area estimate modelling techniques. Sources of bias and methodological limitations exist in both methods but differ between population-based health surveys and EHR data. Care seeking populations captured in EHR data may differ dramatically from a randomly-sampled survey cohort perfectly representative of a geographic population. More exploration is needed to understand the respective strengths and limitations of each data source. If survey and EHR-based estimates are different, how can those differences be interpreted? Should EHR-based estimates replace population-based health surveys or should they be considered a complementary information source? Validation studies have been recommended to guide LPHAs in the appropriate use and interpretation of local EHR-based estimates.

Public health agencies in New York City, Massachusetts, and Indiana who have EHR-based surveillance systems have led the development validation methods with a focus on comparisons between EHR- and survey-based estimates. These studies have found measure by measure variation on validity. Two studies have determined that diabetes and hypertension prevalence estimates are valid for population based surveillance purposes.

Small geographic area health assessment is a major advantage of EHR data. Sub-county health information in the form of census tract estimates can identify communities with higher disease burden to target resources and evaluate local program and policy-based intervention impact. However, few routine surveillance datasets have sufficient sample to produce stable census tract level estimates. Many important questions remain about how
to estimate, interpret, and validate census tract estimates using EHR data. How accurate and precise are EHR-based census tract estimates? What do differences between census tracts mean? How does population coverage impact the precision or the interpretation of variation in census tract estimates? How do EHR-based estimates compare to modelled census tract estimates from telephone surveys? What analytic approaches can be used to minimize bias?

EHR data can produce local information on priority health topics that are difficult, if not impossible, to measure using health surveys. Routine clinical quality indicators measure control status for common chronic diseases; these could inform prevention and outreach interventions when applied to geographic or demographic groups. Defining and validating methods for EHR-based measures of local chronic disease control can broaden and enhance existing population health surveillance methods. These cross-institution initiatives require effort and momentum to build governance structures and relationships that support access to and use of EHR data by LPHAs. New and innovative approaches to measuring population health are on the horizon as stakeholders have growing access to EHR-based tools to realize this goal.

**Purpose of the Study**

The purpose of this research was to: 1) characterize local public health agency beliefs, practices, and overall barriers related to local population health surveillance, 2) generate county and census tract estimates of chronic disease prevalence from EHR data aggregated across multiple healthcare institutions, to compare with survey-based
estimates, and 3) estimate and describe temporal trends, socio-demographic disparities, and geographic patterns in chronic disease control.

**Specific Aims and Hypothesis**

Research Aim 1: Characterize local public health agency beliefs, practices, and barriers related to local population health surveillance. Conduct key informant interviews with experienced public health surveillance stakeholders to:

a) Assess LPHA approaches to population health surveillance including data sources, dissemination strategies, stakeholders, ideas about validation of a new data source, and unmet surveillance needs,

b) Explore organizational structure and investment in surveillance including roles and responsibilities related to surveillance work,

c) Explore how and why LPHA population health surveillance practices have evolved due to greater focus on population health, and

d) Explore each informant’s perspective on EHR-based surveillance including barriers and facilitators affecting local population health surveillance capacity.

Research Aim 2: Generate estimates of chronic disease prevalence from EHR data aggregated across multiple healthcare institutions to compare with survey-based estimates.

a) Temporal Analysis: Use EHR data to estimate annual (2011-2015) diabetes, and hypertension prevalence (based on ICD-9 and ICD-10 diagnosis codes) across Denver and for demographic subgroups (e.g., age, race and ethnicity, gender).
b) Accuracy: Apply population weights to EHR-based Denver diabetes and hypertension prevalence estimates to compare unweighted and weighted measures.

c) Geospatial Analysis: Combine EHR data from 2014 and 2015 to estimate adult diabetes and hypertension prevalence at the census tract level.

d) Comparison: Compare EHR-based annual estimates to survey-based annual estimates at the county (e.g., BRFSS survey) for each year from 2011 to 2015.

Research Aim 3: Estimate and describe community-level trends in chronic disease (i.e., diabetes and hypertension) control.

 a) Temporal analysis of Hypertension Control: Use blood pressure measures to assess annual rates of hypertension control among a cohort of hypertensive patients for 2011 to 2015.


 c) Geospatial Analysis: Combine EHR data from 2014 and 2015 to estimate adult diabetes and hypertension control at the census tract level.

 A mixed methods approach including qualitative and quantitative analysis was used to ensure progression of learning, integration, and cohesion between the research aims. For this study, mixed methods have been shown to answer new or emerging questions related to both research inquiry and practice\textsuperscript{57-61}. Qualitative data collected in aim 1 described the landscape of LPHA population health surveillance activities and provide context to aims 2.
and 3. Interview questions were designed to assess LPHA readiness for population health surveillance, define current activities and data sources, and describe the strengths and limitations of current data tools. Understanding assets and barriers clarified if and how EHR data can be used by LPHAs, framed the analysis of aims 2 and 3.

Aim 2 addressed the need for validation of EHR-based estimates, tested the efficacy of population weights to reduce bias (the primary limitation of EHR data), and demonstrated capacity to generate sub-county measures of disease burden. Because prevalence does not provide a complete picture of a community’s chronic disease burden, aim 3 explored the feasibility of using EHR data to produce local measures of chronic disease control. Because local telephone surveys do not collect measures to assess disease control and national control estimates cannot provide local insights, addressing this gap with EHR data showcased its value to measure population health.

**Significance of the Study**

While the capacity to conduct local population health surveillance is critical, LPHAs have limited tools and resources to support these activities. This research depicts the national landscape of local population health surveillance beliefs, practices, and barriers from the LPHA perspective. Findings from key informants were used to define promising practices, describe barriers, and discuss lessons learned that could be translated to other locales to increase LPHA readiness conducting EHR-based surveillance. Barriers could be used to set policy priorities, inform systems change, or recommend allocation of resources to accelerate adoption of EHR-based surveillance.
Limited guidance exists for how to conduct population health surveillance and LPHAs grapple with if and how traditional data sources can identify communities where the burden of disease is the highest to target of prevention and intervention resources \(^{17,62-64}\).

While gold standard health surveys easily and efficiently provide national and state level information on health and wellbeing, they perform poorly at the community level (e.g., sub-county) and are constrained to information that can be collected through a survey. Despite survey efforts to over sample, model small area estimates, and combine data across years, the capacity of surveys to meet local health data needs remains limited. The Public Health 3.0 model recommends that LPHAs begin a culture shift to prioritize making high quality granular health data available to local population health stakeholders.

Recently, EHRs have emerged as a promising data source for local surveillance efforts. However, secondary EHR data applications are relatively new and methods to validate EHR data for local surveillance and interpret EHR-based estimates for sub-county geographic areas and socio-demographic disparities are needed. LPHAs have encountered many challenges related to EHR data access, use, and interpretation. This validation study evaluated the fitness for use of EHR data for two priority health indicators through systematic comparisons with population-based surveys. This research defined and tested a method for EHR-based disease prevalence estimation at the census tract level. While other studies have focused on validating EHR-based prevalence estimates, this research will define and test a method for county and census tract estimation of hypertension and diabetes control from EHR data. Because the value of EHR data will only be realized when
LPHAs readily see information generated from digital access, validation findings from this research demonstrate to other LPHAs that validation is within reach and a worthy undertaking.

**Summary**

Providers, public health, and policy makers have a strong and mutual goal in improving population health. Understanding the current landscape of population health surveillance and ensuring that data, systems, and capacity exist to measure progress towards this goal are pressing infrastructure and information needs. How EHR data can contribute to population health surveillance and what value it adds to the existing repertoire of data sources is unclear. This research contributes knowledge about the landscape of population health surveillance from the perspective of LPHAs. A comparison of survey and EHR-based population health estimates, including the application of population-adjusted weights, expands the body of knowledge defining how to validate EHR-based estimates including validation findings for comparison with similar studies.
CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter presents a comprehensive review of literature related to the three research aims. This research draws from the disciplines of population health, chronic disease, public health surveillance, and public health and clinical informatics. A literature review was conducted to understand the:

- history, definition, and known frameworks of population health
- importance and burden of chronic disease within population health
- justification and recommendations for measurement of population health
- relationship between public health surveillance and population health surveillance
- existing surveillance data sources including strengths and limitations
- role of electronic health records (EHRs) in population health surveillance
- current approaches to EHR-based surveillance
- strengths and limitations of EHR data
- process for assembling cases and population health indicators from EHR data
- findings from validation studies comparing EHR-based and survey-based estimates

The chapter concludes with a summary of known literature and identified gaps filled by this research.
Literature Search Methods

This literature review was conducted to identify relevant peer-reviewed publications from multiple domains: population health, chronic disease, public health surveillance, and EHR-based surveillance. Because these concepts are complex, interconnected, and ever evolving, a narrative review (instead of a systematic review which is conducted using rigourous methodological approaches) was conducted using a myriad of search terms to ensure that all relevant information could be identified. The goal of a narrative review is to describe and discuss the state of the evidence on a specific topic from a contextual perspective. Population health search terms included but are not limited to:

- ‘population health definition’
- ‘population health framework’
- ‘population health surveillance’
- ‘population health monitoring’
- ‘population health measurement’
- ‘public health’ and ‘population health’
- ‘public health surveillance’ and ‘population health surveillance’
- ‘population health surveillance’ and ‘chronic disease’
- ‘population health surveillance’ and ‘diabetes’
- ‘population health surveillance’ and ‘hypertension’

Public health surveillance search terms included but are not limited to:

- ‘public health surveillance’ and ‘chronic disease’
- ‘public health surveillance’ and ‘diabetes’
- ‘public health surveillance’ and ‘hypertension’
- ‘limitations’ and ‘public health surveillance’
- ‘limitations’ and ‘public health data’
- ‘limitations’ and ‘survey data’ and ‘public health surveillance’
When searching within the EHR literature, the terms electronic health record, electronic medical record, EHR, and EMR were each used for separate searches. EHR surveillance search terms included but are not limited to (here the example is restricted to ‘electronic health records’):

- ‘public health surveillance’ and ‘electronic health records’
- ‘population health surveillance’ and ‘electronic health records’
- ‘surveillance’ and ‘electronic health records’
- ‘limitations’ and ‘electronic health records’ and ‘surveillance’
- ‘diabetes’ and ‘electronic health records’ and ‘phenotype’
- ‘hypertension’ and ‘electronic health records’ and ‘phenotype’
- ‘diabetes’ and ‘electronic health records’
- ‘hypertension’ and ‘electronic health records’

PubMed and Journal Author Name Estimator were used to identify peer-reviewed publications. Using exclusion criteria, abstracts and titles were reviewed for relevance to this research. The following exclusion criteria were used across topics:

- The article did not include a clear definition for population health
- The article was based on a definition or concept of population that differed greatly from the Kindig definition that guided this research
- The article focused exclusively on population health management
- The article did not mention measurement or monitoring of population health
- The article did not present any new or relevant information
- The article did not discuss data sources for public health surveillance
- The article did not present any new or relevant information
- The article was focused on global, national, or state level surveillance and did not discuss local surveillance
- The article discussed only national examples or approaches of EHR surveillance
- The article focused on electronic case reporting, case finding, or infectious disease surveillance
- The article focused on other chronic disease surveillance without translatable findings for diabetes and hypertension
First, abstracts were evaluated according to the exclusion criteria. If an abstract was not rejected by any exclusion criteria, the entire article was reviewed and considered. Due to the expansive body of knowledge in each of these topics, articles published within the past 5 years were prioritized for review. The bibliographies of included articles were used to identify additional peer-reviewed publication for review.

**Literature Search Results**

Because population health is a term of great interest and chronic disease is a topic of intense focus, the body of literature that was identified was vast. Between March and December 2016, over 300 peer-reviewed publications were identified and evaluated for relevance. Approximately 150 publications were reviewed in full and almost all offered concepts, content, and findings to inform this work. Throughout the development of this research, the literature review was refreshed as publications became available that introduced concepts or opinions not already present.

**Population Health Concept, Definition, and Framework**

The foundation of this research is the concept of population health, a term that has gained momentum over the past several decades\textsuperscript{11,12,14}. Population health, as defined here, is one third of Berwick’s Triple Aim Framework, a priority for the U.S. Department of Health and Human Services (HHS), a goal of the Centers for Medicare & Medicaid Services (CMS) and Institute for Healthcare Improvement (IHI), and a strategic objective of many national
and international health foundations, such as the Robert Wood Johnson Foundation and The Commonwealth Fund\textsuperscript{1,10,65-67}.

Consensus about the exact meaning of population health is unlikely as multiple perspectives have been put forth multiple definitions. This research subscribes to the Kindig definition, “the health outcomes of a group of individuals, including the distribution of such outcomes within the group”\textsuperscript{2-4,8,9,68}. Kindig clarified that the subjects of population health are a group of people within a geographic area (e.g., a neighborhood or a community) which suggests population health activities and measurement should occur at very localized levels. Kindig differentiated population health which was the measurement of outcomes, from population health management which included activities or interventions to improve the health of a sub-population (e.g., empaneled individuals)\textsuperscript{3}. When making population health measurements, Jacobson and Teutsch also defined the total population as the residents of a geopolitical area\textsuperscript{24}. This research is focused on population health surveillance, meaning efforts to monitor a range of health outcomes that together capture key factors influencing the health and well-being of a population within a geographic area.

Many frameworks have been developed to improve the understanding of population health which vary by level of detail, audience, and venue, but convey similar themes and ideas\textsuperscript{2,8,69}. The Evans and Stoddart framework was one of the earliest population health frameworks which illustrated the influence of the social and physical factors on individual behaviors and health outcomes\textsuperscript{9,69}. This framework showed that community factors played a significant role in individual behaviors and outcomes; measurement at the community
level can uncover insights and correlates that an individual’s level of disease, genetics, and biology alone cannot. Subsequent population health frameworks expanded to consider related concepts like measurement, prevention, and intervention but retained a focus on community factors and their association with racial, ethnic, and/or gender health disparities. This research was intended to incorporate the association of place and disparities by conducting analysis at a granular geographic level and make comparisons by demographic subgroup to detect health disparities between and within groups.

**The Role of Chronic Disease in Population Health**

Most national population health agendas recognize chronic disease as an important topic specifically prioritizing hypertension and diabetes which are diseases linked to the 1st, 5th, and 7th leading causes of death nationwide. The burden of diabetes and hypertension is high and growing. For both diseases, many cases are undiagnosed. At-risk populations consistently and rapidly progress to disease. Diabetics and hypertensives suffer serious complications, and disease-related care is a costly burden to the healthcare system. Diabetes and hypertension are inextricably linked. Many individuals manage both diseases, and their risk for contracting both increases once diagnosed with one. These diseases share some risk factors, socio-demographic incidence and prevalence trends, prevention and disease control lifestyle modifications, and treatment recommendations.

In 2014, more than 29.1 million Americans had diabetes, representing about 9.3 percent of the population. One in four people with diabetes were undiagnosed,
underestimating true diabetes population burden. One third (37%) of American adults are pre-diabetic; 30% of adults with pre-diabetes will develop diabetes within five years. Diabetes is a leading driver of health care costs nationwide. In 2012, direct and indirect costs associated with diabetes care were estimated to be more than $245 billion. A majority of diabetics (90 to 95 percent of cases) have type II diabetes, a chronic health condition caused by insulin resistance and poor blood glucose regulation. While type I diabetes is important, this research is focused exclusively on type II diabetes and the use of the term diabetes from this point forward references type II diabetes.

Many diabetes prevalence trends were similar to hypertension or high blood pressure trends. According to the American Heart Association, about 80 million adults nationwide had hypertension in 2016. Like diabetes, health disparities were observed with hypertension. After controlling for the impact of age, disparities (i.e., higher rates) of hypertension prevalence persisted among males, Blacks, and lower income adults. Hypertension treatment and control efforts are impacted by underdiagnosis as well; one in five hypertensive adults are undiagnosed and unaware of their disease until a related health consequence occurs. Hypertension screening by measuring blood pressure is now a recommended prevention activity at every clinic visit. High blood pressure is the primary risk factor for stroke, which is the fifth leading cause of death. Those with high blood pressure were three times more likely to die from heart disease and four times more likely to die from a stroke. The risk of severe health consequences from hypertension increased with increasing age, a family history of heart disease, male gender, being overweight or
obese, smoking, high cholesterol, diabetes, and being physically inactive. With high hypertension prevalence and severe health consequences, associated healthcare costs were high. In 2014, national hypertension-associated healthcare and societal costs were $47.5 billion including health care services, medications, and missed days of work.\(^\text{73}\)

Diabetes and hypertension are population health priorities because many cases, and more importantly, severe health complications are preventable.\(^\text{73,80}\) A number of clinical and community-based interventions have been proven effective at preventing individuals from advancing to develop these diseases or helping individuals achieve better disease control. The Guide to Community Preventive Services recommends many evidence-based community interventions for diabetes and hypertension including case management and physical activity programs.\(^\text{82}\)

The prevalence, severity, and compounded effect of concomitant diabetes and hypertension suggest monitoring these two diseases is an important population health activity. Local surveillance of disease prevalence to better understand disease burden, identify populations disproportionally affected, and monitor prevalence and control trends over time are immediate needs. Identifying communities where the burden of disease is the highest can improve the targeting of limited prevention and intervention resources. For diabetes and hypertension incidence, prevalence, and control, local population health information at or below the county level is essentially nonexistent.
Population Health Surveillance

Strong consensus exists about the importance of population health measurement\(^ {12,17,18,20,24,25,83-85}\). The absence of a standardized collaborative framework and infrastructure has impeded development of efficient population health information systems, and coordinated surveillance efforts across organizations. Many organizations, “struggle with what to measure related to the execution of the Triple Aim and with accessing the needed data”\(^ {17,18,67}\). In 2016, an expert panel published a national research and development agenda for population health informatics that described population health monitoring as both crucial and very difficult\(^ {86}\).

How to conduct population health surveillance relies heavily on the knowledge and proficiency of public health surveillance stakeholders\(^ {21,23,84,87}\). Assessment (i.e., capacity to conduct surveillance and use the information) is one of three core public health functions\(^ {28,88,89}\). Public health has over a century of surveillance experience, defined as:

*...the ongoing systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link in the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs.*\(^ {88}\)

In recent decades, public health has embraced the term population health and surveillance activities have expanded to include health behaviors and chronic disease\(^ {90,91}\). As national concerns about surveillance and data privacy has grown during the past decade, some
public health agencies have selected euphemisms for surveillance such as evaluation or monitoring. Mounting evidence about the importance of place and the impact of social determinants has led LPHAs to reorient analytic efforts for comparison between geographic and demographic subgroups. LPHAs have begun to investigate how non-traditional data sources like clinical measures of height and weight or combining data from multiple sources (e.g., Medicaid enrollment and median home values) can improve the quality and utility of local health information. Presenting surveillance data stratified by individual and community determinants like poverty and education is now a public health best practice\textsuperscript{91}.

Considering their experience with surveillance methods and focus on local geographic areas, LPHA are ideally suited for population health surveillance. A 2016 article describes a reimagined view of public health agencies to include an active role in population health surveillance\textsuperscript{21}. Many LPHAs have built solid foundations through population-based, data-driven activities like community health assessments\textsuperscript{17,26-28} and the development of informatics departments (informatics is a field devoted to the science and information systems of population-based surveillance)\textsuperscript{21,25,92,93}. Limited funding for staff and infrastructure, misalignment between informatics workforce required skills and academic training programs, and dependency on state health agencies has kept local informatics resources from maximizing their contribution\textsuperscript{94}. Population health surveillance use cases have introduced new data and analytic requirements that have challenged a frail public health information infrastructure\textsuperscript{93}. A National Association of City and County Health Officials (NACCHO) survey found that LPHAs need resources and support to improve their
informatics capacity and meet health stakeholder’s rapidly growing health information needs\textsuperscript{93}. The public health informatics workforce needs support in interpreting quantitative data, designing and running reports from information systems, using and interpreting qualitative data, using statistical or other analytical software, project management, and using geographical information systems\textsuperscript{95}. The U.S. Assistant Secretary of Health and Human Services, highlighted the need to expand surveillance data, metrics, and methods so that LPHAs have “timely, locally relevant health information system”\textsuperscript{21}. While LPHAs may be engaged in population health surveillance, little has been catalogued of successful surveillance activities. What data sources are being used and how is this information being shared? What is working well and what barriers have LPHA encountered?

**Indicators of Population Health**

Selecting a unified set of population health metrics can and should be measured has been difficult\textsuperscript{17,62,64}. What metrics best describe population health? Measures of population health should be valid and reliable, easily understood, measureable over time and for specific geographic or demographic groups, quantifiable from available data source, and sensitive to socio-economic, environmental, or political changes\textsuperscript{62,64}. In 2012, the Institute for Healthcare Improvement published a guide to measuring the triple aim which showcased the complex relationship between determinants (including upstream and individual factors), health outcomes and quality of life that must be considered when measuring the picture of health\textsuperscript{18}. IHI recommended that population health measurement include intermediate outcomes of disease burden and injury in addition to traditionally
public health outcomes like mortality and quality of life. Equity and the impact of social determinants of health were identified as key factors impacting population health which suggests that measurement should compare health outcomes between socio-demographic and geographic groups to highlight disparities. Stiefel et al. recommended metrics within each (Table 1) measurement domain. This research is focused on how to measure disease burden, the concept identified as an intermediate outcome.

Health outcomes should include traditional summary measures such as mortality and life expectancy, which predominate in population health monitoring\textsuperscript{62,63}. Disease burden, specifically chronic disease, is identified as a valuable and important intermediate outcome to study the impact of individual and upstream factors on health before death. Monitoring chronic disease incidence and prevalence on an annual basis is a recommended population health surveillance activity\textsuperscript{3,8,17,18}. Chronic disease prevalence is measured annually by the County Health Rankings and the Big Cities Health Inventory, two leading national population health measurement initiatives\textsuperscript{96,97}. However, prevalence and incidence fail to address the severity of disease among individuals diagnosed with those diseases, a key indicator for local disease management activities. Including measures that assess rates of disease control and severity, when measurable, would provide actionable information for population health programs. Disease management registries, claims data, EHR data, paper health records, and population-based surveys are alternative sources of information for that purpose\textsuperscript{17,18,24,44,98}. This research tests whether EHRs can generate
accurate, local disease prevalence and disease control estimates to inform practice and interventions.

Measuring disease prevalence for small geographic areas is a new challenge and opportunity for LPHAs. Public health agencies in New York City, Massachusetts, Indianapolis and Denver have leveraged large infrastructure investments through advanced informatics capacity to explore aggregation of EHR data through organically developed systems and methods for local chronic disease surveillance$^{25,48,53,99-101}$. While exciting, these opportunities are rare, with differing infrastructures, often built over extended periods of time, at great expense, with difficulty to maintain$^{101}$. Many LPHAs will need significant support to use EHR data to monitor and report on the relationship between social determinants, place-based factors, and chronic disease prevalence at more granular community levels.

**Review of Traditional Surveillance Data Sources**

Because data are foundational for effective surveillance, traditional public health surveillance data sources seem like an obvious population health data resource$^{17,18,63,88,89,91}$. For this research, traditional surveillance data sources included birth and death certificates, the Census and American Community Survey, and health surveys$^{30-32,96,97,102-105}$. Thacker describes traditional and non-traditional data sources (Figure 2) that might make up a local surveillance system$^{91}$. Thacker's definition of information systems includes data from medical records. Because Figure 2 includes both traditional and non-traditional public
health data sources as well as a category for other data resources, Thacker suggested that public health should use all available data to carry out robust public health surveillance.

Death certificate data are systematically collected, uniformly completed, widely available, and effectively track general population mortality trends at the national, state, and local level\(^63\). Most of these sources are supported by statute (e.g., required collection of death information prior to burial or communicable disease case reports to registries) and are accessible without a fee.

To measure chronic disease locally, LPHAs have typically relied on county level estimates of chronic disease deaths and county estimates of disease prevalence from health surveys\(^88-91\). Surveys and death certificates provide data no more frequently than once per year and is often delayed by up to 18 months after data collection. While mortality data effectively monitor general trends in causes of or ages at death, mortality data provide little insight about health status among the living. Death information for rare conditions or demographic subgroups of small size may need to be combined across years or aggregated to larger geographic units to produce stable estimates or protect individual privacy.

The Behavioral Risk Factor Surveillance System (BRFSS) and the National Health and Nutrition Examination Survey (NHANES) are the gold standard annual public health surveys used to understand chronic disease\(^71,96,97,104,106-108\). NHANES is funded by the Centers for Disease Control and Prevention (CDC) and administered by the National Center for Health Statistics. Because NHANES is conducted in person; some indicators are self-reported while others are measured by survey staff\(^30\). Published annually, NHANES produces national
results on a variety of chronic disease measures while BRFSS is used to produce state and county level results. Because NHANES produces information for neither states nor counties, it is not frequently used by LPHAs.

BRFSS is “an ongoing, state-based, random-digit–dialed telephone survey of the non-institutionalized U.S. population aged >18 years” that is also funded by CDC and administered annual by the state health department or survey administration vendor. Both surveys use weighting methods to adjust survey responses to represent the underlying demographic population. In 2016, the CDC funded the 500 Cities Project which combined BRFSS responses with demographic and geographic data to model estimates at the census tract level for every BRFSS question. While these data are publicly available, LPHAs have had too little time to assess its utility as an operational surveillance tool.

Using survey data for population health surveillance presents a number of benefits and challenges. The greatest strength of health surveys is that the sample is collected randomly and systematically. No bias is introduced from the way that individuals were selected, the survey administrator, or the way survey data were captured and stored. Health surveys have their limitations due to high cost, reliance on self-reporting, restriction to respondents with telephones, and questionnaire burden. The high cost to administer surveys often limits the sample size. The Denver BRFSS sample includes roughly 1,100 adult (18 years and above) respondents per year. Survey-based estimates drawn from a small sample are imprecise with wide confidence intervals that hinder LPHAs ability to detect
significant changes over time or differences between groups. Frequently, multiple years of BRFSS data must be combined to produce more precise estimates but reduces capacity to measure how health changes over time\textsuperscript{21,104,111}. Sample sizes dramatically limit capacity to estimate for subpopulations like racial and ethnic minorities or geographic analyses and representation of those estimates. Smaller minority populations in Colorado, such as native Hawaiians, are often such a small part of a sample that they are either a) combined with other groups (e.g., Asian/Pacific Islanders) to create a more stable estimate for a less specific group, or b) not reported because of suppression rules.

Survey measures from BRFSS are self-reported and may be biased by self-report or recall error\textsuperscript{35}. For example, prevalence estimates only capture those with diagnosed conditions who recall a provider discussed that diagnosis. Trouble remembering information such as when blood pressure was last checked may distort estimates, an example of recall bias. Surveys are a snapshot of health, providing information at a single point in time\textsuperscript{29,30}. When conducted annually, surveys produce serial cross sectional datasets that describe changes in a behavior- or disease-specific population over time. Studying change in a defined cohort of individuals over time is impossible with these methods. While BRFSS estimates prevalence trends at the county level by demographic group, BRFSS cannot track individuals over time, evaluate control or change in control, or compare trends at the individual level, or assess severity within the prevalence group.

Using BRFSS and NHANES, this review was able to find national, state, and county estimates of diabetes and hypertension prevalence and national estimates of disease
control, some of which are used for comparison in this research\textsuperscript{30,32,102,104,108}. BRFSS produced annual county estimates of diabetes and hypertension prevalence, which could be stratified by socio-demographic group. In Denver, survey-based annual prevalence estimates for diabetes had confidence intervals spanning 5 percentage points (5.76 to 10.46 in 2011). Disease prevalence estimates for minority race and ethnic groups were based on insubstantial samples: 260 Hispanic and 91 black non-Hispanic adults in 2013. No local estimates of chronic disease control could be identified. While effective at tracking national trends, both surveys provide limited information at the county level and no information at the census tract level.

**Electronic Health Records as a Surveillance Data Source**

As the field of population health gains greater clarity and focus, health data collected and stored in EHRs has been identified as a promising source of surveillance information\textsuperscript{25,38,39,98,112-114}. The Health Information Technology for Economic and Clinical Health (HITECH) Act was passed in 2009 with a goal of nationwide adoption of certified electronic health record technology for exchange of health information through the Meaningful Use (MU) incentive program\textsuperscript{42,115}. HITECH incentivized providers and hospitals through major national investments to adopt and implement EHRs to improve clinical care and patient outcomes, with a secondary benefit that EHR data could be used for research and surveillance efforts. The incentive program defined MU standards which have improved the quality of EHR data collected and the ability to use it for surveillance\textsuperscript{40,115}. In 7 years,
the percent of hospitals using an EHR grew from 9% in 2007 to 75% in 2014. Health data describing diagnoses, lab test results, procedures, utilization, and use of pharmacotherapy are now available electronically for most healthcare provided in the United States.

Accelerated EHR implementation has created a vast health data resource, but LPHAs must overcome many challenges to operationalize EHR-based surveillance. LPHAs have difference levels of informatics capacity. Recent studies suggest that that staff knowledge and capacity, financial constraints, dependency on the state health agency, and size of a LPHA pose major challenges to efficient informatics practices. Clinical data are larger and more complex than traditional surveillance data (e.g., health survey data). The legal and regulatory requirements for storing and using EHR data are stringent and relatively unfamiliar to the public health workforce. It is unlikely that existing LPHAs information systems can be easily expanded or modified to accommodate EHR data. To prepare raw EHR data for surveillance, LPHAs may have to partner with healthcare organizations to carry out advanced data extraction, manipulation, and integration analytic tasks. Geocoding is often needed to link patient data with place-based information and few LPHAs have robust infrastructure and capacity for geospatial analytics. In 2015, the Public Health Informatics Institute published an EHR planning toolkit including recommendations for how LPHAs may begin to consider an EHR-based surveillance system. This toolkit suggested strategies to mitigate the many implementation challenges and identified the federated query network model as a promising best practice.

As implementation of EHRs has progressed, healthcare organizations have also been
driven to interoperate health data across systems; health information exchanges with centralized or distributed architectures have established EHR networks that assemble electronic patient information across multiple providers. Distributed networks have been implemented nationally, regionally and locally to facilitate access to multisite EHR datasets assembled for research and surveillance\textsuperscript{117}. Distributed network structures exchange data using federated query software, ensuring that participating healthcare organization retain full control over their data, at all times. Distributed networks were initially pioneered by researchers and then adopted by public health for surveillance (e.g., FDA Mini-sentinel) with the goal of assembling a population representative dataset with the least amount of bias\textsuperscript{41,54,118-124}. Participating healthcare providers agree to contribute data to a network for a common purpose, usually research or public health surveillance. No central repository of multisite data is created. A handful of networks have demonstrated that distributed queries to identify, retrieve, and aggregate EHR data from multiple organizations in a multi-site EHR dataset are a viable and useful resource for surveillance and research information\textsuperscript{117,120,122-125}.

Existing EHR networks and related surveillance efforts span a continuum of success and maturity. National surveillance networks were the first to be developed and are the most mature, having benefited from the most resources, expertise, and experience. The first investments in distributed surveillance were provided in response to the need for national post-marketing surveillance for drugs and vaccine adverse\textsuperscript{117,123,126}. The Mini Sentinel Project is a U.S. Food and Drug Administration funded national distributed network...
that uses clinical, administrative and claims data to conduct post marketing drug surveillance. Between 2000 and 2011, over 300 million person-years, 2.4 billion encounters, 38 million inpatient hospitalizations, and 2.9 billion dispenses of medication were collected\textsuperscript{126}. Since its inception in 2009, Mini Sentinel has conducted rapid cycle surveillance on countless drugs. Mini Sentinel data have also been used to develop and validate phenotypes for common diseases like acute myocardial infarction\textsuperscript{126}. A national observational study used Mini Sentinel data to assess changes in cardiac risk among a cohort of individuals using smoking cessation drugs\textsuperscript{127}.

The Vaccine Safety Datalink is a CDC led national network that uses EHR data to monitor the safety of vaccines among 9.2 million individuals in eight U.S. medical care organizations\textsuperscript{117}. VSD is also used for observational and comparative effectiveness research studies. VSD data was used to study the safety of the H1N1 influenza vaccine\textsuperscript{128} and detect added risk for pertussis in children participating a delayed vaccine schedule\textsuperscript{129}.

The success of national networks created a foundation for local efforts to pilot EHR-based surveillance solutions. Massachusetts Department of Public Health Network (MDPHnet) is a network of selected ambulatory care providers who provide data to conduct routine public health surveillance on Massachusetts residents\textsuperscript{55,130}. MDPHnet conducts surveillance for communicable diseases, chronic conditions, and health behaviors such as flu vaccination coverage, opioid prescriptions, and diabetes prevalence\textsuperscript{41}.

New York City Department of Health and Hygiene has created a similar distributed network infrastructure known as the Hub developed as part of the Primary Care
Information Project where EHR data from care providers are used to improve clinical quality measures. The Macroscope Project conducts EHR-based surveillance and focuses on validation activities\textsuperscript{48,131}. The Macroscope Electronic Health Record Surveillance System uses data from the Hub system to study trends in chronic disease and make comparisons with population based surveys\textsuperscript{53}. The Hub infrastructure includes data from over 2500 providers that tracks over 2.5 million New Yorkers and produces information at the zip code level. Query functionality and public health alerts have been used to study local trends in emerging public health issues such as primary care utilization following Hurricane Sandy\textsuperscript{132}. Both MDPHnet and the Hub use only ambulatory EHR data, assembled from multiple healthcare providers to measure the health of their respective populations.

The Colorado Health Observation Regional Data Service (CHORDS) is a Denver area distributed network that generates local surveillance information through a distributed network and federated query structure\textsuperscript{100}. CHORDS has successfully generated local prevalence estimates of childhood and adult obesity, tobacco use prevalence, and depression prevalence. While CHORDS queries may be restricted to ambulatory data only, data from emergency department visits and hospital stays are also available. Currently CHORDS includes data from 11 healthcare providers that inform about the chronic disease status for approximately one third of Denver’s population each year.

These examples demonstrate the growing presence of EHR data and networks housed in public health organizations to provide data in the population health surveillance space. Tools and methods from national distributed data networks, financed by national
agencies, have been adapted and implemented for state and local use. Each network has effectively queried EHR data across institutions, for public health surveillance use providing sub-county estimates.

**Strengths of EHR Data for Surveillance**

Compared to information from a self-reported telephone survey (e.g., BRFSS), EHR data represent a larger sample with objective, clinician-measured, accurate, precise and detailed clinical, geographic, and socio-demographic information. One of the greatest advantages of EHR data is volume. More patients may be seen in a single day, by a single provider or healthcare organization than the number of individuals in an annual survey sample. The number of individuals represented in EHR data improves the precision of EHR-based estimates, valuable for assessing population trends over time or differences between demographic groups. Unlike surveys, EHR-based surveillance systems have the population coverage to produce detailed health estimates on important subpopulations like children and minorities as well as small geographic areas.

**Data collection and management requirements related to Health Information**

Portability and Accountability Act, billing for services provided, Meaningful Use standards, and clinical practice guidelines result in a rich body of clinical, demographic, and geographic EHR data collected for each patient at each healthcare encounter in one record. EHR data systems are designed to collect and longitudinally link together multiple observations for an individual. With few exceptions, a patient’s demographic information, address, height, weight, temperature, smoking status, and blood pressure are collected at almost every
primary care visit. Detailed clinical information (e.g., medications ordered or dispensed, laboratory test results, or procedures) is not easily collected through surveys. While both BRFSS and EHRs may estimate hypertension prevalence, only EHRs identify individuals moving from uncontrolled to controlled high blood pressure or those who moved from normal blood pressure to hypertension. Accurate self-reported hemoglobin A1C test results and diabetes control status is hampered by recall bias, yet this information is easily extracted from an EHR. EHRs capture and link more complex and detailed information for individuals than can be collected from a brief telephone survey, including repeated observations so that health markers like disease control or severity can be monitored over time.

Because most EHR data are objectively measured, EHR-based information is more accurate compared to survey-based estimates which may be biased by self-report. When self-reporting, individuals are likely to under estimate their weight and overestimate their height leading to inaccurate estimates of BMI and obesity prevalence. Heights, weights, lab results, and blood pressures collected in a clinical setting reflect reading from calibrated instruments (e.g., scales and blood pressure monitors); disease prevalence estimates based on diagnosis codes reflect accepted diagnostic criteria applied by clinically trained professionals.

EHR data are timelier than surveys in two ways. First, EHR data can be accessed and made available soon after collection compared to than extended latency while survey data collection is completed, weighted, and analyzed. When appropriate, EHR data could be used
to track weekly, monthly, or quarterly health trends. Second, when a survey must be expanded to assess an emerging health issue, multiple years may be spent developing and adding a survey question, conducting the survey, and analyzing data compared to EHRs which may have already collected relevant data. Timeliness is essential when tracking emerging health threats like opioids or Zika virus and when monitoring the impact of a policy or intervention. In an emergency, delayed surveillance information is a substantial public health threat. Syndromic surveillance systems that includes chief complaint and diagnosis data from EHRs in hospital emergency rooms, has shown how EHR data can be used as soon as it becomes available to detect rapidly emerging health threats and provide rapid-cycle (e.g., daily or weekly) feedback to local jurisdictions. EHR data can be used in near real-time to generate local information with frequent updates per year to monitor change.

**Limitations of EHR data for surveillance**

EHR data are not without limitations; clinical data from EHRs are collected for clinical and billing uses, and EHR systems were not explicitly designed for population health surveillance. An article by Hoffman and Podgurski stated that, “researchers and analysts who rely on EHR data must proceed with caution and understand the potential limitations of EHRs. Public health findings based on EHR data can be tainted by selection, confounding, and measurement bias.” This review identified a number of limitations to EHR data related to completeness, correctness, concordance, bias, plausibility, and currency as related to surveillance.
EHR data, collected by many different providers and clinical staff within many treatment and care settings for documentation purposes, may not be standardized in method\textsuperscript{135}. Each EHR platform collects and stores patient data differently. The many workflows within a healthcare system and between healthcare organizations cause missing, fragmented, and incomplete data in every EHR system\textsuperscript{137}. Behavioral data such as diet, tobacco use, substance use and contextual information like patient employment and income are often inconsistently collected\textsuperscript{25}. Screening and collection of sensitive data in a standardized manner even within a single organization remains challenging. The quality of patient demographic (e.g., race, language, and ethnicity) information is notoriously poor\textsuperscript{49,135,139}. When combining EHR data across institutions, patterns in missing data and completeness vary by organization\textsuperscript{137}.

EHR data are biased in multiple ways that may introduce systematic and random error into EHR-based estimates\textsuperscript{50,52,135,137}. First, EHRs contain health data collected from a care seeking population that does not necessarily represent the underlying geographic population\textsuperscript{135,137,140,141}. Studies have shown that individuals seen in care are likely to be older, female, non-Hispanic, and insured compared with individuals not seeking healthcare\textsuperscript{141}. While these findings about over-represented groups may be generally true, including EHR data from federally qualified health centers can improve the representativeness by capturing uninsured individuals and minorities. Healthy adults are underrepresented in an EHR dataset, a concept referred to as sick bias\textsuperscript{25}. Individuals often seek out healthcare because of disease or illness while generally well individuals are likely
to forgo a well check-up; this is especially true in young adults\textsuperscript{140,141}. Survey data have shown that diabetes, hypertension, and other chronic diseases are more common among a care-seeking population compared to individuals not visiting a provider\textsuperscript{141}. Second, some self-reported data stored in the EHR, such as tobacco use or nutritional habits, are prone to the same self-reporting bias as surveys\textsuperscript{33,34,51,142}. Third, EHR data from a single healthcare institution are biased by a number of organizational factors including which insurance is accepted, what services are provided, and where the site is located. The lower income patient population of a safety net institution differs dramatically from patients seen at a private for-profit healthcare system providing care only to privately insured patients.

This review identified multiple methods to address EHR bias. To make data more representative and reduce the impact of a single healthcare institution, EHR data can be blended across multiple healthcare institutions\textsuperscript{25}. Bias can further be minimized by restricting EHR data for surveillance to information collected in ambulatory outpatient settings\textsuperscript{25}. Excluding hospital and emergency care can restrict the sick bias by only using data from primary care settings where an individual may seek care if ill or well. Patient data can be geocoded to a small geographic area to better reflect the community populations. Individuals over-represented are weighted to contribute less to population-level estimates while under-represented observations are weighted to contribute more.

**Electronic Health Record-Based Case Definition and Case Finding**

For EHR data to be a viable chronic disease surveillance data source, chronic disease cases must be identified from the care seeking population and aggregated to estimate rates
of population estimates of incidence, prevalence, and control\textsuperscript{143}. A case definition is implemented within EHR data using a computable phenotype which is defined as ‘a clinical condition, characteristic, or set of clinical features that can be determined solely from the data in EHRs and ancillary data sources and does not require chart review or interpretation by a clinician’\textsuperscript{140,143}. A computable phenotype uses structured data commonly collected within an EHR such as diagnosis codes, procedure codes, and lab test results and logic expressions (AND, OR, NOT) to identify cases with no or little manual steps. Phenotypes exist for both case definitions and indicators, where the phenotype defines the appropriate numerator and denominator for a given population health measures (e.g., diabetes prevalence or control)\textsuperscript{144}. A computable phenotype is advantageous because it is reproducible and can be used across healthcare organizations to identically implement a clinical concept or metric. Phenotypes can be designed to overcome the limitations and know bias inherent in EHR data while also standardizing the way patient data are used for population health\textsuperscript{137,144}. Without phenotypes, EHR-based measures are not comparable across settings\textsuperscript{144}. Many phenotypes have been developed for common health conditions and metrics\textsuperscript{140,144}; however, the process and level of rigor applied to developing and validating phenotypes varies. Figure 3 displays the process that the Center for Medicare and Medicaid Services (CMS) follows for developing and implementing phenotypes\textsuperscript{145,146}. A single phenotypes for one condition will not be appropriate for all applications. For a defined use (e.g., quality improvement, surveillance, or research), a computable phenotype should be explicit, reproducible, reliable, and valid\textsuperscript{140}. A valid phenotype
correctly identifies true positives (i.e., individuals with the condition or disease) and does not identify false negatives (i.e., individuals without the disease)\textsuperscript{137,140}. Selecting the best phenotype is important; subtle differences between a phenotype for the same measure can have massive implications at the population health scale\textsuperscript{147}.

The national focus on using EHR data for population health has accelerated the development and validation of many computable disease phenotypes for widespread use. No single comprehensive repository of phenotype information exists\textsuperscript{144}. The NIH Collaboratory is a publicly available electronic clearing house of research tools and information that includes a repository of phenotypes and associated validation results\textsuperscript{148}. The Phenotype Knowledge Base or PheKB is an online collaborative environment organized by the Electronic Medical Records and Genomics Network to publish and disseminate validated phenotype definitions\textsuperscript{149}. CMS maintains an online repository of case and measure phenotypes that have been systematically developed, rigorously vetted, and are frequently reviewed to ensure optimal functioning\textsuperscript{145,146,150}. These repositories provide access to validated and reproducible phenotypes for population health surveillance applications\textsuperscript{137,140}.

Phenotypes are applied to this research for identification and definition of diabetes and hypertension cases, prevalence, and control. CMS measures, having undergone consistent and systematic development, are used whenever possible for this research project. Details about the phenotypes used for this research are described in the methods section.
Validation of EHR-Based Estimates

Validation studies are needed to expand knowledge of the strengths and limitations of EHR data for population health surveillance\(^25,50,135,143\). A review from Paul et al. posed an important question, “To what extent can EHR data support population health surveillance and provide reliable data to set priorities, plan interventions, and monitor progress towards goals”\(^25\)? How are differences between survey and EHR-based estimates similar or dissimilar across health indicators? Though the comparison between BRFSS and EHR are imperfect, exploring possible explanations for differences can offer insights into how EHR data can be effectively used. Do differences between survey and EHR-based estimates shrink as population coverage increases? How do survey-based and EHR-based estimates compare for demographic subgroups? Validation efforts should demonstrate and measure the representativeness of EHR populations, test the efficacy of population weights to overcome misrepresentativeness, and make comparisons with gold standard health survey data to understand differences\(^143\). For measures without a health survey comparison, validation studies may use the literature or alternative comparison data sources such as hospitalization or emergency discharge datasets as a proxy measure for the EHR-based indicator.

Studies that have compared EHR-based estimates to those collected from health surveys have had mostly positive validation results\(^56,138,143\). A 2013 study from Spain found that health surveys detected some conditions at a higher rate (chronic allergies, neck pain, and headaches) and multi-morbidity was higher in health surveys than the EHR (60% vs.
43%; p≤0.001)\(^{151}\). One study found that an EHR-based estimate and NHANES estimate of childhood obesity prevalence were equivalent\(^{152}\). Tobacco use prevalence varied by less than 1% between surveys and primary care EHR data\(^{36}\).

The New York City Macrooscope recently completed validation studies comparing EHR data to an expanded local version of the NHANES, known as NYC-HANES, and a local community health survey (NY CHS) for ten health indicators\(^{47,48}\). Investigators proposed five criteria for validation of EHR based estimates. A EHR measure is valid if its within five percentage points of the survey estimate (equivalence margin), the p-value of a one-sided test of equivalence between and EHR and survey-based estimates is significant, and if a ratio of EHR-based estimate to survey-based estimate is between (0.85-1.15)\(^{47}\). Primary care data from 2013 and 2014 was extracted through the Hub System\(^{43,48}\) and used to estimate prevalence, treatment and control of hypertension, high cholesterol and diabetes, prevalence of obesity, depression and smoking, and influenza vaccination rates. Estimates of hypertension prevalence were comparable between the Macroscope, NY CHS, and NYC HANES (30.7% vs. 30.9% vs. 30.9%)\(^{43,44}\). Estimates of obesity prevalence were 29.5% for Macroscope, 25.4% for NY CHS, and 28.2% for NYC HANES\(^{43,53}\). EHR-based diabetes prevalence estimates were statistically equivalent to estimates from both health surveys. Estimates of hypercholesterolemia prevalence, diabetes and hypertension treatment, and diabetes and hypertension control were not equivalent between data sources. The directionality of differences (e.g., which source produce higher estimates) varied by indicator. Additional studies are needed to better understand inconclusive findings, build
consensus about validation methods, and delve into coverage, weights, and small area estimation strategies.

The Marion County Health Department has engaged in validation work\textsuperscript{153} to understand that coverage and representativeness of EHR data in collaboration with the local health information exchange. EHR data from 2011-2013 were extracted for county residents and coverage rates were calculated to measure the percent of residents represented in aggregated EHR data through the exchange. By census tract, coverage ranged from 50\% to over 120\% of the 2010 census tract population\textsuperscript{153}. Unreasonably high coverage rates could reflect double counting of patients or rapid population growth. A Colorado collaborative investigating the impact of county and census tract level population weights on estimates of obesity prevalence including using a hot decking method to impute missing race values\textsuperscript{154}. For most census tracts, crude estimates were very similar to adjusted estimates (Pearson $R = 0.981$, $P < 0.01$), and the median absolute difference in crude versus adjusted adult obesity prevalence estimates was 0.05 and varied by less than 1 percentage point for most (interquartile range: -0.54–0.80)\textsuperscript{154}.

**Summary of Gaps**

Much of the population health literature is vague and few concrete measurement recommendations could be located. Though LPHAs are ideally suited to carry out surveillance activities to produce local health status information, their understanding of population health surveillance measurement has not been assessed and their interest in and readiness for EHR-based surveillance are unclear. Little could be found describing how
often LPHAs are using EHR for surveillance and scant guidance has been published recommending best practices and lessons learned for success in population health surveillance. How do population health surveillance practices vary across LPHAs? What data are being used and what indicators are being tracked?

Though population health measurement frameworks recommended monitoring chronic disease prevalence and severity and examine trends and disparities between groups at the community level, traditional public health data sources are not designed to provide timely trend analysis or examine sub-county estimates. Many limitations of population-based surveys make it unlikely or impossible that gold standard surveys will feasibly measure chronic disease cost-effectively. EHRs emerge as a promising data source, well positioned to meet the recommendations of population health measurement where traditional data sources like surveys are more challenged. Multiple sources identified insufficient validation evidence as a gap and defined a critical need for validation of EHR-based estimates through comparison with population health surveys and evaluation of the impact of population-based weights to resolve misrepresentativeness.

As population health surveillance matures and evolves, the cadre of recommended indicators will undoubtedly grow and expand. New indicators may focus on disparities, disease severity and comorbidities, utilization and transitions of care, healthcare costs, access to care, and quality of life. Meaningful clinical indicators such as disease control will be translated for population level measurement. The cost limitations and complexity of clinical concepts will challenge health surveys to gather this information. Little has been
published exploring how detailed clinical information from EHR data (e.g., lab test results, prescriptions, and procedures) could be used to generate more detailed population health insights not assessed by surveys.
Table 1: Recommended Population Health Measures by Domain

<table>
<thead>
<tr>
<th>Population Health Measurement Domain</th>
<th>Recommended Population Health Measures: example data sources or collection tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Outcomes</td>
<td>• Mortality: Years of Potential Life Lost; Life expectancy; standardized mortality ratio</td>
</tr>
<tr>
<td></td>
<td>• Health and Functional Status: Single-question assessment or multi-domain assessment</td>
</tr>
<tr>
<td></td>
<td>• Healthy Life Expectancy: Combines life expectancy and health status into a single measure, reflecting remaining years of life in good health</td>
</tr>
<tr>
<td>Intermediate Outcomes</td>
<td>• Incidence (yearly rate of onset, average age of onset) of major chronic conditions</td>
</tr>
<tr>
<td></td>
<td>• Prevalence of major chronic conditions</td>
</tr>
</tbody>
</table>

Figure 1: Conceptual Framework for Public Health Surveillance
Figure 2: The Measures Management System Blueprint

- Conduct environmental scans
- Perform Gap Analysis
- Develop Business Cases
- Convene a Technical Expert Panel
- Submit to CMS for Approval

- Post for public review
- Perform initial feasibility studies
- Convene a Technical Expert Panel
- Refine the business case
- Submit to CMS for approval

- Develop test plans
- Post test results for public review
- Convene a Technical Expert Panel
- Submit to CMS for approval

- Develop implementation plans and harmonize
- Publish MUC List for public review
- MAP review
- Submit to CMS for final approval
- Refine measure specification
- Consensus endorsement

- Collect measure data
- Aggregate measure data and create reports
- Assess impact, effectiveness, and efficiency
- Publish reports and assessments for public review
- Evaluate reports and make assessments
- Submit revision and new measure recommendations to CMS
CHAPTER III

METHODS

This dissertation employed a mixed methods design to achieve the research aims: qualitative methods for aim 1 and quantitative methods for aims 2 and 3. According to mixed methods experts, mixing qualitative and quantitative methods can minimize the limitations of each method\textsuperscript{59,60}. Engaging in an initial phase of qualitative data collection (aim 1) to inform subsequent quantitative methods is characteristic of sequential exploratory design (aim 2 and aim 3). Sequential exploratory design is recommended to explore a phenomenon as well as when developing and testing a new instrument. This design is well suited for this research because it explores EHR-based surveillance as a novel approach and tests EHR for population based measurement. The use of mixed methods adds value to this research by generating evidence about ability of LPHAs to use EHR data in a population health surveillance system combined with assessments about the validity of EHR data for population monitoring\textsuperscript{59,60}.

Aim 1

To gain insight into the landscape of local population health surveillance, key informant interviews with LPHAs were conducted. A qualitative approach provided vital information on a diverse landscape of LPHA perspectives. Interviews were focused on the following questions:
The purpose of the interview was to understand each organization’s approach to local population health surveillance and define key issues, challenges, and barriers LPHAs are facing. Findings from aim 1 were used to develop recommendations that were vetted with LPHA experts and used to tailor aims 2 and 3 to reflect the LPHA perspectives and capacity.

Study Population and Setting

Aim 1 targeted surveillance stakeholders at LPHAs nationwide. Local was defined as public health agencies working at the county or city level. National public health agencies or state public health departments were not included in this analysis.

Sampling

We identified surveillance stakeholders from 25 LPHAs to participate in key informant interviews. LPHAs representing larger geographic populations (400,000 and above) were prioritized. Though the experiences of smaller LPHAs are important, the focused inquiry about EHR data was unlikely to be relevant to smaller agencies where limited resources often prohibit the development of robust surveillance infrastructure.

LPHAs were identified from national population health collaboratives such as the Big Cities Health Initiative, NACCHO, and the Academy Health Electronic Data Methods
Population Health Community of Practice. Individual contacts for each LPHA were identified from a review of websites, publications, reports, and through membership lists from the collaborative groups above. Key informants were contacted via email between April and May 2017 and invited to participate in a one hour phone interview. Only one individual was interviewed from each organization. Invitations included the option to select another representative from the organization to participate in their place.

Interview Guide

Prior to conducting interviews, a detailed interview guide was developed and piloted with two LPHAs (Appendix A). The guide included the key informant protocol and questions used for interview.

Interview

Phone interviews with 25 key informants were scheduled at a time that was convenient for the informant. A description of the study was provided in the invitation and informants provided verbal consent before the interview began. A semi-structured format was used to allow informants to talk for the amount of time they wanted and to answer questions in the manner and style they chose. Interviews were recorded and analyzed for key themes. The interviewer reconciled interviews within one week of conducting them to ensure rapid analysis. This approach allowed for quick assessment of the content and gave a sense of variation or gaps in information.
Analytic Approach

At the completion of data collection, the interviewer used open coding and integrated interview data using matrix analysis\textsuperscript{59,60}. The interviewer conducted two rounds of coding to ensure complete capture and definition of major and minor themes. During initial coding, the interviewer identified major themes that were captured in a code book. Subsequent rounds of coding were used to assess alignment between interviews and major themes, as well as identifying secondary or emergent themes. A second coder listened to and coded 3 of 25 interviews. Results from both coders were compared to ensure inter rater reliability. Coding continued until no new concepts were identified (i.e., thematic saturation). Results were organized to align with the structure of the interview and identified themes within each interview question and a summary of results that informed subsequent aims. To conclude analysis, facilitators and barriers were described and used to generate recommendations. To ensure validity, recommendations were vetted with experienced public health stakeholders not participating in key informant interviews for feedback and then finalized in the discussion section of this document.

Ethics

A protocol describing aim 1 methods was submitted and approved by the Colorado Multiple Institution Review Board and determined to be exempt (Protocol 17-0324) and required no further review.
Aims 2 and 3

The goal of aim 2 was to generate estimates of chronic disease prevalence from EHR data aggregated across multiple healthcare institutions for comparison with survey-based estimates. The goal of aim 3 was to explore what additional insights EHR could provide by generating estimates of chronic disease control from EHR data aggregated across multiple healthcare institutions. Methods required for aims 2 and 3 are described together based on numerous similarities.

Study Population and Setting

This dissertation focused on Denver County population health between 2011 and 2015. Denver County is a rapidly growing urban county in central Colorado and home to roughly 12.5% of the state’s 5.46 million residents\(^\text{155}\). Since 2010, Denver’s population has grown by approximately 15% from 600,158 to 693,060 in 2016\(^\text{155}\). Most of Denver’s residents are of white race (53%) and one in three (31%) identify as Hispanic\(^\text{155}\). The median age of a Denver resident was 34 years in 2014, slightly younger than median age nationally (37 years) and for Colorado (36 years)\(^\text{155}\).

Data Sources

Three data sources were used for aims 2 and 3 of this research: county and census tract population estimates to assess coverage, county prevalence estimates from health survey data, and EHR data. Population estimates for Denver County and its 144 census tracts were collected from the American Community Survey\(^\text{105}\). Health survey data from
BRFSS were used to make comparisons at the county level\textsuperscript{30,102}. Health information from EHRs was extracted from three EHR data sources, two integrated healthcare systems and one managed care network, and were used to estimate county and census tract hypertension and diabetes prevalence and control\textsuperscript{100}.

American Community Survey

The American Community Survey (ACS) samples approximately 10,000 Denver residents each year and produces annual 1) county population estimates and 2) rolling 5-year census tract population estimates (i.e., 2010-2014)\textsuperscript{105}. At the county level, annual chronic disease prevalence rates used the corresponding year’s annual ACS population estimates. Denver has 144 census tracts. Within each census tract, the total population ranges from 705 to 10138 and adult population ranges from 698 to 7680. One Denver census tract has a population of 0. Geospatial analysis of prevalence at the census tract level used the 5-year census tract (2010-2014) population estimates.

Denver County overall, age group-specific, and gender-specific population estimates from the American Community Survey were used to assess representativeness of EHR datasets and to calculate population-based weighting methods. While the 2010 census captures exact population counts, Denver has experienced rapid population growth since 2010. Using ACS population estimates for the observation period improved the accuracy and utility of weighted estimates.
Behavioral Risk Factor Surveillance System

Data from the 2011-2015 Denver BRFSS samples were used to compare survey and EHR-based estimates. The Colorado BRFSS began in 1990 as a joint project of the Colorado Department of Public Health and Environment (CDPHE) and the CDC. Each year, the Survey Research Unit at the CDPHE administers the BRFSS survey and uses random digit dialing to sample respondents. Colorado residents with telephones or who speak a language other than English or Spanish are surveyed. In Colorado, BRFSS employs a stratified sampling design based on 21 health statistics regions to ensure that a sufficient sample is achieved for stable HSR and, as is true in Denver County, county estimates. Survey respondents confirm their county of residence as a part of survey administration.

BRFSS includes questions about patient demographic characteristics, healthcare utilization, and general health status. Some questions are administered every year and some on more ad hoc schedules (e.g., every two or three years). BRFSS questions used in this research included:

- Have you ever been told by a provider that you have diabetes? (included 2011-2015)
- Have you ever been told by a provider that you have high blood pressure? (included 2011, 2013, and 2015)

BRFSS collects information about respondent’s race and ethnicity, age, and gender that were used to produce estimates for each demographic subgroup. This research used respondent counts, weighted disease prevalence point estimates, and confidence intervals from BRFSS annual survey results provided by CDPHE.
Electronic Health Record Data

EHR data from three healthcare organizations providing care to Denver residents was included in this analysis: Denver Health and Hospital Authority (DH), Kaiser Permanente of Colorado (KPCO), and University of Colorado Health and Hospital (UCH). Each entity provided information on patient encounters, demographic characteristics, geocoded residence (to the census tract level), vital signs, diagnoses, and lab test results for patients residing in Denver County and receiving care between 2011 and 2015. A custom data extraction program was developed and executed manually at each site to assemble population health datasets that meet the regulatory standards for limited datasets. Site datasets were stripped of site identifiers and aggregated.

These sites were selected to maximize the population coverage. DH is Denver’s integrated healthcare safety-net system which provides inpatient, outpatient, and emergency care to a majority of Denver’s uninsured and Medicaid populations. DH takes care of approximately 200,000 individuals each year. KPCO is Colorado’s largest managed care provider with approximately 600,000 members, most of whom reside in the Denver Metro region. UCH is a large integrated health system with locations across Colorado’s Front Range.

Geolocation Data

Aims 2 and 3 required geospatial analysis and census tract level estimates based on a patient’s residence. Each healthcare organization has collected patient address, geolocated the address to a Federal Information Processing Standard or FIPS code, and stored
that information at their site. DH and KPCO utilized an internal geocoder (e.g., ESRI product) and UCH partnered with an external service to geocode their data. Because historical location information is not uniformly available, this research used a patient’s most recent geolocation. Some addresses could not be geocoded to a specific census tract but were attributable to a county; these were included in county-level analysis, but excluded from census tract analysis. Individuals who reported experiencing homelessness were attributed to Denver County and included in the county analysis.

**Inclusion Criteria**

The study population for this dissertation was adults (18 years and above) residing in Denver County between 2011 and 2015. Age was calculated based on the date of healthcare encounter (for EHR data) or the date of survey administration (for survey data). Individuals less than 18 years of age or those reporting a residence outside of Denver County were excluded from this analysis.

**Analysis**

Aim 2 analyses used EHR data to generate and compare diabetes and hypertension prevalence estimates using three different measure phenotypes. Diabetes and hypertension prevalence was evaluated within a care population defined as the number of adults seen in an ambulatory care setting during that year. Three prevalence phenotypes, which combined case definitions for the care population and diabetes and hypertension cohorts, were evaluated and one phenotypic approach was selected for the remainder of the analysis.
Using the selected phenotype, annual Denver prevalence rates for diabetes and hypertension for each year (2011-2015) and demographic group were estimated. Annual diabetes and hypertension cohorts were weighted to correct for imperfect representativeness by age group and gender and adjusted annual diabetes and hypertension prevalence estimates were generated at the county level. County crude and adjusted EHR-based and survey-based estimates of disease prevalence were compared. Aim 3 use the diabetes and hypertension cohorts identified in aim 2 to calculate rates of diabetes and hypertension control for 2011 to 2015.

Prevalence and control analysis described above were replicated at a smaller geographic unit: the census tract. To improve the stability of census tract estimates, two years of EHR data (patients receiving care in 2014 and 2015) were aggregated and deduplicated by census tract to estimate tract-level diabetes and hypertension prevalence and control rates and generate prevalence and control maps.

Data extraction and analysis were performed in Structured Query Language (SQL) through SQL Server 2016 and SAS 9.3. Maps were generated using the ESRI ARC GIS software.

**Case Definitions and Phenotypes**

Case definitions and phenotypes were used throughout this analysis. Case definitions were implemented using computable EHR phenotypes to identify disease cases, calculate prevalence, and estimate rates of control. Because this research was not focused on using unstructured EHR data like notes, only rule-based phenotypes that use structured
data were considered. Rule-based phenotypes based on structured data are relatively easy to implement, easily explainable and interpretable, and are reproducible\textsuperscript{156}.

\textit{Care Population Phenotype}

A case definition and phenotype were created to assemble the care population which is the population that acts as the denominator for prevalence analysis. The care population included adults (18 years of age and older on the date of their healthcare encounter) seeking primary or specialty care in an ambulatory setting between 2011 and 2015 at any of the three data contributors. Emergency and inpatient care were excluded to reduce the likelihood that a patient would be represented in multiple systems and counted as a duplicate, an analytic practices of other similar studies\textsuperscript{25,48,55}. This study estimated and compared three care population phenotypes: 1 encounter in 1 year, 2 encounters in 1 year, and 2 encounters in 2 years.

The care population was compared to 2013 ACS population to consider representativeness. Population coverage estimates overall and by demographic group were calculated.

\textit{Hypertension and Diabetes Case Phenotypes}

Case definitions for diabetes and hypertension cases informed aim 2 and aim 3 analyses and are described in Table 2. Hypertension and diabetes case definitions focused on the most common types of these diseases and excluded manifestations of these disease related to other factors such as comorbid conditions (e.g., pregnancy-induced hypertension) or alternative causal pathway (e.g., gestational diabetes and type I diabetes).
A phenotype for a hypertension case was used to identify a hypertension cohort and assess prevalence and control. Hypertension is defined as the condition of abnormally high blood pressure and can lead to damage to the heart, blood vessels, and brain such as a stroke. Blood pressure is considered high when it rises above 140 mm Hg systolic and 90 mm Hg diastolic. A diagnosis of hypertension typically requires multiple high blood pressure readings. Because many factors such as pain, stress, anxiety, or physical activity can cause blood pressure to rise, one high blood pressure observation is not recommended to identify hypertension cases.

Multiple hypertension case phenotypes are in use. The New York City Macroscope identified a hypertension case as individuals with one or more hypertension diagnosis codes while CMS identified a hypertension case as an individual with two or more hypertension diagnosis codes. Based on these examples, this research used hypertension ICD-9 and ICD-10 diagnosis codes to identify hypertension cases (Appendix B) and compared combinations of 1 and 2 diagnosis code phenotypes. Only final diagnoses from an outpatient setting were used to evaluate cases.

Type 2 diabetes mellitus may be indicated in the EHR by the presence of diabetes ICD-9 and ICD-10 diagnosis codes, evidence of the use of pharmacotherapy to treat diabetes, and common laboratory tests often used to screen or monitor diabetes (e.g., Hemoglobin A1C or glucose). Multiple diabetes case phenotypes are in use; one study identified 23 distinct diabetes phenotypes. The New York City Macroscope identified a case as individuals having one or more diabetes diagnosis codes while CMS identified a case...
as an individual with two or more diabetes diagnosis codes\textsuperscript{47,99,150}. Diabetes ICD-9 and ICD-10 diagnosis codes (Appendix B) were used to identify diabetes cases and combinations of 1 and 2 diagnosis code phenotypes were assessed and compared. Only final diagnoses from an outpatient setting were used to evaluate cases.

\textit{Hypertension and Diabetes Control Phenotypes}

Best practice guidelines recommend that hypertensives and diabetic patients be seen by a provider at least annually\textsuperscript{158,159}. Depending on the intended use, thresholds used to assess hypertension and diabetes control vary across and within institutions. The diabetes and hypertension control thresholds for this research were selected by reviewing the literature to identify the most common thresholds and selecting the threshold for each diaseae that was used by national measurement frameworks (e.g., HEDIS and NCQA) or what was used by similar validation studies\textsuperscript{47,160,161}.

Hypertension control was assessed from systolic and diastolic recorded blood pressure levels among a cohort of hypertensive patients using measure definitions shown in Table 3. A literature review failed to identify a consistently used control threshold\textsuperscript{47,73,79}. While many providers use the HEDIS hypertension control threshold of systolic < 140 and diastolic < 90 for reporting, other providers use slightly different thresholds for older adults and other subpopulations\textsuperscript{161}. Though hypertension control thresholds differ by age and comorbidity, this study defined an individual controlled if both systolic and diastolic blood pressures were below recommended levels (systolic=140 and diastolic=90). Only blood pressure readings collected during an ambulatory encounter were included. Measures
where one but not both systolic and diastolic blood pressure readings were present were excluded. If multiple blood pressures from multiple dates for the same individual were present, the last blood pressure during the observation period was included. Hypertensive patients without recorded blood pressures were excluded. A similar study completed in NYC by the Macroscope found that very few hypertensive patients had to be excluded from control analysis because of missing blood pressure measures.47

Control of diabetes was assessed using measured hemoglobin A1C (HGBA1C) laboratory test results among a cohort of individuals previously diagnosed with diabetes. HGBA1C is the primary tool providers use to monitor diabetes control and diabetic patients are recommended to receive at least one HGBA1C test per year. Much like hypertension, no one best control threshold could be identified from the literature.47 For diabetes control, HEDIS defines a HGBA1C over 9% as poorly controlled and recommends organizations track what percent of patients are above this threshold. However, many healthcare organizations strive to keep diabetes HGA1Cs lower, below 8% or 7%, to achieve a higher level of diabetes control. For this research, an individual’s diabetes was defined as controlled if their HGBA1C was less than 9%.72,75 Individuals without an HGBA1C test result were excluded from the control denominator. Records of an HGBA1C where results were inconclusive or the sample was contaminated were excluded. All HGBA1C lab tests regardless of care settings were included in diabetes control estimates. If multiple lab tests were present during an observation period, the latest HGBA1C test result during the observation period was included.
Phenotype Comparison

This research considered and compared three phenotypes for hypertension prevalence and three phenotypes for diabetes prevalence using data from 2 healthcare providers and one year of data (2013) with the goal of selecting one phenotype to be used for the research. Case definitions identifying the care population and cases were paired to create each prevalence phenotype. Two of three prevalence phenotypes reflect what is currently used by CMS (Augmented) and the New York City Macroscope project (Annual) for observational research and surveillance\(^47,150\). The Extended phenotype expands the structure of the Annual phenotype over a longer observational time period.

Prevalence phenotypes combined specifications regarding the number of diagnoses and time period during which those diagnoses were collected as well the number of encounters and the time period during which those encounters occurred. The ideal care population phenotype identified the largest and most representative care population. The ideal case phenotype balanced sensitivity and specificity to identify the most number of true positives and the least number of false positives. Combined, case and care population phenotypes predict the most accurate population prevalence estimate.

Prevalence estimates were compared to 2013 county prevalence estimates from BRFSS. Selection of one phenotype was based upon the size and representativeness of the care population as well as the alignment between EHR-based and survey-based prevalence estimates.
Disease Prevalence Estimation

Using the selected care population and case phenotype, adult diabetes and hypertension annual prevalence were estimated for 2011 to 2015. Each year, every individual in the care population was evaluated using case criteria. Individuals who were seen multiple times during that year, were counted only once during each calendar year and contributed one observation for each year for which they had an encounter. Cochran-Armitage test of trends was used to assess the significance of temporal trends.

Adjusted Prevalence Estimation

Population weights were used to adjust county-level prevalence estimates for misrepresentativeness. Weighting is based on the assumption that some age and gender groups are under or over represented and the application of weights can correct the impact of over and under represented groups. For this research, weights will be applied to make cases from overrepresented groups count less and cases from underrepresented groups count more. Annual ACS population estimates and standardized census population proportions were used to adjust the prevalence numerator for each age and gender group to estimate the projected number of cases in each strata for each year. Adjustment was done in Microsoft Excel and adjusted prevalence was then calculated as the percent of the ACS population represented by the projected number of cases. Seven age groups (i.e., 18-19, 20-24, 25-34, 35-44, 45-54, 55-64, and 65+) and two genders resulted in 14 age and gender strata. Individuals with an unknown gender were exclude from adjusted estimates.
Adjusted estimates were compared to unadjusted estimates and differences were quantified and described.

**Chronic Disease Control Estimation**

The purpose of chronic disease control estimation was to assess county and census tract level rates of diabetes and hypertension control. Adult diabetes and hypertension control rates were estimated based on two-year aggregated disease cohorts identified in aim 2. Markers of disease control (i.e., blood pressure levels and HGBA1C levels) were extracted from EHR data. Diagnosed hypertensive patients who had no blood pressure observations recorded and diagnosed diabetic patients who had no HGBA1C test results during that year, despite evidence of a medical visit, were excluded.

**Geospatial Analysis**

Population differences can lead to different epidemiological profiles of disease within each census tract. The purpose of geospatial analysis was to visualize the variation in health status within Denver County at the census tract level that cannot be detected through county level measurement. Past community health assessments have documented how several health outcomes such as obesity and depression can vary between Denver census tracts. For each census tract, population coverage, the proportion of census tract population represented by CHORDS data, was calculated and used to describe the range of coverage and measure the association between coverage and EHR-based health estimates. Similar coverage tools do not exist to understand the geographic representation of the BRFSS sample.
Four maps, showing diabetes and hypertension prevalence and control at the census tract level, were created. For each map, two years of data (2014 and 2015) were aggregated to improve the performance and precision of tract-level estimates. A single observation was selected for each individual to ensure that was only counted on time on a map. Individuals who could not be geocoded to a Denver census tract, despite being attributed to Denver county, were excluded from the geospatial analysis.

Though the care population was large, some exclusions and warnings were used to draw attention to information that should be interpreted with caution. Prevalence estimates based on a care population of less than 50 individuals may be unstable and were flagged as ‘insufficient data’. Census tracts with estimates based on insufficient data were included on maps but flagged to encourage the user to interpret and use with caution.

**Ethics**

A protocol describing aim 2 and 3 methods was submitted to the Colorado Multiple Institution Review Board and did not meet the definition of research involving human subjects and did not require IRB review and approval.
Figure 3: Overview of Sequential Exploratory Research Design

Table 2: Diabetes and Hypertension Case Definitions

<table>
<thead>
<tr>
<th>Disease</th>
<th>Case Definition</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes Mellitus</td>
<td>Diabetes mellitus type 2, controlled or uncontrolled</td>
<td>Gestational diabetes, type I diabetes, chemically induced (secondary) diabetes, neonatal diabetes, polycystic ovarian syndrome, hyperglycemia, pre-diabetes, or similar states or conditions (such as impaired fasting glucose or glucose intolerance)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Essential hypertension, hypertensive heart disease, hypertensive chronic kidney disease, hypertensive heart and kidney disease</td>
<td>Myocardial infarction, congestive heart failure, angina, portal hypertension, secondary hypertension, gestational hypertension, or pregnancy-induced hypertension (unless there was a preexisting diagnosis of hypertension), borderline hypertension, or similar nonspecific comment</td>
</tr>
</tbody>
</table>
### Table 3: Overview of Control Measure Case Definition

<table>
<thead>
<tr>
<th>Control Measure</th>
<th>Numerator</th>
<th>Denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Control</td>
<td>Hypertension Control: systolic &lt; 140 and diastolic &lt; 90</td>
<td>Hypertension case from basic, augmented, or extended hypertension prevalence phenotype in Table 2 who had a blood pressure taken.</td>
</tr>
<tr>
<td>Diabetes Control</td>
<td>Diabetes Control: HGBA1C &lt; 9%</td>
<td>Diabetes case from basic, augmented, or extended hypertension prevalence phenotype in Table 2 who had a HGBA1C test result.</td>
</tr>
</tbody>
</table>

### Table 4: Overview of Case and Prevalence Phenotypes

<table>
<thead>
<tr>
<th>Prevalence Phenotype</th>
<th>Case (Numerator) Definitions</th>
<th>Care Population (Denominator) Definitions</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>at least 1 hypertension diagnosis code in 12 months</td>
<td>at least 1 healthcare visit in 12 months</td>
<td>NYC Macrooscope project&lt;sup&gt;47&lt;/sup&gt;</td>
</tr>
<tr>
<td>Annual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmented</td>
<td>at least 2 hypertension diagnosis codes in 12 months</td>
<td>at least 2 healthcare visit in 12 months</td>
<td>CMS Chronic Disease Warehouse&lt;sup&gt;11&lt;/sup&gt;</td>
</tr>
<tr>
<td>Extended</td>
<td>at least 2 hypertension diagnosis codes in 24 months</td>
<td>at least 2 healthcare visit in 24 months</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>at least 1 diabetes diagnosis code in 12 months</td>
<td>at least 1 healthcare visit in 12 months</td>
<td>NYC Macrooscope project&lt;sup&gt;47&lt;/sup&gt;</td>
</tr>
<tr>
<td>Annual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmented</td>
<td>at least 2 diabetes diagnosis codes in 12 months</td>
<td>at least 2 healthcare visit in 12 months</td>
<td>CMS Chronic Disease Warehouse&lt;sup&gt;150&lt;/sup&gt;</td>
</tr>
<tr>
<td>Extended</td>
<td>at least 2 diabetes diagnosis codes in 24 months</td>
<td>at least 2 healthcare visit in 24 months</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER IV

RESULTS

Results are presented by research aim. This research was conducted between January 2017 and April 2018.

Aim 1

The objective of key informant interviews was to characterize LPHA beliefs, practices, and barriers related to local population health surveillance. Key informant interviews were used to:

a) Assess LPHA approaches to population health surveillance including data sources, dissemination strategies, stakeholders, ideas about validation of a new data source, and unmet surveillance needs,

b) Explore organizational structure and investment in surveillance including roles and responsibilities related to surveillance work,

c) Explore how and why LPHA population health surveillance practices have evolved due to greater focus on population health, and

d) Explore each informant’s perspective on EHR-based surveillance including barriers and facilitators affecting local population health surveillance capacity.

Key Informants

Twenty eight LPHAs were recruited to participate in interviews. LPHAs were recruited from the Big Cities Health Initiative, NACCHO, and the Academy Health Electronic
Data Methods Population Health Community of Practice. Three LPHAs declined to participate in the interview because they did not feel the topic was relevant to their work or did not have staff time to commit to the interview and did not significantly differ from those who did participate in an interview. Twenty five key informants participated in an interview from January to August 2017 (Table 5).

Informants had varying depth of knowledge and experience. Some interview questions were outside of the knowledge or experience of the individual interviewed. Informants in high level leadership positions, like directors, had less detailed information about day to day surveillance activities while informants in operational and analytic roles, like epidemiologists, had less insight into the reason behind organizational decisions.

The following section is organized to align with the structure of the interviews and describes responses to interview questions (See Appendix A for interview guide with full questions). Appendices B-G include supporting quotes from key informants for each results section. Double coding was used to validate the findings from a subset of key informant interviews. A second coder analyzed 3 of 25 interviews (12%) to ensure that the primary researcher’s interpretation of interviews were accurate; coders achieved over 90% agreement on themes, barriers, facilitators.

**Approaches to Population Health Surveillance**

LPHAs were asked to describe their approach to population health surveillance including data sources, dissemination strategies, stakeholders, unmet surveillance needs, and methods for validating a new data source. Informants cautioned that surveillance activities...
were spread throughout the department and no one person could be aware of all activities. Approaches to population health surveillance were diverse and were assigned into three categories: standard, active, and advanced.

- A **standard** approach entailed using only traditional surveillance data sources like birth and death information and the BRFSS survey to produce surveillance information required by statutes and grants.

- An **active** approach included all standard activities plus limited efforts to seek out additional non-traditional sources of population data (e.g. hospital discharge data) to enrich information from traditional data sources.

- An **advanced** approach included all standard activities and many efforts described within an active approach with an extensive portfolio of population health data, multiple parallel efforts to acquire and use new data sources, and complex analytic activities outside of traditional public health epidemiology.

Table 6 displays the distribution of informants by surveillance category with examples of surveillance activities in that category.

Sixteen percent of LPHAs were categorized as standard. LPHA with a standard approach were generally located in smaller geographic areas, had smaller surveillance teams, less infrastructure, and were more limited by resources.

*We have focused very specifically on getting out the data that people most need and not developing new surveillance systems...We made a decision that we were not going to go there [syndromic surveillance]. We felt it was much more important for our local area to have a good solid foundation of*
morbidity, mortality, behavior, and environmental indicators and determine which one of those we could provide to the community at a sub-geography level. We believe that was a role we needed to play. Quote from Director of Epidemiology and Preparedness at county LPHA with a Standard Surveillance Approach

Almost half of LPHAs (44%) were categorized as having an active surveillance approach.

LPHAs in the active category were conducting all of the standard surveillance activities and had made a modest investment in surveillance resources to gain access to more detailed data.

*Our approach to population health surveillance begins with acquiring as much population level data as possible. We spend quite a bit of time acquiring datasets that we know exist but also investigating other sources of information.* Quote from Director of Research and Evaluation at city LPHA with an Active Surveillance Approach

*Our approach is to help produce data that drives action. That requires us to have data at the most disaggregated level possible.* Quote from Director of Assessment, Policy, and Evaluation at county LPHA with an Active Surveillance Approach

Forty percent of LPHAs were categorized as using an advanced surveillance approach and were generally located in dense urban centers, responsible for larger populations, and had larger surveillance teams who often specialized in specific data types, datasets, software, or analytic methods.

*When we become aware of data that’s relevant to the population, we try and get it. We actively look for data that’s relevant to issues of higher importance to the health of the population.* Quote from Director of Epidemiology at county LPHA with Advanced Surveillance Approach

*It’s been a population health approach of fits and starts, depending on who’s in charge and the perspective of the individual appointed... The energy,*
resource commitment, and attention devoted to population health surveillance is different.... a lot depends on how the local government decides to structure the work. Quote from Director of Health Information at city/county LPHA with Advanced Surveillance Approach

LPHAs with an advanced approach could pinpoint a past event that caused extreme growth in their surveillance capacity including: a large grant from a foundation or federal agency, a partnership with a mayor or policy maker that led to an increase in stable surveillance funding, a partnership with an EHR vendor or health information exchange, or a strong coalition of partners, often related to local community health improvement planning efforts.

LPHAs differentiated between surveillance activities required by statute (i.e., ‘what we have to do’) and other surveillance activities that they related to population health. Surveillance activities supported by a statute were associated with traditional surveillance data sources (e.g. death certificate data and BRFSS) with better developed systems that staff felt are easy to use, understand, interpret, and describe to stakeholders.

What we do at a local health department is really driven by what we have to do. I don’t think you can look more broadly at population health from a health department perspective without also looking at the requirements and activities that we have do to, whether it is associated with funding or local and state law....We also have some additional activities that we’ve looked at broadly to bring information in. Syndromic surveillance is something we invested in early on that our state partners at the time didn’t choose to put their investment. Quote from Director of Epidemiology at City LPHA

Consistency in expectations, datasets, and deliverables for statutory surveillance responsibilities led them to feel confident and certain in their ability to do this work.

Informants felt that population health surveillance use cases were more challenging than traditional public health epidemiology because they are not mandated by statute or policy,
are rarely funded or supported by national public health organizations, and require more complex analytic techniques using larger and complex datasets (including combining multiple datasets). Informants recognized a false assumption exists that skills from tradition surveillance tasks (e.g., BRFSS survey analysis) can be easily translated for population health uses with non-traditional data (e.g., diabetes prevalence estimation using EHR data).

When asked why population health surveillance was important, every LPHA described surveillance as critical to understanding the needs of the population that they served and cited surveillance as an essential public health service.

*With population health you can really get into what it means for the people you are supposed to be serving. It’s not just statistics, its characteristics, neighborhoods, its zip codes, its census tracts and blocks. Quote from Director of Health Information at city/county LPHA*

*If you are going to keep the population healthy, you need to know what ails them. Quote from Senior Epidemiologist at county LPHA*

Informants felt that LPHAs are a neutral and impartial party and more equipped to do this work than any other organization.

**Data Sources**

LPHAs were asked to describe the data being used by their organization for population health surveillance. LPHAs are using multiple data sources to respond to surveillance needs and identified between 5 and 21 traditional and non-traditional surveillance data sources listed in Table 7, noting that they would be challenged to remember them all. Multiple informants described their responsibility to “acquire as much
data as possible” and went on to clarify that the realm of data is no longer restricted to health data.

Informants stated that using traditional surveillance data sources is part of population health surveillance but that additional non-traditional data sources are needed to meet the demands and expectations of internal and external stakeholders. Informants access traditional data sources through public websites or the state health department but they forge their own relationships with other partners to procure non-traditional data sources. While one relationship with the state health department provided multiple data sources to an LPHA, each non-traditional data source requires one or more stakeholder relationship. Roughly half of LPHAs interviewed conducted their own community survey(s) to gather additional information or obtain a larger sample from priority subpopulation while others invested in oversampling BRFSS for priority populations that they identified.

Informants identified a number of non-traditional data sources and described the pursuit of non-traditional data sources for population health surveillance as high value as well as high risk, challenging, and rarely funded. LPHAs felt that non-traditional data recruitment, exchange, and analysis was challenging because the systems housing non-traditional data were often antiquated or ill designed for easy extraction and analysis.

*We spend a lot of time getting it [population health data]. It takes a lot of effort to acquire and maintain. The built environment data changes all the time and every time we use it we have to revalidate it before I use it again.*

*Quote from Director of Research and Evaluation at City LPHA*
Without funding, informants are challenged to commit staff time and resources to bring in and mastering a new data source and expressed concern about the inability to accurately estimate how much time and resources would be required to get a new dataset up and running.

**Dissemination**

Informants recognized dissemination as a critical to showing the value of population health surveillance and a key role for LPHA staff.

> We've tried to create an environment where the data that we collect and share as well as what partners collect and share informs all the work throughout the city. From that perspective, doing the actual surveillance work and sharing is very important. I feel like one of our main missions is to disseminate. Get it out there as much as possible so that people can use it. We want all the people involved in the public health system to have the information they need to make decisions based on as much data as possible. 

*Quote from Director of Research and Evaluation at City LPHA*

Stakeholders expect to be able to request and access data and information at multiple levels including raw datasets for public use, the data resulting from analysis, and information about the findings of an analysis and are constantly receiving data requests across datasets for each of these products. Many LPHAs feel like they have been unable to meet the expectations of their community partner in a timely way, especially since the push for data-driven decision making has increased the frequency, complexity, and detail of stakeholder demands. For example, requests for the top ten leading causes of death have been replaced by requests for life expectancy or tobacco use prevalence by neighborhood.
LPHAs disseminated data and information through presentations, ad-hoc reports, and web-based tools but they identified the community health assessment (CHA) as the primary mechanism to share local population health information. LPHAs felt that CHAs had increased the volume and engagement of external partners. The frequency of assessments ranged from every 1, 2, 3, or 4 years. While some LPHAs described CHA activities as a standalone report, others named their CHA as central to their department’s strategic planning, community engagement, and prioritization of work.

Stakeholders

LPHAs provided extensive lists of internal and external stakeholders for population health data and spoke about the importance of enriching and deepening the relationship with both groups of stakeholders through data sharing and surveillance activities. LPHAs felt that data were one of the most powerful tools to bring new stakeholders into a partnership. Stakeholders were also discussed as sources for new data and forging new partnerships was part of searching for data. Internal stakeholders included public health programmatic staff and leadership. For LPHAs closely tied to a city or county government, internal stakeholders included other government agencies. External stakeholders included community partners, students and faculty at academic institutions, insurance companies, foundations, healthcare professional organization such as the American Heart Association, healthcare providers, media, city or county agencies, and lay community members. The number of internal and external surveillance stakeholders has grown substantially as a
A culture of data-driven decision making has been adopted across the health sector and the value of public health surveillance has increased.

LPHAs felt that the state health department was one of the most crucial surveillance stakeholders. Most felt that the state was the most important surveillance data source and partner as well as a significant challenge in their daily work.

*Our relationship with the state health department is good but there are so many roadblocks in place to sharing record level data. Each department has its own set of rules and policies around sharing. Some may be easier than others regardless of sensitivity of data, for no good reason. The state houses so much data that if, well integrated, could provide a great picture of population health. Quote from Director of Epidemiology at City LPHA*

Informants felt that states were struggling to balance statutory responsibility and respond to LPHA needs causing delays and denials to data requests. LPHAs described that state health departments are overwhelmed by their statutory and data stewardship burden and are unable to initiate or participate in novel surveillance activities.

LPHAs discussed a growing interest in population health surveillance from hospitals and other healthcare institutions. The mandate for non-profit hospitals to conduct a community health needs assessment has increased the level of collaboration between LPHAs and integrated care systems. A national movement away from fee for service toward value-based payment has increased provider and payer’s interest in community-level health trends and had led some payers to share data, provide funding, or participate in other surveillance activities like CHA.
Academic institutions were identified as an important stakeholder for population health surveillance, universities with a public health school in particular. LPHA partnerships with public health schools provided supplementary surveillance capacity through student time and methodological expertise through collaboration with professors.

**Validating a new data source**

LPHAs described data validation activities as central to their surveillance role and expressed concern that guidelines for validation are not more formalized and structured. LPHAs feel intense pressure to ensure that their data and information are correct and described validating non-traditional data sources such as EHR data, claims, and 911 call data as very challenging. LPHAs did not define a consistent validation method but spoke about the importance of comparisons to other sources, extensive data quality discovery, and detailed documentation of limitations to feel confident in health information using new data sources. Informants suggested that validation should include:

- Reviewing what is known about the data source from the literature,
- Understanding the primary intention for data collection and use,
- Studying data source documentation describing strengths, limitations, and collection methodology,
- Assessing the dataset for completeness and missingness,
- Identify what useful public health information could be extracted,
- Developing and testing multiple use cases (preferably use cases where comparisons to other sources can be made),
• Engage with topical or clinical experts to discuss estimates, and

• Comparing estimates to established data sources that produce the same measure (e.g., EHR-based hypertension prevalence to survey-based hypertension prevalence), or

• Comparing temporal trends and socio-demographic patterns to established data sources that produce similar proxy measures (e.g., hypertension control compared to stroke death or emergency department estimates).

LPHAs noted that frequently, no perfect comparison for validation exists and that using proxy measures for clinical related concepts or using national estimates were suitable alternatives.

Unmet Surveillance Needs

Informants describe many unmet surveillance needs related to both population health data and systems including delayed data, inadequate sub-county estimates, and lack of data on many priority health conditions. LPHAs identified that the limitations of traditional data sources kept them from effectively:

• Detecting changes from year to year

• Measuring an indicator more often than once per year

• Measuring health among a critical minority population like non-Hispanic blacks, Asians, or native Americans

• Measuring the health of targeted demographic groups like children
• Measuring health at a sub-county or neighborhood level

• Cultivating a detailed understanding of priority issues by measuring multiple indicators for a given disease

When discussing unmet surveillance needs, LPHAs were aware that EHR data had the potential to address almost all of these gaps. LPHAs felt that surveillance information about causes and conditions of death and disability were less helpful than data describing the impact of social determinants and community factors on the incidence and prevalence of common health conditions. LPHAs identified delays in tradition data sources procured from the state as the most challenging limitation.

*Getting data 1-2 years later doesn’t give us any ability to engage with the community about real time health issues. And it [data] was at such a macro level that is did not tell us the story of our neighborhoods. Quote from Director of Public Health at city/county LPHA*

Specific to chronic disease, LPHAs identified surveys as the gold standard data source but felt that surveys did not provide enough information to inform local community level chronic disease interventions. LPHAs used estimates of county prevalence for diabetes and hypertension from BRFSS with some complementary use of emergency department and hospital discharge datasets. LPHAs felt uneasy about using hospitalization or emergency department data as a proxy for disease prevalence.

Key informants felt that more timely detailed health data could 1) increase community buy in and commitment to place-based prevention, 2) expedite program
implementation, 3) reduce the waste of resources for intervention efforts in low burden areas, and 4) increase the frequency of feedback on impact of the intervention.

When LPHAs spoke with community partners and staff about data, data limitations and insufficient surveillance information dominated their conversations. Community partners felt that traditional data was not local enough and too old to be relatable and desired real time data for a specific community to detect change quickly and assess the significance of disparities between groups. Communities desired indicators that show progress upstream from major health events like death or hospitalization. Community stakeholders requested neighborhood health information and find little use in county rates. While data gaps varied by jurisdiction, every LPHA interviewed could quickly identify many health topics and indicators where more data were needed.

Informants identified a need for precise estimates of chronic disease burden and estimates of incidence and severity for common chronic illnesses like diabetes and hypertension.

One of the biggest barriers that we face in surveillance is in chronic disease and injury data. When you don’t have these state mandated processes in place to report, it is very difficult to get data. We know chronic conditions are what is ailing our population, but we don’t have regulated required data systems. EHRs, because of HITECH, are very appealing; its promising that one day we will be able to mine and use that data for many different purposes: 1) to fill this surveillance void for chronic disease and injury information and 2) the ability of a reportable disease specialist to log directly into a system to look up labs and clinical information that are not faxed to us. Quote from Director of Epidemiology and Informatics of county LPHA.
LPHAs wanted the ability to measure chronic disease more frequently than once per year and without extended delays. LPHAs desired robust chronic disease incidence and prevalence estimates for socio-demographic subgroups such as small minority populations. LPHAs aspired to conduct surveillance on more chronic diseases like asthma, chronic obstructive pulmonary disease, and heart disease.

**Organizational Structure and Investment in Population Health Surveillance**

Surveillance capacity was closely linked to how the surveillance resources were organized within the health department. Informants included a description of team size, reporting structure, and orientation within the LPHA when describing their approach to population health surveillance. Some LPHAs housed surveillance resources within topic-specific public health teams such as an epidemiologist within the cancer, chronic disease, or HIV program team. Other LPHAs house a surveillance and/or informant team that worked across topic areas housed within the strategic planning bureau, in a standalone special projects branch, or within the office of the director or health commissioner.

*We realigned the office of epidemiology under the department of strategic planning. So rather than it sitting under the medical officer with the other surveillance units, we moved the unit so that we sit in the midst of the planning and policy groups to create synergy and promote use of the data.*

*Quote from Director of Epidemiology at City LPHA*

Several LPHAs had re-organized the surveillance team as part of an effort to rebrand surveillance under the guise of population health, often moving the team to report to a higher level of public health leadership.
When discussing organization structure, LPHAs describe the broad continuum of surveillance roles and responsibilities as a data consumer, data steward, methodological expert, and information provider. As a data consumer, they were responsible for seeking out new data including cultivating partnerships, putting data sharing agreements in place, and validate the data source for use. As an information provider, LPHAs were responsible for transforming, manipulating and analyzing data including telling a story with data to reach the communities they serve. Informants described their role as the both interpreter of data and the liberator of data. LPHAs felt that no members of the public health workforce had the training and experience to carry out all of these tasks and that effective data stewardship responsibilities for each data source required a coordinated effort by multiple individuals.

Informants felt that the time and resources required for governance, data exchange, and data stewardship was not sustainable and many LPHAs were over-leveraged having more data sources than they had time to use responsibly. LPHAs accessing data from multiple external sources described an overwhelming burden to maintain those partnerships and data exchange agreements that at times felt overly cumbersome. Further, the resources required to take on a new data source are daunting and LPHAs described them, as time costs are poorly calculated or hard to understand by leadership.

**Investment**

Surveillance practices were heavily influenced by how the organization had invested in personnel, infrastructure, technology, data, and partnerships. LPHAs were most likely to
invest in personnel, both adding new positions and training existing staff. LPHAs invested in surveillance by conducting their own community health survey on topics not addressed by BRFSS or to clarify the findings from BRFSS. Survey funding was provided by foundations, city or county funding, and other community partners such as insurance payers.

For LPHAs using many non-traditional data sources, informants described an investment in governance and stewardship as well as web-based platforms to make data available to external partners. Some LPHAs had positions exclusively committed to data stewardship. LPHAs with many data sources also described an investment in time to maintain and revalidate data sources, especially true of geospatial data.

Informants discussed blending hard and soft funding to support surveillance personnel and infrastructure which includes 2 to over 10 full time epidemiologists. Few LPHAs had core funding for part of or their entire surveillance team.

*It was hard to advocate for more resources because it was hard to demonstrate what we were delivering beyond the very traditional public health services. This part [population health surveillance] of what we do has never been known because we haven't actually done it. Quote from Director of Public Health at city/county LPHA*

LPHAs were able to add personnel by using grant funds and writing new proposals to maintain personnel funding or taking advantage of healthy budget periods to request a budget expansion for surveillance positions. LPHAs also grew the surveillance team by converting other programmatic positions through attrition (like retiring nurses or dietitians) to epidemiologists or informaticians.
Other surveillance investments included purchasing datasets from the state health department or other data providers, developing and maintaining a web portal, setting up and maintaining surveillance electronic infrastructure, purchasing additional questions on the BRFSS survey, and purchasing a larger sample for BRFSS or YRBS surveys in their jurisdiction.

**Evolution of Population Health Surveillance**

LPHAs felt that the volume, diversity, frequency, and complexity of surveillance data expectations has increased coincident with the growing interest in population health and movement toward data-driven decision making.

*It used to be that you [public health] were trying to explain that there was a problem. Now we are trying to explain what we are doing about it and what services are you providing.* Quote from Epidemiology Manager at county LPHA

*I felt it as a secular change in terms of what the role of public health is. The arcane ways of public health from the old days in terms of producing life tables and vital statistics rates in rote fashion. Those days are over. While we do still produce those, we primarily recognize that the role of PH has changed to embrace the role of all of these social determinants. The epidemiology and surveillance unit cannot exist without strong cross sector collaboration from other city agencies and outside partners who have access to data not collected by the health department.* Quote from Director of Epidemiology at City LPHA

Growing interest in surveillance data had many positive implications on LPHAs who perceive the field of public health as more reputable than in the past. Public health staff are increasingly viewed as subject matter experts in local sources of data and their use to explain health trends. LPHA surveillance capacity has expanded by hiring more staff,
training existing staff in new skills, working with new and more complex information sources, and engaging in more advanced analytic efforts such as small area estimation or blending multiple data sources to measure complex or nuanced health factors. LPHAs have seen existing stakeholders more interested in data in recent years. LPHAs have been able to use data to enter into partnerships with existing stakeholders and identify new stakeholders using local data.

**Perspectives on EHR Data for Population Health Surveillance**

LPHAs believed that EHR data could address many unmet surveillance needs and felt that community partners expected them to have access to real time EHR data already. EHR data were desirable because it was more frequently available and could be used for small area or subgroup analysis. Most LPHAs could identify many ways they would use EHR data to complement traditional data sources and monitor health conditions or indicators where no surveillance data is currently available.

*We don’t need this [EHR] data to tell you what the problem is. We need this data to show progress in small geographic areas through individual level data. Quote from Manager of Epidemiology at county LPHA*

When asked about their ability to use EHR data, LPHAs level of readiness and capacity spanned a broad continuum. Responses to EHR-based surveillance questions were used to create four categories of EHR-based surveillance functionality: awareness, pursuit, evaluation, and operational.
Table 8 includes a description of each of these categories and the distribution of LPHA by utilization and Table 9 compares LPHA jurisdiction size by level of EHR-based surveillance activity. Though one objective of these interviews was to understand how public health stakeholders are using EHR data, few (16%) LPHAs interviewed were conducting EHR-based surveillance. LPHAs operational with EHR-based surveillance were serving populations under 1,000,000 and all LPHAs in the evaluation category were serving the largest population (>1,000,000). Despite grave concerns about their capacity, 36% of LPHAs were actively building systems and partnerships to pursue access to EHR data (EHR data did not include syndromic surveillance data or platforms).

Operational LPHAs were accessing EHR data through a distributed EHR network or through a partnership with a health information exchange, health system, or EHR vendor. How EHR data were being used varied based on funding mechanism, interest and political will of data owner(s), and technical capacity for data transformation and analysis. Some LPHAs had limited access to EHR data for one health topic while others had access to the entirety of EHR data to monitor their health priorities. LPHAs were using EHR data for surveillance for exclusively communicable or chronic conditions including tuberculosis, HIV, hepatitis, asthma, depression, diabetes, falls, and opioids. No operational LPHAs had more than three surveillance indicators that they were tracking with EHR data. LPHAs felt that pilot activities to demonstrate the feasibility of EHR based surveillance combined with the value of related local population health data were crucial.
Barriers

LPHAs identified many barriers impacting their ability to use EHR data for surveillance that can be grouped into technical, resource, political, and organizational barriers.

**Technical barriers** included complexity of data, access to IT support, cost of software, cost of infrastructure, and insufficient technical skills within public health staff. Public health staff were ill-equipped to manage complex surveillance datasets. LPHAs were challenged to efficiently consume external datasets, store data securely, and make the data accessible to surveillance stakeholders in an appropriate format. Informants described multiple instances of technical barriers delaying or prohibiting their ability to access EHR data.

Insufficient resources for surveillance were a recurring challenge. **Resource barriers** included on time, personnel, cost of software and IT support, and cost of infrastructure. Time is needed for existing staff to apply for grants, conduct data discovery and analysis, validate new data sources, translate findings into visualization and develop key messages, respond to data requests from external partners, and provide technical assistance to surveillance data customers. Informants felt like time and funding needed to train existing personnel for more complex data sets and analysis was insufficient. Pursuing access to EHR data required LPHAs to develop partnerships and governance for data exchange and was described as a long unfunded process that yielded slow progress. Informants called out the lack of alignment between the analytic and data management training from MPH programs.
and what is required to deal with real world public health datasets. This experience was compounded by the growing availability of new data. The complexity of EHR data required LPHAs to invest more time and resources upfront. EHR-based surveillance activities were taxing for analytic resources because EHRs yield huge dirty dataset that required extensive discovery, cleaning, transformation, and interpretation.

_The minute that you step into it [EHR data], you realize how complicated it is. It’s not how we think. As epidemiologists we think about events and we are much better at figuring out what constitutes an event and having an appropriate denominator or population at risk. Quote from Director of Assessment and Planning at county LPHA_

LPHAs felt that many epidemiologists or analysts did not have the capacity to analyze an EHR dataset without additional training. LPHAs felt challenged to understand the underlying structure and meta-data behind EHR databases and were hesitant about transforming clinical data into surveillance information without strong clinical data experts. LPHAs struggled to find the resources to purchase Tableau, ArcGIS, SAS, and SQL Server, software packages needed to use and disseminate data effectively. When software or infrastructure could be purchased, LPHAs encountered challenges within procurement, installation, and maintenance. LPHAs were distressed that resources had not increased to meet the growing demand for surveillance, analytic, and data expertise.

**Political barriers** related to the local political landscape were directly linked to difficulty prioritizing and allocating an investment in surveillance resources and data sharing. LPHAs without robust surveillance capacity identified a lack of political support for public health activities and a main contributor. Lacking political will was a key barrier that
blocked cultivation of new data sharing partnerships or expand existing partnerships.

LPHAs identified circumstances where competing demands for political support caused a surveillance project to take a back seat.

LPHAs identified the primary organizational barrier as bureaucracy and described many ways that bureaucratic processes negatively impacted surveillance. Legal rules, processes, and red tape delay and attenuate the process of developing and executing data sharing agreements. Procurement processes create delays that can slow, challenge, or totally block hardware and software requests, even when funds were not an issue. Informants felt that external data sharing partners had become discouraged or frustrated and had abandoned data sharing projects with LPHA due to these challenges.

Facilitators

Despite many barriers, LPHAs identified facilitators which had enhanced their overall surveillance capacity for progress toward EHR-based surveillance.

Strong, diverse partnerships with traditional and non-traditional public health partners were a key aspect of successful and high functioning surveillance teams. LPHAs recommended cultivating a broad base of partners to efficiently recruit data sources, identify funding opportunities and resources, increase the reach of surveillance information, and advocate for the value of surveillance outside of public health circles. LPHAs using EHR for surveillance had unique partnerships with an EHR vendor, a health information exchange, a clinical data research network, or a clinical translational science awarded institution.
An advanced population health surveillance approach was often attributed to public health leadership with a deep appreciation for data combined with awareness about the time and resources required to use data appropriately. LPHAs recommended an intentional effort to educate public health leaders about both the value of population health data and the resources required to monitor health effectively. Data savvy public health leaders were more likely to advocate for and prioritize data and surveillance resources and invest in surveillance workforce and infrastructure.

LPHAs using EHR data for surveillance had invested in infrastructure; the required storage infrastructure and software was a major concern for LPHAs pursuing EHR data. LPHAs that invested in infrastructure or collaborated with another agency to share infrastructure were functioning at a higher level. Because public health operates in a resource constrained environment, LPHAs suggested thinking creatively to identify any opportunity to invest in data and technology infrastructure. Insufficient information technology infrastructure and support held LPHAs back from accessing valuable novel surveillance datasets.

LPHAs made a connection between responding to emerging surveillance trends and keeping public health surveillance and data relevant. Several LPHAs recalled that participating in open data initiatives, data collaboratives, and novel and usually unfunded surveillance projects had opened the door to new partnerships, data sources, and opportunities. Relevance helped LPHA successfully advocate for resources. LPHAs
recommended maintaining a public health presence in emerging opportunities to stay relevant, even when resources and time are scarce.

Summary of Findings

The purpose of these interviews was to characterize the population health surveillance landscape and understand if and how LPHAs were using EHR data for population health surveillance. LPHAs were positioning themselves to lead population health surveillance in their communities and were experiencing greater expectations from public health and community stakeholder for more detailed and timely data, especially data at the neighborhood level.

*It [Surveillance] is one of the core parts of the science of public health. It’s linked to epidemiology, the basic science of public health. If you are going to keep the population healthy, you kind of have to know what ails them. Quote from Senior Epidemiologist at County LPHA*

When LPHAs were unable to respond to requests from community stakeholder with traditional data sources, some felt overwhelmed by unmet surveillance needs while others focused on messaging what can be measured to manage community partner expectations. As LPHAs struggled to analyze data available to them, they continued to pour resources into accessing non-traditional data sources to assess emerging issues, health disparities, local geospatial trends, and the link between place and outcomes. Because external funding is rare, LPHAs were attempting to do more with less including recruiting data, building infrastructure, and preparing their staff for robust population health surveillance activities.
LPHAs believed that EHR data presents a rare opportunity to examine small geographic areas and demographic subgroups, more disease states, multiple indicators for a disease state, and be combined with information about place and social determinants to gain a deeper understanding of health.

*Population health surveillance, especially incorporating EHRs, is an elephant. We’ll be nibbling one bite at a time.* Quote from Director of Epidemiology and Informatics of county LPHA

Less than 20% of LPHA were actively using EHR data; most LPHA felt unprepared to take on EHR-based surveillance and its many challenges. Accessing EHR data had multiple compounding political, partnerships, resource, and technical barriers. Those who could access EHR data struggled to define indicators and analytic approaches or identify staff to carry out the analysis. Those who could analyze EHR data were often confused about how to interpret it, especially in the absence of clear guidance documents. Those who could interpret findings were challenged to clearly communicate the assumptions and limitations to EHR-based surveillance with end users. Despite all this, almost every informant was actively trying to increase their institutional readiness for EHR-based surveillance, by addressing barriers identified in pursuit of obtaining more information about the communities they serve.

**Aim 2**

Results for aim 2 include a comparison of chronic disease phenotypes and description of county and census tract diabetes and hypertension prevalence.
Phenotype Comparison

Phenotypes were compared using one year of data (2013) from two of the three EHR data contributing healthcare organizations. Comparisons were made between phenotypes for the care population, diabetes cohort, hypertension cohort, and associated prevalence estimates.

Comparison of Care Population

Three care population phenotypes were compared: extended, annual, and augmented. The extended phenotype required individuals to have 2 or more ambulatory encounters in two years (2012 and 2013). The annual phenotype required individuals to have one ambulatory encounter in 2013. The augmented phenotype required individuals to have two ambulatory encounters in 2013. Results from each phenotype were compared in the tables and figures below.

Using data from 2 healthcare providers, the annual care population was the largest and included 120,303 adults which is equivalent to 25% coverage of the Denver adult population in 2013. The extended care population selected 119,804 for 25% percent coverage. The augmented care population selected 83,588 individuals and had the lowest population coverage at 17%. Population coverage and counts were similar when comparing the extended and annual care populations. The annual care population had the highest population coverage for seven of twelve age, gender, and racial ethnic sub-groups.

When making comparisons by age group, population coverage was higher for older adults (55-64 and 65+) and lower for young adults (18-24 and 25-34 years). Over 30% of
Denver adults 55-64 and 65+ were present in the annual care population while only 23% of 18-24 year olds and 20% of 25-34 year olds were captured. All three care populations had higher population coverage for women (29%, 29%, and 21%) compare to men (20%, 20%, and 13%).

For racial ethnic subgroups, more of Denver’s black and Hispanic adult population was represented compared to the white population. In 2013, the annual care population contained data from 28% of Denver black non-Hispanic and Hispanic adults but only 17% of Denver white non-Hispanic adults. The meaning and impact of missing racial ethnic information is difficult to interpret and a limitation of EHR data. The annual and extended care populations included over 10,000 Denver adults with a missing race and ethnicity while the smaller augmented care population had roughly 7,000 individuals with a missing race and ethnicity. No population estimates exist for individuals with a missing or unknown race making it impossible to calculate a coverage estimate. While these individuals are not missing from the age group and gender coverage estimates, it is impossible to know how population coverage estimates for the extended, annual, and augmented care population would be different if this information was available.

The distribution of ACS population estimates in 2013 was compared to distribution of care population and relative differences were calculated. Based on population estimates, adults 25-34 years, females, and white non-Hispanics are the largest population groups. Within extended and annual cohorts, adults 25-34 years, females, and white non-Hispanics were also the largest groups; 10% of adults had an unknown race and ethnicity. Using the
augmented care population adults 65+, females, and white non-Hispanics were the largest
group with 8% of adults having an unknown race and ethnicity. When calculating the
relative difference, all three cohort definitions under represent younger adults and over
represent older adults. However, the under representation of 25-34 year olds, despite being
the largest group by age, remained problematic with over 5% difference. All cohorts over
represented females and under represented males by roughly 10%. White non-Hispanic
adults were under represented by care populations by 17% and 18%..

Comparison of Diabetes and Hypertension Phenotypes

Based on the three care population phenotypes defined above, three diabetes and
hypertension cohorts were identified with a parallel structure. Diabetes and hypertension
cohorts were paired with care populations to create three estimates of diabetes and
hypertension prevalence. Prevalence was calculated as cases divided by the care
population. Phenotypes for diabetes and hypertension cases mirrored the care population
phenotypes but used diagnosis codes in lieu of ambulatory encounters. Cases were selected
from the care population to exclude individuals who had diagnosis codes without any
evidence of healthcare utilization. Extended cases had two or more diagnosis codes (not on
the same day) during 2012 or 2013. Annual cases had 1 or more diagnosis codes in 2013.
Augmented cases had two or more diagnosis codes from separate encounters in 2013.
Table 12 displays a comparison of cases and prevalence estimates. Annual hypertension and
diabetes cohorts identified the most cases. Extended and annual cohorts identified very
similar counts of individuals. Augmented case cohorts were the smallest; annual
hypertension cases exceeded augmented hypertension cases by 79% and annual diabetes cases exceeded augmented diabetes cases by 37%.

For hypertension, annual prevalence estimates were the highest (23.79%) and augmented prevalence estimates were the lowest (19.09%). Estimates of diabetes varied from 12.13% using the extended phenotype to 12.85% using the augmented phenotype but were all within one percentage point.

This comparative analysis identifies extended and annual care population and prevalence phenotypes as the most representative of the underlying geographic population. The annual phenotype produced the largest care population. Though it was only slightly larger than the extended, the annual care population allowed for annual measurement which was timelier and allowed for year over year comparison. The annual phenotype definition was selected from these two options to ensure clarity when producing annual estimates and studying trends over time.

**Care Population**

Annual estimates for adults in care ranged from 131,060 to 147,098 unique individuals. Table 13 compared the care population each year. In all five years, white non-Hispanic individuals (>40%), females (≈60%), and adults 25-34 years of age (≈22%) were the largest racial ethnic, gender, and age subgroups represented in the care population.

Representativeness of groups within the care population compared with population estimates was consistent across years. Care population members 65+ expanded from 17% in 2011 to 19% of the care population in 2015. Females decreased from 60% of the care
population in 2011 to 59% in 2015. White non-Hispanic individuals increased from 43% in 2011 to 47% in 2015 and unknown race ethnicity decreased from 10% in 2011 to 7% in 2015. Table 14 compares the care population to population estimates from the ACS in for one year of data (2015); 27% of the Denver county adult population was observed in the care population. Population coverage was highest among 18-19 year olds (34%), 55-64 year olds (33%), 65+ year olds (37%), females (32%), and black non-Hispanic individuals (36%). Population coverage of those with an unknown race was not assessed.

Two years of data (2014 and 2015) were used to assess population coverage for each Denver census tract and is displayed on the map below (Figure 5). One census tract was excluded and is shaded with hash markings because it had no adult residents, according to the ACS. Each census tract had between 706 and 8852 adult residents and between 92 and 3570 adults with an EHR observation. On average, each census tract had 1333 adults in the EHR data for 3593 adult residents. No census tracts were excluded because they had less than 10% coverage.

Based on EHR data from 2014 and 2015 and population estimates from 2015, coverage varied from 11% to 109% of the adult population with a mean of 38% coverage and a median of 34% coverage. One census tract had coverage over 100% meaning that more individuals were present in the care population than were thought to live in that census tract. Explanations for high coverage include temporary addresses that assign invididuals to a census tract where they do not reside or that ACS could not assess such as correctional facilities, healthcare providers, or transitional housing. Additionally, address
values such as ‘homeless’ or ‘transient’ combined with a zip code could be erroneously assigned to a census tract in central Denver.

Figure 6 was created to visualize and assess the relationship between adult population and population coverage to detect if more populated census tracts had higher or lower coverage in general. While the scatter plot suggests a slight inverse relationship, no significant positive or negative linear relationship existed between the two factors ($R^2=0.0152$).

**Diabetes Prevalence**

Between 2011 and 2015, over 13,000 prevalent diabetes cases were identified from the care population each year. Cases had between 1 and 151 diabetes-related diagnoses codes for visits during in a given year. In each year, 20-25% of diabetes cases had one encounter where only 1 diagnosis code for diabetes was recorded. When used to calculate prevalence, 10% of the care population met the case definition for diabetes in 2011, increasing to 11% in 2012 through 2015 ($p \leq 0.0001$; Table 15). Diabetes was most common among adults 65+ (≈25%), men, black non-Hispanic, and Hispanic adults. Diabetes was least common among young adults 18-24 years of age and white non-Hispanic adults.

Age and gender population weights were applied to crude EHR-based prevalence estimates and are displayed in Table 16. Both crude and adjusted diabetes prevalence estimates were higher than survey-based estimates for all five years. The age and gender adjustment created a slightly higher adjusted prevalence in 2011 and 2012 and a slightly
lower adjusted prevalence in 2013 to 2015. Only 2013 adjusted diabetes prevalence differed from the crude prevalence by more than 1 percentage point.

Figure 7 displays visual trends in crude and adjusted EHR-based diabetes prevalence compared to survey-based prevalence. Survey-based estimates of diabetes prevalence varied substantially year over year with a dramatic decline between 2014 and 2015 from 7.4% to 5.3%. Crude and adjusted EHR-based estimates were relatively flat between 2011 and 2015 and above every annual survey-based prevalence estimate, falling outside of the confidence intervals of survey-based estimates. Applying the population weights minimized the appearance of a significant diabetes prevalence increase between 2011 and 2015 from a 1 percentage point increase to a 0.1 percentage point increase. Between 2012 and 2013, adjusted EHR-based prevalence dropped from 11.1% to 9.6%. While crude EHR-based estimates increased 1 percentage point between 2011 and 2015, adjusted EHR-based estimates increased 0.1 percentage points during the same time period.

Care population and diabetes case data from 2014 and 2015 were combined to estimate diabetes prevalence by census tract (Figure 8). Across 143 tracts, 9.2% of the care population had a diagnosis of diabetes with a median of 9.3%. Diabetes prevalence ranged from 17.9% to 2.3%. Figure 10 displays the geographic distribution of diabetes prevalence in Denver. Because no census tracts had a care population of less than 50 individuals or coverage below 10%, no census tracts are marked with insufficient data. Diabetes prevalence was highest in the southwest and northern census tracts in Denver. Twenty one census tracts had diabetes prevalence below 5%.
**Hypertension Prevalence**

In 2011 and 2015, 26,839 and 29,118 hypertension cases, respectively were identified and 51,734 unique individuals were identified with hypertension across all five years. Hypertension cases had between 1 and 86 diagnoses codes in a given year. Annually, between 44 and 46% of cases had one encounter with only 1 diagnosis code for hypertension. When used to calculate prevalence, between 19% and 22% of the care population annually met the case definition for hypertension (Table 16). Hypertension was most common among adults 65+ (~25%), men, black non-Hispanic, and Hispanic adults. Hypertension was least common among young adults 18-24 years of age, white non-Hispanic adults, and adults of an unknown race. Hypertension prevalence decreased between 2011 and 2015 from 20% to 19%. Though hypertension prevalence changes overall and within groups were subtle, p-values suggest that changes were statistically significant for all groups except young adults, and adults of an unknown or other race.

Age and gender population weights were applied to crude prevalence estimates to produce adjusted EHR-based prevalence estimates for hypertension. Table 18 displays crude and adjusted EHR-based hypertension prevalence for 2011 to 2015 and makes comparisons to survey-based prevalence for years that the BRFSS survey included a hypertension question (2011, 2013, and 2015). Like diabetes prevalence, population weighting increased the adjusted hypertension prevalence in 2011 and 2012 and decreased estimates for 2013-2015.
Figure 19 displays crude and adjusted EHR-based hypertension prevalence and survey-based estimates of hypertension prevalence in 2011, 2013, and 2015. Weighting pushed the adjusted estimate toward the survey-based estimate in 2011 and away from the survey-based estimate in 2013 and 2015. Both crude and adjusted EHR-based hypertension prevalence estimates were below survey-based estimates and outside the 95% confidence interval in all three comparisons. Hypertension prevalence varied by census tract (Figure 10) from 4.3% to 41.7%. The mean and median hypertension prevalence across 143 census tracts was 16.4%. Because no census tracts had a care population of less than 50 individuals or coverage below 10%, no census tracts are marked with insufficient data. Hypertension prevalence was not highest in the same census tracts as diabetes; hypertension was most common in the southeast, southwest, and north central Denver census tracts.

**Summary of Findings**

Analysis from aim 2 identified that the annual prevalence phenotype maximized population coverage, timeliness, and representativeness. The annual phenotype was more timely than the extended phenotype, which identified a similar size care population. The augmented phenotype identified an older and less representative care population compared to the annual care population. EHR data suggests that diabetes prevalence significantly increased from 10% in 2011 to 11% in 2015 ($p \leq 0.0001$) and is higher than survey’s suggest (7.4% in 2015). EHR-based hypertension prevalence declined slightly from 20% in 2011 to 19% in 2015 ($p \leq 0.0001$) and below survey based estimates (26.8% in 2015). Adjusted prevalence estimates were very similar to crude prevalence estimates and
minimized the diabetes prevalence increase and hypertension prevalence decreases seen between 2011 and 2015. Census tract prevalence estimates varied widely; hypertension prevalence spanned 4.3% to 41.7% and diabetes prevalence spanned 2.3% to 17.9%.

Aim 3

Results for aim 3 include a description of county and census tract diabetes and hypertension control rates.

Diabetes Control

Diabetes control was assessed based on the last hemoglobin A1C test results for the year being examined. Table 19 shows the breakdown of all diabetes cases compared to those with a lab test by year. Cases without an HBGA1C lab test were excluded from estimates of diabetes control.

Approximately 90% of diabetes cases had an HGBA1C lab test result between 2011 and 2015. While the loss of 10% of diabetes cases is concerning, the NYC MacroScope also found that 10% of diabetes cases were missing an HGBA1C lab test result and could not be assessed for control\(^\text{47}\). A retrospective analysis of the 10% of diabetes cases without a HGBA1C in 2015 shows that 823 (49%) had no HGBA1C test ever recorded and 51% had records of HGBA1C tests recorded in previous years. Diabetes control in Denver increased from 80% in 2011 to 81% in 2012 and 2013 and then decreased to 79% in 2014 and 78% in 2015 (Table 20). Diabetes control did not increase in any age gender or racial/ethnic group between 2011 and 2015. Diabetes control estimates in younger age groups with smaller
diabetes cohorts were erratic from year to year with big increases and decreases; the
changes were not statistically significant. Diabetes control decreased among those 45-54
years, 55-64 years, and 65+ years by 2 to 5 percentage points. Diabetes control decreased
among Hispanics, white non-Hispanics, and those with an unknown race.

Two years of data were combined to estimate diabetes control by census tract. On
average, 127 diabetes cases with HGBA1C tests were identified in each census tract to
estimate control; HGBA1C test results from 100 patients were controlled for a mean control
rate of 81%, higher than county control rates of 79% in 2014 and 78% in 2015 (Figure 11).
Diabetes control varied widely between census tracts from 69% to 100%. In two Denver
census tracts, all diabetes patients had a controlled HGBA1C test result.

**Hypertension Control**

Hypertension control was assessed using systolic and diastolic blood pressure
readings among adults diagnosed with hypertension. Individuals with no blood pressure
readings during the year were excluded from control estimates for that year (Table 21). Less
than 5% of cases were excluded from control estimates because a blood pressure was not
recorded despite having an encounter.

Hypertension control decreased among Denver adults between 2011 and 2015 from
74% to 71% \((p \leq 0.0001; \text{ Table 22})\). All demographic subgroups saw a decrease in
hypertension control during the five year observation. In younger age groups (e.g., 18-19
and 20-24) where the number of hypertension cases was low, hypertension control was
volatile with sharp increases and decreases year over year. For example, hypertension
control among 20-24 year olds dropped from 72% to 55% between 2011 and 2015 but p-value do not suggest that such a major decline is significant. Hypertension control was highest among adults over 45 years, white non-Hispanic and Hispanics. Hypertension control did not vary between females and males; both genders show a statistically significant decrease between 2011 and 2015 that mirrors what was seen in the overall trends. Hypertension control was lowest among young adults under 35 years, black non-Hispanic adults, and adults with an unknown race and ethnicity.

Two years of data were combined to estimate hypertension control by census tract. On average, 252 hypertension patients with blood pressures were identified in each census tract to estimate control; blood pressure observations from 181 patients were controlled for a mean control rate of 72% (between 2014 and 2015, county control rates were 73% and 71%, respectively). The range of hypertension control census tract control estimates varied from 57% to 84%. Hypertension control was lowest in central and north central Denver.

To explore how prevalence rates and control rates are associated, Figure 13 visualizes a patterns and the linear trend between the two indicators for diabetes (red) and hypertension (orange). At the census tract level, diabetes prevalence had a statistically significant negative linear relationship with chronic disease control meaning that census tracts where diabetes prevalence was lower were more likely to have high rates of diabetes control (R2=0.2239). Unlike diabetes, hypertension prevalence did not have a significant positive or negative linear relationship with hypertension control.
Summary of Findings

Approximately 90% of diabetes cases had an HGBA1C lab test result between 2011 and 2015. Denver adult diabetes control increased from 80% in 2011 to 81% in 2012 and 2013 and then decreased to 79% in 2014 and 78% in 2015. Diabetes control varied widely between census tracts from 69% to 100% but diabetes control was higher in census tracts where diabetes prevalence was lower. Less than 5% of cases were excluded from hypertension control estimates because a blood pressure was not recorded despite having an encounter. Hypertension control decreased among Denver adults between 2011 and 2015 from 74% to 71%. The range of hypertension control census tract control estimates varied from 57% to 84%. Hypertension control was lowest in central and north central Denver.
<table>
<thead>
<tr>
<th>ID</th>
<th>Role</th>
<th>Type**</th>
<th>Jurisdiction Size***</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Director of Assessment, Policy, and Evaluation</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>2</td>
<td>Director, Epidemiology and Preparedness</td>
<td>County</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>3</td>
<td>Director of Epidemiology</td>
<td>County</td>
<td>700,001-1,000,000</td>
</tr>
<tr>
<td>4</td>
<td>Director of Epidemiology and Informatics</td>
<td>Multi-county</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>5</td>
<td>Senior Epidemiologist</td>
<td>County</td>
<td>700,001-1,000,000</td>
</tr>
<tr>
<td>6</td>
<td>Director of Assessment, Planning, and Special Projects</td>
<td>City</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>7</td>
<td>Director of Assessment, Planning, and Special Projects</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>8</td>
<td>Director of Epidemiology</td>
<td>City</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>9</td>
<td>Director of Epidemiology</td>
<td>City</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>10</td>
<td>Epidemiologist</td>
<td>City/County</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>11</td>
<td>Director of Assessment, Planning, and Evaluation</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>12</td>
<td>Epidemiology Manager</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>13</td>
<td>Director Assessment and Epidemiology</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>14</td>
<td>Director of Research and Evaluation</td>
<td>City</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>15</td>
<td>Epidemiologist</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>16</td>
<td>Community Health Improvement Planning Supervisor</td>
<td>County</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>17</td>
<td>Epidemiology Manager</td>
<td>County</td>
<td>700,001-1,000,000</td>
</tr>
<tr>
<td>18</td>
<td>Director of Epidemiology, Informatics, and Surveillance</td>
<td>County</td>
<td>700,001-1,000,000</td>
</tr>
<tr>
<td>19</td>
<td>Director of Community Health</td>
<td>Multi-county</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>20</td>
<td>Director of Health Information</td>
<td>City/County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>21</td>
<td>Chief Epidemiologist</td>
<td>County</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>22</td>
<td>Chief Epidemiologist</td>
<td>City</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>23</td>
<td>Chief Epidemiologist</td>
<td>City</td>
<td>&gt;1,000,000</td>
</tr>
<tr>
<td>24</td>
<td>Director of Strategic Information</td>
<td>City</td>
<td>400,001-700,000</td>
</tr>
<tr>
<td>25</td>
<td>Director of Public Health</td>
<td>City/County</td>
<td>400,001-700,000</td>
</tr>
</tbody>
</table>
Table 6: Categories of LPHA Approaches to Population Health Surveillance

<table>
<thead>
<tr>
<th>Surveillance Category (n and %)</th>
<th>Size of Jurisdiction</th>
<th>Examples of Surveillance Activities as described by LPHA in Surveillance Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>400,000-700,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>700,001-1,000,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;1,000,000</td>
<td></td>
</tr>
<tr>
<td><strong>Standard</strong> (4 of 25 or 16%)</td>
<td>3</td>
<td>• Producing a community health assessment</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>• Using data from birth and death certificates to assess leading causes of death and mortality rates overall and by socio-demographic group</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>• Using BRFSS data to produce county level estimates of chronic disease prevalence</td>
</tr>
<tr>
<td><strong>Active</strong> (11 of 25 or 44%)</td>
<td>3</td>
<td>• Using hospitalization, syndromic surveillance data</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>• Conducting local health survey to supplement BRFSS data or oversampling BRFSS survey</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>• Combining traditional public health data sources such as merging survey data with ACS or place-based built environment data.</td>
</tr>
<tr>
<td><strong>Advanced</strong> (10 of 25 or 40%)</td>
<td>1</td>
<td>• Using EHR or claims data from a single institution</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>• Using an EHR data network</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>• Conducting modelling analysis or other methods to produce estimates for small area analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Combining two or more data sources to link risk and protective factors to health outcomes</td>
</tr>
</tbody>
</table>
Table 7: Traditional and Non-Traditional Surveillance Data Sources

<table>
<thead>
<tr>
<th>Traditional Data Sources</th>
<th>Non-Traditional Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 Cities census tract level BRFSS estimates</td>
<td>311 calls</td>
</tr>
<tr>
<td>Air quality and water quality monitoring</td>
<td>911 calls</td>
</tr>
<tr>
<td>American Community Survey and Census BRFSS*</td>
<td>All Payer Claims Database</td>
</tr>
<tr>
<td>Cancer registry*</td>
<td>Arrests and other data from police department</td>
</tr>
<tr>
<td>Child Health Survey*</td>
<td>Criminal justice system</td>
</tr>
<tr>
<td>Childhood lead testing*</td>
<td>Department of Motor Vehicles self-reported height and weight data</td>
</tr>
<tr>
<td>Communicable disease reporting systems* (including NEDSS,</td>
<td>Drug use and disciplinary data collected by schools</td>
</tr>
<tr>
<td>electronic lab, and case reporting)</td>
<td>EHR data (from a health information exchange, a single healthcare provider, or group of</td>
</tr>
<tr>
<td>Electronic HIV/AIDS reporting system</td>
<td>healthcare providers)</td>
</tr>
<tr>
<td>Emergency department discharge dataset</td>
<td>Emergency Transport data from Emergency Medical Service</td>
</tr>
<tr>
<td>Hospitalization dataset</td>
<td>Fitnessgram data collected by schools</td>
</tr>
<tr>
<td>Locally administered surveys (often referred to</td>
<td>Syndromic surveillance (e.g., Biosense, Essence, and NSSP)</td>
</tr>
<tr>
<td>as community health assessment)</td>
<td>HIV testing</td>
</tr>
<tr>
<td>National HIV Behavior Survey</td>
<td>Housing eligibility and status from housing authority</td>
</tr>
<tr>
<td>National Drug Use in Health survey</td>
<td>Income tax data</td>
</tr>
<tr>
<td>Outbreak management system (e.g., Maven)</td>
<td>Medicaid eligibility</td>
</tr>
<tr>
<td>Pregnancy Risk Assessment Monitoring Survey*</td>
<td>Medicaid claims</td>
</tr>
<tr>
<td>Trauma registry*</td>
<td>Medical Examiner or Coroner autopsies</td>
</tr>
<tr>
<td>Violent death reporting system</td>
<td>Medicare claims</td>
</tr>
<tr>
<td>Vital Statistics (birth and death records)*</td>
<td>Open data repositories</td>
</tr>
<tr>
<td>Youth Risk Behavior Survey*</td>
<td>Parks and recreation department</td>
</tr>
<tr>
<td></td>
<td>Public health department EHR</td>
</tr>
<tr>
<td></td>
<td>Public transportation use datasets</td>
</tr>
<tr>
<td></td>
<td>Retail food purchase monitoring systems</td>
</tr>
<tr>
<td></td>
<td>Ryan White program</td>
</tr>
<tr>
<td></td>
<td>Walkability scores and sidewalk datasets</td>
</tr>
<tr>
<td></td>
<td>WIC program enrollment and participation</td>
</tr>
<tr>
<td></td>
<td>Zoning and land use datasets</td>
</tr>
</tbody>
</table>

*Accessed via the state health department
Table 8: Level of EHR-based Surveillance Activity

<table>
<thead>
<tr>
<th>Level of EHR-based Surveillance Activity</th>
<th>Number and % of LPHAs in each category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Awareness</strong>: Aware of EHR data but not engaged in activities to access it for surveillance</td>
<td>8 of 25 (32%)</td>
</tr>
<tr>
<td><strong>Pursuit</strong>: Currently engaged in activities to access EHR data for surveillance</td>
<td>9 of 25 (36%)</td>
</tr>
<tr>
<td><strong>Evaluation</strong>: Have access to EHR data and evaluating data for surveillance</td>
<td>4 of 25 (16%)</td>
</tr>
<tr>
<td><strong>Operational</strong>: Have access to and actively using EHR data for local surveillance (including sharing it with the community)</td>
<td>4 of 25 (16%)</td>
</tr>
</tbody>
</table>

Table 9: Level of EHR-based Surveillance Activity by Jurisdiction Size

<table>
<thead>
<tr>
<th>Jurisdiction Size</th>
<th>Level of EHR-based Surveillance Activity (n)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Awareness</td>
<td>Pursuit</td>
</tr>
<tr>
<td>400,000-700,000</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>700,001-1,000,000</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>&gt;1,000,000</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Strata</td>
<td>Denver Population*</td>
<td>Extended 2012-2013</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td>%</td>
</tr>
<tr>
<td>Total Population</td>
<td>619297</td>
<td>-</td>
</tr>
<tr>
<td>Adult Population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in Years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>487825</td>
<td>78.8%</td>
</tr>
<tr>
<td>25-34</td>
<td>130639</td>
<td>26.8%</td>
</tr>
<tr>
<td>35-44</td>
<td>94141</td>
<td>19.3%</td>
</tr>
<tr>
<td>45-54</td>
<td>73689</td>
<td>15.1%</td>
</tr>
<tr>
<td>55-64</td>
<td>64407</td>
<td>13.2%</td>
</tr>
<tr>
<td>65+</td>
<td>64879</td>
<td>13.3%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>243002</td>
<td>49.8%</td>
</tr>
<tr>
<td>Female</td>
<td>244823</td>
<td>50.2%</td>
</tr>
<tr>
<td>Race and Ethnicity**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black non-Hispanic</td>
<td>48621</td>
<td>10.0%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>141085</td>
<td>28.9%</td>
</tr>
<tr>
<td>White non-Hispanic</td>
<td>291400</td>
<td>59.7%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td>6719</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

*2013 Population Estimates for age and gender groups from American Community Survey

**2013 Population Estimates for racial ethnic groups from Department of Local Affairs
<table>
<thead>
<tr>
<th>Strata</th>
<th>Column %</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td>Extended</td>
</tr>
<tr>
<td>Age in Years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>12.3%</td>
<td>11.0%</td>
</tr>
<tr>
<td>25-34</td>
<td>26.8%</td>
<td>21.3%</td>
</tr>
<tr>
<td>35-44</td>
<td>19.3%</td>
<td>17.4%</td>
</tr>
<tr>
<td>45-54</td>
<td>15.1%</td>
<td>17.0%</td>
</tr>
<tr>
<td>55-64</td>
<td>13.2%</td>
<td>16.8%</td>
</tr>
<tr>
<td>65+</td>
<td>13.3%</td>
<td>16.5%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49.8%</td>
<td>39.8%</td>
</tr>
<tr>
<td>Female</td>
<td>50.2%</td>
<td>60.2%</td>
</tr>
<tr>
<td>Race and Ethnicity*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black non-Hispanic</td>
<td>10.0%</td>
<td>11.3%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>28.9%</td>
<td>32.7%</td>
</tr>
<tr>
<td>White non-Hispanic</td>
<td>59.7%</td>
<td>41.8%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>9.7%</td>
</tr>
<tr>
<td>Other</td>
<td>1.4%</td>
<td>4.6%</td>
</tr>
</tbody>
</table>

*2013 Population Estimates for racial ethnic groups from Department of Local Affairs
Figure 4: Care Population Waterfall, Denver, 2013

Healthcare Provider 1
- 7,520,530 encounters between 1/1/2013 and 12/31/2013
  - 1,556,560 ambulatory outpatient clinic encounters between 1/1/2013 and 12/31/2013
  - 633,709 ambulatory encounters between 1/1/2013 and 12/31/2013 among adults 18+ on the date of the encounter
  - 364,361 adult (18+ on the date of the encounter) individuals with at least one ambulatory encounter between 1/1/2013 and 12/31/2013
  - 58,407 adult (18+ on the date of the encounter) individuals with at least one ambulatory encounter between 1/1/2013 and 12/31/2013 residing in Denver County

Healthcare Provider 2
- 836,773 encounters between 1/1/2013 and 12/31/2013
  - 624,921 ambulatory outpatient clinic encounters between 1/1/2013 and 12/31/2013
  - 440,388 ambulatory encounters between 1/1/2013 and 12/31/2013 among adults 18+ on the date of the encounter
  - 54,444 adult (18+ on the date of the encounter) individuals with at least one ambulatory encounter between 1/1/2013 and 12/31/2013
  - 61,645 adult (18+ on the date of the encounter) individuals with at least one ambulatory encounter between 1/1/2013 and 12/31/2013 residing in Denver County
Table 12: Diabetes and Hypertension Prevalence Comparison, Denver, 2013

<table>
<thead>
<tr>
<th>Hypertension Prevalence Phenotype Comparison</th>
<th>Extended Prevalence</th>
<th>Annual Prevalence</th>
<th>Augmented Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Cases</td>
<td>27,100</td>
<td>28,620</td>
<td>15,959</td>
</tr>
<tr>
<td>Care Population</td>
<td>119,804</td>
<td>120,303</td>
<td>83,588</td>
</tr>
<tr>
<td>Prevalence</td>
<td>22.6%</td>
<td>23.8%</td>
<td>19.1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diabetes Prevalence Phenotype</th>
<th>Extended Prevalence</th>
<th>Annual Prevalence</th>
<th>Augmented Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes Cases</td>
<td>14,530</td>
<td>14,678</td>
<td>10,739</td>
</tr>
<tr>
<td>Care Population</td>
<td>119,804</td>
<td>120,303</td>
<td>83,588</td>
</tr>
<tr>
<td>Prevalence</td>
<td>12.1%</td>
<td>12.2%</td>
<td>12.9%</td>
</tr>
</tbody>
</table>
Table 13: Demographic Description of Care Population, Denver, Colorado, 2011-2015

<table>
<thead>
<tr>
<th>Strata</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>131060</td>
<td>133185</td>
<td>135790</td>
<td>143477</td>
<td>147098</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>18-19</th>
<th>20-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>4543</td>
<td>11140</td>
<td>28521</td>
<td>22643</td>
<td>21230</td>
<td>21086</td>
<td>21897</td>
</tr>
<tr>
<td>2012</td>
<td>4455</td>
<td>10591</td>
<td>28812</td>
<td>23217</td>
<td>21204</td>
<td>21512</td>
<td>23394</td>
</tr>
<tr>
<td>2013</td>
<td>4185</td>
<td>10614</td>
<td>29266</td>
<td>23718</td>
<td>21094</td>
<td>22111</td>
<td>24802</td>
</tr>
<tr>
<td>2014</td>
<td>4510</td>
<td>11059</td>
<td>31257</td>
<td>25184</td>
<td>22215</td>
<td>23159</td>
<td>26093</td>
</tr>
<tr>
<td>2015</td>
<td>4627</td>
<td>11061</td>
<td>31562</td>
<td>26186</td>
<td>22448</td>
<td>23355</td>
<td>27859</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>78580</td>
<td>52480</td>
<td>2</td>
</tr>
<tr>
<td>2012</td>
<td>79638</td>
<td>53545</td>
<td>4</td>
</tr>
<tr>
<td>2013</td>
<td>81443</td>
<td>54343</td>
<td>1</td>
</tr>
<tr>
<td>2014</td>
<td>85555</td>
<td>57921</td>
<td>2</td>
</tr>
<tr>
<td>2015</td>
<td>87384</td>
<td>59712</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race and Ethnicity*</th>
<th>Black</th>
<th>Hispanic</th>
<th>Other</th>
<th>Unknown</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>15852</td>
<td>39607</td>
<td>5940</td>
<td>13661</td>
<td>56000</td>
</tr>
<tr>
<td>2012</td>
<td>15797</td>
<td>40195</td>
<td>6117</td>
<td>13269</td>
<td>57807</td>
</tr>
<tr>
<td>2013</td>
<td>16360</td>
<td>40650</td>
<td>6261</td>
<td>12277</td>
<td>60242</td>
</tr>
<tr>
<td>2014</td>
<td>17391</td>
<td>42412</td>
<td>6891</td>
<td>12350</td>
<td>64433</td>
</tr>
<tr>
<td>2015</td>
<td>18361</td>
<td>42116</td>
<td>7050</td>
<td>10671</td>
<td>68899</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strata</th>
<th>Column %</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-19</td>
<td>3%</td>
</tr>
<tr>
<td>20-24</td>
<td>8%</td>
</tr>
<tr>
<td>25-34</td>
<td>22%</td>
</tr>
<tr>
<td>35-44</td>
<td>17%</td>
</tr>
<tr>
<td>45-54</td>
<td>16%</td>
</tr>
<tr>
<td>55-64</td>
<td>16%</td>
</tr>
<tr>
<td>65+</td>
<td>17%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>60%</td>
<td>40%</td>
<td>0%</td>
</tr>
<tr>
<td>2012</td>
<td>60%</td>
<td>40%</td>
<td>0%</td>
</tr>
<tr>
<td>2013</td>
<td>60%</td>
<td>40%</td>
<td>0%</td>
</tr>
<tr>
<td>2014</td>
<td>60%</td>
<td>40%</td>
<td>0%</td>
</tr>
<tr>
<td>2015</td>
<td>59%</td>
<td>41%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race and Ethnicity*</th>
<th>Black</th>
<th>Hispanic</th>
<th>Other</th>
<th>Unknown</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>12%</td>
<td>30%</td>
<td>5%</td>
<td>10%</td>
<td>43%</td>
</tr>
<tr>
<td>2012</td>
<td>12%</td>
<td>30%</td>
<td>5%</td>
<td>10%</td>
<td>43%</td>
</tr>
<tr>
<td>2013</td>
<td>12%</td>
<td>30%</td>
<td>5%</td>
<td>9%</td>
<td>44%</td>
</tr>
<tr>
<td>2014</td>
<td>12%</td>
<td>30%</td>
<td>5%</td>
<td>9%</td>
<td>45%</td>
</tr>
<tr>
<td>2015</td>
<td>12%</td>
<td>29%</td>
<td>5%</td>
<td>7%</td>
<td>47%</td>
</tr>
</tbody>
</table>
Table 14: Care Population Coverage by Demographic Group, Denver, Colorado, 2015

<table>
<thead>
<tr>
<th>Strata</th>
<th>2015 Care Population</th>
<th>2015 ACS Population</th>
<th>Population coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Population</td>
<td>147098</td>
<td>541874</td>
<td>27%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>4627</td>
<td>13630</td>
<td>34%</td>
</tr>
<tr>
<td>20-24</td>
<td>11061</td>
<td>44499</td>
<td>25%</td>
</tr>
<tr>
<td>25-34</td>
<td>31562</td>
<td>152787</td>
<td>21%</td>
</tr>
<tr>
<td>35-44</td>
<td>26186</td>
<td>106206</td>
<td>25%</td>
</tr>
<tr>
<td>45-54</td>
<td>22448</td>
<td>79103</td>
<td>28%</td>
</tr>
<tr>
<td>55-64</td>
<td>23355</td>
<td>70834</td>
<td>33%</td>
</tr>
<tr>
<td>65+</td>
<td>27859</td>
<td>74815</td>
<td>37%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>87384</td>
<td>271490</td>
<td>32%</td>
</tr>
<tr>
<td>Male</td>
<td>59712</td>
<td>270384</td>
<td>22%</td>
</tr>
<tr>
<td>Unknown</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Race and Ethnicity*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>18361</td>
<td>51315</td>
<td>36%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>42116</td>
<td>153493</td>
<td>27%</td>
</tr>
<tr>
<td>Other</td>
<td>7050</td>
<td>38072</td>
<td>19%</td>
</tr>
<tr>
<td>Unknown</td>
<td>10671</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>White</td>
<td>68899</td>
<td>298994</td>
<td>23%</td>
</tr>
</tbody>
</table>

Figure 5: Population Coverage Map by Census Tract, Denver, Colorado, 2014-2015
Figure 6: Comparison of Care Population and Coverage by Census Tract, Denver, Colorado, 2014-2015

![Graph showing comparison of care population and coverage with R^2 = 0.0152](image-url)
<table>
<thead>
<tr>
<th>Strata</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Population</td>
<td>10%</td>
<td>11%</td>
<td>11%</td>
<td>11%</td>
<td>11%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>18-19</td>
<td>1%</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
<td>1%</td>
<td>0.3222</td>
</tr>
<tr>
<td>20-24</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0.3093</td>
</tr>
<tr>
<td>25-34</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>0.0674</td>
</tr>
<tr>
<td>35-44</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>5%</td>
<td>0.5180</td>
</tr>
<tr>
<td>45-54</td>
<td>12%</td>
<td>13%</td>
<td>13%</td>
<td>14%</td>
<td>13%</td>
<td>0.0137</td>
</tr>
<tr>
<td>55-64</td>
<td>18%</td>
<td>19%</td>
<td>19%</td>
<td>20%</td>
<td>19%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>65+</td>
<td>23%</td>
<td>24%</td>
<td>24%</td>
<td>25%</td>
<td>24%</td>
<td>0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Male</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>13%</td>
<td>12%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Unknown</td>
<td>25%**</td>
<td></td>
<td></td>
<td></td>
<td>50%**</td>
<td>0.3105**</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>13%</td>
<td>14%</td>
<td>14%</td>
<td>15%</td>
<td>15%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hispanic</td>
<td>15%</td>
<td>16%</td>
<td>16%</td>
<td>16%</td>
<td>16%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Other</td>
<td>9%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>11%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Unknown</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>White</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>0.0165</td>
</tr>
</tbody>
</table>

*Two sided p-value from Cochrane test for Trends
**Based on less than ten observations
Table 16: Comparison of EHR-Based and Survey-Based Estimates of Annual Diabetes Prevalence, Denver, Colorado, 2011-2015

<table>
<thead>
<tr>
<th>Data Source/Method</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>(95% Confidence Interval)</td>
<td>(95% Confidence Interval)</td>
<td>(95% Confidence Interval)</td>
<td>(95% Confidence Interval)</td>
<td>(95% Confidence Interval)</td>
</tr>
<tr>
<td>Crude EHR-based</td>
<td>10.2%</td>
<td>10.8%</td>
<td>11.0%</td>
<td>11.3%</td>
<td>11.2%</td>
</tr>
<tr>
<td></td>
<td>(10.05-10.38)</td>
<td>(10.63-10.97)</td>
<td>(10.83-11.16)</td>
<td>(11.11-11.44)</td>
<td>(11.00-11.33)</td>
</tr>
<tr>
<td>Adjusted EHR-based</td>
<td>10.7%</td>
<td>11.1%</td>
<td>9.6%</td>
<td>11.0%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Difference (percentage points)</td>
<td>+0.5</td>
<td>+0.3</td>
<td>-1.4</td>
<td>-0.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>Survey-Based</td>
<td>8.1%</td>
<td>9.5%</td>
<td>7.3%</td>
<td>7.4%</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>(5.8-10.5)</td>
<td>(7.5-11.6)</td>
<td>(5.7-9.0)</td>
<td>(5.7-9.1)</td>
<td>(3.2-7.3)</td>
</tr>
</tbody>
</table>

Figure 7: Comparison of Denver Diabetes Prevalence by Data Source, 2011-2015
Figure 8: Diabetes Prevalence Map by Census Tract, Denver, Colorado, 2014-2015
### Table 17: Trends in Denver EHR-Based Hypertension Prevalence by Demographic Group, 2011-2015

<table>
<thead>
<tr>
<th>Strata</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Population</td>
<td>20%</td>
<td>22%</td>
<td>21%</td>
<td>20%</td>
<td>19%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.3%</td>
<td>1%</td>
<td>0.3271</td>
</tr>
<tr>
<td>20-24</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0.6220</td>
</tr>
<tr>
<td>25-34</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>0.0036</td>
</tr>
<tr>
<td>35-44</td>
<td>9%</td>
<td>9%</td>
<td>9%</td>
<td>8%</td>
<td>8%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>45-54</td>
<td>23%</td>
<td>24%</td>
<td>23%</td>
<td>21%</td>
<td>19%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>55-64</td>
<td>36%</td>
<td>36%</td>
<td>35%</td>
<td>35%</td>
<td>31%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>65+</td>
<td>53%</td>
<td>56%</td>
<td>54%</td>
<td>52%</td>
<td>49%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>19%</td>
<td>20%</td>
<td>20%</td>
<td>19%</td>
<td>18%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Male</td>
<td>22%</td>
<td>24%</td>
<td>24%</td>
<td>22%</td>
<td>21%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Unknown</td>
<td>50%**</td>
<td>25%**</td>
<td></td>
<td></td>
<td></td>
<td>0.2049**</td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>30%</td>
<td>33%</td>
<td>32%</td>
<td>31%</td>
<td>29%</td>
<td>0.0028</td>
</tr>
<tr>
<td>Hispanic</td>
<td>20%</td>
<td>21%</td>
<td>21%</td>
<td>20%</td>
<td>19%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Other</td>
<td>19%</td>
<td>20%</td>
<td>20%</td>
<td>19%</td>
<td>18%</td>
<td>0.3864</td>
</tr>
<tr>
<td>Unknown</td>
<td>17%</td>
<td>18%</td>
<td>19%</td>
<td>18%</td>
<td>18%</td>
<td>0.1588</td>
</tr>
<tr>
<td>White</td>
<td>19%</td>
<td>20%</td>
<td>19%</td>
<td>18%</td>
<td>17%</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Two sided p-value from Cochrane test for Trends
**Based on less than ten observations

### Table 18: Comparison of EHR-Based and Survey-Based Estimates of Annual Hypertension Prevalence, Denver, Colorado, 2011-2015

<table>
<thead>
<tr>
<th>Source/Method</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Prevalence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(95% Confidence Interval)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHR-based</td>
<td>20.5% (20.26-20.70)</td>
<td>21.7% (21.52-21.96)</td>
<td>21.3% (21.05-21.49)</td>
<td>20.3% (20.09-20.50)</td>
<td>19.1% (10.86-19.26)</td>
</tr>
<tr>
<td>Adjusted EHR-based</td>
<td>21.2%</td>
<td>22.2%</td>
<td>18.7%</td>
<td>19.7%</td>
<td>18.1%</td>
</tr>
<tr>
<td>Difference (percentage point)</td>
<td>+0.7</td>
<td>+0.5</td>
<td>-1.6</td>
<td>-0.6</td>
<td>-1.0</td>
</tr>
<tr>
<td>Survey-Based</td>
<td>24.9 (21.6-28.3)</td>
<td>Not asked</td>
<td>24.7 (22.0-27.4)</td>
<td>Not asked</td>
<td>26.8 (22.2-31.3)</td>
</tr>
</tbody>
</table>
Figure 9: Comparison of Denver Hypertension Prevalence by Data Source, 2011-2015

Figure 10: Hypertension Prevalence Map by Census Tract, Denver, Colorado, 2014-2015
### Table 19: Persons with Diabetes Seen with Hemoglobin A1C Lab Test Status, Denver, Colorado, 2011-2015

<table>
<thead>
<tr>
<th>Hemoglobin A1C Test Results</th>
<th>Diabetes Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011</td>
</tr>
<tr>
<td>None</td>
<td>13,385</td>
</tr>
<tr>
<td>At least 1 Hemoglobin A1C Test</td>
<td>12,119</td>
</tr>
</tbody>
</table>

### Table 20: Denver EHR-Based Diabetes Control Trends by Demographic Group, 2011-2015

<table>
<thead>
<tr>
<th>Strata</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>80</td>
<td>81</td>
<td>81</td>
<td>79</td>
<td>78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>45</td>
<td>20</td>
<td>27</td>
<td>64</td>
<td>13</td>
<td>0.2330</td>
</tr>
<tr>
<td>20-24</td>
<td>68</td>
<td>53</td>
<td>58</td>
<td>60</td>
<td>57</td>
<td>0.4032</td>
</tr>
<tr>
<td>25-34</td>
<td>67</td>
<td>73</td>
<td>71</td>
<td>65</td>
<td>69</td>
<td>0.6101</td>
</tr>
<tr>
<td>35-44</td>
<td>70</td>
<td>70</td>
<td>71</td>
<td>68</td>
<td>68</td>
<td>0.0854</td>
</tr>
<tr>
<td>45-54</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>71</td>
<td>68</td>
<td>0.0002</td>
</tr>
<tr>
<td>55-64</td>
<td>79</td>
<td>79</td>
<td>80</td>
<td>77</td>
<td>77</td>
<td>0.0400</td>
</tr>
<tr>
<td>65+</td>
<td>89</td>
<td>89</td>
<td>89</td>
<td>88</td>
<td>87</td>
<td>0.0005</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>81</td>
<td>81</td>
<td>82</td>
<td>80</td>
<td>79</td>
<td>0.0021</td>
</tr>
<tr>
<td>Male</td>
<td>79</td>
<td>80</td>
<td>80</td>
<td>78</td>
<td>77</td>
<td>0.0044</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>80</td>
<td>80</td>
<td>82</td>
<td>80</td>
<td>80</td>
<td>0.8432</td>
</tr>
<tr>
<td>Hispanic</td>
<td>76</td>
<td>76</td>
<td>77</td>
<td>74</td>
<td>74</td>
<td>0.0011</td>
</tr>
<tr>
<td>White</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>85</td>
<td>83</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Other</td>
<td>84</td>
<td>86</td>
<td>85</td>
<td>84</td>
<td>84</td>
<td>0.5245</td>
</tr>
<tr>
<td>Unknown</td>
<td>80</td>
<td>82</td>
<td>80</td>
<td>79</td>
<td>78</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

*Two sided p-value from Cochrane test for Trends
Figure 11: Diabetes Control Map by Census Tract, Denver, Colorado, 2014-2015

Table 21: Persons with Hypertension Seen with Blood Pressure Recorded, Denver, Colorado, 2011-2015

<table>
<thead>
<tr>
<th>Blood Pressure Observations</th>
<th>Hypertension Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011 N=26,839</td>
</tr>
<tr>
<td></td>
<td>2012 N=28,958</td>
</tr>
<tr>
<td></td>
<td>2013 N=28,886</td>
</tr>
<tr>
<td></td>
<td>2014 N=29,118</td>
</tr>
<tr>
<td></td>
<td>2015 N=28,042</td>
</tr>
<tr>
<td>None</td>
<td>1,271</td>
</tr>
<tr>
<td>At least 1</td>
<td>25,568 (95%)</td>
</tr>
<tr>
<td></td>
<td>27,718 (96%)</td>
</tr>
<tr>
<td></td>
<td>27,992 (97%)</td>
</tr>
<tr>
<td></td>
<td>28,227 (97%)</td>
</tr>
<tr>
<td></td>
<td>27,460 (98%)</td>
</tr>
</tbody>
</table>
Table 22: Denver EHR-Based Hypertension Control Trends by Demographic Group, 2011-2015

<table>
<thead>
<tr>
<th>Strata</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>74</td>
<td>75</td>
<td>75</td>
<td>73</td>
<td>71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>73</td>
<td>76</td>
<td>69</td>
<td>54</td>
<td>68</td>
<td>0.4315</td>
</tr>
<tr>
<td>20-24</td>
<td>72</td>
<td>65</td>
<td>55</td>
<td>65</td>
<td>55</td>
<td>0.0685</td>
</tr>
<tr>
<td>25-34</td>
<td>67</td>
<td>64</td>
<td>65</td>
<td>62</td>
<td>62</td>
<td>0.0351</td>
</tr>
<tr>
<td>35-44</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>70</td>
<td>64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>45-54</td>
<td>72</td>
<td>73</td>
<td>73</td>
<td>71</td>
<td>69</td>
<td>0.0004</td>
</tr>
<tr>
<td>55-64</td>
<td>75</td>
<td>76</td>
<td>75</td>
<td>74</td>
<td>72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>65+</td>
<td>75</td>
<td>76</td>
<td>76</td>
<td>75</td>
<td>74</td>
<td>0.0013</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>74</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Male</td>
<td>74</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Unknown</td>
<td>0**</td>
<td>0**</td>
<td>0**</td>
<td>0**</td>
<td>0**</td>
<td>0.0004</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>67</td>
<td>70</td>
<td>70</td>
<td>69</td>
<td>66</td>
<td>0.0093</td>
</tr>
<tr>
<td>Hispanic</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>White</td>
<td>76</td>
<td>78</td>
<td>77</td>
<td>76</td>
<td>75</td>
<td>0.0073</td>
</tr>
<tr>
<td>Other</td>
<td>77</td>
<td>76</td>
<td>76</td>
<td>75</td>
<td>72</td>
<td>0.0021</td>
</tr>
<tr>
<td>Unknown</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>68</td>
<td>64</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

*Two sided p-value from Cochrane test for Trends

**based on less than ten observations
Figure 12: Hypertension Control Map by Census Tract, Denver, Colorado, 2014-2015
Figure 13: Comparison of Census Tract Hypertension and Diabetes Prevalence and Control, Denver, Colorado, 2014-2015

R² = 0.2239

R² = 0.0173
CHAPTER V

DISCUSSION

This research combined the science of surveillance, the concept of population health, the experience of public health, and an emerging data source in support of local-level health assessment. Our qualitative study among LPHAs confirmed that EHR-based surveillance remains exploratory and generally a longer term goal. Quantitative analysis found agreement between EHR- and survey-based estimates demonstrating their complementary value. EHR-based prevalence and control estimates detected year to year changes and socio-demographic disparities and produced novel census tract level information. Lacking local comparison data posed a challenge to interpreting disease control estimates and suggested the need for broader array of reference data that extends beyond telephone surveys. To achieve more widespread adoption of EHR-based surveillance, LPHAs must allocate resources to support use of local EHR data and instill confidence in staff and stakeholders that EHR data are an appropriate, reliable, and accurate source for surveillance information. LPHAs can benefit from the unprecedented opportunity that EHR data hold to deepen their local health knowledge, but this research reaffirmed that LPHAs have much to learn and accomplish to derive maximum value from EHR data for the jurisdictions they serve.
Implications for Public Health Practice

For any health indicator, an exact measurement at the population level is unknowable. Public health compiles the best available surveillance data to generate the most accurate estimates while appreciating the inherent limitations of any data source. Public health stakeholders must be clear about what criteria constitute actionable surveillance data. To be optimally useful, surveillance data should be sensitive to changes in temporal trends, variation between communities, and disparities between demographic groups while also being financially sustainable and organizationally manageable. Any data captured through human means will have some degree of error associated with its estimates. Even telephone surveys, which are considered by some to be the gold standard of population level health measurement, face increasing challenges collecting data.

This research appreciates that BRFSS has been a mainstay in public health practice. BRFSS data are easy to understand, easy to explain, and have a rich history. Since 1984, BRFSS has provided robust national estimates, useful state estimates, and very rough county estimates. As needs for local data have intensified, LPHAs clearly stated that BRFSS data are no longer meeting their local surveillance data needs. But BRFSS response rates have declined and land lines are no longer used compounding the impact of recall, self-reported, and non-response bias and possibly exceeding what statistical methods can overcome\(^\text{33,35,51,162}\). Funding to support BRFSS is declining and changes in technology are amplifying the numerous challenges and limitations associated with telephone surveys\(^\text{111,163,164}\). Now that EHR data is becoming accessible and being used for surveillance,
public health stakeholders must clarify what applications of BRFSS remain useful and refocus on identifying local data alternatives like EHR to respond to emerging community health priorities.

Compared to BRFSS, our findings clearly showed that EHR data provided more robust local data on topics within and beyond the BRFSS survey. As communities gravitate toward addressing health concerns that surveys poorly track, identifying new data sources like EHR to measure prevalence and progress becomes a local imperative. Public health has historically provided leadership in survey data validation and expertise in appropriate use of data sources despite their limitations: this capability is directly translatable to EHRs as a data source. When compared, the limitations of EHR were no greater than the limitations of the BRFSS survey. In fact, because each one’s strengths and limitations differed, BRFSS and EHR can be excellent complementary data sources. Local population health efforts stand to benefit from a more complete picture of health if BRFSS health information, designed to limit systematic error through random sampling, were combined with large EHR datasets, which have minimal random error. Application and interpretation of EHR data for surveillance will be an ongoing jurisdiction-specific process and LPHAs can combine their intimate “shoe-leather” knowledge of local health trends with methodological experience studying the accuracy of a data source to validate the use of EHR for surveillance in their jurisdiction.

Historically, healthcare institutions have been reticent to share clinical data with other potentially competitive organizations; consequently, public health has struggled to
forge partnerships where providers share data across institutions. Local EHR-based health information can help public health stakeholders gain traction with providers, healthcare systems and payers who are engaging in novel collaborative partnerships to become a learning health system (LHS)\textsuperscript{165,166}. Intensive data needs are central to the LHS model including community level data to identify opportunities for intervention and improvement and rapid cycle process and operational data within a feedback loop for learning and improvement\textsuperscript{165,166}. While the EHR-based surveillance described here may not contain the needed operational and process data, this research is an example of community level data that could inform LHS efforts to address chronic disease at the population level. LPHAs could add value to LHS partners by providing population health data to orient partners to the local landscape in their jurisdiction thus improving their contextual understanding of their patient population\textsuperscript{167-169}.

Census tract information helps LPHAs think spatially about the relationships between disease burden, disease control and social determinants. Because LPHAs are using ACS data to identify how social determinants like race, education, and income are geographically distributed, the ability to examine geographic patterns in disease burden and visualize how pockets of disease align with social determinants adds value to local surveillance efforts.

Local geospatial analysis of diabetes and hypertension prevalence and control detected many communities where disease burden and severity was higher (i.e., hotspots) and lower (i.e., coldspots). Hotspotting and coldspotting are analytic approaches that
identify statistically significant geographic disparities to guide decision making on which higher prevalence communities might benefit from intervention\textsuperscript{170-172}. Coldspots, where protective factors may be associated with lower disease burden, are explored to understand how those protective factors could be applied to hotspots\textsuperscript{173}. To date, little data has been available for LPHAs to operationalize hotspotting and coldspotting for priority chronic diseases like hypertension and diabetes. EHR-based census tract estimates of disease prevalence and severity allow LPHAs to identify hotspots and coldspots, specific to each disease, and assess to what degree targeted interventions cause hotspots to cool over time.

Because sub-county estimates of disease burden and severity are novel, public health stakeholders have limited options to use complementary datasets for comparison to triangulate local health trends, as was done for county estimates in this research. Modelled census tract BRFSS estimates from the 500 Cities dataset may be a usable reference data source that allows LPHAs to conduct the comparisons, apply weighting protocols and perform trend analyses at the census tract level.

\textbf{Implications for Usability of EHR Data for Surveillance}

Results from comparisons between survey and EHR-based estimates were inconclusive and differed by measure. Unlike several recent studies\textsuperscript{47,56,174-176}, annual EHR-based prevalence estimates were statistically different than survey-based estimates and temporal trends diverged between data sources and from national estimates\textsuperscript{177}. Between 2011 and 2015, BRFSS-reported diabetes prevalence in Denver declined by 34% and
hypertension prevalence increased 7% while EHR-based estimates were opposite: a 1% increase in diabetes prevalence and a 14% decline in hypertension prevalence. We found that EHR-based diabetes prevalence (11% in 2015) was closer to the national NHANES estimate (9% in 2015) than Denver BRFSS (5% in 2015)\(^{157}\). Our study suggested Denver hypertension prevalence was lower compared to both BRFSS and NHANES 2015 estimates (18% vs. 27% vs 29%, respectively)\(^{79,177}\). Distinct differences between the care-seeking population in the EHR and those not seeking care (but possibly sampled by BRFSS) could contribute to these discrepancies. For example, populations not in the EHR may have low diabetes prevalence and high hypertension prevalence. Other studies showed the burden of undiagnosed diabetes likely caused both BRFSS and EHR to underrepresent true prevalence\(^{157,178}\).

This research has important implications for future efforts to validate EHR-based estimates. Recent publications from New York City and Massachusetts have proposed a resource-intensive validation methodology that relies on rigorous statistical comparisons between EHR and surveys\(^{47,56}\). These studies assert that if EHR-based estimates and survey-based estimates are not equivalent, then EHR-based estimates are invalid and unfit for use, which runs counter to quality improvement and LHS principles and could have grave long-term consequence on adoption of EHR-based surveillance and data-driven health improvement efforts. Instead, a thorough review of the literature found that fundamental differences between these two data sources point to a number of valid reasons that estimates understandably differ between data sources. Self-reported data are known to be
affected by recall bias and self-reporting bias. Regarding chronic disease prevalence, the BRFSS questions are not oriented to a specific time period thus are less clear in their meaning. Clinical data are biased by the fact that individuals who seek care are different from those who do not. EHR and surveys, while conceptually linked to a theme or domain, differ in important ways.

The proposed validation approaches from New York City and Massachusetts did not consider the basic differences between these data sources. This approach is unwise to pursue because it is based on a problematic and unrealistic assumption that two very different sources should produce the same data. Further, this approach is too resource intensive to be feasible for most public health agencies and is not a viable long-term strategy for extracting the maximum value from EHRs to measure and improve population health. Restricting the use of EHR data for surveillance to indicators that can be compared to a BRFSS survey question would be a missed opportunity to gain valuable insights and inform community-based interventions. While current efforts to validate EHR data through comparisons with surveys offer useful insights, a broader focus on minimizing the concern about data limitations and developing strategies and methods to optimize EHR information for widespread use may be a path to greater adoption by communities and LPHAs to consider.

If telephone surveys should not be used as a comparator, how can LPHAs be certain that EHR-based estimates are correct? How public health can balance its parallel pursuits of usability and accuracy merits more focused thought, development and study. Public health
stakeholders could benefit from conceptuating this work as implementation science and accessing local health expertise through community engagement to understand local estimates. Clinical quality improvement principles could be translated to support and study EHR-based surveillance applications including a focus on patients (adapted to communities for this work), focus on being part of the team (referring to a team of learners interpreting EHR data together), and a focus on applied use of the data\textsuperscript{179,180}. A key indicator of utility of EHR-based estimates would be the extent to which public health stakeholder and community expectations for local data are met. Using this framework as a guide, local EHR-based estimates could be evaluated and interpreted by the community that each estimate represents.

**Recommendations**

An opportunity exists to develop elements of a model practice and augment existing resources to accelerate adoption of EHR-based surveillance by eliminating the uncertainty of these processes through a standardized governance model and analytic practices. While developing methods for this study, many gaps in current resources and unknowns were noted. One recommendation is to augment existing public health EHR tools\textsuperscript{45,98} to include practical resources that could be reused such as sample data use agreements, phenotype repositories and frameworks for phenotype comparison, simplified population weighting methods, and many more. Each component represents a model practice element for EHR-based surveillance that could be used by and beyond public health, some components of which could be standardized while others must be locally determined. LPHAs could benefit
from guidance that defines which elements of EHR surveillance systems can be borrowed and which must be tailored specifically to their needs.

A second and related recommendation is to formalize a community of learning that captures lessons learned and best practices, both technical and non-technical, from those practicing EHR-based surveillance. Key informant interviews found that public health stakeholders working with EHR data are constantly learning and improving their systems, methods, and techniques, but no mechanism exists to compile that knowledge. To date, public health stakeholders have borrowed heavily from clinical research and reporting, but qualitative findings revealed that the LPHA appetite and need for a dedicated EHR-based surveillance learning community is strong. Institutes like the Public Health Informatics Institute, consortiums of technical professionals like the Council of State and Territorial Epidemiologists or the American Medical Informatics Association, or national support organizations like NACCHO or CDC could all be well positioned to convene and lead this effort.

Though a unified national platform for EHR-based surveillance is likely infeasible in the current environment\textsuperscript{143}, a third recommendation is to consider how federal agencies could drive adoption of EHR-based surveillance nationwide. While the landscape of syndromic surveillance and lab reporting was at one time fragmented, federal support for Biosense and the Digital Bridge created standards that allowed many public health agencies to easily participate and a foundation for learning communities to share use cases, applications, and best practices\textsuperscript{181,182}. 

134
If progress toward adoption of EHR-based surveillance is our goal, LPHAs need a larger, better prepared workforce connected to a learning community to share knowledge and resources\(^{94,95,183,184}\). Transforming raw EHR data into county and census tract prevalence and control estimates was resource intensive, required advanced technical tools, and used skills typically outside of traditional public health activities. Population health surveillance builds upon traditional surveillance skills but requires staff to create relationships with partners, use new data sources, and measure new indicators at a more detailed and granular level\(^{12,13,17,44,185}\). Decades of focus on traditional surveillance data have poorly prepared most of the public health workforce to work with EHR data\(^{94,95}\). LPHAs are searching for a skill set that extends beyond traditional epidemiology training to include clinical informatics, business analysis, and data science\(^{94,95}\). This research recommends that MPH programs address this workforce void by expanding their curriculum to impart a broader set of technical skills, expose students to large clinical datasets, and orient them to: message standards (e.g., semantics, structure and transport), project management, stewardship, and governance of data and information systems. Additionally, the Council on Education for Public Health could direct MPH programs to focus on applied informatics by expanding accreditation requirements to include well-defined competencies around the use of non-traditional data\(^{186}\). Shared challenges across LPHAs suggest that a learning community and LHS could help discuss barriers, identify resources, develop standards, and share solutions.

135
Future Research

This research identified many opportunities for future research that could establish and refine model governance, technical, and dissemination best practices to aid LPHAs in adoption of EHR-based surveillance.

Birkhead’s 2017 EHR-based surveillance review pointed out the importance of thoughtful and well defined disease phenotypes in appropriate use of EHR data\textsuperscript{143}. Additional research is needed to assess, compare, and recommend a method for selecting the best phenotypes for EHR-based surveillance. This is especially important considering the complexity of phenotype comparisons which was an unanticipated aspect of this study and challenging to execute.

This study detected significant decreases in Denver diabetes and hypertension control between 2011 and 2015. Local EHR-based estimates of diabetes control showed similar patterns to NHANES where control decreased slightly from 72\% in 2011/2012 to 71\% in 2013/2014\textsuperscript{187}. Local EHR hypertension control trends were similar to NHANES which showed a decrease from 53\% in 2013/2014\textsuperscript{188} to 48\% in 2015/2016\textsuperscript{177}. Our study measured much higher rates of hypertension control (over 70\% for all five years) likely because of involvement in care vs the NHANES survey with includes those not seen in a healthcare system and missing from the EHR dataset. Both NHANES and EHR-based estimates detected higher hypertension control among older adults compared to younger and higher control among non-Hispanic whites compared to non-Hispanic blacks and Hispanics\textsuperscript{188}. Unlike NHANES which measured lower hypertension control among men compared to women\textsuperscript{188},
EHR-based control estimates between genders were comparable (72% for women vs 71% among men in 2015). Because measurement of control required an annualized utilization construct for each annual cohort based on patient and provider stratification, the annual phenotype may poorly fit the needs for annual measurement. Further research that might replicate this study using an array of HGBA1C and blood pressure control thresholds and observation periods could build consensus and evidence about the appropriate methods for population level disease control surveillance.

Census tract estimates of control do not tell the complete story of the individuals challenged to manage their chronic illness. The association between census tracts with lower levels of control and higher or lower levels of disease prevalence was not conclusive and merits further research. An individual’s ability to achieve disease control is likely to be significantly impacted by personal-decision making, provider behaviors, healthcare systems, policy changes, barriers to accessing care, and social determinants\textsuperscript{189-193}. As local control estimates are used to target interventions and conduct outreach, LPHAs should engage community members to understand how local factors (e.g., financial, health literacy) contribute to adverse health outcomes and what community resources might help reverse measured health inequities.

**Challenges**

Policy changes between 2011 and 2015 influenced who was seen for care; 2014’s Colorado Medicaid expansion expanded access to care for a population previously disengaged from care which may have increased the representativeness of the EHR care
population. Our study suggested that local context regarding policy and environmental changes may impact results of estimates from EHR data. Understanding population coverage, meaning how those in care differ from the underlying geographic population, and using coverage data when interpreting local surveillance findings is essential to building confidence in EHR-based estimates.

The EHR care population is a non-random population sample and EHR data are less complete compared to data obtained from surveys. Missing demographic, clinical and geographic data in EHR and misrepresentativeness of care seeking population impacted our prevalence and control estimates. However, missingness was minimal within the research dataset and improved during the observation period. Comparisons did not suggest that individuals with missing data differed from those with complete data in any notable way. Additionally, national efforts to use EHR data to quantify healthcare quality and link to performance and payment suggest that EHR data quality will continue to improve. Future research should assess strategies to address missing values like imputation would be a valuable addition. While EHRs have data completeness issues, they contain a broad range of data elements and types not otherwise available from any other sources.

For prevalence, population weights corrected point estimates for under and overrepresented groups by generating a more representative prevalence estimate. However, weights were not applied to estimates of disease control and how to weight population estimates that are not based on a denominator that can be easily adjusted remains unclear. Methods to address missing data and weight EHR-based estimates across
indicators and for sub-county estimates will be critical tools to drive adoption of EHR-based surveillance.

Some patients may have been misclassified as a diabetes or hypertension case (false positive) because of data entry errors. Since a diabetes case required only one diagnosis code, providers entering a diagnosis code to order a HGBA1C to rule out diabetes could create false positives that inflated EHR-based estimate of diabetes prevalence. Other studies recommend using more specific phenotype of 2 diagnosis codes or expanding criteria for a diabetes case to include HGBA1C and glucose test results or diabetes medications to improve both sensitivity and specificity\textsuperscript{56,175,200,201}. Individuals could be erroneously categorized as a non-case (false negative) because of missing data or coding errors. Undiagnosed hypertension or diabetes patients would appear as false negative in the dataset. While this research devoted much attention to crafting, calculating, and comparing multiple prevalence phenotypes, it is possible that different phenotypes would have produced different results. There is much work to be done to build up an evidence-base to inform phenotype selection.

Missing HGBA1C test results or blood pressures could have caused individuals to be misclassified as controlled or not controlled, depending on the missing lab test value or blood pressure reading. The 10% of diabetes cases who were excluded from control analysis because no HGBA1C test result had been collected in 2015 was concerning. Of those 10%, half had no HGBA1C test result ever recorded which may warrant additional exclusion criteria from the phenotype for a diabetes case. Additional research to consider phenotypes
that use laboratory tests and clinical observations to identify hypertension or diabetes cases or categorize control could improve the sensitivity of prevalence and control estimates.

**Limitations**

This study has many limitations. Interviews were limited to a convenience sample of 25 urban LPHAs from large cities and counties. The attitudes and practices described may not represent other urban LPHAs not included or the perspectives of rural LPHAs nor state health departments which were not sampled; these represent selection bias. Subsequent efforts to compare our findings to the perspectives of state and rural public health stakeholder could enhance the value of this research. Only one individual from each LPHA was interviewed; their personal contribution or knowledge may incompletely represent population health surveillance activities and attitudes across that organization; these represent self-reporting or response bias. As informants remarked, population health surveillance is evolving rapidly and the attitudes and practices described when interviews were conducted may have changed or be less relevant today. Establishing a mechanism for frequent and routine collection of LPHAs perspectives on surveillance could be led by national public health organizations to further explore findings from this research.

Analyses of prevalence and control were performed using EHR data which by definition only captured information about individuals who received healthcare. Individuals not having received care between 2011 and 2015 were excluded; the observed population may differ from the population not receiving care leading to selection bias in the results. However, two of three data contributors were safety net care providers who see many low
income and uninsured patients, which may have minimized the impact of this type of bias. If this analysis had included individuals not seeking care, or seeking care at other organizations, results may have varied.

Prevalence and control were estimated using data from three large healthcare providers in Denver. Patient populations were complementary as their patient base differs in socio-demographic and geographic characteristics. However, data contributors may include some of the same patients who were counted more than once in the prevalence control and denominator biasing the results. Emergency department and inpatient encounters, a common source of duplication due to irregular care seeking behaviors, were excluded from this analysis to limit the impact of duplicates. Additionally, structuring this analysis to consider repeated annual cross sections of data restricts the risk of duplication and double counting to individuals seeking outpatient care at multiple institutions within one same year.

**Summary of Contributions and Implications**

This research assessed LPHAs’ capacity for EHR-based surveillance, conducted a comparison of EHR-based estimates, and explored more granular, EHR-based, population health insights. Key informant interviews revealed a highly diverse population health surveillance landscape and variable progress to incorporate EHR data in their surveillance efforts. Few LPHAs were using EHR data, but consensus prevailed about potential value of EHR data and need for resources to support more widespread adoption. Because differences between EHR and survey-based prevalence estimates were statistically
significant, this study did not confirm equivalence between data sources but demonstrated that EHR-based estimates were more stable and precise over time and detected statistically significant disparities between demographic groups which aligned with national data. Population weighting only slightly adjusted crude prevalence estimates, suggesting concerns about misrepresentativeness of EHR data may be overestimated. EHR-based estimates showed declining hypertension and diabetes control between 2011 and 2015, but finding local references for comparison would be challenging. To our knowledge, this study produced the first and only census tract level estimates of hypertension and diabetes prevalence and control, which revealed substantial variation at the community level and highlighted the value of EHRs to detect communities with high disease burden, target resources, and evaluate impact.
REFERENCES


26. Definitions of Community Health Assessments (CHA) and Community Health Improvement Plans (CHIPs). 
http://archived.naccho.org/topics/infrastructure/chaip/.


44. Thorpe L. Population Health Data from Primary Care Records. Council of State and Territorial Epidemiologists; 2015; Boston, MA.


50. Hoffman S, Podgurski A. Big Bad Data: Law, Public Health, and Biomedical Databases. *Journal of Law, Medicine, and Ethics*. 2013;Spring:56-60.


Behavioral Risk Factor Surveillance System In. Prevention CfDCa, trans: Centers for Disease Control and Prevention.


APPENDIX

APPENDIX A

INTERVIEW GUIDE AND QUESTIONS FOR KEY INFORMANTS

Pre-interview: Send out an introductory email (follow-up with call, if necessary) requesting a date and time for the interview (about 30-45 minutes). Briefly explain the purpose of the interview, the dissertation design and focus, and my role. Try to schedule the interviews for fall 2016.

During the interview: Briefly explain the purpose of interview and the dissertation. My name is Emily McCormick and I am a PhD student in Health Information Technology at the University of Colorado. Thank you so much for agreeing to speak with me today. You were identified as a public health stakeholder experienced in local population health surveillance. I would like to focus our discussion on local population health and chronic disease surveillance. I am defining population health as ‘the health outcomes of a group of individuals, including the distribution of such outcomes within the group’ and local as surveillance at the county level or smaller.

The purpose of this interview is:

- to understand your organization’s approach to local population health surveillance
- to understand the key issues, challenges, and barriers your department faces in population health surveillance

While this research will have no direct benefit to you, your participation will be used to improve local population health.

This interview should take 30 to 45 minutes. Your decision to speak with me is voluntary. If you agree to participate, I’d like to record our interview to capture all of the information that you share with me. You may decline to answer any questions or stop the interview at any time. If, at a later time, you’d like to retract certain statements, you may do so. Your information will be stored on a password-protected computer. What you share will be grouped together with the information we gather from the other people we’re interviewing.

Now, I have three questions for you. The first is, are you willing to participate in this interview? Do you give me permission to record this interview? And, do you give me permission to contact you at a later time if I have additional questions?

Once you have consent, start the interview and begin recording.

Background information (3 minutes)

- What is your name and what organization do you work for?
- Please explain your role – what do you do?
Probing question: Do you consider yourself a LPHA director, an epidemiology leader, or front line epidemiologist?

- How long have you been at the organization?
- Are you knowledgeable in your organization's local population health surveillance activities?
  - If, no: Is there anyone else in your organization who might know more about this effort?
    - If no: Thank you very much for taking the time to speak with me today.
    - If yes: Thank you very much for taking the time to speak with me today. Could you connect me with this person via email to set up an interview?
  - End the interview.
- Is your organization a population health data consumer, information provider, or both?

**General Local Population Health Questions (15 minutes)**

- Why is it important for public health to engage in population health surveillance?
- Can you describe your organization's approach to local population health surveillance?
  - If no: Why not?
    - Have you received requests for local surveillance information?
    - **Probing question about CTSA and non-profit hospitals**
- What types of population health data is your organization working with?
- What internal and external stakeholders are important to your local population surveillance work?
  - **Probing questions: Who are the stakeholders and how are they involved?** Probe question about CTSA and non-profit hospitals?
  - How are stakeholders using this information?
- Has your organization committed funds to local population health surveillance activities?
  - If so, how much and on what? (e.g., FTE's, software, buying data?)
  - When did this funding begin?
  - Was there an impetus for the initial investment?
- How would you describe your organization's capacity to monitor population health?
- If you had unlimited resources, what additional population health data and or informational activities would you pursue?
- What factors have been critical to your success?
  - **Probing Question: Leadership? Partnerships? Data? Grant funding?**

**Chronic Disease Surveillance Questions (20 minutes)**

I'd like to focus our conversation to local chronic disease surveillance. We are specifically interested in hypertension and diabetes; however, please feel free to share experience with other common chronic disease.

- Can you give some examples of chronic disease surveillance at your organization?
  - **Probing questions: What data do you use for this specific chronic disease?**
  - **Probing questions: What indicators or measures are you able to track?**
Probing questions: How easily can you report on demographic or geographic subgroups?

- Are there any additional indicators or data element you would like to be tracking but cannot?
  - Probing questions: If so, what are they?
  - Probing questions: Do these data elements exist elsewhere in electronic form? If so, what are the barriers to obtaining them for your use?
- EHR is an emerging data source. What role do you see for EHR data in the chronic disease surveillance space?
- How are you making local chronic disease information available to stakeholders?
- Do you currently have any plans to provide community level chronic disease information to clinical care providers?
  - Probing questions: If so, what are they?

Barriers and Limitations Questions (15 minutes)

Related to local population health surveillance:

- Can you describe some of the partnership barriers your organization has faced (e.g., privacy)?
- Can you describe some of the organizational barriers your organization has faced (e.g., prioritization or executive sponsorship)?
- Can you describe some of the resource or technology barriers your organization has faced (e.g., people, technology, knowledge)?
- Are there other barriers or limitations that you would like to share?

That is all of the questions that I have to ask you. Is there anything else you would like to add? Are there any questions that you wish I had asked?

End of Interview. Stop recording.

Thank you very much for taking the time to speak with me today. If you have questions about this dissertation, you may contact me at 303-602-3604 or Emily.mccormick@dhha.org or the local ethics review board COMIRB at 303-724-1055 or COMIRB@ucdenver.edu.
APPENDIX B

DIABETES AND HYPERTENSION ICD-9 AND ICD-10 DIAGNOSIS CODES

Hypertension diagnosis codes used to identify cases include:

'362.11', '401.0', '401.1', '401.9', '402.00', '402.01', '402.1', '402.10', '402.11', '402.9',
'402.90', '402.91', '403', '403.0', '403.00', '403.01', '403.1', '403.10', '403.11', '403.9', '403.90', '403.91', '404', '404.0', '404.00', '404.01', '404.02', '404.03', '404.1', '404.10', '404.11', '404.12', '404.13',
'404.9', '404.90', '404.91', '404.92', '404.93', '437.2', 'H35.031', 'H35.032', 'H35.033', 'H35.039', 'I10', 'I11.0', 'I11.9', 'I12.0', 'I12.9', 'I13.0', 'I13.1', 'I13.10', 'I13.11', 'I13.2', 'I67.4'

Diabetes diagnosis codes used to identify cases include:

'249', '249.0', '249.00', '249.01', '249.1', '249.10', '249.11', '249.2', '249.20', '249.21',
'249.3', '249.30', '249.31', '249.4', '249.40', '249.41', '249.5', '249.50', '249.51', '249.6', '249.60',
'249.61', '249.7', '249.70', '249.71', '249.8', '249.80', '249.81', '249.9', '249.90', '249.91', '250',
'250.0', '250.00', '250.01', '250.02', '250.03', '250.1', '250.10', '250.11', '250.12', '250.13', '250.2',
'250.20', '250.21', '250.22', '250.23', '250.3', '250.30', '250.31', '250.32', '250.33', '250.4',
'250.40', '250.41', '250.42', '250.43', '250.5', '250.50', '250.51', '250.52', '250.53', '250.6', '250.60',
'250.61', '250.62', '250.63', '250.7', '250.70', '250.71', '250.72', '250.73', '250.8', '250.80', '250.81',
'250.82', '250.83', '250.9', '250.90', '250.91', '250.92', '250.93', '357.2', '362.0', '362.01', '362.02', '362.03', '362.04', '362.05', '362.06', '362.07', '366.41', '648.0', '648.00', '648.01', '648.02', '648.03',
'648.04', 'E08.00', 'E08.01', 'E08.10', 'E08.11', 'E08.21', 'E08.22', 'E08.29', 'E08.311', 'E08.319', 'E08.321', 'E08.329', 'E08.331', 'E08.339', 'E08.341', 'E08.349', 'E08.351', 'E08.359', 'E08.36', 'E08.39', 'E08.40', 'E08.41', 'E08.42', 'E08.43', 'E08.44', 'E08.49', 'E08.51', 'E08.52', 'E08.59',

163
APPENDIX C

QUOTES DESCRIBING APPROACH TO POPULATION HEALTH SURVEILLANCE

We think of surveillance in two buckets: population health surveillance and communicable disease or disease control surveillance... Disease control surveillance is really following protocols established by the (state health department) to comply with laws and regulations around reportable conditions...with regard to population health surveillance, we are very invested in looking at the characteristics of the communities we serve. Quote from Director of Epidemiology and Informatics at multi-county LPHA

If we believe in data-driven policies, programs, and services, then having population health surveillance data is really critical. And then ultimately, we need that data because if our goal is optimal health, which is generally close to everyone's goal in some way shape or form, as well as reducing disparities. The only way to know that is to have the data. Quote from Director of Assessment, Policy, and Evaluation at county LPHA

We are the neutral agency. We are the ones who can collect, analyze, and interpret the data and then provide it to everyone else. I think that even though we have a lot of universities and healthcare corporations, none of those are impartial parties. Researchers have their own research questions. We are that middle ground, we have the ability and the skillset to access certain datasets that no one else has access to in the same way we do. We also have analytic expertise in surveillance. Quote from Epidemiologist at city/county LPHA

Quotes Describing Data Sources

There are a lot of new public health data sources coming out. There are a lot more data availability and there are things like 500 cities where people are getting access to sub-county level health behavior estimates and health outcomes, things that people haven't had access to before. Since we've traditionally been the organization that's provided those types of data to stakeholders in the community, we are now thinking about ok, now that people have access to data and these types of modeled estimates, what is it that we provide that no one else does. Quote from Director of Assessment, Planning, and Special Projects at city LPHA

We look at the interactions between all different data sources to describe health outcomes in the city. Quote from Director of Epidemiology at city LPHA

Quotes Describing Dissemination

It's hard work to explain (data) when you have a willing audience and sometimes you don't even have a willing audience. Quote from Senior Epidemiologist at county LPHA

One of the things that is really difficult is to communicate to people who control the budgets at the highest level. People who don't deal with data every day think its magic. Oh get me those numbers. That must be easy. They don't really understand the manpower and software or the analyst's time that it...
takes to pull the data out and assemble it.....it’s very difficult to explain the intense work it can take. Quote from Senior Epidemiologist at county LPHA

For mortality data, it’s hard to get folks to understand that it not going to change greatly year to year. But they get impatient if you don't have something. It’s hard to look at trends in real time. Deaths are the tip of the ice berg and we need to be showing a lot of other factors too. If we can supplement them with hospital data. Quote from Epidemiology Manager at county LPHA

Quotes Describing Stakeholders

We get most of our datasets from the state.....We have a state that is at times difficult to deal with in terms of getting data that we should just have access to all the time. It’s a process of negotiation to get data from the state for surveillance purposes. Quote from Epidemiologist at city/county LPHA

We want data from the state. They don’t want to give it us because they are afraid we'll expose it somehow. It’s less of a risk not to give it to us than to give it to us. Quote from Director of Epidemiology at County LPHA

There can be a lot of time put into maintaining these relationships to make sure surveys are conducted and we have access to the data and we can produce the data. Quote from Director of Assessment, Policy, and Evaluation at county LPHA

They [healthcare corporations] all have an interest in health data and population health surveillance. And so it’s a little bit tricky here. The context is because there are a lot of companies with deep pockets interested in investing, but no one can decide what to invest in or who's doing to own the data. So, there is a lot of interest and a lot of money but not a lot of commitment. Quote from Epidemiologist at city/county LPHA

There seems to be a shift in our community and a willingness to look at hospital involvement in population health. One of the hospitals has actually hired a population health person and created a population health department. Some of this might be a change in healthcare reimbursement and the penalties associated with readmission to the hospital. Having to do a health assessment has made them [healthcare providers] want to work together and think about data more. Quote from Epidemiology Manager at county LPHA

Quotes Describing Unmet Surveillance Needs

People want it [chronic disease data] at the sub-geography level. There aren’t enough numbers to be able to get it at the sub-geography that they want. People want you to get right down to the neighborhood. And that’s one of our biggest struggles. We [LPHA] aren’t small but we aren’t huge. People seem to want something at a block level and it’s hard to explain to them that your numbers are too variable or you don’t have enough numbers. Quote from Director, Epidemiology and Preparedness at county LPHA

Even if we get county level estimates, that’s not especially helpful for us. We really need chronic disease
indicators at sub-county geographies. That’s really our biggest missing piece. Quote from Epidemiologist at city/county LPHA

For something that causes so much burden of illness in the population, we have very little information about it [chronic disease]. We lean on what we can glean from ED and hospitalizations and death data to get a handle on chronic disease…. How many people have diabetes? And what are their demographic characteristics and what’s the severity of their diabetes and what services do they need? We’ve got so little information about chronic disease, and I want to know everything about it. Quote from Director of Epidemiology at county LPHA

I don’t have a lot of confidence in some of these [traditional] data sources and they don’t give me data at granularity I need. I cannot get neighborhood data, even through oversampling our big surveys. That was the driving force behind this [EHR data exploration]. Quote from Director of Assessment, Planning, and Special Projects at county LPHA

There are a number of pop based surveys at the federal level that are pushed down to the state….that information is frustrating because its driven at the federal level down to the state and then the states down to the county, it’s not designed in such a way that we can get information that is localized to the population we serve directly. Quote from Director of Epidemiology at city LPHA

Quotes Describing Unmet Surveillance Needs

If the information smells real to them [stakeholders], if it appears to reflect reality, they’ll be happy with it… Whatever data we can get that is reasonable and informative; I think people are fine with. Quote from Director of Epidemiology at County LPHA

Validation becomes very condition specific.....If that was me and we were validating hypertension control. We might look at who is on blood pressure medication and that’s something we could look at from our BRFSS data to get some sense of who is doing what in that area. I’d use that to get some sense of how far apart they are and why they might be far apart. In situations where there isn’t another dataset or a set of sources that you could validate, it’s a challenge. We dig around in the evidence and we look for a proxy measure for the indicator we are looking at. EMS data is a good example. Part of the reason I haven’t used it is because we haven’t really done what we needed to do for validation. Quote from Director of Research and Evaluation at City LPHA
APPENDIX D

QUOTES DESCRIBING ORGANIZATIONS STRUCTURE AND INVESTMENT IN POPULATION HEALTH SURVEILLANCE

I restructured the department to create a division of policy planning and evaluation to start building the infrastructure around data and technical skills for policy analysis, program planning, and evidence informed action so that we could start integrating and partnering more effectively with folks inside and outside of government... We were very intentional in saying we need more than epidemiologists on our team and we want to focus on applied epidemiology and look at data outside of health data to help inform our decision making and the profile of health. We weren't doing it well and we weren't being responsive as we could. Quote from Director of Public Health at city/county LPHA

We have five county supported epidemiology positions that haven't been threatened. We don't have to spend time in supporting ourselves. We only help others chase grants, but we do not do any for ourselves. I don't have to spend any time on grants. There is a great deal of freedom in stable funding. Quote from Epidemiology Manager at county LPHA

The challenge for our [population health surveillance] strategy is sustainability. You end up spending a fair amount of your time scrambling. Quote from Director of Epidemiology at city LPHA

The commissioner advocated for increased capacity for this unit and we have been success at garnering those resources form the city. With that we've been to hire more staff and divide the unit up into specialists..... We have successfully convinced the funding of a health survey which we use to track health utilization, risk behaviors. I would say that there has been an expansion in resources but we still need more. Quote from Director of Epidemiology at city LPHA
APPENDIX E

QUOTES DESCRIBING EVOLUTION IN POPULATION HEALTH SURVEILLANCE

The RWJ culture of health perspective has brought lots of partners to our [public health] door. The more diverse these partnerships become, the more we [epidemiology] are able to offer. If people decide that it’s important to understand how a change in city zoning could impact health including how to work with real estate development and underserved communities. The health department is now sitting at the table at this conversation that needs data management and monitoring and so it has put population health surveillance in the middle of all of these discussions we’ve never been in before. Quote from Director of Epidemiology at city LPHA

We began creating this distinction between what we [surveillance group] do and what we do best is population health surveillance. And where you get more into healthcare, which is sometimes, can have a health economics or cost benefit component to it. We just don’t do it. We are not healthcare services researchers. And we sometimes know the right questions but we don’t know how to answer them. We would really need more resources..... We somehow have maintained that distinction and when pressed, we fall back to that place of “we have limited staff.... and this is what we can provide.” Quote from Director of Assessment, Planning, and Evaluation at county LPHA

I am constantly thinking about our relevance as a public health agency. We have all these healthcare institutions and academic institutions and we have to constantly demonstrate our relevance to them to remain at the table. Because of where we are, I dedicate a lot of resources to new data sources and new analytic methods... I err on the side of pursuing new data sources and analytic methods and balance that with folks who need the stuff we’ve always done. One way we demonstrate our value is bringing this [data and analytic techniques] to the table. Quote from Director of Research and Evaluation at City LPHA

It [public health] has changed and I think we were naïve about the expectations and asks. We didn’t understand the scale of the problem or the scale of the data we could potentially have to work with. It definitely has increased and we haven’t been able to deliver on those increased expectations. Quote from Director of Public Health at city/county LPHA
APPENDIX F

QUOTES DESCRIBING PERSPECTIVES ON EHR DATA FOR POPULATION HEALTH SURVEILLANCE

I’m attracted to the idea of building a [EHR] surveillance system from scratch which is how I see it. Being a new area [ehr data], I would still consider it exploratory at this point. I know others have made progress validating EHR data against other metrics..... The scale of it [EHR data] is so attractive, it’s not a survey with 2500 responses, its medical records on millions of patients which would make any kind of inferences so much more robust and could potentially help us truly describe what’s happening with our populations and health..... It would be huge for the health department and planning to have something beyond self-reported data for chronic disease prevalence. Quote from Director of Epidemiology at City LPHA

I’m sure there are some things I could never get out of EHR data because there are some people aren’t using healthcare. But most people are and there so much potential to get good information out of it... To use it with confidence, I need to have a better understanding of its biases. Quote from Director of Epidemiology at County LPHA

There is an intuitive sense that more information about what happens inside the healthcare system can help us [public health] as a partner of the healthcare system. It’s always a bit funny for me because I believe that a major function of public health is to keep people out of the medical care system or to use it only for prevention. So it’s a bit at odds to say let’s look inside.....what’s going on in there. And as we’ve looked and had more access to that kind of information through the EHR climate, we see it’s a big mess. Quote from Director of Epidemiology and Informatics at multi-county LPHA

If I can demonstrate that the data that I have [EHR] is superior, because I feel in my gut that getting individual data is going to give you a better picture of a small geography than breaking out aggregate data, then I think I can demonstrate it. Quote from Director of Assessment, Planning, and Special Projects at county LPHA

With this data [EHR] that is giant convenience sample data, what is it really telling us? Because it’s not the population; it’s not representative. How do I come up with methods to analyze this data when, in this zip code I have 4% of the people, they are not randomly selected. This other zip code, I have 82% of the people. What can I say about this data as a whole? What does it tell me about my community? Quote from Director of Assessment, Planning, and Special Projects at county LPHA

There are so many records. With such large sample sizes, we can get better health estimates for smaller geographic areas. We could get information on diseases we have no information about currently like Alzheimer’s or epilepsy. It could complement the survey work we are doing now. Maybe there are some survey questions we wouldn’t need to ask because we could get that information from EHRs. Quote from Director of Assessment, Planning, and Special Projects at city LPHA

It’s important for those in public health to work with EHR vendors to encourage recording and collection of data in a consistent way, and to discuss inclusion of fields that are important to public health even though some may not be typical in EHR software. Quote from Director of Assessment, Planning, and Special Projects at city LPHA
Quotes Describing Barriers

It costs a lot of money and is all very new. There is an early adopter penalty. People who try and adopt a system early on pay a price to try and update those systems moving forward. There is also a lot of complexity with the public health surveillance population health based needs compared to patient level EHR data structures. Quote from Senior Epidemiologist at county LPHA

Sometimes it’s not about the access to the data. It’s about having enough people power and knowledge and experience to do it. Quote from Director of Assessment, Policy, and Evaluation at county LPHA

We are in the information age and sometimes public health hasn’t quite caught up. Our sophistication for using it (data) is really underdeveloped....what they are teaching in schools of public health is a little unconnected from what skills you need to be in a public health department. Quote from Senior Epidemiologist at county LPHA

Having someone who has the idea and having someone who can see it through is sometimes the issue. Sometimes people don’t have any ideas, and that’s a problem. Sometimes the idea is there but the execution element is not. So you need people with the technical knowledge and you need people with some project management skills to see something through from start to finish. When trying to get some of these other datasets, they need to be able to negotiate and understand the other party a little bit. It’s a little more complex set of skills than just “here's some data; I need some frequencies on it.” Quote from Director of Strategic Information at city LPHA

The expectation is that we have access to EHR information that can give us real time information on population health. And we just don’t [have it]. We don’t have systems that talk to each other. Quote from Director of Public Health at city/county LPHA

Quotes Describing Facilitators

We’ve really tried to develop a team that knows how to work with population data. We’ve made a commitment to understand how to do these analytic methods that we are using.... What I cannot do is create a culture where people values data and then not be able to fulfill that need. I don’t want to have to hire consultants or work with a partner who knows how to do it. We continue to build in house capacity. Because that is the way that you can continue to help people feel that the data are available is by responding to their needs and I cannot respond to their needs if I cannot do the analysis myself. Quote from Director of Research and Evaluation at City LPHA

I think has been really important is to have leadership who are really supportive and believe in data driven work... Much of our data work is supported with flexible dollars which means county or state dollars. In order to have that allocation of those extremely tight county and state dollars, folks need to see a benefit from what we are doing. Quote from Director of Assessment, Policy, and Evaluation at county LPHA

We are finally moving in a strategic direction with our agency on data. Our current strategic plan includes a section on information technology with objectives specific to data. We found that every single
program area collects some sort of data that is geolocated. We are aligning our technology with our organizational needs.... Once data and technology made its way to the strategic plan, I have a platform to advocate for investment. Quote from Director of Epidemiology and Informatics of county LPHA

Population Health surveillance has really benefited from the emphasis on meaningful use and now MACRA and MIPS, the new value-based structure that health services payment is implementing. All of that requires practices and corporations to pay attention to their data. If you can align your local grant-funded surveillance efforts with the same sort of measurement and data collection that is being emphasized by the payment system then you can get a lot of traction. We've tried to do that. Quote from Director of Health Information at city/county LPHA
APPENDIX G

DIABETES PHENOTYPE COMPARISON BY AGE, GENDER, AND RACE ETHNICITY

<table>
<thead>
<tr>
<th>Strata</th>
<th>Survey Based Prevalence</th>
<th>Expected Patients</th>
<th>Extended Phenotype Count</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1.10%</td>
<td>661</td>
<td>143</td>
<td>21.64%</td>
</tr>
<tr>
<td>25-34</td>
<td>2.08%</td>
<td>2722</td>
<td>517</td>
<td>19.00%</td>
</tr>
<tr>
<td>35-44</td>
<td>2.68%</td>
<td>2520</td>
<td>1463</td>
<td>58.06%</td>
</tr>
<tr>
<td>45-54</td>
<td>7.75%</td>
<td>5711</td>
<td>3164</td>
<td>55.40%</td>
</tr>
<tr>
<td>55-64</td>
<td>15.43%</td>
<td>9937</td>
<td>4217</td>
<td>42.44%</td>
</tr>
<tr>
<td>65+</td>
<td>23.36%</td>
<td>15153</td>
<td>5026</td>
<td>33.17%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7.21%</td>
<td>17532</td>
<td>6697</td>
<td>38.20%</td>
</tr>
<tr>
<td>Female</td>
<td>7.43%</td>
<td>18202</td>
<td>7833</td>
<td>43.03%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>9.59%</td>
<td>4665</td>
<td>2099</td>
<td>45.00%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9.63%</td>
<td>13582</td>
<td>6695</td>
<td>49.29%</td>
</tr>
<tr>
<td>White</td>
<td>4.79%</td>
<td>13952</td>
<td>3888</td>
<td>27.87%</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>1252</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>9.59%</td>
<td>4665</td>
<td>2099</td>
<td>45.00%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9.63%</td>
<td>13582</td>
<td>6695</td>
<td>49.29%</td>
</tr>
<tr>
<td>White</td>
<td>4.79%</td>
<td>13952</td>
<td>3888</td>
<td>27.87%</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>1306</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>8.78%</td>
<td>590</td>
<td>596</td>
<td>101.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strata</th>
<th>Survey Based Prevalence</th>
<th>Expected Patients</th>
<th>Annual Phenotype Count</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1.10%</td>
<td>661</td>
<td>164</td>
<td>24.82%</td>
</tr>
<tr>
<td>25-34</td>
<td>2.08%</td>
<td>2722</td>
<td>544</td>
<td>19.99%</td>
</tr>
<tr>
<td>35-44</td>
<td>2.68%</td>
<td>2520</td>
<td>1469</td>
<td>58.30%</td>
</tr>
<tr>
<td>45-54</td>
<td>7.75%</td>
<td>5711</td>
<td>3050</td>
<td>53.40%</td>
</tr>
<tr>
<td>55-64</td>
<td>15.43%</td>
<td>9937</td>
<td>4170</td>
<td>41.96%</td>
</tr>
<tr>
<td>65+</td>
<td>23.36%</td>
<td>15153</td>
<td>5281</td>
<td>34.85%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7.21%</td>
<td>17532</td>
<td>6728</td>
<td>38.38%</td>
</tr>
<tr>
<td>Female</td>
<td>7.43%</td>
<td>18202</td>
<td>7950</td>
<td>43.68%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>9.59%</td>
<td>4665</td>
<td>2112</td>
<td>45.28%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9.63%</td>
<td>13582</td>
<td>6709</td>
<td>49.40%</td>
</tr>
<tr>
<td>White</td>
<td>4.79%</td>
<td>13952</td>
<td>3964</td>
<td>28.41%</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>1306</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>9.59%</td>
<td>4665</td>
<td>2112</td>
<td>45.28%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9.63%</td>
<td>13582</td>
<td>6709</td>
<td>49.40%</td>
</tr>
<tr>
<td>White</td>
<td>4.79%</td>
<td>13952</td>
<td>3964</td>
<td>28.41%</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>1306</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>8.78%</td>
<td>590</td>
<td>587</td>
<td>99.47%</td>
</tr>
</tbody>
</table>

185
<table>
<thead>
<tr>
<th>Strata</th>
<th>Survey Based Prevalence</th>
<th>Expected Patients</th>
<th>Augmented Phenotype Count</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Population</strong></td>
<td>7.32%</td>
<td>35709</td>
<td>10739</td>
<td>30.07%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1.10%</td>
<td>661</td>
<td>99</td>
<td>14.98%</td>
</tr>
<tr>
<td>25-34</td>
<td>2.08%</td>
<td>2722</td>
<td>325</td>
<td>11.94%</td>
</tr>
<tr>
<td>35-44</td>
<td>2.68%</td>
<td>2520</td>
<td>956</td>
<td>37.94%</td>
</tr>
<tr>
<td>45-54</td>
<td>7.75%</td>
<td>5711</td>
<td>2192</td>
<td>38.38%</td>
</tr>
<tr>
<td>55-64</td>
<td>15.43%</td>
<td>9937</td>
<td>3074</td>
<td>30.93%</td>
</tr>
<tr>
<td>65+</td>
<td>23.36%</td>
<td>15153</td>
<td>4093</td>
<td>27.01%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7.21%</td>
<td>17532</td>
<td>4835</td>
<td>27.58%</td>
</tr>
<tr>
<td>Female</td>
<td>7.43%</td>
<td>18202</td>
<td>5904</td>
<td>32.44%</td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>9.59%</td>
<td>4665</td>
<td>1559</td>
<td>33.42%</td>
</tr>
</tbody>
</table>
### APPENDIX H

#### HYPTERTENSION PHENOTYPE COMPARISONS BY AGE, GENDER, AND RACE ETHNICITY, 2013

<table>
<thead>
<tr>
<th>Strata</th>
<th>Survey Based Prevalence</th>
<th>Expected Patients</th>
<th>Extended Phenotype Count</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Population</strong></td>
<td>24.74%</td>
<td>120688</td>
<td>27100</td>
<td>22.45%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>11.71%</td>
<td>7032</td>
<td>89</td>
<td>1.27%</td>
</tr>
<tr>
<td>25-34</td>
<td>11.09%</td>
<td>14482</td>
<td>658</td>
<td>4.54%</td>
</tr>
<tr>
<td>35-44</td>
<td>12.37%</td>
<td>11645</td>
<td>2077</td>
<td>17.84%</td>
</tr>
<tr>
<td>45-54</td>
<td>29.90%</td>
<td>22031</td>
<td>5039</td>
<td>22.87%</td>
</tr>
<tr>
<td>55-64</td>
<td>47.98%</td>
<td>30899</td>
<td>7659</td>
<td>24.79%</td>
</tr>
<tr>
<td>65+</td>
<td>57.09%</td>
<td>37039</td>
<td>11578</td>
<td>31.26%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24.88%</td>
<td>60466</td>
<td>12096</td>
<td>20.00%</td>
</tr>
<tr>
<td>Female</td>
<td>24.59%</td>
<td>60193</td>
<td>15004</td>
<td>24.93%</td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>49.77%</td>
<td>24198</td>
<td>2435</td>
<td>10.06%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>18.48%</td>
<td>26075</td>
<td>8629</td>
<td>33.09%</td>
</tr>
<tr>
<td>White</td>
<td>25.27%</td>
<td>73637</td>
<td>10159</td>
<td>13.80%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
<td>4754</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>13.76%</td>
<td>925</td>
<td>1123</td>
<td>121.47%</td>
</tr>
<tr>
<td><strong>Strata</strong></td>
<td><strong>Survey Based Prevalence</strong></td>
<td><strong>Expected Patients</strong></td>
<td><strong>Annual Phenotype</strong></td>
<td><strong>Coverage</strong></td>
</tr>
<tr>
<td>Adult Population</td>
<td>24.74%</td>
<td>120688</td>
<td>28620</td>
<td>23.71%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>11.71%</td>
<td>7032</td>
<td>120</td>
<td>1.71%</td>
</tr>
<tr>
<td>25-34</td>
<td>11.09%</td>
<td>14482</td>
<td>804</td>
<td>5.55%</td>
</tr>
<tr>
<td>35-44</td>
<td>12.37%</td>
<td>11645</td>
<td>2301</td>
<td>19.76%</td>
</tr>
<tr>
<td>45-54</td>
<td>29.90%</td>
<td>22031</td>
<td>5379</td>
<td>24.42%</td>
</tr>
<tr>
<td>55-64</td>
<td>47.98%</td>
<td>30899</td>
<td>7873</td>
<td>25.48%</td>
</tr>
<tr>
<td>65+</td>
<td>57.09%</td>
<td>37039</td>
<td>12143</td>
<td>32.78%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24.88%</td>
<td>60466</td>
<td>12963</td>
<td>21.44%</td>
</tr>
<tr>
<td>Female</td>
<td>24.59%</td>
<td>60193</td>
<td>15657</td>
<td>26.01%</td>
</tr>
<tr>
<td><strong>Race and Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>49.77%</td>
<td>24198</td>
<td>4886</td>
<td>20.19%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>18.48%</td>
<td>26075</td>
<td>9032</td>
<td>34.64%</td>
</tr>
<tr>
<td>White</td>
<td>25.27%</td>
<td>73637</td>
<td>10866</td>
<td>14.76%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
<td>2645</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>13.76%</td>
<td>925</td>
<td>1191</td>
<td>128.82%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strata</th>
<th>Survey Based</th>
<th>Expected</th>
<th>Augmented Phenotype</th>
</tr>
</thead>
</table>

187
<table>
<thead>
<tr>
<th>Age</th>
<th>Prevalence</th>
<th>Patients</th>
<th>Count</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Population</td>
<td>24.74%</td>
<td>120688</td>
<td>15959</td>
<td>13.22%</td>
</tr>
<tr>
<td>18-24</td>
<td>11.71%</td>
<td>7032</td>
<td>45</td>
<td>0.64%</td>
</tr>
<tr>
<td>25-34</td>
<td>11.09%</td>
<td>14482</td>
<td>327</td>
<td>2.26%</td>
</tr>
<tr>
<td>35-44</td>
<td>12.37%</td>
<td>11645</td>
<td>1142</td>
<td>9.81%</td>
</tr>
<tr>
<td>45-54</td>
<td>29.90%</td>
<td>22031</td>
<td>2925</td>
<td>13.28%</td>
</tr>
<tr>
<td>55-64</td>
<td>47.98%</td>
<td>30899</td>
<td>4372</td>
<td>14.15%</td>
</tr>
<tr>
<td>65+</td>
<td>57.09%</td>
<td>37039</td>
<td>7148</td>
<td>19.30%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24.88%</td>
<td>60466</td>
<td>7061</td>
<td>11.68%</td>
</tr>
<tr>
<td>Female</td>
<td>24.59%</td>
<td>60193</td>
<td>8898</td>
<td>14.78%</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>49.77%</td>
<td>24198</td>
<td>2844</td>
<td>11.75%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>18.48%</td>
<td>26075</td>
<td>5369</td>
<td>20.59%</td>
</tr>
<tr>
<td>White</td>
<td>25.27%</td>
<td>73637</td>
<td>5610</td>
<td>7.62%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
<td>1537</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>13.76%</td>
<td>925</td>
<td>599</td>
<td>64.79%</td>
</tr>
</tbody>
</table>
**COMIRB Protocol**

COLORADO MULTIPLE INSTITUTIONAL REVIEW BOARD

CAMPUS BOX F-490   TELEPHONE: 303-724-1055   Fax: 303-724-0990

---

**Project Title:** Harnessing Electronic Health Record Data for Local Population Health Monitoring: Key Informant Protocol

**Principal Investigator:** Emily McCormick, MPH

**Description:** Key informant interview to assess knowledge, attitudes, and practices related to population health surveillance among local public health stakeholders.

---

**I. Hypotheses and Specific Aim:**

**Short Narrative Summary:**

Local public health departments have an important role in population health surveillance. However, little is known describing how local health stakeholders are approaching population health.

For this exempt research, we propose to conduct key informant interviews with local public health departments nationwide to: 1) assess local public health department approaches to population health surveillance, 2) describe technology, partnership, and resource barriers impacting population health surveillance, 3) explore key stakeholders to local public health...
department population health activities, and 4) gather opinions about the role of EHR data in population health surveillance.

**Specific Aim:** Conduct key informant interviews with public health practitioners to assess knowledge, attitudes, and practices related to population health surveillance.

- **Hypothesis 1:** Local public health departments believe population health surveillance is aligned with the public health mission and scope of work.
- **Hypothesis 2:** Approaches to population health surveillance including indicators and data sources vary between local public health departments.
- **Hypothesis 3:** Public health stakeholders believe that EHR data can be a valuable population health data source.

**II. Background and Significance:**

Population health, defined as the health outcomes of a group of individuals, including the distribution of such outcomes within the group, requires local clinical and public health stakeholder to collaboratively address a community’s greatest health needs. Improving local population health relies on robust sources of local health information and experience analyzing health data at the community level. Local public health departments will leverage experience in public health surveillance to measure and monitor the health of populations. However, little is known describing how local health stakeholders are approaching population health.

Widespread adoption of electronic health records (EHRs) has created libraries of electronic health information at healthcare institutions nationwide. Patient health data extracted from electronic health records (EHRs) and aggregated across healthcare organizations into multi-institutional datasets is ideally suited for population health surveillance. However,
the capacity of local public health departments to access, analyze, and transform EHR data into population health information is unknown.

We propose to conduct key informant interviews with local public health departments nationwide to: 1) assess local public health department approaches to population health surveillance, 2) describe technology, partnership, and resource barriers impacting population health surveillance, 3) explore key stakeholders to local public health department population health activities, and 4) gather opinions about the role of EHR data in population health surveillance.

III. Preliminary Studies/Progress Report:

Describe qualitative feedback from key informants regarding knowledge, attitudes, and practices related to population health surveillance in general and specific to chronic disease. This study will also seek to understand attitudes and beliefs about the role of EHR data in population health surveillance.

IV. Research Methods

A. Outcome Measure(s):

The primary outcome will be to understand local public health department population health surveillance knowledge attitudes and practices.

B. Description of Population to be Enrolled:

Study Population: The study population will include up to 30 leaders from local public health departments. Contacts at each health department have been identified but
each informant will be given the option to suggest another member of their
health department more aligned with the subject matter.

<table>
<thead>
<tr>
<th>Chart 1: Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Informants</strong></td>
</tr>
<tr>
<td><strong>Public Health Practitioners</strong></td>
</tr>
<tr>
<td><strong>POTENTIAL TOTAL</strong></td>
</tr>
</tbody>
</table>

Inclusion Criteria: Public health departments engaged in population health activities from large urban areas are included. Public health departments were identified from literature, publication, or health reporting.

Exclusion Criteria: State and national public health organizations are excluded. Rural health departments unlikely to be engaged in population health were excluded.

C. Study Design and Research Methods:

A convenience sample of leaders from local public health departments will be recruited to participate in 45-60 minute key informant interviews. Key informants will be invited to participate through an introductory email that describes the purpose and goals of the study. During the interview, informant name, role and organizations will be collected. Interview responses will be used to:

- to understand how LPHAs approach local population health surveillance
- to understand the key issues, challenges, and barriers LPHAs face in population health surveillance?

Interview questions include but are not limited to:

- Why is it important for public health to engage in population health surveillance?
- Can you describe your organization’s approach to local population health surveillance?
- What types of population health data is your organization working with?
• What internal and external stakeholders are important to your local population surveillance work?
  o How are stakeholders using this information?
• How would you describe your organization’s capacity to monitor population health?
• If you had unlimited resources, what additional population health data and or informational activities would you pursue?
• What factors have been critical to your success?

Each interview will be recorded and transcribed. Transcriptions will be used for qualitative analysis of key themes.

Informed Consent/Agreement: For this exempt research, a formal informed consent document or process is not required. However, each participant will still be asked for affirmative agreement to participate before beginning each interview. Specifically, the Investigator or member of the research team will read the Invitation to Participate text to describe the study. Participants will be informed that their participation is voluntary, and that they can refuse to answer any question, or withdraw from the interview, at any time without consequence. Participants will be provided with the opportunity to ask questions before providing verbal agreement.

D. Description, Risks and Justification of Procedures and Data Collection Tools:

Study Risks:

a. It is highly unlikely that participants could perceive interview questions as sensitive or uncomfortable. Further, disclosures of the participants’ responses outside of the research could not reasonably place the participants at risk of criminal or civil liability or be damaging to the participants’ financial standing or reputation.

b. The primary risk of this study is the potential for loss of confidentiality of a participant through voice identification during the transcription process. We believe this risk is very low.

c. There are no physical risks to participants.

Protection against Risks:
a. To reduce possible participant discomfort, participants can choose to not answer specific questions or withdraw from the interview at any time without consequence.
b. Voice recordings will be stored on a secure Denver Health (DH) network drive and destroyed after transcription.

**Study Benefits:** Participants will receive no direct benefit from their participation in this study. However, this project may improve population health surveillance capacity among public health stakeholders.

**Risk-Benefit Justification:** This investigation involves minimal risk to any participant. The benefits, to improve local public health capacity to engage in population health surveillance, outweigh the minimal risks to participants.

**Privacy and Confidentiality:**

**Privacy:**

a. The interviews will collect the individuals name and organization. However, this information will not be used during the analysis.

**Confidentiality:**

a. **Plan to protect identifiable information from improper use and disclosure:** Participant responses will be aggregated to assess themes. Individual responses will not be attributed to specific participants. Interviews will be stored on the Denver Public Health secure servers, which require unique user IDs and password combinations. Only study team members will have access to study data. Data are backed up nightly.

b. **Data destruction plan:** All interview recordings will be kept on a secure DH network drive and permanently deleted after transcription.

c. **PHI will not be collected:** No protected health information will be collected in this study.
Data and Safety Monitoring Plan:

**Oversight:** The PI will oversee monitoring of the safety of the research and report any unexpected adverse events to IRB.

**Unanticipated Problems Related to the Research:** Should an unanticipated problem occur, it will be assessed, addressed and reported to COMIRB according to each's respective policies.

**E. Potential Scientific Problems:**

None anticipated.

**F. Data Analysis Plan:**

All interviews will be audio recorded and thematically coded. Common themes will be identified. Digital recordings will be stored on a secure access limited DH network drive and deleted from the recording device. Once analysis has been completed, the digital recording will be deleted from the network drive.

**G. Summarize Knowledge to be Gained:**

Results from of this study could have significant impact on the practice of local population health surveillance by public health department. This study will describe the current and aspirational capacity of local public health to use EHR data for local population health monitoring. This study will serve as an environmental scan of local public health capacity and population health surveillance activities, data sources, and indicators.

**COMIRB PROTOCOL Reference List**
• Stoto M. *Population Health in the Affordable Care Act Era:* Academy Health;2013.