The Geographic Distribution of Cardiovascular Health in SPHERE

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Public Health in the Graduate School of The Ohio State University

By

Caryn Roth

Graduate Program in Public Health

The Ohio State University

2014

Master’s Examination Committee:

Dr. Philip R.O. Payne, Advisor

Dr. Randi E. Foraker

Dr. Albert M. Lai
Copyright by

Caryn Roth

2014
Abstract

**Introduction:** Coronary heart disease and stroke are leading causes of morbidity and mortality in the United States (US) and cost the healthcare system an estimated $315.4 billion annually. Area-level factors which impact health may affect the distribution of cardiovascular disease (CVD), given its multifactorial and chronic nature.

**Objective:** We sought to characterize the geographic distribution of cardiovascular health (CVH) among women over 65 years of age involved in the ongoing Stroke Prevention in Healthcare Delivery EnviRonmEnts (SPHERE) Study.

**Methods:** Data derived from electronic health records (EHRs) were collected from all women over 65 years of age in two Midwest primary care clinics. We characterized CVH according to the American Heart Association’s *Life’s Simple 7* campaign, which scores modifiable risk factors for CVH and attributes a higher score to better health. We characterized these factors using the collected patient data, and geocoded patient addresses to the level of the US census tract. We integrated census tract population characteristics into the analysis to determine if they were associated with CVH. We conducted sensitivity analyses for missing data on the overall CVH score, as well as for some of the individual components of the score. We calculated the mean fractional score, the actual CVH score divided by the maximum possible. The association between US
Results: The mean fractional score was .63 across both clinics. Few patients were in ideal CVH, and the distribution of overall CVH and individual factors differed geographically. Only weekly per capita expenditure on fruits and vegetables was associated with CVH score at the 0.05 significance level. Imputing missing values had little effect on overall patient CVH classification.

Discussion: Our patient population has significant room for improvement in modifiable behaviors and factors that contribute to CVH. By empowering patients to understand their health with their primary care physicians, we can encourage behavior changes and shift the population distribution of CVH. In order to do this most effectively, we must deploy interventions that are relevant and actionable for diverse patient and provider populations, which can vary within one medical center. Augmenting EHR-based data with available census tract-level data provides additional opportunities to understand our patient population and consider the lifestyle and socioeconomic factors that impact their health.
Dedication

This work is dedicated to my father, Ilan Roth, who always encouraged me to pursue my master’s degree.
Acknowledgments

I wish to sincerely thank my advisors, Philip Payne and Randi Foraker, for their guidance and support throughout my time working and learning at The Ohio State University. I would also like to thank the other faculty, staff and students in the Department of Biomedical Informatics who have greatly enhanced my learning and graduate experience.
Vita

2006.........................................................Berkeley High School, Berkeley, CA

2010..........................................................B.S. Computational and Systems Biology,
University of California Los Angeles, Los Angeles, CA

2010-2011 .............................................CDC Public Health Associate, Delaware General Health District, Delaware, OH

2011-2012 .............................................Research Administrator, Department of Biomedical Informatics, The Ohio State University, Columbus, OH

2013-2014 .............................................Graduate Research Associate, Department of Biomedical Informatics and College of Public Health, The Ohio State University, Columbus, OH
Publications


Fields of Study

Major Field: Public Health
Specialization in Biomedical Informatics
Table of Contents

Abstract .......................................................................................................................... ii

Dedication ....................................................................................................................... iv

Acknowledgments ........................................................................................................... v

Vita................................................................................................................................ vi

Publications.................................................................................................................... vii

Fields of Study ............................................................................................................... viii

List of Figures .................................................................................................................. xii

Chapter 1: Introduction & Background ........................................................................ 1

Cardiovascular Health in the US .................................................................................... 1

Life’s Simple 7™ .............................................................................................................. 2

Learning Healthcare System .......................................................................................... 3

SPHERE .......................................................................................................................... 4

Geographic Information Systems .................................................................................... 8

Chapter 2: Methods ......................................................................................................... 10

Specific Aims and Hypothesis ......................................................................................... 10
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Patient Data</td>
<td>14</td>
</tr>
<tr>
<td>Patient Address Data</td>
<td>15</td>
</tr>
<tr>
<td>Census Tract-Level Population Characteristics</td>
<td>17</td>
</tr>
<tr>
<td>Chapter 3: Results</td>
<td>19</td>
</tr>
<tr>
<td>SPHERE baseline data</td>
<td>19</td>
</tr>
<tr>
<td>Geocoded Data</td>
<td>24</td>
</tr>
<tr>
<td>Community Level Data</td>
<td>32</td>
</tr>
<tr>
<td>Chapter 4: Discussion</td>
<td>37</td>
</tr>
<tr>
<td>Baseline Patient Data</td>
<td>37</td>
</tr>
<tr>
<td>Area-Level Factors</td>
<td>39</td>
</tr>
<tr>
<td>Challenges and Limitations</td>
<td>41</td>
</tr>
<tr>
<td>Future Directions</td>
<td>43</td>
</tr>
<tr>
<td>Conclusion</td>
<td>44</td>
</tr>
<tr>
<td>References</td>
<td>45</td>
</tr>
<tr>
<td>Appendix A: Data Provenance for Analysis</td>
<td>51</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Measures of CVH in categories of poor, intermediate, and ideal ................................. 3
Table 2: Demographics and continuous CVH score components (mean, SD) by clinic:
SPHERE baseline.................................................................................................................. 19
Table 3: Overall CVH score by category, missing and imputed values .............................. 20
Table 4: Patients classified in poor, intermediate and ideal CVH categories, missing and
imputed values .................................................................................................................. 21
Table 5: Diabetes classification at Worthington clinic using fasting glucose (FG) and
diabetes medications (DM meds) versus DM meds alone............................................... 23
Table 6: Diabetes classification at Worthington clinic using hemoglobin A1c (HbA1c)
and diabetes medications (DM meds) versus DM meds alone .................................... 23
List of Figures

Figure 1: Information flow during patient encounter ........................................ 6
Figure 2: SPHERE application embedded in the EHR ........................................ 7
Figure 3: Analysis workflow and specific aims .................................................... 13
Figure 4: Cohort selection for univariate linear regression and mapping patients by CVH ................................................................. 18
Figure 5: Proportion of older female patients in categories of each CVH factor ....... 22
Figure 6: CarePoint East vs. Worthington clinic patient addresses ...................... 25
Figure 7: Patients with no missing data, plotted by CVH score ............................. 26
Figure 8: BMI score .......................................................................................... 27
Figure 9: Cholesterol Score ............................................................................... 28
Figure 10: Smoking Score .................................................................................. 29
Figure 11: Blood pressure score ......................................................................... 30
Figure 12: Diabetes score ................................................................................... 31
Figure 13: Franklin county weekly per capita fruit and vegetable expenditures by census tract ........................................................................................................... 33
Figure 14: Franklin county percent population > 25 unemployed by census tract .... 34
Figure 15: Franklin county weekly per capita sugar and artificial sweeteners expenditures by census tract ................................................................. 35

Figure 16: CVH score and weekly per capita expenditure on fruits and vegetables ....... 36
Chapter 1: Introduction & Background

Cardiovascular Health in the US

Each year, an estimated 831,000 Americans die from cardiovascular disease (CVD).¹ According to the American Heart Association’s strategic impact goals through 2020, coronary heart disease, stroke and related vascular deaths will continue to be leading causes of death and disease in the United States (US).² Currently, CVD accounts for the largest percentage of US healthcare expenditure at an annual estimate of $315.4 billion in direct and indirect costs.³

There has been substantial progress in decreasing the population burden of CVD, due to reduction in risk factors such as high cholesterol, hypertension and smoking. More effective treatments and pharmacological therapies have contributed to this reduction, leading the age-standardized risk of CVD to decrease.⁴ However, the prevalence of obesity and diabetes, two pivotal risk factors for CVD, are steadily increasing, as 36% of adults and 18% of children and adolescents in the US are now considered obese.⁵,⁶ Similarly, diabetes prevalence has been increasing, and is projected to continue to rise among both adults and adolescents.⁷,⁸ Not surprisingly, data from the National Health and Nutrition Examination Surveys (NHANES) show that less than 1% of US adults achieve ideal cardiovascular health (CVH).

¹

1
Additionally, individuals are living substantially longer than they did 40 years ago, meaning that there are more older adults at risk for CVD. This age group also suffers from other chronic disease comorbidities, such as diabetes and chronic obstructive pulmonary disease, which increase the risk of developing CVD.\textsuperscript{9,10} Since this age group is most at risk for the disease to begin with, the overall prevalence of CVD continues to rise. Researchers estimate that as the population distribution in the US continues to shift, elderly and minority rates of stroke will increase the most.\textsuperscript{11} For these reasons, it is crucial to focus especially on these groups in terms of reducing overall stroke and CVD prevalence.

**Life’s Simple 7\textsuperscript{TM}**

In an attempt to focus efforts in tackling this disease, the American Heart Association introduced the CVH metric in 2010 and the corresponding *Life’s Simple 7\textsuperscript{TM}* prevention campaign.\textsuperscript{2} This metric classifies CVH factors into categories of poor, intermediate, and ideal (Table 1).

These factors have been consistently associated with CVD risk and are all modifiable.\textsuperscript{2,12,13} Thus, this metric emphasizes prevention of risk factors, as well as promotes changes to reach ideal CVH. In this way, it serves as a public health tool to assess the health of Americans in a meaningful and actionable way.\textsuperscript{14}
Table 1: Measures of CVH in categories of poor, intermediate, and ideal (2)

<table>
<thead>
<tr>
<th>Modifiable factor</th>
<th>Poor Health</th>
<th>Intermediate Health</th>
<th>Ideal Health</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking status</strong></td>
<td>Yes</td>
<td>Former ≤ 12 months</td>
<td>Never or quit &gt; 12 months</td>
</tr>
<tr>
<td><strong>Body mass index</strong></td>
<td>≥30 kg/m²</td>
<td>25 - 29.9 kg/m²</td>
<td>&lt;25 kg/m²</td>
</tr>
<tr>
<td><strong>Total cholesterol</strong></td>
<td>≥240 mg/dL</td>
<td>200-239 mg/dL or treated to goal</td>
<td>&lt;200 mg/dL</td>
</tr>
<tr>
<td><strong>Blood pressure</strong></td>
<td>SBP ≥140 mmHg or DBP ≥90 mmHg</td>
<td>SBP 120-139 mmHg or DBP 80-89 mmHg or treated to goal</td>
<td>SBP &lt;120 mmHg and DBP &lt;80 mmHg</td>
</tr>
<tr>
<td><strong>Fasting glucose</strong></td>
<td>≥126 mg/dL</td>
<td>100-125 mg/dL or treated to goal</td>
<td>&lt;100 mg/dL</td>
</tr>
<tr>
<td><strong>Physical activity</strong></td>
<td>0 minutes of moderate or vigorous activity/week</td>
<td>1-149 min moderate, 1-74 min vigorous, or 1-149 min moderate/vigorous</td>
<td>≥150 min moderate ≥75 min vigorous, or ≥150 min moderate/vigorous</td>
</tr>
<tr>
<td><strong>Healthy diet score</strong></td>
<td>0 – 1 components</td>
<td>2 – 3 components</td>
<td>4 – 5 components</td>
</tr>
</tbody>
</table>

Learning Healthcare System

The US healthcare landscape is changing, which provides opportunities to address CVD in new ways. In the past decade, the use of electronic health records (EHRs) in the US has skyrocketed. This increase was driven by the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, a federal stimulus bill which provides incentive payments for health care providers to offset costs associated with EHR implementation.15 With the passage of HITECH, EHR adoption has rapidly expanded, increasing from 9.4% of hospitals in 2008 to 44.4% in 2012.16 Most of the stimulus bill funds are allocated as Centers for Medicare and Medicaid Services (CMS)
incentive payments for “meaningful use” of health information technology, mandating that EHRs contribute to enhanced quality, efficiency and engagement in healthcare.\textsuperscript{17,18}

The increase in adoption of EHRs and emphasis on meaningful use of them present significant opportunities to utilize EHRs for clinical interventions and research purposes.\textsuperscript{19} Clinical data are now collected and recorded in the EHR and are available for these “secondary uses”. As such, research costs lie only in the analysis, and patient-level clinical data can be aggregated for research that impacts the larger population.\textsuperscript{20,21} Utilizing EHR data, researchers can compare costs, benefits and risks associated with various aspects of clinical practice in a systematic and large-scale manner. This knowledge generated as a byproduct of routine clinical care can then serve to inform future practice, resulting in higher quality outcomes.\textsuperscript{22} As we work towards this type of learning healthcare system where clinical practice and research build on each other, studies of interventions delivered through the EHR can truly help us realize its full potential and benefit. Ultimately, EHRs can allow scientists to surmount translational research barriers by capitalizing on these rich data sources, fully enabling patient-centered outcomes and comparative effectiveness research.\textsuperscript{23,24}

**SPHERE**

With its serious consequences and numerous modifiable factors, CVD warrants far-reaching interventions that utilize new technologies to promote public health. Rather than just focus on the most high-risk patients, a disease as far-reaching as CVD is amenable to a population-based prevention approach.\textsuperscript{25} For CVD and other chronic
diseases, changing the focus to primary and primordial prevention can be cost-effective and benefit health. We sought to use existing EHR infrastructure to deliver an evidence-based intervention and visualization to facilitate patient-provider communication and understanding of CVH risk, in order to empower patients to improve their health. We designed SPHERE (Stroke Prevention in Healthcare Delivery EnviRonmEnts), a data-driven health assessment instrument that enables seamless point-of-care interactive risk profiling for primary care providers and their patients. We piloted SPHERE among academic medical center primary care providers serving an urban, low socio-economic status patient population.

In order to incorporate the public health guidelines into our commercial EHR, we employed the composite CVH score described above, modified from the AHA’s Life’s Simple 7™ public health prevention campaign. We garnered feedback from providers to verify that the tool would incorporate health factors that were relevant for their patients, as well as accurately display information collected in routine clinic visits. Similarly, we queried historical data in the EHR to check which factors had a high proportion of missing values, in an attempt to use more complete data to populate SPHERE.

We designed SPHERE to pull current and historical data from the EHR (Figure 1) to populate a visualization delivered through a best practice alert within the patient record. This alert presents a visual display of modifiable risk factors for CVD and the patient’s score for each factor (0, 1 and 2 corresponding to poor, intermediate and ideal health, respectively), as well as an overall CVH score (0-100) encompassing all the factors. Diet and physical activity, the two factors not currently available in our academic
medical center’s EHR, are optional in the overall score calculation, and a different calibration is used depending on whether or not these values are present. If any values are missing or outdated, the provider can order new tests directly from the EHR alongside SPHERE.

![Figure 1: Information flow during patient encounter](image)

The application is intended to generate conversation between the patient and provider about these modifiable factors as a way to empower and engage the patient in his or her health care. Importantly, the application is interactive, so patients can immediately see the impact of small changes on overall CVH. The standard color-coding scheme we employed (green, yellow and red represent ideal, intermediate and poor health, respectively) is simple and clear for both provider and patient, making it a useful tool for quick understanding and communication (Figure 2).

We are iteratively refining and updating SPHERE as we evaluate its impact on providers and our patient population. In this way, the project contributes to the learning healthcare system’s goals, intertwining clinical care and research for better patient outcomes.
Figure 2: SPHERE application embedded in the EHR
Geographic Information Systems

Current EHRs lack easily accessible diet and physical activity data, which are key in chronic disease management and prevention. The overall SPHERE CVH score accesses only the available EHR data, so the absence of behavioral and lifestyle information gives us an incomplete picture of the patients in the SPHERE study. We do, however, have access to patient address data, which we may be able to use as a proxy to enhance our understanding of our patient population using information from their area of residence.

While the causes of CVD are multifactorial, previous research has shown that socioeconomic, demographic, and built environment factors may influence the development of the disease. Geographic Information Systems (GIS), a spatial analysis method, allows researchers to map and explore the link between area of residence and health outcomes. In public health, it is widely accepted that where you live affects your health, and social determinants of health are increasingly associated with location. By patient location according to US Census tracts and linking these areas to community-level data, we may be able to study the effect of area of residence on disease. Characterizing the neighborhoods where our patients reside may help us better understand our patients, and therefore provide more relevant and effective care to improve CVH. Furthermore, we may be able to identify subsets of patients for whom interventions are more effective, thereby strengthening our learning healthcare system even further. By augmenting patient data with geographic context, we may be able to
explore how community-level factors impact disease and health, including the
distribution of CVH.33
Chapter 2: Methods

Specific Aims and Hypothesis

Towards our overall goal of characterizing the geographic distribution of cardiovascular health among patients in SPHERE, we laid out the following specific aims and hypotheses related to the three broad areas of this analysis: baseline patient data, patient address data, and census tract level population characteristics. These aims are also summarized in Figure 3.

Baseline Patient Data

1. Characterize the demographics and baseline cardiovascular health (CVH) data for patients in the intervention and control clinics for SPHERE, according to the American Heart Association’s Life’s Simple 7 classification system. Calculate means and standard deviations of the continuous variables, as well as frequencies for scores within CVH categories.

Hypothesis: We hypothesize that few patients will be in ideal health on all 5 factors. Furthermore, we hypothesize that the two clinics will differ in terms of
demographics, means, and frequencies within CVH categories. Specifically, we anticipate that the control clinic will be healthier than the intervention clinic.

2. Impute missing EHR-based data based on means and modes to determine whether patients with imputed values fall into different categories of overall CVH, which we will classify based on natural cut points in the data.

Hypothesis: We hypothesize that imputing missing EHR-based data will not significantly affect our aggregate baseline analysis results.

3. Conduct sensitivity analysis using Hemoglobin A1c, Fasting Glucose and Diabetes medication data from the control clinic to compare the CVH classification across the three different scoring systems.

Hypothesis: We hypothesize that scoring only using diabetes medications will make our patient cohort appear healthier than reality, and that patients will be classified in worse health if they are scored with either Hemoglobin A1c or Fasting Glucose data.

Patient Address Data

4. Geocode patient address to map patient location and color-code coordinates based on clinic location, overall CVH and individual parameters in the CVH score
Hypothesis: We hypothesize that patients from the control clinic will cluster in the North-Western portion of the county, while patients from the intervention clinic will cluster in the South-Eastern portion of the county, based on clinic location. We also hypothesize that we will see clear differences in the distributions of overall CVH as well as individual components of the score across the county.

Census Tract Level Population Characteristics

5. Characterize census tracts according to expenditure and socioeconomic data and plot these data to visually assess their distribution across the county.

Hypothesis: We hypothesize that the distributions of these census-tract level characteristics will not be uniform across the county. Specifically, we hypothesize that the North-Western portion of the county will have higher expenditures on healthy foods, lower expenditures on unhealthy foods, and higher levels of the various socioeconomic indicators than the South-Eastern portion of the county.

6. Link population characteristics to health data by census tract and perform univariate linear regression, clustered by census tract to determine which, if any, of these variables are associated with CVH at the 0.05 significance level.

Hypothesis: We hypothesize that some of the population-level characteristics will be significantly associated with CVH. In particular, healthy behaviors and factors
will be associated with better CVH, and adverse behaviors and factors will be associated with worse CVH.

**Figure 3**: Analysis workflow and specific aims

- **Aim 1**: Characterize demographics and CVH
- **Aim 2**: Impute missing data for overall CVH
- **Aim 3**: Conduct sensitivity analysis for diabetes scoring
- **Aim 4**: Geocode patient addresses and map distributions of CVH
- **Aim 5**: Characterize and map census tracts according to population characteristics
- **Aim 6**: Link to CVH data by census tract and perform univariate linear regression
Baseline Patient Data

We queried retrospective data from The Ohio State University Wexner Medical Center (OSUWMC) Information Warehouse (IW) for female patients 65 or older seen at two primary care clinics from May 1 through July 31, 2013. We examined data from the CarePoint East clinic at OSUWMC, the pilot site for the SPHERE intervention, and used data from the Worthington clinic as a control for the overall analysis of the intervention’s efficacy. In both locations, we limited the intervention and data collection to patients seen in primary care. For these patients, we queried basic demographic data such as age and race, address data and health data for the previous year collected through the EHR. These data included values for total cholesterol, fasting glucose, hemoglobin A1c, height, weight, systolic and diastolic blood pressure and smoking status. We also requested data on hypertension, diabetes and lipid medications.

Since our data was comprised of multiple visits for many of the patients in our cohort, we chose to use the earliest date within the year-long query window. We merged these data with the medications and laboratory assessments and discarded data where the prescription or order date was after the initial visit, in order to capture an accurate snapshot of our patients at the baseline time point. We also standardized data across patients (for example different ways of recording feet and inches for height), and converted numeric data stored as strings into the correct format. We used height and weight to calculate body mass index [BMI (kg/m²)] and assigned scores for each factor (BMI, cholesterol, blood pressure, blood sugar and smoking) based on the AHA criteria for Life’s Simple 7™ (Table 1). Each factor was rated “poor,” “intermediate” or “ideal,”
and scored as 0, 1 and 2 respectively. We calculated an overall CVH score by summing across individual factors, for each patient, ranging from 0 (all health factors poor) to 10 (all health factors ideal). We calculated means and standard deviations of the continuous variables, as well as frequencies for scores within CVH categories. We performed this analysis using Stata (Version 12.1, StataCorp LP, College Station, Texas).

Beyond the factors that Life’s Simple 7\textsuperscript{TM} uses to calculate the CVH score, we analyzed patient data for alternative health markers relating to blood sugar and diabetes. In addition to the recommended fasting glucose value, we also considered using hemoglobin A1c (HbA1c) values or diabetes medications to classify patients in regard to their diabetes status. We compared the score (0–2) for this factor using the different data values, in order to conduct a sensitivity analysis for missing fasting glucose data in the Worthington clinic cohort.

We also addressed missing data in the overall CVH score by imputing values that were missing. We imputed mean values for the continuous variables of BMI and total cholesterol, and the mode for the categorical value smoking. We then determined how imputing these values affected where patients fell in categories of overall CVH, which we classified as poor (0–4), intermediate (5–7) and ideal (8–10), based on natural cut points in the data.

Patient Address Data

We used the patient address data to geocode our patients’ location in both the intervention and control clinics. We utilized ArcGIS (Version 10.2, ESRI, Redlands,
California) to find the latitude and longitude of each address in our dataset. In order to perform this geocoding, we had to import a shapefile into the software containing geographies of interest. In this case, we used a Tiger/Line® shapefile all of the census tracts in the state of Ohio, which we acquired from the US Census website. Since most of our patients were from Ohio, we did not geocode the one patient who resided outside the state. We imported our patient address data into ArcMap, and geocoded at the level of roof top coordinates. We used the world geocoding service address locator, an ArcGIS dataset of address attributes, indexes and queries to convert our patient addresses to coordinates. We manually reviewed each match, and corrected or removed addresses that the software could not accurately geocode. We removed patients (n=7) for whom the address listed in the EHR was a P.O. Box, as this would not give us data on their home location.

We joined our patient data to the US census tract shape file based on spatial location, which adds attributes from each census tract (for example the census tract number) to the latitude and longitude data associated with each patient. In this way, we were able to plot the patient coordinates on a map, as well as associate patients with their respective census tracts.

Using ArcMap, we plotted patient address on the map of Ohio. We grouped patient by clinic location (Worthington versus CarePoint East) to visualize the catchment area of patients for each clinic. For the patients who had data available on all CVH health factors, we separated patients by overall CVH score, categorizing them into poor health (CVH score 0-4), intermediate health (5-7) or ideal health (8-10). We also plotted
patients by the individual health factors that make up the CVH score, in order to assess which, if any, might have spatial distributions that differed from the overall score.

Census Tract-Level Population Characteristics

In order to enhance our understanding of our patient population, we acquired area-level data to describe the different census tracts in Franklin County. We used Nielsen PrimeLocation, a database of consumer marketing behavior data collected through online, phone and mail surveys, as well as through smartphone applications and barcode scanners. These data included weekly per capita expenditures on various food groups, as well as socioeconomic data on education and unemployment. We also collected data on median household income from the 2008-2012 American Community Survey 5 year estimates, conducted by the US Census. Further detail on data sources and data provenance for all variables used throughout this analysis is described in Appendix A. We mapped the socioeconomic and expenditure data in ArcGIS to visually assess the distribution of these factors.

Finally, we linked the area-level data to our patient health data, using census tract number as the common variable between datasets. From our initial cohort of 222 patients, we removed those for whom we had missing or incomplete data (Figure 4). These included patients outside of Ohio (n=1), patients with a P.O. box as their address (n=7), patients outside of Franklin County (n=26) and patients with missing values for some components of the CVH score (n=35). We performed univariate linear regression,
clustered by census tract, to determine which area-level characteristics were most correlated with CVH score.

Figure 4: Cohort selection for univariate linear regression and mapping patients by CVH
Chapter 3: Results

SPHERE baseline data

Patients in our combined cohort had a mean age of 73.8 (SD=7). 30.9% of the total cohort was black; however, the percentage was much higher at CarePoint East than at Worthington (35.4% vs. 19.4%). The mean fractional score, the actual CVH score divided by maximum possible score, accounting for missing values, was nearly identical across the two clinics, averaging 63% (Table 2).

Table 2: Demographics and continuous CVH score components (mean, SD) by clinic: SPHERE baseline

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Cohort</th>
<th>CarePoint East</th>
<th>Worthington</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 222</td>
<td>n = 160</td>
<td>n = 62</td>
</tr>
<tr>
<td>Age</td>
<td>Mean 73.8</td>
<td>Mean 74.2</td>
<td>Mean 72.8</td>
</tr>
<tr>
<td></td>
<td>SD 7</td>
<td>SD 6.7</td>
<td>SD 7.5</td>
</tr>
<tr>
<td>Black (%)</td>
<td>30.9</td>
<td>35.4</td>
<td>19.4</td>
</tr>
<tr>
<td>BMI</td>
<td>Mean 30.2</td>
<td>Mean 30.7</td>
<td>Mean 28.9</td>
</tr>
<tr>
<td></td>
<td>SD 6.3</td>
<td>SD 6.5</td>
<td>SD 5.8</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Mean 182.7</td>
<td>Mean 182.9</td>
<td>Mean 182</td>
</tr>
<tr>
<td></td>
<td>SD 37.7</td>
<td>SD 39.9</td>
<td>SD 31.5</td>
</tr>
<tr>
<td>Blood pressure, systolic</td>
<td>Mean 125.2</td>
<td>Mean 126.1</td>
<td>Mean 122.7</td>
</tr>
<tr>
<td></td>
<td>SD 16.6</td>
<td>SD 16.9</td>
<td>SD 15.7</td>
</tr>
<tr>
<td>Blood pressure, diastolic</td>
<td>Mean 71.6</td>
<td>Mean 72.1</td>
<td>Mean 70.3</td>
</tr>
<tr>
<td></td>
<td>SD 10.3</td>
<td>SD 11.1</td>
<td>SD 8</td>
</tr>
<tr>
<td>Mean Fractional Score (actual</td>
<td>Mean 0.63</td>
<td>Mean 0.63</td>
<td>Mean 0.64</td>
</tr>
<tr>
<td>score/maximum possible)</td>
<td>SD 0.15</td>
<td>SD 0.14</td>
<td>SD 0.15</td>
</tr>
</tbody>
</table>

In the CarePoint East cohort, 12.5% (n=20) of data on BMI was missing from the EHR. Eleven percent (n=18) of cholesterol data was missing for CarePoint East patients,
compared to 12.9% (n=8) from the Worthington clinic. One patient in the CarePoint East cohort lacked smoking data.

According to our three cut points for overall CVH score, 9%, 77% and 14% of our cohort had poor (0–4), intermediate (5–7) and ideal (8–10) CVH, respectively. When we imputed the mean value for BMI (30.2) and cholesterol (182.7), as well as the mode for smoking score (2, i.e. the patient never smoked or has not smoked for at least one year), these percentages changed to 8.5%, 72.5% and 19% for poor, intermediate and ideal CVH, respectively (Table 3).

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
<th>Percent</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>20</td>
<td>9.01</td>
<td>19</td>
<td>8.56</td>
</tr>
<tr>
<td>Intermediate</td>
<td>171</td>
<td>77.03</td>
<td>161</td>
<td>72.52</td>
</tr>
<tr>
<td>Ideal</td>
<td>31</td>
<td>13.96</td>
<td>42</td>
<td>18.92</td>
</tr>
</tbody>
</table>

A matrix showing the differences between the scores using missing data versus imputed data is shown in Table 4. One patient who was previously categorized as poor health was now categorized as intermediate, and 11 patients previously categorized as intermediate were bumped up to ideal, using the imputed values.
Table 4: Patients classified in poor, intermediate and ideal CVH categories, missing and imputed values

<table>
<thead>
<tr>
<th>Missing Data</th>
<th>Poor</th>
<th>Intermediate</th>
<th>Ideal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0</td>
<td>160</td>
<td>11</td>
<td>171</td>
</tr>
<tr>
<td>Ideal</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>161</td>
<td>42</td>
<td>222</td>
</tr>
</tbody>
</table>

For the five factors that comprise the CVH score, the breakdown of different categories (ideal, intermediate and poor) is shown in Figure 5. We initially planned to assess diabetes according to Life’s Simple 7™, which reflects the practice of using fasting glucose for to assess for diabetes in population-based cohort studies. However, when we explored the baseline patient data, we found that 100% of patients in our intervention clinic were missing this value. User acceptance testing with providers in that clinic revealed that they had already adopted the use of hemoglobin A1c (HbA1c) as the metric to assess for diabetes, as recently recommended by the American Diabetes Association. In light of this, we modified the SPHERE application to use HbA1c to calculate the diabetes parameter score during the patient encounter.

However, further inquiry with the physicians revealed that due to reimbursement restrictions, even the HbA1c tests were inconsistently ordered. Thus for the baseline cohort characterization, we scored patients as either treated by a glucose-lowering medication (intermediate) or not (ideal).
Our control clinic had a higher proportion of patients with both FG and HbA1c values, so we conducted a sensitivity analysis to determine how much the decision to classify baseline patient data with only diabetes medication affected the score. This allowed for comparative diabetes scoring with the three different classification systems (Tables 5 and 6). For the 29 patients with data available for FG, 41% (n = 12) were misclassified when we only used diabetes medications. One of the two patients (50%) originally classified as intermediate with only diabetes medications was now classified as poor health on that factor, and 11 of the 27 patients (41%) who were initially classified as ideal were bumped down one category to intermediate. Using the available HbA1c data for 14 patients, we saw that seven (50%) were misclassified using the DM meds alone. These included four of the six patients (66%) in the intermediate category who were reclassified as poor, one of eight patients (13%) in the ideal category who was reclassified as poor, and two of the eight (25%) who were reclassified as intermediate. Of note, both the FG scoring (Table 1) and HbA1c scoring (≥6.5 poor; 5.7-6.4 poor).

Figure 5: Proportion of older female patients in categories of each CVH factor

5a. CarePoint East Clinic  
5b. Worthington Clinic
intermediate; \(<=5.6\) ideal) systems still utilized diabetes medication to determine if the patient was being treated for DM, which lowered their calculated score for that factor.

This guideline is taken directly from *Life’s Simple 7*, where the rational is that those who are currently at ideal levels for blood sugar, blood pressure and cholesterol only with the help of medication were in the past at non-ideal levels, and therefore have been exposed to these adverse risk factors which accumulate.

Table 5: Diabetes classification at Worthington clinic using fasting glucose (FG) and diabetes medications (DM meds) versus DM meds alone

<table>
<thead>
<tr>
<th>Diabetes score using FG and DM meds</th>
<th>Missing</th>
<th>Poor</th>
<th>Intermediate</th>
<th>Ideal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Ideal</td>
<td>27</td>
<td>0</td>
<td>11</td>
<td>16</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>1</td>
<td>12</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Diabetes classification at Worthington clinic using hemoglobin A1c (HbA1c) and diabetes medications (DM meds) versus DM meds alone

<table>
<thead>
<tr>
<th>Diabetes score using HbA1c and DM meds</th>
<th>Missing</th>
<th>Poor</th>
<th>Intermediate</th>
<th>Ideal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Ideal</td>
<td>46</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

For patients with data available for FG (n=29), 3\% (n=1), 38\% (n=11) and 55\% (n=16) were in poor, intermediate and ideal health, respectively. Using available data for HbA1c (n=14), 36\% (n=5), 29\% (n=4) and 36\% (n=5) patients were classified as poor, intermediate and ideal, respectively, for this health factor. These numbers differ greatly
from the classification using only diabetes medications, which categorizes these same patients, plus those with missing values for FG and HbA1c as 13% (n=8) intermediate and 87% (n=54) ideal for the diabetes factor.

Geocoded Data

When we plotted patients by clinic location (Figure 6), we saw that patients from the South-East portion of the county generally sought primary care at CarePoint East (shown in blue), while patients from the North-West portion of the county tended to see providers at the Worthington clinic (shown in yellow).
Figure 6: CarePoint East vs. Worthington clinic patient addresses

We plotted patients by overall CVH score (Figure 7), with red, yellow and green corresponding to poor (0-4), intermediate (5-7) and ideal (8-10) health, respectively. It appears that patients on the Southern portion of the map, especially on the Eastern side, have higher numbers in the poor (red) category, while the Northern part of the map depicts more patients in ideal (green) or intermediate (yellow) CVH.
Figure 7: Patients with no missing data, plotted by CVH score

We also examined the distributions of the individual factors that make up the CVH score, plotting 0 (poor health), 1 (intermediate health) and 2 (ideal health) as red, yellow and green, respectively in Figures 8-12. Poor BMI scores, like overall CVH appear to be more prevalent in the South-East portion of the county (Figure 8). Cholesterol values are mostly at ideal and intermediate levels across the entire county (Figure 9), as are smoking scores (Figure 10). The blood pressure score (Figure 11) is more variable across the county, with ideal and intermediate levels more prevalent across
the North-West portion. Finally, for diabetes (Figure 12), classified here only as ideal or intermediate, more patients in the South-East portion of the map are in the intermediate category than in the North-West area, where almost all patients are classified as ideal.

Figure 8: BMI score
Figure 9: Cholesterol score
Figure 10: Smoking score
Figure 11: Blood pressure score
Figure 12: Diabetes score
Community Level Data

We plotted data for the community-level factors by census tract on a map of Franklin County. Maps depicting weekly per capita expenditures on fruits and vegetables, percent of population over 25 years old unemployed and weekly per capita expenditures on sugar and artificial sweeteners by census tract are shown in Figures 13, 14, and 15. Of note, the census tract containing the airport shows no data (white in the maps). We saw that the distributions of these area level factors varied widely across Franklin County. Fruit and vegetable expenditures (Figure 13) were lowest in the central and South-Eastern portion of the county. Percent unemployment (Figure 14) and sugar and artificial sweetener expenditures (Figure 15) were highest in these same areas, showing an inverse effect.
Figure 13: Franklin county weekly per capita fruit and vegetable expenditures by census tract
Figure 14: Franklin county percent population > 25 unemployed by census tract
Figure 15: Franklin county weekly per capita sugar and artificial sweeteners expenditures by census tract
When we performed univariate linear regression, clustered by census tract, only the variable representing weekly per capita expenditures on fruits and vegetables was significantly associated with CVH score at the 0.05 significance level. We plotted this variable overlaid with the patient health data to visualize the distributions together (Figure 16). We see more patients in the ideal category of CVH in the Northern portion of the map, where fruit and vegetable expenditures are highest.

Figure 16: CVH score and weekly per capita expenditure on fruits and vegetables
Chapter 4: Discussion

Baseline Patient Data

We found that few patients in our cohort were in ideal CVH on all five of Life’s Simple 7™ categories that we were able to query. While the mean fractional score was nearly identical across both the CarePoint East and Worthington clinics, score distributions for the individual components of CVH differed across the clinics. In the CarePoint East clinic, consisting of a higher proportion of black patients, the component of CVH health where most patients were categorized as ideal was smoking. In the Worthington clinic, the CVH factor with the highest percentage of patients falling into the ideal category was diabetes. Since we used diabetes medication to classify the diabetes variable, some of these differences may be attributable to different prescription practices at the two clinics, even though they are part of the same medical system.

Similarly, we saw substantial differences in testing for hemoglobin A1c and fasting glucose across the two clinics, and widely different scores for the diabetes factor, depending on which value (FG, HbA1c, DM meds) we used in the assessment. Even if patient populations differ somewhat, it seems likely that the two clinics have differing patterns of care in terms of which laboratory assessments they order and use. Differences
could also be due to patients with high FG or HbA1c levels who are not yet treated or diagnosed at one or both clinics.

When we consider the appropriate data to use for EHR-based interventions such as SPHERE, it is crucial to consider the target audience in terms of not only our patient population, but also the providers who will use the application. Our work showed that provider practice can differ substantially even across one health system, so it is crucial to study workflow and practice at the clinic-level before deploying such interventions. Likewise, while SPHERE was the first application of this type at our institution, researchers are already eager to automate other public health campaigns or risk scores in a similar way. While this offers great potential to integrate these prevention campaigns into primary care, it is necessary to fully consider which ones are most amenable EHR-based interventions that will actually be relevant and actionable for patients and providers alike. As we make these deliberations, we must carefully think through what data needs to be pulled to populate them, and will it be available, accurate and relevant.

Our results also showed that missing EHR data could change CVH categorization using our scoring criteria. When we imputed mean and mode values for our patient data, 5.4% of patients (12 out of 222 patients) changed to a different category. While some of these patients likely do fall into the higher score categories, others probably do not have the mean values for these factors. However, when we look at the population in aggregate for the ongoing SPHERE evaluation, this very small percentage will likely not impact our analysis and calculations.
On the other hand, it may not make sense to impute data for individual patients when they see their personal CVH score, as we would not want them to underestimate their own risk with an artificially enhanced CVH score. For this reason, SPHERE currently shows missing data in grey, prompting the provider to order needed laboratory assessments or add in known data that is not already present in the EHR.

Area-Level Factors

For the SPHERE intervention, we chose to focus on factors that patients could modify in order to improve their health. We felt that the primary care clinic presented an excellent venue for this type of prevention activity, as primary care providers are highly trusted by patients and see them at regular intervals. Thus, these doctors are able to truly focus on prevention of disease, rather than only prescribe treatment once a patient is already sick. This emphasis on primary and primordial prevention has shown substantial benefit for CVD, and can be cost-effective in our current healthcare system.\textsuperscript{26}

Many of the modifiable factors we focused on in SPHERE may be linked to area-level characteristics that impact health. For example, healthy diet and physical activity, two of the factors in \textit{Life’s Simple 7}, may be more accessible in neighborhoods with higher numbers of grocery stores, farmers markets, parks or recreation centers.\textsuperscript{35} Similarly, socioeconomic factors like income and education have been shown to impact CVH, and are important to consider as risk factors for this multifactorial disease.\textsuperscript{36,37} These socioeconomic factors may also contribute to an individual’s ability to purchase and prepare healthful foods.\textsuperscript{38} thus impacting BMI, diabetes, cholesterol and blood
pressure, and promoting poorer CHV. All of these health factors are clearly intertwined, but we can more deeply understand them, as well as target appropriate interventions, by considering where our patients live.

For these reasons, we enhanced our analysis with census-tract level data describing our patients in terms of characteristics not contained in the EHR. Although only one of the area-level factors we analyzed, weekly per capita expenditures on fruit and vegetables, was significantly associated with CVH, extensions of this work will continue to explore additional area-level characteristics thought to be associated with CVH. Furthermore, by mapping and analyzing these characteristics together, we may be able to identify patterns and clusters across our patient population, and use this area-level data to gain better insight into the clinical and non-clinical factors that contribute to CVH.

Long term, if we find that SPHERE’s efficacy varies by patient location or cluster, we may be able to tailor the intervention or focus it primarily on patients with specific area-level characteristics. Furthermore, we could systematically bring data into our EHR which describes these census tract characteristics, in order to allow physicians to get a sense of a patient’s living environment and make more relevant health recommendations. For example, if medical data in the record showed that the patient had an elevated BMI but census-tract level data showed that the patient’s neighborhood was high in crime with a low walkability score, the doctor could tailor a physical activity recommendation to focus on exercises the patient could do at home. In this way, understanding the diverse environments of our patients can enhance comparative
effectiveness research, provide insight into why certain treatments or interventions are more effective for specific populations, and offer ways to tailor interventions to make them more meaningful to a patient’s life. Thus incorporating community data with clinical data can promote a more robust learning health system that incorporates a holistic representation of each patient, rather than just a snapshot.

Challenges and Limitations

This study has several limitations, given the nature of our intervention and analysis. Through this work, we discovered that the Worthington clinic has a very different patient population than that at CarePoint East. This will bias comparisons between intervention and control clinics during our pilot period to evaluate the overall effect of SPHERE. However, we will evaluate SPHERE’s impact on CVH by comparing baseline data at each clinic to data collected over the same time period one year later. Thus, having the control clinic may allow us to understand overall trends across the medical center which impact CVH, irrespective of the SPHERE application.

Furthermore, knowing that patient populations, as well as provider practices vary across these clinics can help us tailor future interventions across our entire health system and others.

It is also important to consider the limitations of data derived from EHRs, which are not recorded with this type of secondary use in mind. For this reason, differences in EHR usage at the two clinics could also lead to the differences we saw in various CVH score factors. The low percentage of smokers in the data also suggests that missing data
could affect our analysis, causing us to classify our patient population as healthier than their actual CVH state.

While illuminating to the analysis, geocoding patient data presents its own set of challenges. Some patients had missing or inaccurate data, forcing us to drop them from this portion of the analysis. Similarly, other patients received their mail at P.O. boxes, making their address data meaningless for our purposes. These decisions potentially add bias to the analyses. Finally, since we focused on women over 65, many of them are living at senior homes or with caretakers, so census tract level averages for expenditures or income may not truly represent their individual characteristics or reflect their ability to make these choices. We still believe the methodology of augmenting patient data with area-level data to be useful, but it may be less relevant with this older patient population.

Finally, while we accounted for clustering within census tract, we did not account for the characteristics of neighboring census tracts in our analysis. Thus, we did not explore trends across segments of the county, which could be important for those who live on census tract borders and might have characteristics more similar to neighboring census-tract averages.

We also wish to highlight a methodological limitation due to the cross-sectional nature of this study. The health data for the baseline cohort, as well as the area-level data that describe the population characteristics of census tracts both capture a snapshot in time, and thus we cannot prove causation. While we can show that some of these data are associated, further longitudinal analyses are required to conclude that higher expenditures on fruits and vegetables actually lead to improved CVH.
Future Directions

We plan to analyze patient data collected one year following the baseline patient data in an identical manner to compare overall CVH and see if any improvements may be attributable to SPHERE. We are currently expanding the SPHERE application to other clinics, all the while being cognizant of local practice and norms to make this expansion successful. These additional sites will also bring SPHERE to patients of other ages, as well as males, so augmenting the area-level data and averages may prove more useful at that time. We also plan to add additional data collected at the census-tract level to this analysis, and explore clusters and patterns that emerge.

Currently, we are adding functionality in MyChart, the patient portal to our EHR, which will enable patients to fill out a short survey to populate the diet and physical activity parameters in Life’s Simple 7™. These data will flow directly into SPHERE, and be available for physicians as part of the CVH visualization. Once we have these data for some of our patients, we can start to correlate census-tract level characteristics with these intermediate factors (e.g. expenditure on fruit and vegetables to cups of fruits and vegetables consumed) and see if clear links emerge. If so, we may be able to use census-tract level factors as a proxy for physical activity and diet data in the application for patients who do not complete the survey.

Finally, one of the most important lessons learned through this work is the rapidly changing nature of clinical guidelines and practice, as demonstrated by the discrepancy in
diabetes assessment. As the healthcare landscape and EHR usage continue to evolve, we must employ an iterative process to keep SPHERE and other such applications useful. Reimbursement changes, new guidelines and technological updates are just some of the many factors that can impact physician usage of SPHERE, so in order to keep it relevant and meaningful, we are continuously evaluating and refining the application.

Conclusion

We characterized the baseline CVH of older female patients at two primary care clinics in our academic medical center. We used the AHA’s Life’s Simple 7™ campaign to characterize CVH, only slightly modifying the diabetes classification due to provider adoption of national, clinical best-practice guidelines. We enhanced these baseline data with community-level attributes, linked to the patients by census tract, in order to study associations between these factors and overall CVH. We further mapped CVH by overall category, within individual components of the score, and in conjunction with area-level data, in order to visually comprehend the geographic distribution of CVH in our patient cohort. This work paves the way for future integration of community and EHR-based data, as well as novel ways to gain insight into the many factors that affect CVH and other chronic diseases.
References


Appendix A: Data Provenance for Analysis

I. SPHERE application
   a. Patient encounter – updated in real-time
      i. Blood pressure
      ii. BMI - height and weight
   b. Enterprise Data Warehouse SPHERE Data Mart – updated every 24 hours from EHR
      i. Smoking status
      ii. Total cholesterol
      iii. Hemoglobin A1c
      iv. Medications

II. Baseline Analysis
   a. Enterprise Data Warehouse
      i. All variables

III. Address Geocoding
   a. Enterprise Data Warehouse – updated every 24 hours
i. Patient address

b. US Census
   i. 2013 Tiger/Line Shapefile – updated annually

c. ArcGIS Desktop 10.1 SP1
   i. World geocoding service address locator – updated with new ArcGIS
      software versions and service packages

IV. Census Tract Characterization

a. Nielsen PrimeLocation – software license updated by Ohio Department of
   Health annually
   i. Weekly per capita expenditures on various food groups
   ii. Unemployment and education variables

b. 2008 – 2012 American Community Survey 5 year estimates – updated
   annually
   i. Median household income