Accelerating Data Value Across a National Community Health Center Network (ADVANCE):

Community Data Report

July, 2015

ROBERT GRAHAM CENTER

AAFP Center for Policy Studies

The Robert Graham Center: Policy Studies in Family Medicine and Primary Care
1133 Connecticut Avenue N.W., Suite 1100, Washington, DC 20036
Phone: (202) 331-3360 ▪ Fax: (202) 331-3374
Email: policy@aafp.org ▪ www.graham-center.org

Table of Contents

1. Background: Community Data and Patient Centered Outcomes Research

Place matters to both personal and population health. Given increasing access to population "big data" and geospatial technologies, social determinants of health have begun to shape public health and policy interventions, but there are few formal mechanisms for incorporating information about social determinants into the study of Patient Centered Outcomes Research (PCOR). With the increasing mining and aggregation of patient-level information from electronic health records, the time has arrived to integrate population and personal health data for PCOR, and at the point-of-care. For PCOR to positively impact patient health, investigators must understand the social and environmental influences on the health of patients being studied. Patient Centered Outcomes Research Institute (PCORI) efforts to create Clinical Data Research Networks (CDRNs) provide the perfect opportunity to test the aggregation not only of a patient's clinical milieu – hospital, emergency, testing and outpatient records – but also of community contextual information about each and every patient.

Recent Institute of Medicine and Office of the National Coordinator for Health Information Technology calls for the incorporation of social and contextual data into clinical recordkeeping makes the efforts we describe below even more timely and relevant.[1]

OCHIN charged the Robert Graham Center/HealthLandscape team to identify relevant and useful community and neighborhood characteristics to be included in the ADVANCE CDRN Data Warehouse (Acuere). Then based on their findings, create a written plan that can be incorporated in the ADVANCE data diversity milestone and ADVANCE longitudinal data milestone approach. The plan will be submitted and reviewed by PCORnet, the National Patient-Centered Clinical Research Network. Once PCORnet has approved the plan, the team will execute the community-level data approach by curating identified community-level data, geocoding ADVANCE patient records and appending with these community-level data, and then assisting OCHIN as they integrate the community level data. Finally, the team will design and demonstrate an application that illustrates how geographic information systems can expand patient centered outcomes research efforts via use of the Acuere tool.

1.1 The Opportunity

The Robert Graham Center & HealthLandscape

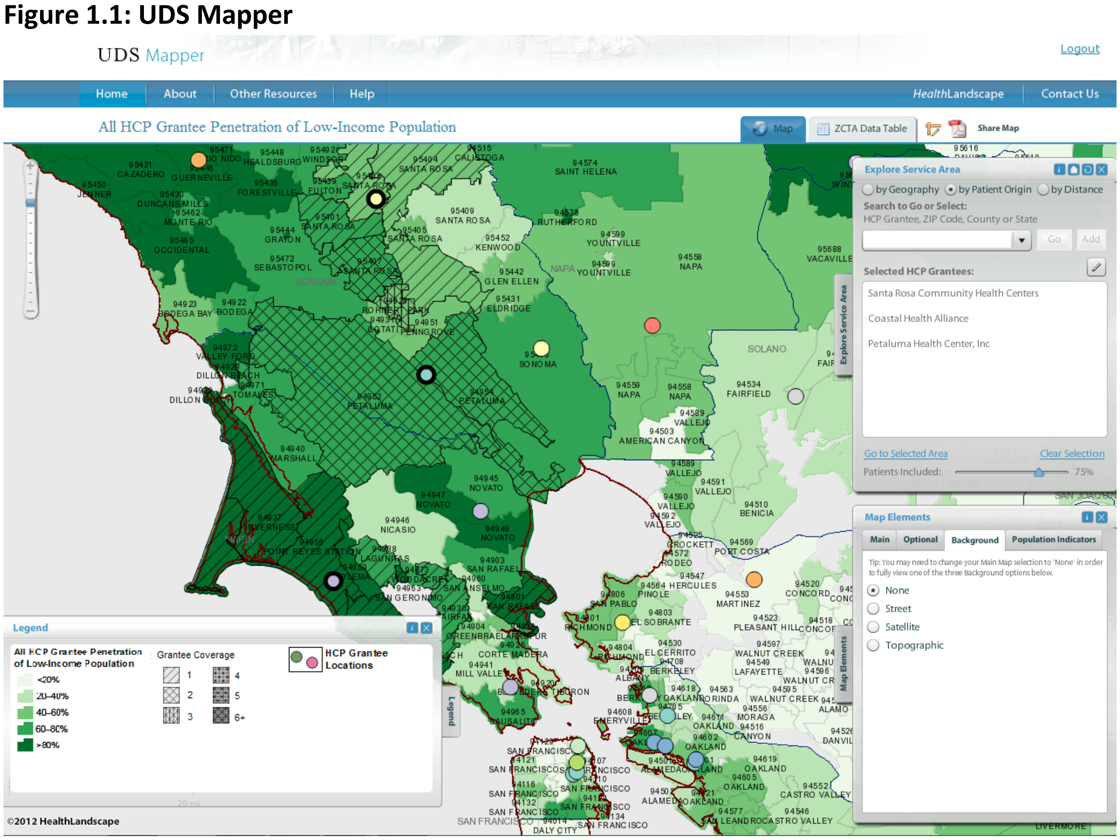
The Robert Graham Center (Graham Center) is a Policy Research Center based in Washington, DC, and an editorially independent division of the American Academy of Family Physicians. It exists to improve individual and population health by enhancing the delivery of primary care, aiming to achieve this vision through the generation or synthesis of evidence that brings a family medicine and primary care perspective to health policy deliberations from the local to international levels. In addition to original evidence making and publication, the Graham Center and its HealthLandscape partners have demonstrated the ability to build 1) online mapping platforms that integrate clinical and population data which allow users to map key patient and service area characteristics, and 2) analytic functions that reveal the richness of those data for planning purposes.

HealthLandscape is an innovation of the American Academy of Family Physicians, born in partnership with the Graham Center. In existence since 2008, HealthLandscape designs, develops, and deploys interactive web applications that allow public users, policy makers, academic researchers, and planners to combine, analyze, and display information in ways that promote understanding. HealthLandscape also acquires, catalogs, and routinely updates an extensive data library of social, behavioral, economic, and health related data as part of the HealthLandscape Platform. HealthLandscape's comprehensive data library includes over 10,000 national, regional, county, and small area measures ranging from health economics, healthcare workforce, population estimates, education, vital statistics, criminal justice, migration, healthcare quality indicators, demographics, population estimates, poverty, social environment, physical environment, mental health, and substance abuse and prevention. HealthLandscape also maintains and supports a detailed spatial database of the nation's health infrastructure and built environment.

Some of the tools built collaboratively by the Graham Center and HealthLandscape include:

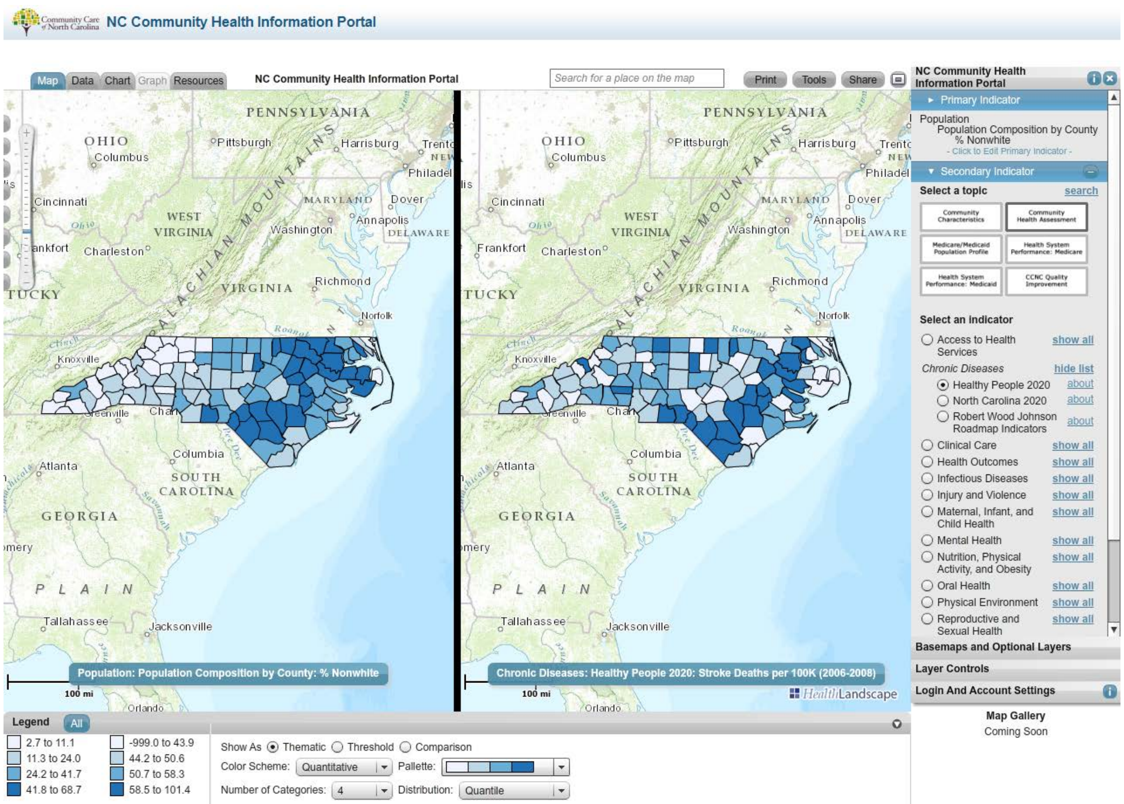
1) The UDS Mapper[2]: This tool was developed to assist HRSA, the Bureau of Primary Health Care, and health center organizations to evaluate the geographic reach, penetration, and growth of the Health Center Program and its relationship to other federally linked health resources. As of the end of June 2014, the UDS Mapper has more than 10,000 registered users from a variety of backgrounds. The development of the UDS Mapper occurred at a time when federal health data were just beginning to be liberated. More importantly, the UDS Mapper brings together never before seen community-level patient reporting with community data- combining both patient and context in a single tool improving decision making regarding whether communities should apply for federal funding and support in deciding which applications the federal government should fund. The tool has been responsive to users' needs, growing in both functionality and data included.

Figure 1.1: UDS Mapper



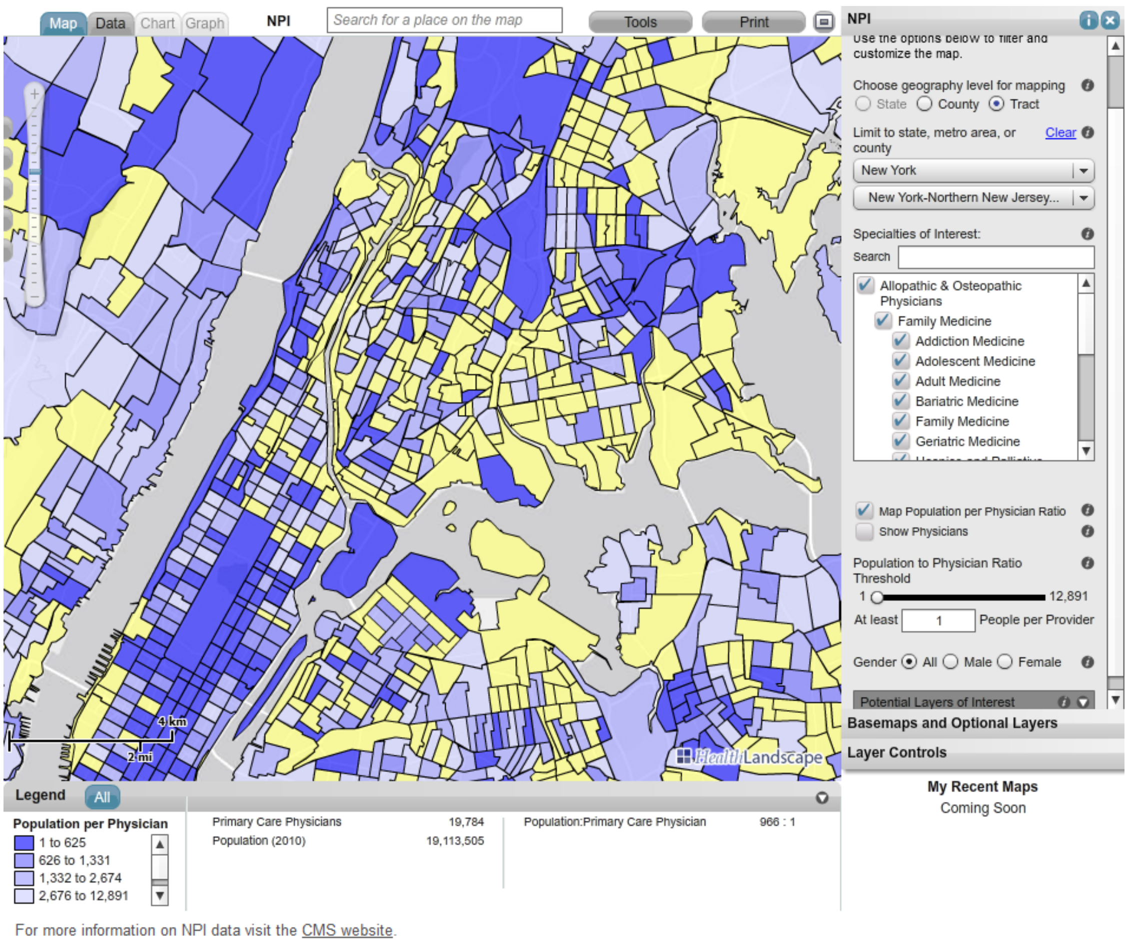
2) The North Carolina Community Health Information Portal[3]: The Graham Center's geospatial analytic staff & HealthLandscape partnered with the state Medicaid organization Community Care of North Carolina and the Southern Piedmont Beacon Community to construct the North Carolina Community Health Information Portal. Through this growing knowledge management tool, stakeholders are able to access and visualize information about Medicaid costs and utilization, social determinants of health, and population characteristics in a single data interface. The next phase of the project will integrate workforce data into this dynamic milieu to assist with prospective workforce planning by state and local stakeholders.

Figure 1.2: North Carolina Health Information Portal



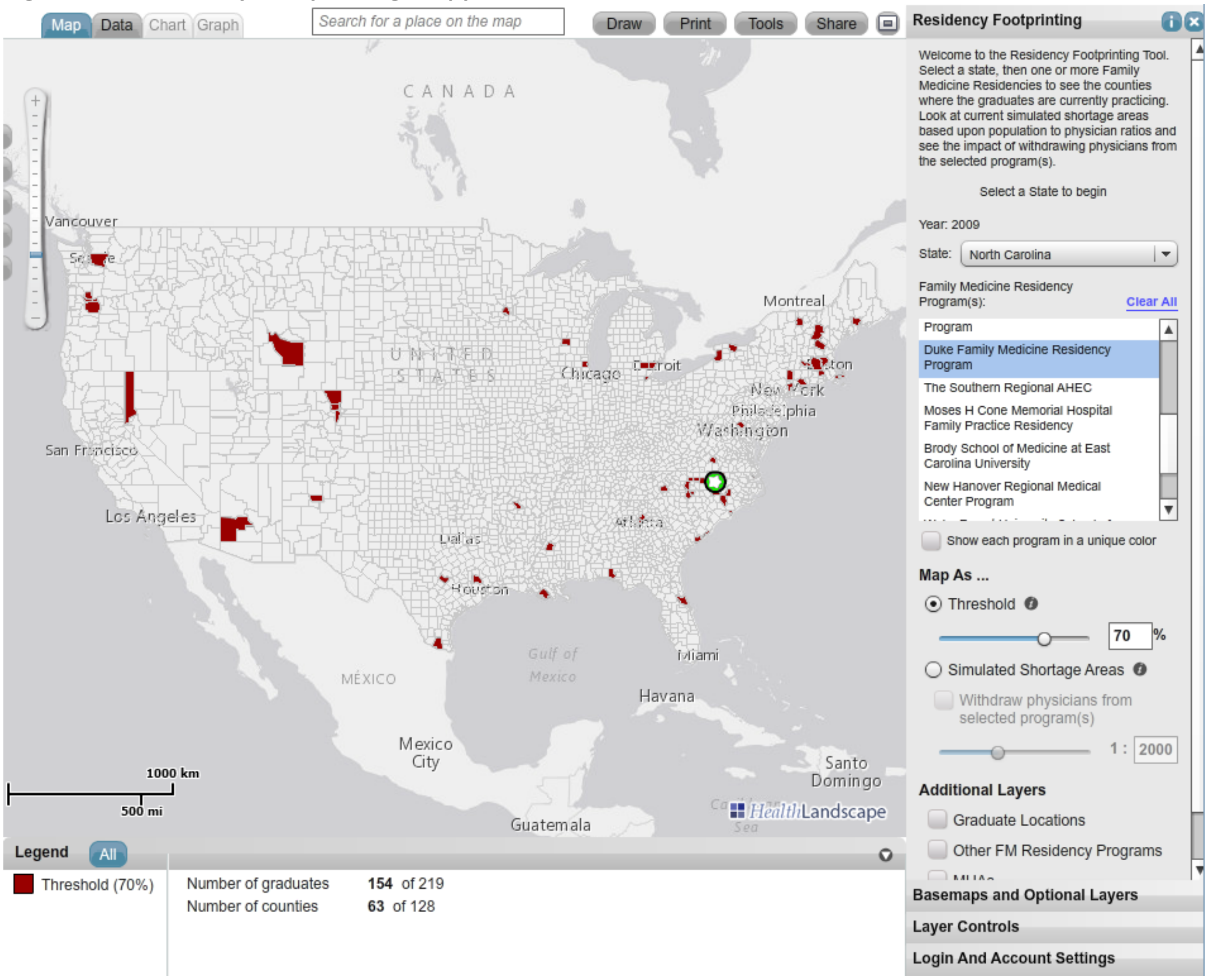
3) The Primary Care Physician Mapper[4] was completed and launched in 2013, allowing users to explore the distribution of primary care physicians by state, county, or census tracts in metropolitan areas. The mapper allows users to dynamically set the threshold at which map areas with a certain physician to population ratio are displayed.

Figure 1.3: Primary Care Physician Mapper



4) The Residency Footprinting Mapper[5] depicts the relationship between a residency program and its community, region and state by displaying where residency program graduates practice after graduation. Users can explore counties that could be considered shortage areas based on the population-to-physician ratios they have selected and see what would happen if graduates from selected programs were no longer practicing there.

Figure 1.4: Residency Footprinting Mapper



Project Aims

In this project, we will append community and social determinants data to the ADVANCE CDRN to enrich investigations of patient-centered outcomes. This effort will allow investigators to better understand clinical and specific disease-oriented service areas, clinical 'hotspots' for chronic illness, and 'coldspots' where community socio-demographics, social deprivation, and other population risk factors predict high likelihood of challenges to achieving success in patient-provider care objectives.

1.2 Community Data

The linkage between social and behavioral factors to the onset and progression of disease has been well established in the literature, as have disparities in morbidity and mortality among socioeconomic lines. The US spends more on health care than other developed nations, but still experiences higher levels of negative health outcomes. Given this observation, it is important to consider social and behavioral determinants of health when trying to visualize and affect the health of the nation overall. This is not a new approach to the public health discipline, which has long advocated that we address social determinants of health, or "circumstances in which people are born, grow up, live, work, and age, as well as the health systems they utilize."[6] The place-health link is necessarily one driven by data and as such has been driven largely in the field of public health. While traditional public health data sets and reporting are typically available at the state or county level, research has shown that data collection should be done at the individual and neighborhood level in order to truly understand the individual in the context of where they live. But not only should data about the patient be collected, they should be viewed in the context of socioeconomic factors for the neighborhood as a whole.[7]

Because of the split between public health and medical care, providers often are not taught how to view patients in the community health context, have few opportunities to engage in population health management, and for those that have an interest in putting patients into the community health context, they often lack access to ready data to support their endeavors.[8] Because of provisions in the Health Information Technology for Economic and Clinical Health (HITECH) Act and the Patient Protection and Affordable Care Act, there is renewed interest in the medical field on how to improve patient care and health outcomes while reducing costs.[9]

Many of the solutions proposed by these two Acts, not surprisingly, are data driven, calling for access to more data, but also to collect and use the data "meaningfully." In an age of access to big data, it is easy to just point to a data source as a potential solution to any given problem. However, while more and more facts are becoming available to assist with decision-making, human cognitive capacity has not similarly grown. There are billions of facts that are available potentially to a doctor at the point of care to help diagnose and plan for treatment, thousands of which are actually relevant to the situation being presented, but human cognitive capacity is limited to handling only four or five of those facts at time.[10] Therefore the addition of data to the decision making process must consider these limitations.

1.3 Community Vital Signs

We propose appending curated community-level data to patient records. The appended data would amount to community vital signs which provide context to clinicians, public health officials and researchers looking for ways to improve individual and population health outcomes.

Meaningful Data

While there are countless variables that could be considered in the depiction of this relationship between patient and community, not all variables are meaningful nor are all useful. The Centers for Medicare and Medicaid Services (CMS) is working to establish Meaningful Use Stage 3 regulations in order to expand the data contained in electronic health records beyond traditional medical information. To that end, they have tasked a committee with defining criteria for identifying "meaningfulness" in order to inform the domains, or conceptual variables, to be considered by the Office of the National Coordinator for inclusion in EHRs.

A domain is considered to be "meaningful" if it has been shown to be associated with health outcomes. The IOM report[11] on capturing social and behavioral domains lays out six criteria for domains to be considered for inclusion in EHRs. A domain (and the variables used to represent the domain) must be:

- evidence-based,
- useful,
- feasible to document and collect,
- sensitive to the patients' privacy,
- have an available, reliable, and valid measurement system, and
- not be easily accessible from other means of collection or data sources.

Domains were evaluated on their utility as well as their meaning. Usefulness was considered from three different standpoints:

- providers treating the individual patient,
- those interested in the health of the general population, and
- researchers.

In other words, domains must be considered useful for clinician decision making, describing and monitoring population health, and informing policy decisions and clinical research.

Using this set of criteria, seventeen domains were identified as meaningful. There are five broad categories into which these domains can be categorized:

1. *Sociodemographic domains* include sexual orientation, race/ethnicity, country of origin, education, employment, and financial resource strain.
2. *Psychological domains* include health literacy, stress, negative mood and affect, and psychological assets.
3. *Behavioral domains* include dietary patterns, physical activity, nicotine use and exposure, and alcohol use.
4. *Individual-Level Social Relationship domains* include social connections and isolation, as well as exposure to violence.
5. Finally, the *Neighborhoods and Communities domain* includes information to be linked through geocoding - socioeconomic and racial/ethnic characteristics of the place in which they live.

The committee is currently working to lay out specific measures to quantify each of these domains. Including this type of rich information in an EHR can provide crucial data to those working toward improving the national health status.

The first four categories of domains listed above capture patient-reported information about the patient. The last domain is intended to capture information about neighborhoods and communities where patients live. The current project focuses specifically on the last domain - information about the geographic context of a patient's record. This is a critical addition to the EHR, and represents data that cannot be collected directly from the patient. Patients may not be able to provide specific information about their communities, but geocoding patient addresses will allow us to link place-based data at various geographic levels with the individual medical record.

While the neighborhood-level measures included in the Neighborhoods and Communities domain help to illustrate community characteristics along the lines of the types of data contained in the patient-level Sociodemographic domains category there are similarly available, community-level, geo-referenced data related to the Psychological, Behavioral, and Individual-Level Social Relationship domains. Community-level data from all four of these categories speak to the social and physical environments of a patient, all of which are linked to health outcomes. Population characteristics describe the geography using basic demographic information, including measures of age distribution and racial and ethnic composition. The social environment takes into account measures such as socioeconomic status, unemployment rate, and level of education. The physical environment would include measures such as walkability, access to recreational facilities, pollution indexes, and crime rates.

These types of place-based measures are useful to the three groups of interest described above - providers, the public health community, and health researchers. Individual clinicians could use the data in the course of treatment - knowing the neighborhood conditions could help them to individualize treatment protocols for their patients. Public health agencies can use the linked data to monitor population health by social groupings and to target communities for preventive care and awareness campaigns. Researchers can dive deeper into the data, investigating the relationship between these indicators and health outcomes and evaluating the effectiveness of health care interventions.

Feasibility

In order to have a meaningful impact, the data chosen to be included in the Community Vital signs must have certain characteristics. Therefore the social and behavioral measures that comprise the Community Vital Signs all should be:

- Population based, representing conditions at the community level and not at the programmatic or clinical level;
- Valid measures of concepts broadly outlined in IOM Social and Behavioral Domains;
- Easily understandable to both patient and clinical practitioners;
- Produced and disseminated by a trusted, reputable source;
- Available consistently over time, at intervals no greater than every five years; and
- Population or patient health outcomes that can be improved via provided public health interventions or clinically actionable recommendations.

Availability

We already have access to a multitude of variables that are meaningful and feasible. Data available through the HealthLandscape Community Vital Signs Geocoding API (described in detail below) have been culled from multiple national sources and are organized by domain (see Table 1).

American Community Survey (ACS): The American Community Survey (ACS) is an ongoing survey that samples a small percentage of the population. ACS data cover a wide range of topics, including housing, education, income, poverty, and demographics. More information can be found at http://www.census.gov/acs/www/.

U.S. Department of Housing & Urban Development (HUD), Neighborhood Stabilization Program (NSP): HUD's Neighborhood Stabilization Program data include indicators for the percent of foreclosure starts and the percent of vacant addresses. More information can be found at http://www.huduser.org/portal/datasets/NSP.html.

CDC's National Environmental Public Health Tracking Network: The Centers for Disease Control and Prevention's National Environmental Public Health Tracking Network contains data from a variety of sources, including the CDC and EPA. The Tracking Network contains data on a wide range of topics, including air quality, toxic substances, health conditions, and community design. More information can be found at http://ephtracking.cdc.gov/showHome.action.

Robert Wood Johnson County Health Rankings: RWJ County Health Rankings measure the health of nearly all counties in the nation and rank them within states. The Rankings contain data from a variety of sources, including the EPA's Safe Drinking Water Information System (SDWIS). More information can be found at http://www.countyhealthrankings.org/our-approach.

U.S. Census Bureau's County/ZIP Code Business Patterns: County Business Patterns (CBP) is an annual series that provides county and ZIP Code economic data, including the number of establishments by industry (NAICS codes). More information can be found at http://www.census.gov/econ/cbp/.

The USDA Food Access Research Atlas: The Food Access Research Atlas provides census tract level data for a variety of indicators related to access to healthy food. More information can be found at http://www.ers.usda.gov/data-products/food-access-research-atlas/documentation.aspx.

The USDA Food Environment Atlas: The objective of the Food Environment Atlas is to provide county-level data for a variety of indicators related to the food environment, including access to supermarkets and groceries, food and nutrition assistance programs, and community characteristics. More information can be found at http://www.ers.usda.gov/foodatlas/.

The Center for Disease Control & Prevention's Modified Retail Food Environment Index (mRFEI): The mRFEI measures the number of healthy food retailers divided by the total number of food retailers by census tract. Higher values for the mRFEI indicate better access to healthy food options. More information can be found at http://www.cdc.gov/obesity/resources/reports.html.

The Dartmouth Atlas of Health Care: The Dartmouth Atlas provides access to a wide variety of Medicare indicators, including those related to Primary Care Access & Quality. These indicators are useful for showing how Medicare indicators vary geographically across the U.S. More information can be found at http://www.dartmouthatlas.org/.

The Centers for Medicare & Medicaid Geographic Variation Dashboard: The Centers for Medicare & Medicaid released a Geographic Variation public use file that contains a wide range of Medicare indicators on the following topics: demographics, spending, utilization, and quality indicators for the Medicare fee-for-service population. More information can be found at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/.

Census Bureau 2012 Urban/Rural Classification: The Census Bureau's urban-rural classification is a system that delineates geographical areas into two types of urban areas or as a rural area. To qualify as an urban area, the territory identified according to criteria must encompass at least 2,500 people, at least 1,500 of which reside outside institutional group quarters. The Census Bureau identifies two types of urban areas: Urbanized Areas (UAs) of 50,000 or more people; Urban Clusters (UCs) of at least 2,500 and less than 50,000 people. “Rural” encompasses all population, housing, and territory not included within an urban area. More information can be found at https://www.census.gov/geo/reference/ua/urban-rural-2010.html.

The Agency for Toxic Substances & Disease Registry’s (ATSDR) Geospatial Research, Analysis & Services Program (GRASP): GRASP created the Social Vulnerability Index (SVI) to help identify vulnerable communities that will most likely need support before, during, and after a hazardous event. More information can be found at http://svi.cdc.gov/.

The United States Department of Agriculture (USDA) Economic Research Service (ERS) Rural-Urban Continuum Codes: The ERS’s 2013 Rural-Urban Continuum Codes delineates metropolitan counties by population size (of their metro area) and nonmetropolitan counties by the degree of urbanization and their location relative to metro areas. More information can be found at http://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation.aspx.

Based on the IOM Report (cite) and data availability, indicators were grouped into seven domains (Table 1). A few of the indicators are available at the specified geographies but will require additional downloading, calculations, and/or formatting. These include the Residential Segregation indicators in the Neighborhood Race/Ethnic Composition Domain and the Foreclosure indicator in the Neighborhood Economic Conditions Domain.

Table 1 lists the domains that have data available for the listed geographies.

Table 1.1: Domains

Domain	Example of types of data	Geography
Neighborhood Socioeconomic Composition	Income, Poverty, Unemployment	County, ZCTA, census tract
Neighborhood Race/ Ethnic Composition	Black, White, Hispanic	County, ZCTA, census tract
Neighborhood Economic Conditions	Inequality, Uninsured	County, ZCTA, census tract
Environmental Exposures	Air pollution, water quality	County
Built Environment	Land use Population density	County Census tract
Neighborhood resources	Access to parks, recreational facilities Access to healthy foods	County Census tract
Medicare **Preventive Care** **Clinical Care** **Hospital Utilization**		County

2. Geocoding

2.1 Definitions

Geocoding is the assignment of an identifier, most frequently a numeric code, to a geographic location. For example, one might take a patient record and, using the patient's home address, determine and append the latitude and longitude (or X and Y coordinate) of that location to it.

Once a geocode is assigned, others may be figured out and appended as well. For example, using the patient's latitude and longitude, one could determine the census block group, ZIP Code, metropolitan area, zoning area, school district, county, state and many other identifiers associated with that address. Affixing any of these to the patient's record is a form of geocoding.

2.2 The Process of Geocoding

In the US there are numerous geocoding databases that have addresses matched to latitude and longitude coordinates. These databases generally normalize the addresses by putting everything in capital letters and removing any extra spaces. Once this cleaning process has been completed, it will try to match to the City and State, ZIP Code, and Street Name, and Street Address. Some geocoding databases will even try to match to the rooftop of the structure at the street address where available. If it cannot match to the street address it will match to the street name, or segment of the street where it believes that address would exist if it were in the database. If the street name cannot be matched then it will try and match to the longitude and latitude of the geographic centroid of the ZIP Code +4 extension; if not the ZIP+4, it will match to the geographic centroid of the five-digit ZIP Code. In some instances if it cannot find a match to the ZIP Code, it will match to the geographic centroid of the City and State.

2.3 Geographic Considerations

To effectively use contextual information in PCOR, it is important to first know more about the range of small area geographies found in publically-available data (e.g., census geographies) as well as how they interact with the postal addresses and geographies most typically associated with patient records. In a clinical setting, data must be provided in a geography that is understandable and meaningful to the patient. In a research setting, it is important to find a common geography between variables in order to compare them.

Most national health outcomes data sets are available at the state or county level, but methodologies have been pioneered to impute data from larger geographies to smaller geographies. These methods allow more flexibility in data availability and more freedom in choosing a common geography. However, limitations exist and the common geography that is selected can increase or reduce these limitations.

On the other hand, patient data are available at the address level. Addresses can be geocoded and then be aggregated to almost any geography, including state or county, or a variety of other, smaller and perhaps more community-context-relevant geographies.

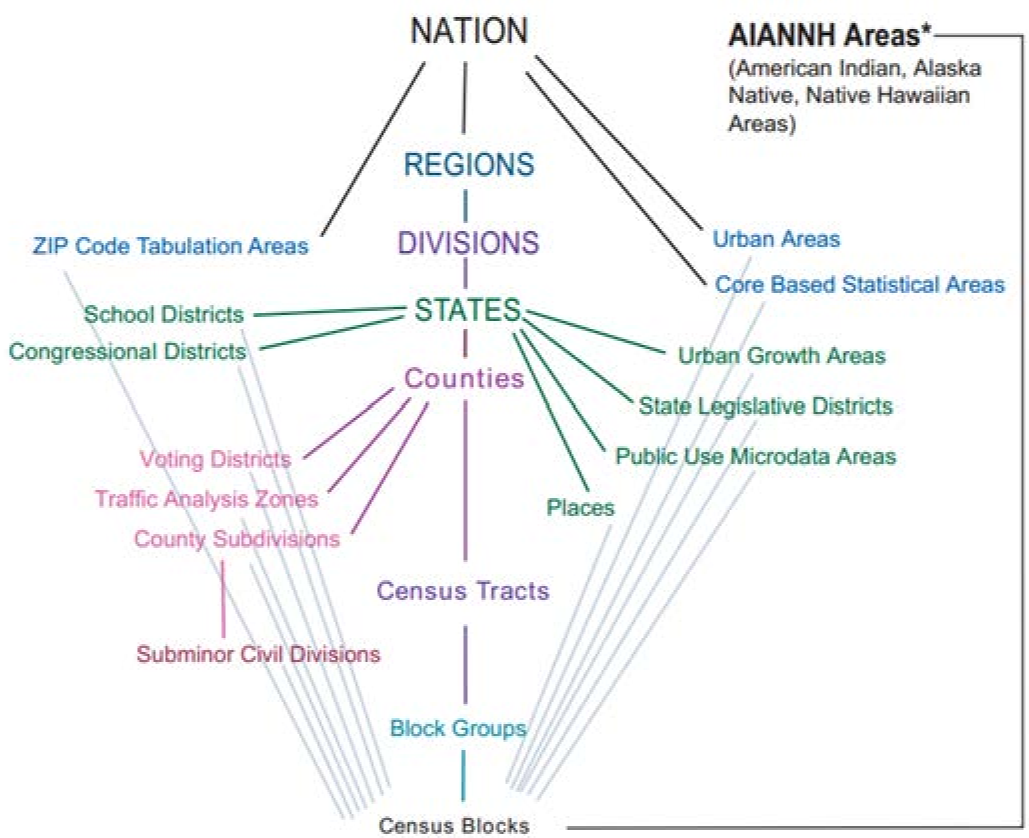
US Census Geographies

One of the main considerations for both display purposes and interpolation methods is how each level of geography interacts with the others. In Figure 2.1, from the US Census Bureau, the US Census hierarchy details how Census geographies relate to each other. The middle line in the diagram shows the most commonly used census geographies, with the largest at the top and the smallest at the bottom.

Some of the lesser known, but policy relevant, geographies appear on the outer edges of the diagram. It is important to note that these geographies align with and respect the boundaries of the parent geography and the census block, only. They do not always respect the boundaries of the smaller geographies between the parent and census block. For example, a congressional district cannot cross state lines, but it can cross county boundaries.

The base Census geography from which every other geography is built is the census block. Unfortunately, the only data reported at this level are population counts and housing units. In order to gain additional information on income, race and ethnicity and a variety of other data collected from the US Census, you have to look at the next level, the census block group.

Figure 2.1: US Postal and Census Geography



Prior to 2005, every ten years the US Census collected both the standard form, including population counts and household information, and the long form. The long form, collected from approximately one in every thirty households, contained detailed racial and ethnic breakdowns, income and other socioeconomic variables. In 2005, the US Census began the American Community Survey to replace the long form. This survey reaches fewer households in a given year, but over a five-year period, claims to reach approximately one in every twenty households. This sample allows the Census to create estimates down to the block group for the more detailed information.

Patient Geography

Patient data are recorded and reported in different ways depending on the setting where the patient is seen. In most cases, a home and a work address are requested, but the information provided and recorded is influenced by the patient providing the information and the medical administrators inputting the information into the patient record. If an accurate address is collected and recorded, then the address can be geocoded and then be aggregated to any geographic level.

Caveats/Limitations

Geocoding can be a challenging process due to the fact that you have multiple levels of potential errors. First, you are dealing with multiple layers of potential human error, and even in cases where there is no human error, geocoding databases themselves may introduce error.

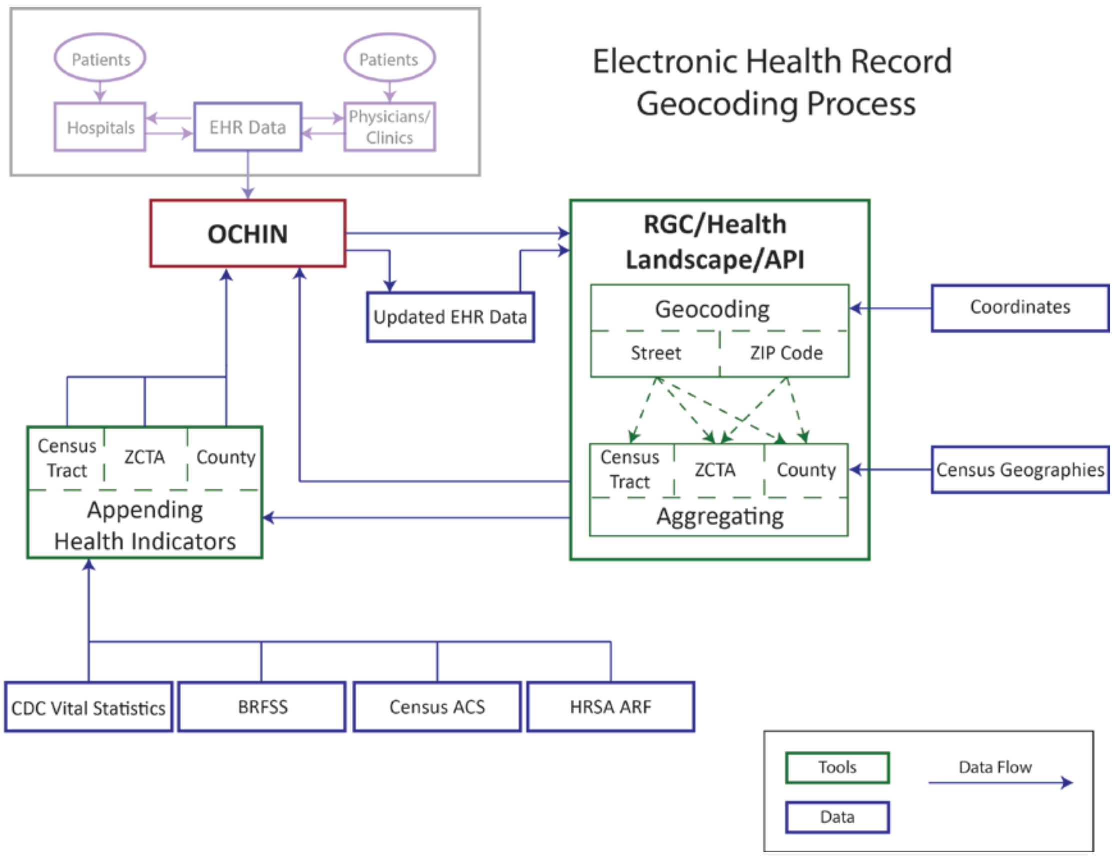
There are at least two people involved in the reporting and recoding of addresses. These are the first two levels where error can be introduced. If I live at 1234 East Main Street (valid, correct address), I might provide that address as 1234 E. Main (valid, but incomplete address), because that is how I am used to saying it. If there is a 1234 E. Main Ave, or 1234 E. Main Terrace, the geocoding database will have a hard time deciding where that address should be assigned. Also, if the person who is recording the address isn't familiar with the area they could easily make a mistake and drop the E. or record the E. after the street name (valid, correct address reported; incomplete or invalid address recorded). Those are just a few examples of how there could be inconsistencies in the reporting and recording of the addresses. Additionally, the geocoding databases were compiled by the collection of data by computers and humans and sometimes there are multiple entries in the databases for each address. Sometimes these entries are extremely similar, but perhaps one is on the left side of the street and one is one the right side of the street. In many cases this can be the difference between being in one census tract, or in another. Also, as described above, by default the geocoding process does everything to find a match and often they are matched to the center of larger geographies such as ZIP Codes, cities, or states. This may place the point in a location that is not where the patient actually resides.

For this and any geocoding project, the limits of acceptable geocodes must be set. In this project, it is recommended that records that have been geocoded to the street address be aggregated to the county, census tract and ZCTA, but as a result of the above listed limitations, it is recommended that the addresses geocoded to the ZIP Code level only be matched to the county and ZIP Code or ZCTA level. In certain cases, even aggregating to geographies larger than a ZIP Code could result in errors since ZIP Codes and their Census equivalent, ZIP Code Tabulation Areas, do not respect county, or even state, boundaries. In the case of a ZIP Code match, if you are applying attributes that are assigned at a smaller, or different, geography than ZIP Code, you may be introducing spatial error that can skew your results. For example, with a ZIP Code match, the patient address point will use coordinates of the geographic center of the ZIP Code. If aggregating points to census tracts, that point will simply be placed in the census tract closest to the center of the ZIP code area. As a result, it will appear that there are more records in that census tract than in reality. In relation to patient data, this occurs when a number of Post Office Box addresses are reported.

Through interpolation methods, variables reported at larger geographies, such as state or county, can be interpolated and estimated at lower geographic levels, such as ZIP Code, ZCTA or census tract. When deciding at which level certain variables should be displayed and used for analysis, it is a good practice to determine what geography will best address the questions you are asking. In many cases, the best answer is to choose the smallest level of geography that is available, but due to imputation and geocoding inconsistencies described above, accuracy should also be considered. In some cases approximately 20 percent of the addresses received and recorded for patients will only be geocodable to the ZIP Code level and therefore can skew results when comparing those data to any smaller geographies. For the remaining 80 percent of the patients, you can consider comparing them to smaller geographies such as census tracts. However, due to geocoding and census geography nuances, the point may be placed in the wrong geography. These inconsistencies can be reduced by considering larger geographies such as census tracts.

One consideration for working with data at various Census geographies is the margin of error potentially introduced through the survey and further imputation. Most data published at the census block group also provide margins of error. The margin of error is most likely spread and increased after imputing the data to other geographies.

2.4 ADVANCE Geocoding process



2.5 Online Geocoding Demonstration Tool

The HealthLandscape Geocoding API provides web services for address geocoding, assigning geographic identifiers, and appending the pool of Community Vital Signs. Services support both HTTP and HTTPS protocols, as well as Cross-Origin Resource Sharing (CORS). All API methods return JavaScript Object Notation (JSON).

The HealthLandscape Geocoding API platform is built on a combination of Linux and Windows Servers, Python scripting, and ArcGIS geocoding processing.

Figure 2.2: Geocoding API Steps

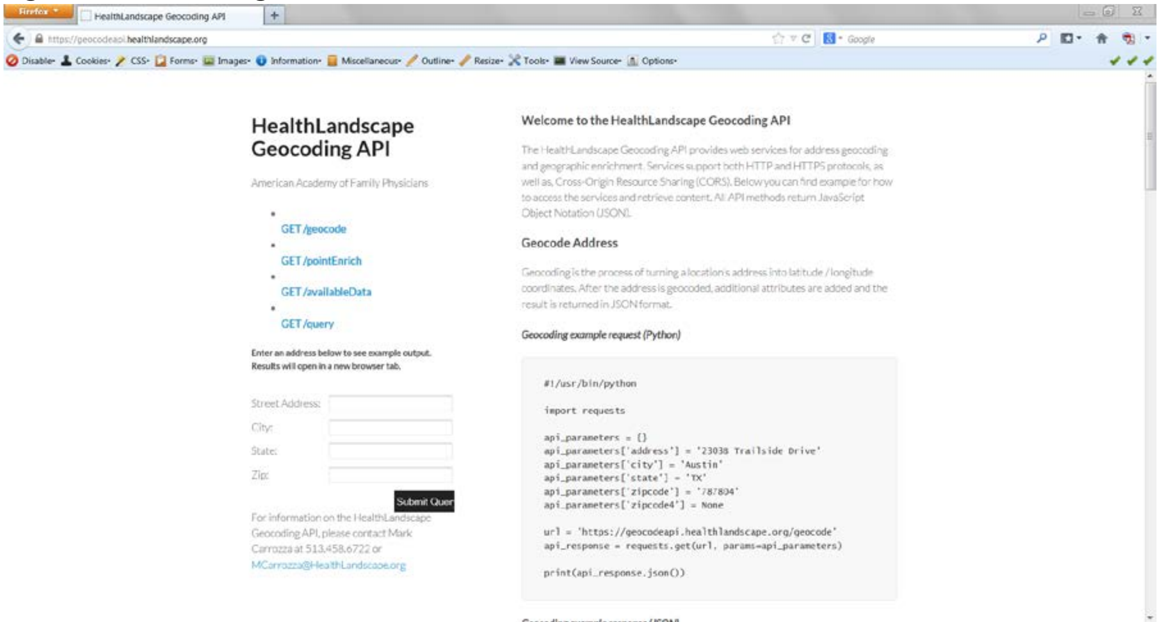
OCHIN	1 → 4 ←	**Geocoding API** Ubuntu Linux, Nginx web server, Python/Gunicorn WSGI server	2 → 3 ←	**ArcGIS Geocoding Services** Windows Server, IIS, ArcGIS Server, NavTEQ Geocoding Reference
All network communications protected by 2048-bit encryption				

The Geocoding API (Figure 2.2) performs four broad steps.

1. The initiating system (OCHIN) submits a JSON string that includes a unique identifier, street, city, state, and ZIP Code;
2. The API routes the address to the geocoding service;
3. The geocoding service returns an x/y (latitude and longitude) coordinate pair to the API;
4. The API performs a spatial join to determine geographic identifiers (county, tract, ZCTA, HPSA, etc.) and appends community indicators to the record, before returning the unique identifier and collection of data to the initiating system as a JSON string.

In development and testing, the API performs all geocoding and data assignment tasks in slightly less than one second per request. The API has been designed to accommodate growth, by taking advantage of cloud computing and the availability of redundant load-balanced servers which can be deployed rapidly as needed.

Figure 2.3: Geocoding API Interface



The PROTOTYPE HealthLandscape Geocoding API is available for testing and review at https://geocodeapi.healthlandscape.org

We DO NOT recommend using this PROTOTYPE until we've completed tested the reliability, validity, and security of this application.

3. Next Steps

3.1 Geocoding OCHIN data

Preliminary Tests

OCHIN and HealthLandscape conducted a preliminary test of data transfer and geocoding using addresses representing a fictitious patient population. A total of 409 addresses were transferred from OCHIN to HealthLandscape. Table 3.1 shows the results of the geocoding process. As would be expected from unstandardized addresses found in patient registration systems, 300 (or 73%) of the addresses were successfully geocoded at the Street Address level. Ninety-five (23% of the addresses were geocoded to either the 5 or 5+4 ZIP Code area. These addresses were assigned the centroid of the ZIP polygon and are likely less accurate than street-level geocoding. Thirteen (three percent) addresses were geocoded to the street mid-point, which has the potential to be inaccurate and highly misleading, especially for streets that extend for many miles. Table 3.1 includes results for rural, suburban, and urban areas. These results will have implications for the types of data that would be appropriate for appending to each record, with less data being recommended for less accurate geocodes.

Table 3.1: Results of Geocoding Process

Match Type**	Rural	Urban Cluster of at least 2,500 and less than 50,000 people	Urbanized Areas of 50,000 or more people	Unmatched	Grand Total
Postal	25	18	34		77
PostalExt	1	1	16		18
StreetAddress	38	57	205		300
StreetName	3	2	8		13
Unmatched				1	1
Grand Total	67	78	263	1	409

** StreetAddr -- Street address, such as 320 Madison St, that represents an interpolated location along a street given the house number within an address range. StreetName -- Street name only, such as Orchard Road. The street name feature may be a feature of many street segments chained together based on the name. The geocoded location is usually placed on the middle of the street feature. Postal -- Basic postal code, such as 60610. PostalExt -- Full postal code including its extension, such as a ZIP+4 code—91765-4383.

Based on these preliminary results, which included a number of purposefully false and nonsensical addresses, OCHIN and HealthLandscape decided to continue the geocoding process with the full, authentic patient address from the ADVANCE system.

Based on the number of records and the desire for maximum efficiency, OCHIN and HealthLandscape decided to split the geocoding function into two tasks, a manual processing of the initial 1.7 million records, and the development of an automated geocoding API designed to allow single-record geocoding and data enrichment. The geocoding API will be used for ongoing geocoding needs.

Manual Geocoding

HealthLandscape has received 1,745,874 distinct patient addresses derived from the ADVANCE system. The data are being processed for geocoding. Each record includes two street address fields, and city, state and ZIP Code information. Detailed review and data management has been completed on the original data. Specifically, we've interrogated each record to assure that the street address that is used for geocoding includes an actual street address, rather than a PO Box or other non-residential location. For example, there are multiple occurrences of address records with "RETURNED MAIL", "PLZ UPDATE", "BAD ADDRESS", and other clerical annotations, which are unable to be geocoded. By systematic review of the two address fields, we can assure that the most usable information is being included in the geocode process.

The manual process has resulted in 80% Street Address match rate, which is typical for such applications. The remainder of the addresses match as shown in Table 3.2 below.

Table 3.2: Geocoding Match by Type

Match Type **	Number	Percent
PostalExt	5,642	0.3
StreetAddress	1,394,806	79.9
StreetInt	28	0.0
StreetName	63,931	3.7
SubAdmin	259,558	14.9
Unable to Geocode / Incomplete Address Information	21,909	1.3
	1,745,874	

** StreetAddr—Street address, such as 320 Madison St, that represents an interpolated location along a street given the house number within an address range. StreetName—Street name only, such as Orchard Road. The street name feature may be a feature of many street segments chained together based on the name. The geocoded location is usually placed on the middle of the street feature. SubAdmin—A local administrative area, such as a city. PostalExt—Full postal code including its extension, such as a ZIP+4 code—91765-4383. StreetInt—Intersection address that contains an intersection connector, such as Union St & Carson Rd.

After geocoding and appending geographic identifiers (state, county, tract, etc.) to each record, community characteristics were also appended to each record. The complete list of community measures is shown in Appendix A.

3.2 Extending these ideas to other PCORI CDRNs

The steps noted above may be adapted to any of PCORIs 11 CDRNs. The API has been designed to accommodate growth, by taking advantage of cloud computing and the availability of redundant load-balanced servers which can be deployed rapidly as needed. Data that make up the Community Vital Signs data pool are from reliable, replicable national sources, allowing consistent measures across all PCORI CDRNs.

HealthLandscape will be licensing access to the Geocoding API at competitive pricing compared to similar services. PCORI CDRNs interested in using the HealthLandscape Geocoding API will be able to engage with HealthLandscape and the Graham Center, following the same general steps as OCHIN: initiating a Geocoding API services agreement, exchanging necessary data use and HIPAA business associate agreements, bulk or batch geocoding of an initial pool of records, and accessing the API for single-record geocoding and data enrichment.

3.3 Future Steps

Phase 2 Development

The *Community Vital Signs Data Library API* will be a complimentary system to the Geocoding API described above. Geocoding of the clinical record, and the addition of the detailed geographic identifiers, is a necessary step before taking advantage of the Community Vital Signs Data Library API we're discussing in this section.

The *Community Vital Signs Data Library API* will be designed to give researchers broader access to the rich collection of data that is available in HealthLandscape. HealthLandscape and the Graham Center have been regularly and systematically acquiring social, behavioral, economic, and health data from multiple national, state, and local sources for almost a decade. HealthLandscape's comprehensive data library includes nearly 10,000 national, regional, county, and small area measures ranging from health economics, healthcare workforce, population estimates, education, vital statistics, criminal justice, migration, healthcare quality indicators, demographics, population estimates, poverty, social environment, physical environment, mental health, and substance abuse and prevention. HealthLandscape also maintains and supports a detailed spatial database of the nation's health infrastructure and built environment.

The *Community Vital Signs Data Library API* will be designed to allow researchers to append custom extracts of this complete data pool to their deidentified clinical records, to study the impact of social and behavioral factors on functional status and the onset and progression of disease. Detailed, expanded metadata will allow PCORI CDRN's to more easily replicate and extend research using Community Vital Signs by flagging measures as they included in research, allowing complimentary research on similar panels or alternate theories to use the identical pool of community measures as controls. For example, a series of research protocols would be able to sequentially examine differing age/race/sex panels of diabetic or asthma patients, but consistently take advantage of identical Community Vital Signs from the pool of possible indicators.

Clinical Vital Signs

The Graham Center, OCHIN and ADVANCE envision the eventual deployment of community health data into primary care practice Electronic Health Records (EHRs), achieving meaningful use in a HIPAA-compliant, secure fashion. Eventually, providers in all OCHIN-ADVANCE networks would be able to access 'Community Vital Signs', allowing providers and administrators to better understand the characteristics of patients in the context of community population distribution by age, gender, race, ethnicity, poverty level and insurance source, quality of care, health outcomes and disparities, diagnosis types, and other diagnostic relationships for local and regional characteristics.

By linking these tools with the expansive national data resources already captured in the HealthLandscape platform, providers would also be able to take advantage of rapid geocoding and visualization of their own additional data, see what health care providers were available in a patient's immediate area, visualize community health indicators from federal and state datasets. Together, the potential of these tools to impact decision-making and PCOR is immense.

Health centers providers and their patients could use these tools to answer questions such as

- What are the community health risk factors facing patients like the one I'm seeing based on a variety of population measures, and aggregate quality metrics for all patients in his/her area?
  - How does this compare to peers across all OCHIN sites?
- Where are a certain population of my patients clustered, what geographic and social barriers do they face in achieving optimal health and how might outreach, additional services or new strategies serve that population better?
- Where might outreach or services expansion be directed to better address my patient's need?
- What resources are available to my patient within my clinic's core service area?
- What resources are available in my neighborhood to help me achieve better health?

Community Vital Signs "Risk Factor Assessment Box" Prototype:

It's known from commercial marketing research that there may be a vast amount of information that is relevant to consumer behavior, but too much information might cloud decision making. Therefore marketers focus on information control to find the right balance of information.[12] The use of data mining techniques and business intelligence helps to find and combine data into meaningful chunks.[13] Likewise there are many bits of information that could be used by providers at the point of care. For successful integration of these data into a patient centered outcomes research database or in the medical record itself, the information will need to be controlled and to some extent aggregated into useful, meaningful chunks to improve health outcomes.

One way to "chunk" the data is to create indices based on data that are available in the patient record and data that are in the Community Vital Signs API. In order to efficiently integrate these data into clinician/patient interactions, a Social Disadvantage Index could be created for each patient and a Neighborhood Disadvantage Index Score could be calculated for each geography. Multi-dimensional scaling would summarize the measures into a single number for each index, making it easy for a physician to quickly access and process the information. In addition, individual measures on which the index greatly underperforms compared to the national average would be flagged for physician attention.

While the ability to append community characteristics and personal social connectedness measures to the clinical record is an important first step, it's also important to be able to summarize this detailed information in a way that can be integrated into the clinical encounter and guide clinically actionable recommendations.

As an initial prototype, we put forward the following Risk Factor Assessment Box as an elegant solution that quickly and efficiently summarizes the degree to which the patient shows social or community context measures that could impact the direction or success of a treatment plan.

Table 3.3: Social Disadvantage Index

Social Disadvantage Index	H			
	M			
	L			
		L	M	H
		Neighborhood Disadvantage Index		

Social Disadvantage Index Dimension. Social Risk Screening can provide valuable information during the patient-doctor encounter and standardized screening tools are being developed and integrated into the clinical record.[14] This dimension would specifically address the 2014 IOM recommendations for capturing social connectedness and social isolation.[15] The Social Disadvantage Index dimension would be constructed during the clinical encounter based on physician or nurse interviews, using a standardized social history template such as those suggested by Beck (2012).[16] Over time, these templates could be standardized and targeted towards specific sub-population (a Senior Interview, and Adolescent/Parent interview, a COHF Interview) in a way that focuses on best practices measures of social connectedness.

Neighborhood Disadvantage Index Dimension. Previous research including work conducted by the RGC, has shown that a multi-dimensional social deprivation index is positively associated with poor access and poor health outcomes.[17] The measures contained with the Community Vital Signs could be combined, through multidimensional scaling, to create a single index that captures community disadvantage.

Using the Risk Factor Assessment Box. Below are three examples of the Risk Factor Assessment Box, each showing a different risk profile for a patient. Links embedded within the Risk Factor Assessment Box would present the clinician with a detailed summary of the social and community metrics of concern, if the information was needed during the clinical encounter.

The first example (Table 3.4) shows a Risk Factor profile that is low on both social and community risk factors. This individual has multiple social support connections and resides in an area with few or limited neighborhood disadvantages.

Table 3.4: Low Resource Deficit, Low Community Deficit

Social Disadvantage Index	H			
	M			
	L			
		L	M	H
		Neighborhood Disadvantage Index		

Table 3.5 shows a sample patient with HIGH social connectedness risk and HIGH community risk. For example, a patient from a high poverty area with limited public transportation or access to pharmacies. This patient also lacks social or family support in the area to aide in access to care or adhering to a treatment plan. This combination of characteristics suggests a patient requiring additional supports for successful treatment outcomes.

Table 3.5: High Resource Deficit, High Community Deficit

Social Disadvantage Index	H			
	M			
	L			
		L	M	H
		Neighborhood Disadvantage Index		

Finally, Table 3.6 shows a sample patient with LOW social connectedness risk and HIGH community risk. For example, a patient from a high poverty area with limited public transportation or access to pharmacies. The patient's social connectedness rating, however, suggests adequate social support from family, friends, and neighbors.

Table 3.6 Low Resource Deficit, High Community Deficit

Social Disadvantage Index	H			
	M			
	L			
		L	M	H
		Neighborhood Disadvantage Index		

This prototype of a Risk Factor Assessment Box can serve as an initial foray into a simple User Interface that will quickly and intuitively summarize detailed information. We are currently completing a post-hoc simulation of this design using 500 family practice clinical encounters that include patient interviews – to create a limited Social Disadvantage Index, and geo-enriched data using the Community Vital Signs Geocoding API to allow the creation of a Neighborhood Disadvantage Index. Following the assessments of those results, we will be in a better position to discuss next steps in planning the evaluation of this tool.

Conclusions

In an era of increasing evidence about the significant impact of sociodemographic and environmental factors upon health outcomes – many argue an impact greater than the provision of medical care – it is time to integrate community and contextual data into any new Clinical Data Repository, including the Patient Centered Outcomes Research Network's Clinical Data Research Network enterprise (PCORnet CDRN). We have outlined in this report the rationale, means, and blueprint for doing exactly that, and look forward to engaging PCORnet partners and PCORI in discussions about how to scale this endeavor across the CDRN enterprise.

These efforts align with federal calls from the IOM, CMS, and ONC to further integrate clinical and public health data in ways that are ultimately them meaningful to the improvement of personal and population health.

Appendices

Appendix A: 128 Core Community Indicators

A complete detailed list of indicators can be downloaded at

http://www.healthlandscape.org/geocodeapi_listofindicators_V1.pdf

Table A.1: Core Community Indicators

Domain	Description	Source	Year	Geographic Level
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	# with Bachelor's Degree or Higher	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Median Household Income	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	Median Household Income	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Median Household Income	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Number of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	County

Neighborhood Socioeconomic Composition	Number of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Number of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Pct Below 100% of Federal Poverty Level	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	Pct Below 100% of Federal Poverty Level	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Pct Below 100% of Federal Poverty Level	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Pct Below 200% of Federal Poverty Level	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	Pct Below 200% of Federal Poverty Level	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Pct Below 200% of Federal Poverty Level	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Percent of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	County
Neighborhood Socioeconomic Composition	Percent of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Percent of persons in managerial, professional, or executive occupations	American Community Survey	2008-2012	ZCTA
Neighborhood Socioeconomic Composition	Unemployment Rate	American Community Survey	2008-2012	County

Neighborhood Socioeconomic Composition	Unemployment Rate	American Community Survey	2008-2012	Census Tract
Neighborhood Socioeconomic Composition	Unemployment Rate	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# American Indian/Alaskan Native	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# American Indian/Alaskan Native	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# American Indian/Alaskan Native	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# Asians	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# Asians	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# Asians	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# Black Alone	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# Black Alone	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# Black Alone	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# Hawaiian/Pacific Islanders	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# Hawaiian/Pacific Islanders	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# Hawaiian/Pacific Islanders	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# Hispanic	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# Hispanic	American Community	2008-2012	Census Tract

		Survey		
Neighborhood Race/Ethnic Composition	# Hispanic	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# NonWhite	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# NonWhite	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# NonWhite	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	# White Alone	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	# White Alone	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	# White Alone	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% American Indian/Alaskan Native	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% American Indian/Alaskan Native	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% American Indian/Alaskan Native	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% Asians	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% Asians	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% Asians	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% Black Alone	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% Black Alone	American Community Survey	2008-2012	Census Tract

Neighborhood Race/Ethnic Composition	% Black Alone	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% Hawaiian/Pacific Islanders	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% Hawaiian/Pacific Islanders	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% Hawaiian/Pacific Islanders	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% Hispanic	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% Hispanic	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% Hispanic	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% NonWhite	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% NonWhite	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% NonWhite	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	% White Alone	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	% White Alone	American Community Survey	2008-2012	Census Tract
Neighborhood Race/Ethnic Composition	% White Alone	American Community Survey	2008-2012	ZCTA
Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/American Indian, Alaskan Native	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/Asian	American Community Survey	2008-2012	County

Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/Black	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/Multiple Races	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/Native Hawaiian, Other Pacific Islander	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Dissimilarity - White/Other Race	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - American Indian, Alaskan Native	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - Asian	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - Black	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - Multiple Races	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - Native Hawaiian, Other Pacific Islander	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - Other Race	American Community Survey	2008-2012	County
Neighborhood Race/Ethnic Composition	Residential Segregation - Exposure - White	American Community Survey	2008-2012	County
Neighborhood Economic Conditions	Dependency Ratio (Old-Age)	American Community Survey	2008-2012	County

Neighborhood Economic Conditions	Dependency Ratio (Old-Age)	American Community Survey	2008-2012	Census Tract
Neighborhood Economic Conditions	Dependency Ratio (Old-Age)	American Community Survey	2008-2012	ZCTA
Neighborhood Economic Conditions	Dependency Ratio (The age dependency ratio is derived by dividing the combined under 18 years and 65 years and over populations by the 18-to-64 population then multiply by 100)	American Community Survey	2008-2012	County
Neighborhood Economic Conditions	Dependency Ratio (The age dependency ratio is derived by dividing the combined under 18 years and 65 years and over populations by the 18-to-64 population then multiply by 100)	American Community Survey	2008-2012	Census Tract
Neighborhood Economic Conditions	Dependency Ratio (The age dependency ratio is derived by dividing the combined under 18 years and 65 years and over populations by the 18-to-64 population then multiply by 100)	American Community Survey	2008-2012	ZCTA
Neighborhood Economic Conditions	Dependency Ratio (Young)	American Community Survey	2008-2012	County
Neighborhood Economic Conditions	Dependency Ratio (Young)	American Community Survey	2008-2012	Census Tract
Neighborhood Economic Conditions	Dependency Ratio (Young)	American Community Survey	2008-2012	ZCTA

Neighborhood Economic Conditions	Estimated percent of foreclosure starts over the past 18 months through June 2008	HUD, NSP	2008	County
Neighborhood Economic Conditions	Estimated percent of foreclosure starts over the past 18 months through June 2008	HUD, NSP	2008	Census Tract
Neighborhood Economic Conditions	Estimated Percent of vacant addresses in June 2008 (90-day vacancy rate)	HUD, NSP	2008	County
Neighborhood Economic Conditions	Estimated Percent of vacant addresses in June 2008 (90-day vacancy rate)	HUD, NSP	2008	Census Tract
Neighborhood Economic Conditions	GINI - Inequality	American Community Survey	2008-2012	County
Neighborhood Economic Conditions	GINI - Inequality	American Community Survey	2008-2012	Census Tract
Neighborhood Economic Conditions	GINI - Inequality	American Community Survey	2008-2012	ZCTA
Neighborhood Economic Conditions	Overall Percentile Ranking for SVI	Agency for Toxic Substances and Disease Registry (ATSDR)	2006-2010	County
Environmental Exposures	Median Housing Structure Age	American Community Survey	2008-2012	County
Environmental Exposures	Median Housing Structure Age	American Community Survey	2008-2012	Census Tract
Environmental Exposures	Median Housing Structure Age	American Community Survey	2008-2012	ZCTA

Environmental Exposures	Number of person-days with maximum 8-hour average ozone concentration over the National Ambient Air Quality Standard (monitor and modeled data)	EPA, EPHTN	2008	County
Environmental Exposures	Number of person-days with PM2.5 over the National Ambient Air Quality Standard (monitor and modeled data)	EPA, EPHTN	2008	County
Environmental Exposures	Percent of Occupied Housing Units Without Complete Plumbing Facilities	American Community Survey	2008-2012	County
Environmental Exposures	Percent of Occupied Housing Units Without Complete Plumbing Facilities	American Community Survey	2008-2012	Census Tract
Environmental Exposures	Percent of Occupied Housing Units Without Complete Plumbing Facilities	American Community Survey	2008-2012	ZCTA
Environmental Exposures	Percent of population potentially exposed to water exceeding a violation limit during the past year	EPA, SDWIS	2012-2013	County
Built Environment	Fast Food Restaurants per 100,000 population (NAICS 722513)	U.S. Census Bureau, County Business Patterns	2012	County
Built Environment	Fast Food Restaurants per 100,000 population (NAICS 722513) - re-coded to ZCTA level	U.S. Census Bureau, ZIP Code Business Patterns	2012	ZCTA

Built Environment	Liquor Stores per 100,000 population (NAICS 445310)	U.S. Census Bureau, County Business Patterns	2012	County
Built Environment	Liquor Stores per 100,000 population (NAICS 445310) - re-coded to ZCTA level	U.S. Census Bureau, ZIP Code Business Patterns	2012	ZCTA
Built Environment	Population Density	American Community Survey	2008-2012	County
Built Environment	Population Density	American Community Survey	2008-2012	Census Tract
Neighborhood Resources	Fast Food Restaurants per 100,000 population (NAICS 722513)	U.S. Census Bureau, County Business Patterns	2012	County
Neighborhood Resources	Low access tract at 1 mile for urban areas or 10 miles for rural areas	USDA Food Access Research Atlas	2010	Census Tract
Neighborhood Resources	Low access tract at 1/2 mile for urban areas or 10 miles for rural areas	USDA Food Access Research Atlas	2010	Census Tract
Neighborhood Resources	Metro/NonMetro Classifcation Codes	USDA, ERS	2013	County
Neighborhood Resources	Modified Retail Food Environment Index (# of Healthy Food Stores divided by All Food Stores)	Centers for Disease Control and Prevention	2008	Census Tract
Neighborhood Resources	Percent of people in a county living more than 1 mile from a supermarket or large grocery store if in an urban area, or more than 10 miles from a supermarket or large grocery store if in a rural area	USDA Food Atlas	2010	County

Neighborhood Resources	Percentage of population living within half a mile of a park	CDC, EPHTN	2010	County
Neighborhood Resources	Recreation Facilities per 100,000 population (NAICS 713940) - re-coded to ZCTA level	U.S. Census Bureau, ZIP Code Business Patterns	2012	ZCTA
Neighborhood Resources	Urban Classification Code - Rural, Urban Cluster (greater than 10,000 population, less than 50,000 population), Urban Area (greater than 50,000 population)	U.S. Census Bureau; USDA Food Access Research Atlas	2010	Census Tract
Clinical Care	Average annual percent of diabetic Medicare enrollees age 65-75 having blood lipids (LDL-C) test	Dartmouth Atlas	2012	County
Clinical Care	Average annual percent of diabetic Medicare enrollees age 65-75 having eye examination	Dartmouth Atlas	2012	County
Clinical Care	Average annual percent of diabetic Medicare enrollees age 65-75 having hemoglobin A1c test	Dartmouth Atlas	2012	County
Hospital Utilization	% readmissions within 30 days of hospital visit	Centers for Medicare & Medicaid (Geographic Variation)	2012	County
Hospital Utilization	Discharges for ambulatory care sensitive conditions per 1,000 Medicare enrollees	Dartmouth Atlas	2012	County

Hospital Utilization	ED Visits per 1,000 Enrollees	Centers for Medicare & Medicaid (Geographic Variation)	2012	County
Preventive Care	Average annual percent of Medicare enrollees having at least one ambulatory visit to a primary care clinician	Dartmouth Atlas	2012	County
Preventive Care	Average percent of female Medicare enrollees age 67-69 having at least one mammogram over a two-year period	Dartmouth Atlas	2012	County

References

[1] Institute of Medicine (US). Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records. Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1. 2014.

[2] UDS Mapper: http://www.udsmapper.org

[3] NC-HIP: https://nchip.n3cn.org

[4] Primary Care Physician Mapper: http://www.graham-center.org/rgc/maps-data-tools/interactive/primary-care-physician.html

[5] Residency FootPrinting Mapper: http://www.graham-center.org/rgc/maps-data-tools/interactive/residency-footprinting.html

[6] Centers for Disease Control and Prevent (CDC), Social Determinants of Health. http://www.cdc.gov/socialdeterminants/

[7] Krieger, Nancy, David R. Williams, and Nancy E. Moss. "Measuring social class in US public health research: concepts, methodologies, and guidelines." Annual review of public health 18.1 (1997): 341-378.

[8] The Folsom Group. "Communities of solution: the Folsom Report revisited." Annals of family medicine 10.3 (2012): 250.

[9] Health Information Technology and Accountability Act, 42 U.S.C. 300jj et seq; 17901 et seq. 2009.

[10] McClellan, Mark B., et al. Evidence-Based Medicine and the Changing Nature of Health Care:: Meeting Summary (IOM Roundtable on Evidence-Based Medicine). National Academies Press, 2008.

[11] Ibid (1)

[12] Ariely, Dan. "Controlling the information flow: Effects on consumers' decision making and preferences." *journal of Consumer Research* 27.2 (2000): 233-248.

[13] Shmueli, Galit, Nitin R. Patel, and Peter C. Bruce. Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons, 2007

[14] Beck, Andrew F., Melissa D. Klein, and Robert S. Kahn. "Identifying social risk via a clinical social history embedded in the electronic health record." Clinical pediatrics 51.10 (2012): 972-977.

[15] Ibid (1)

[16] Ibid (14)

[17] Butler, Danielle C., et al. "Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery." Health services research 48.2pt1 (2013): 539-559.