# PHX: Overview of Application and Data Processing

## Population Health Explorer (PHX) Application

PHX is an application built using R and Shiny. It provides a user interface designed to enable the rapid exploration and statistical analysis of well-characterized health outcomes.

This tool was designed to expedite prep-to-research, to combine clinical data with geospatial data in an easy to use interface, to put data into the hands of researchers more efficiently, and to enable to the discovery and distribution of novel insights.

Outcomes are precomputed for qualifying patients (and generally assessed annually). These outcomes can then be explored across dimensions (or features about patients). These dimensions include race, sex, age range, comorbidities like mental health diagnosis and substance abuse disorders, and social drivers of health, like food security and housing security. They also include neighborhood-level place-based data, like social vulnerability and access to greenspace.
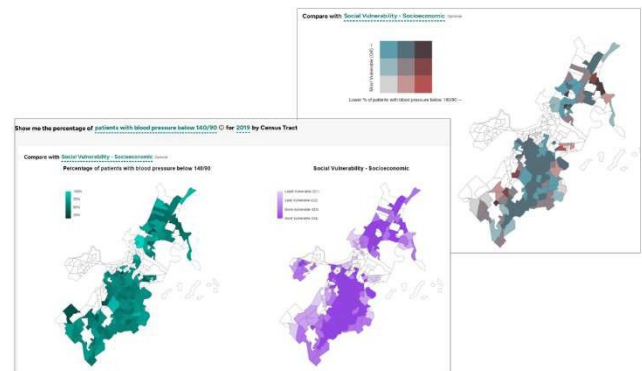


Users can explore changes in health outcomes over time. Dimensions can be added, and data can be filtered to visualize how outcomes differ by patient population, and to focus in on groups of interest.
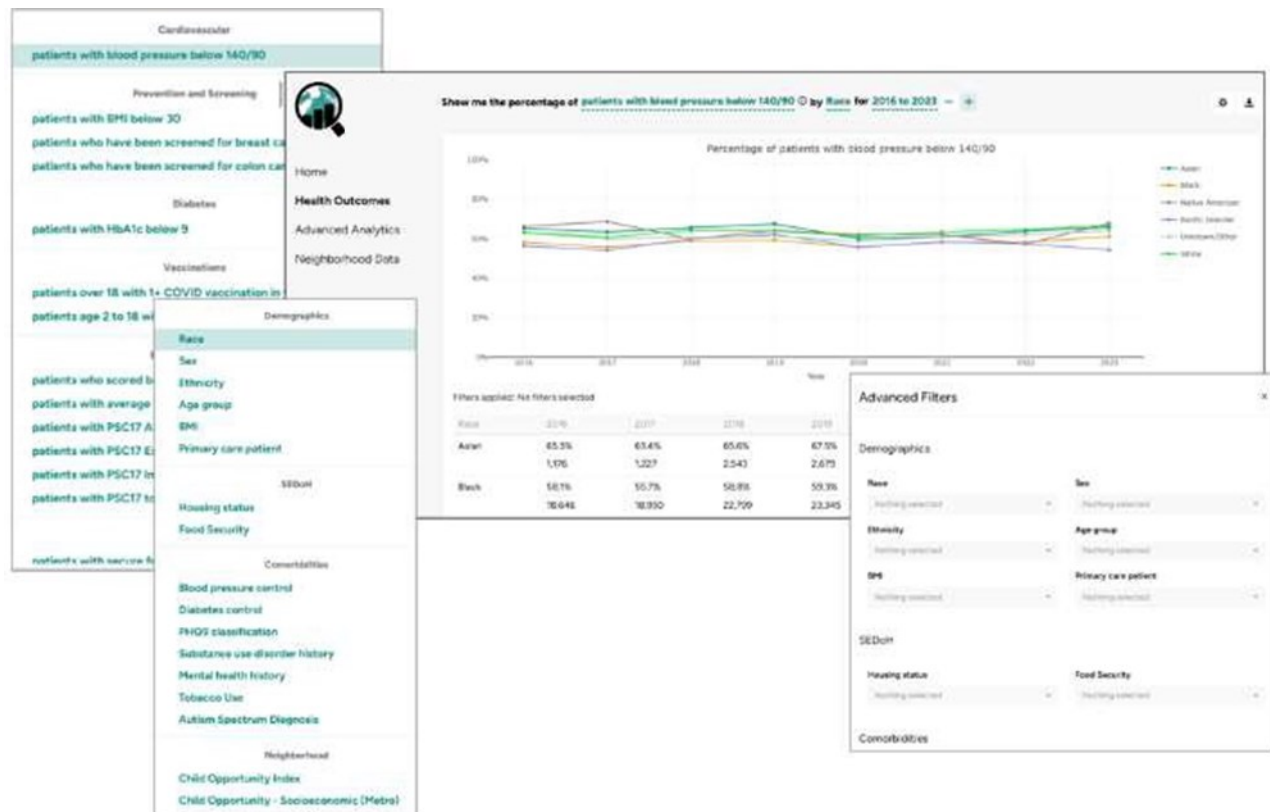
Users are able to explore data, characterize the patient population, and assess data for collinearity. They are then able to dynamically build logistic regression models to better understand the relationship between predictor variables and a health outcome. Model results and performance metrics are available to assist with model refinement.
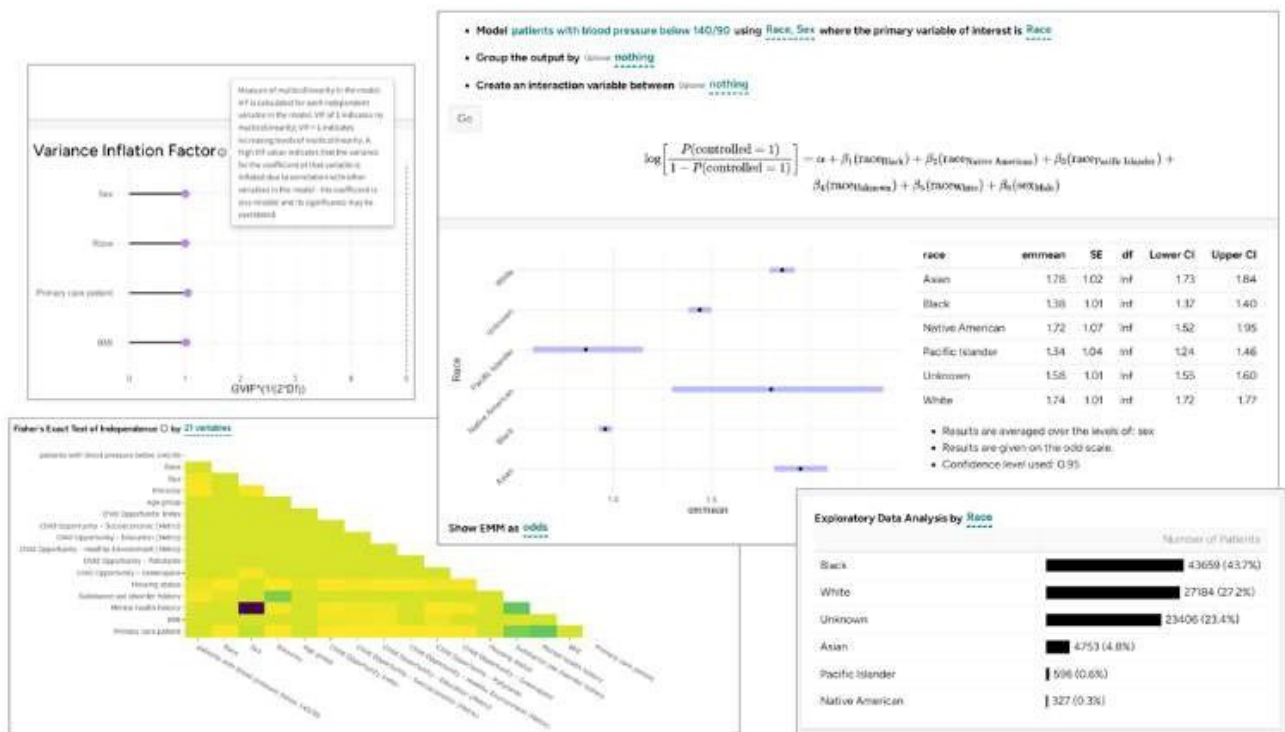


## Data Sources

### OMOP CDM

PHX v1.0 uses health outcomes generated from data stored in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The OMOP CDM is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence.

This standardized data model maps all representations of a data element to a common code, reducing the likelihood of erroneous exclusions and drastically decreasing the time required for researchers to coalesce desired code sets.

The CDM is designed to support research to identify and evaluate associations between interventions (drug exposure, procedures, healthcare policy changes etc.) and outcomes caused by these interventions (condition occurrences, procedures, drug exposure etc.). The CDM, combined with its standardized content (via the Standardized Vocabularies), ensures that research methods can be systematically applied to produce meaningfully comparable and reproducible results.
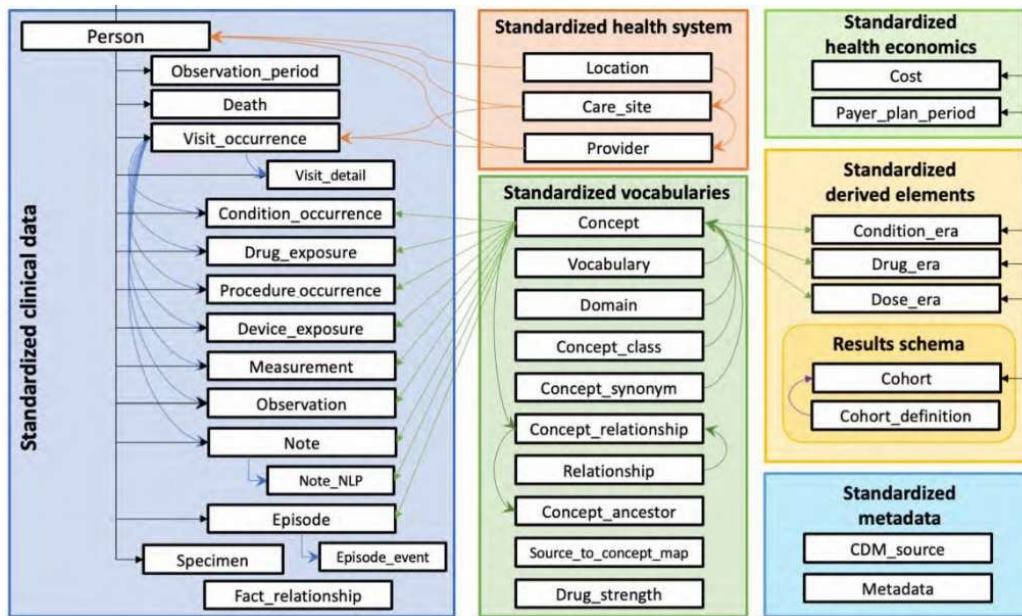
The PHX application utilizes the data standardization of the OMOP CDM to create computable, standardized health outcome generation scripts. For more information on health outcomes generated, please see the Health Outcomes section of this document.

To learn more about OHDSI and OMOP, visit:

- https://www.ohdsi.org/
- https://www.youtube.com/channel/UC2RFIQnptl-nk8GbjFfqztw
- https://ohdsi.github.io/TheBookOfOhdsi/
- https://github.com/OHDSI

*Third party place-based data*

Patient addresses can be geocoded and mapped to the census tract where they live, enabling the incorporation of third-party data describing neighborhood and environmental conditions. This can include information about a neighborhood's walkability, average household income, and level of pollutants in the air and water.

Any information mapped to zip code or census tract can be included in the PHX application. Currently, data sources include the CDC Social Vulnerability Index, and the Child Opportunity Index.

## Application Components

*Health Outcomes Tab*

How to use

Click the "+" icon to add variables to the display. Click the "-" icon to remove variables.

Filter data by clicking on the gear icon on the upper right. To apply the filters, scroll to the bottom of the filter popup and click "Apply".

Text with dashed underlines can be clicked – view and select other options.

Show me the percentage of **patients with blood pressure below 140/90** ⓘ by **Race** for **2016 to 2023**

## Calculations

Data displayed in line chart and table on health outcomes tab is calculated as follows for each health outcome, year and group combination:

**Total Patients Assessed:**

$$COUNT(Patients\ Meeting\ Metric)+COUNT(Patients\ Not\ Meeting\ Metric)$$

**Percent of patients meeting metric:**

$$\frac{COUNT(Patients\ Meeting\ Metric)}{COUNT(Patients\ Meeting\ Metric)+COUNT(Patients\ Not\ Meeting\ Metric)}$$

**Total Patients:**

COUNT(Patients Meeting Metric)+COUNT(Patients Not Meeting Metric)+COUNT(Patients Missing Data)

**Percent of patients missing:**

$$\frac{COUNT(Patients\ Missing\ Data)}{COUNT(Patients\ Meeting\ Metric)+COUNT(Patients\ Not\ Meeting\ Metric)+COUNT(Patients\ Missing\ Data)}$$

### *Advanced Analytics Tab*

### Fisher's Exact Test

Fisher's Exact Test is a statistical test used to determine if there is a relationship between two categorical variables. When two independent variables in a regression model are highly correlated (multicollinearity), model coefficients become unstable and difficult to interpret.

- Unstable coefficient estimates: coefficients may change significantly depending on which independent variables are included in the model.
- Reduced predictive accuracy: model may have difficulty learning the true relationship between the independent and dependent variables.
- Difficulty interpreting the model: it can be difficult to interpret which independent variables have the greatest impact on the outcome variable.

Fisher's exact test works by calculating the probability of obtaining the observed contingency table, or a more extreme table, under the null hypothesis that there is no association between the two variables. The p-value is then the sum of the probabilities of all tables that are at least as extreme as the observed table. If the p-value is less than a pre-specified significance level, such as 0.05, then the null hypothesis is rejected and we conclude that there is a statistically significant association between the two variables.

For interpretability, we have applied a –log10 transformation to the adjusted p-values resulting from this analysis.

If variables appear to be highly correlated, consider carefully which to include in the model – you may want to choose only one of the highly correlated variables, or combine the highly correlated variables into a single variable. There is no specific rule for how correlated is too correlated. Performing additional analysis, and understanding your variables and data will help determine next steps.

## Exploratory Analysis

Use the bar chart in the exploratory data analysis section to better understand the distribution of your data. Significantly unbalanced data will result in unstable or overfit models, or models that perform poorly. Very small cell sizes may result in models that have missing groups. This application enforces the rule of 10 (should have at least 10 events per predictor variable in order to estimate model coefficients accurately).

## Univariate Regression

It is important to perform univariate regression before building a multivariate model for several reasons:

- Identify important predictor variables: understanding which variables seem to have the greatest impact on the dependent variable can guide the selection of variables in a multivariate model, reducing model complexity and improving interpretability.
- Identify potential problems with data: Assessing variables independently can help identify outliers and non-linear relationships.

## Create a Model

Model is generated using glm() function.

Estimated Marginal Means (EMM) provide a more interpretable understanding of the results of a logistic regression analysis. EMMs are calculated by taking the weighted average of the predicted values at each level of the factor, where the weights are determined by the distribution of the data. This means that EMMs are adjusted for the other factors in the model, and are therefore more informative than simply looking at the mean of the outcome variable at each level of the factor.

Data is fit to a logistic regression model to predict the likelihood of a patient meeting the criteria for a given metric. This can help identify potential risk factors or interventions.

Please see Estimated Marginal Means package for more details: https://cran.r-project.org/web/packages/emmeans/index.html

Results can be viewed as probabilities, odds, or logit (log odds).

## Variance Inflation Factor Analysis

Variance inflation factor (VIF) analysis is a statistical technique used to detect multicollinearity in regression models. Multicollinearity is a problem that occurs when two or more independent variables in a regression model are highly correlated with each other. This can cause problems for logistic regression models because it can make the model coefficients unstable and difficult to interpret.

VIF analysis works by calculating a VIF score for each independent variable in the model. The VIF score is a measure of how much the variance of the coefficient estimate for the independent variable is inflated due to multicollinearity. A VIF score of 1 indicates that there is no multicollinearity, while a VIF score greater than 5 indicates that there is severe multicollinearity.

## Health Outcomes

Health outcomes are currently assessed annually for each eligible patient, resulting in one value for each person-year-outcome combination.
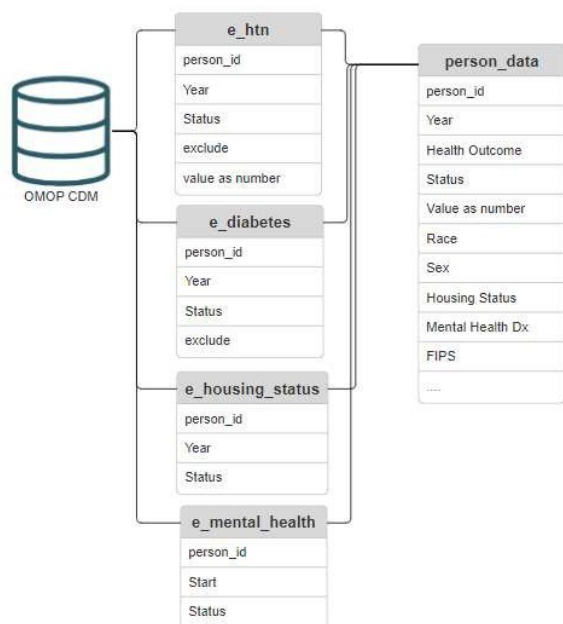


SQL queries to compute each health outcome are run against a site's OMOP CDM using the PHXdata R package. Each health outcome populates a health outcome-specific table stored on the results database. These tables can be reused, or revised to include additional information.

Health outcomes are then combined into a central data table called person_data. This health outcome data is then joined to additional information, including demographic data, clinical data (including comorbidities and data sourced from previously generated health outcomes), and geospatial data (if patient addresses are available).

The sections below include information about how each health outcome is computed, with

### Hypertension - Blood Pressure Control

The outcome for blood pressure control is based on Clinical Quality Measurement NQF 0018.

The denominator for this measure **includes** patients aged 18 to 85 with a diagnosis of essential hypertension made prior to the measurement date. **Excluded** are patients who had documentation of End Stage Renal Disease, renal transplant, dialysis procedures or hospice any time prior to or during the measurement period, and those who were pregnant during the measurement period. Also excluded are patients with diagnoses of frailty, or who have been prescribed medications to treat dementia (see linked measure specifications for more details).

For patients with multiple readings, the last reading during the measurement year is used. If multiple systolic and diastolic pressures are recorded during a visit, the lowest recorded values are used, according to the NQF standard.

Measurements below 140/90 mmHg meet the criteria for controlled blood pressure.

If a patient does not have a record of a blood pressure measurement taken during a reference year after the start date of their diagnosis of hypertension and before the date of their most recent visit, their data considered to be missing for that year.

Values are stored in the e_HTN table.

### HbA1c Control

The outcome for HbA1c Control is based on Clinical Quality Measurement NQF 0059.

The denominator for this measure includes patients ages 18 to 75 who had a diagnosis of diabetes made prior to or during the measurement year. Patients with diagnoses or prescriptions related to fragility and dementia, or who received hospice services prior to or during the measurement period are excluded.

If patients had multiple readings during the measurement year, the most last reading is used.

An HbA1c reading less than or equal to 9 meets the criteria for this outcome.

Data is considered missing if patients have a diagnosis of diabetes, but no HbA1c measurement during a reference year between the earliest condition start date and the patient's most recent visit date.

Values are stored in the e_diabetes table.

### Breast Cancer Screening

The outcome for breast cancer screening is based on the preventative care measure NQF 2372.

The denominator for this measure includes all female patients aged 50 to 74 who had a primary care visit during the measurement year. Patients who have a record of a mastectomy, encounters or diagnoses related to fragility, advanced illness, dementia, or hospice are excluded from this measure.

Criteria is considered met for this metric if a breast cancer screening (mammogram) has been completed within 24 months prior to a patients' primary care appointment date (with a three month grace period).

Missing data is not defined for this measure, as it is equivalent to not meeting the metric.

Values are stored in the e_breastcancerscreening table.

### Colorectal Cancer Screening

The outcome for colorectal screening is based on the preventive care measure NQF 0034.

The denominator for this measure includes all patients ages 50 to 75 who had a primary care appointment during the measurement year. Patients with a history or colorectal cancer, record of a total colectomy, and those who received hospice serviced during the measurement year are excluded from this measure.

Criteria is met for this metric if appropriately colon cancer screening has been completed. This screening can include:

- A fecal occult blood test during the measurement period
- A colonoscopy during or within 9 years prior to the measurement period
- A flexible sigmoidoscopy during or within 4 years prior to the measurement period
- Computed tomography during or within 4 years prior to the measurement period
- FIT-DNA test during or within two years prior to the measurement period

Missing data is not defined for this measure, as it is equivalent to not meeting the metric.

Values are stored in the e_coloncancerscreening table.

### BMI - Adult

All patients ages 18 and above with a body mass index reading during the measurement year are included in this measure.

For patients with multiple readings during the measurement year, the most recent reading will be included. Criteria for this metric is met when most recent body mass index in a given measurement year is less than 30 (the cut-off for obesity).

There is no definition for missing data for this health outcome.

Values are stored in the e_bmi table.

### BMI - Pediatric

All patients below the age of 18 with a body mass index reading during the measurement year are included in this measure.

For patients with multiple readings during the measurement year, the most recent reading will be included. Criteria for this metric is met when most recent body mass index in a given measurement year is below the 95th percentile for the patients' sex and age group.

There is no definition for missing data for this health outcome.

### Food Security

Patients' food security status is based on observations with concept IDs of 36304041 and 36306143 during the measurement year, or in the year prior to the measurement year if there are no records for the reference year. Results are not imputed. Only screenings completed during or one year prior to the measurement year are included.

Patient responses are categorized as "Secure" or "Insecure". If patients indicated insecure food access during or in the one year prior to the measurement year, they are considered to have "insecure" food access. If all responses in the reference year and year prior are "secure", they are considered to have "secure" food access. If there is no recorded response, their food security status is "unknown".

Values are stored in the e_food_security table.

### Housing Security

Patients' housing security status is based on observations with concept IDs of 42869557, 4139934, and 4144274 during the measurement year, or in the year prior to the measurement year if no current value is available. Results are not imputed. Only observations during or one year prior to the measurement year are included.

Patient responses are categorized as "Secure", "Insecure", and "Homeless".

Values are stored in the e_housing_security table.

### COVID Vaccination – Adults

The denominator for this metric includes all patients ages 18 and above who had a primary care appointment during the measurement year. Criteria is met for this metric with at least one COVID vaccination record during the reference year.

There is no definition for missing data for this metric.

Values are stored in the e_covid_vax_adult table.

### COVID Vaccination – Pediatric

The denominator for this metric includes all patients ages under 18 years old who had a primary care appointment during the measurement year. Criteria is met for this metric with at least one COVID vaccination record during the reference year.

There is no definition for missing data for this metric.

Values are stored in the e_covid_vax_peds table.

### PSC-17 Screening (Internalization, Externalization, Attention, Total)

PSC-17 screenings assess for symptoms of anxiety and sadness in children ages 6 to 12 years old. There are three parts to these screenings, resulting in four distinct scores.

The denominator for this metric is all patients age 6 to 12 who completed at least one PSC17 screening during the measurement year.

> **PSC-17 Internalization:** Scores below 5 meet the criteria for this metric.
> **PSC-17 Externalization:** Scores below 7 meet the criteria for this metric.
> **PSC-17 Attention:** Scores below 7 meet the criteria for this metric.
> **PSC-17 Total:** Scores below 15 meet the criteria for this metric.

There is no definition for missing data for this metric.

Values are stored in the e_psc17_all table for all associated measures.

### PHQ-9 - Average

PHQ-9 screenings assess for symptoms of depression and anxiety. Responses are categorized based on total scores – a total score of 10 to 14 indicates moderate symptoms, scores 15 and above indicate severe symptoms.

The denominator for this metric is all patients ages 12 and above who completed at least one PHQ9 screenings during the measurement year. If a patient completed multiple screenings, scores are averaged across screenings.

To meet the criteria for this metric, average score for the measurement year must fall below 10.

There is no definition for missing data for this metric.

Values are stored in the e_phq9_avg table.

### PHQ-9 - Any

PHQ-9 screenings evaluate patients for symptoms of depression and anxiety. Responses are categorized based on total scores – a total score of 10 to 14 indicates moderate symptoms, scores 15 and above indicate severe symptoms.

The denominator for this metric is all patients ages 12 and above who completed at least one PHQ9 screenings during the measurement year.

To meet the criteria for this metric, all screenings that occur during the measurement year must have a score below 10.

There is no definition for missing data for this metric.

Values are stored in the e_phq9_any table.

## Variables
### *Patient-level Variables*
### Race
Source data is generally considered to be information provided by patient.

This field is not exhaustive, and often requires that patients make a choice that may not perfectly describe their race or ethnicity. Racial data is often aggregated for a variety of reasons. We strongly encourage researchers to explore disaggregated racial data when possible during research.

Please note that race and ethnicity do not represent biological or cultural differences between groups. These distinctions are driven by cultural, legal, geographical, and societal influences rooted in structural racism and white supremacy. Significant differences in health outcomes and experiences between groups created in this manner are likely due to the result of structural and individual racism.

Aggregations can be edited in the 1person.sql file. Values are stored in the e_person table.

### Ethnicity
Source data is generally considered to be information provided by patient.

Ethnicity includes the options: Hispanic or Latino, Not Hispanic or Latino, Unknown/Other. Values are stored in the e_person table.

### Sex
In the current iteration of OMOP CDM, the gender_concept_id field has two options: Male and Female. We have renamed this field to "Sex" in the PHX interface in order to better represent its content. To protect patient privacy due to low cell counts, we have filtered out patients with "Unknown" in this field. This filter can be removed by editing the person_data.sql file. Values are stored in the e_person table.

### Primary Care Patient
Primary care patients are currently identified as patients who have a record of at least one preventative service CPT code for new or existing patients during a reference year OR a visit associated with particular care site during the reference year.

The 1primarycare.sql file should be revised to include site-specific definitions of primary care patients when necessary, including care site locations or other identifiers.

Values are stored in the e_primarycare table.

### BMI - Adult
Attribute assigns most recent BMI measurement taken during reference year to "BMI < 30" and "BMI >-30" for patients age 18 and above. This column also contains "Unknown" values when no BMI measurement is available. Values are sourced from the e_bmi table.

### Age Group

Patient age group is calculated by subtracting a patient's birth year from the reference year. To edit age groups, please edit the person_data.sql file.

Patient ages are generally grouped as follows: 0 to 4, 5 to 11, 12 to 17, 18 to 26, 27 to 39, 40 to 55, 56 to 69, 70 to 85, 85 and older.

### Food Security Status

Values are sourced from the e_food_security table to assign a value for each patient and year. Food security status has values of "Secure", "Insecure" and "Unknown".

### Housing Security Status

Values are sourced from e_housing_security table to assign a value for each patient and year. Housing security status has values of "Secure", "Insecure", "Homeless" and "Unknown".

### Comorbidity: Mental Health Diagnosis

Patients with a record of a diagnosis of Depression and/or Anxiety (and descendants of these concepts) with condition_start_date any time during or prior to the measurement year and condition_end_date after the measurement year are considered to have a comorbidity associated with mental health.

Values are stored in e_mentalhealth table and incorporated via the person_data.sql file.

### Comorbidity: Autism Spectrum Diagnosis

Patients with a condition_occurrence record related Autism Spectrum Disorder with a condition_start_date prior to or during the measurement year are considered to have a comorbidity of Autism Spectrum Disorder.

Values are stored in e_asd table and incorporated via the person_data.sql file.

### Comorbidity: Substance Use Disorder

Patients with a condition_occurrence record related to substance abuse and its descendants (including nondependent and dependent, episodic and continuous, misuse or abuse of alcohol, opioids, prescription medications, or illegal substances) any time during or prior to the measurement year are considered to have a comorbidity of Substance Use Disorder.

Values are stored in e_sud_hx table and incorporated via the person_data.sql file.

### Birth Year

Birth Year is stored in the e_person table and is used to assign age groups, as well as determine cut-off criteria for specific measures. Birth Year is not displayed in the PHX interface.

### Death Date

Death Date is stored in the e_person table and is not displayed in the PHX application interface. Death data at many OMOP sites is incomplete or not reliable, so this field is not currently used in any other health outcomes. It can easily be included or converted to a health outcome if desired.

### Disability: Vision (both eyes)

Patients with a condition_occurrence related to a visual impairment impacting both eyes, beginning any time prior to or during the measurement year, are considered to have a comorbidity related to visual impairment.

Values are stored in the e_disability table with an identifier indication applicable disability and start date.

### Disability: Hearing

Patients with a condition_occurrence related to impaired hearing or hearing loss, beginning any time prior to or during the measurement year, are considered to have a comorbidity related to hearing loss

Values are stored in the e_disability table with an identifier indication applicable disability and start date.

### FIPS

FIPS code is an identifier for location ([FIPS County Code](#)). Here, it is associated with census tract; however, as long as data can be mapped to the identifier, this field could contain different levels of granularity.

This field is generated from a multi-step process. Patient addresses must be geocoded and mapped to a FIPS code. Generally, patient addresses are first mapped to a location_id during the OMOP ETL process. Location_ids can be stored in the cdmDatabase.person table and in the cdmDatabase.location_history table. Generally the logic behind location mapping and storage is site-specific. Location_ids should be mapped to FIPS codes.

Please review the location_history.sql file to understand assumptions made in pulling in FIPS code. This file can be edited to account for site-specific requirements and variations.

For additional support with geocoding, please contact the [OHDSI GIS workgroup](#).

*Neighborhood-level/Place-Based  Variables*

Social Vulnerability Index

The Geospatial Research, Analysis, and Services Program (GRASP) created and maintains the CDC/ATSDR Social Vulnerability Index. Social vulnerable populations and communities are at higher risk during public health emergencies due to factors like socioeconomic status, racial and ethnic minority status, housing type, and transportation. While the SVI is often used for disaster planning and emergency preparedness, identifying communities associated with higher risk (for example, greater proportion of individuals without health insurance or access to transportation) can generate insights into the relationship between social vulnerability and health outcomes.

SVI data is sourced from the US Census and American Community Survey responses, weighted, and grouped into representative themes (Socioeconomic, Housing and Transport, Minority Status, and Household Characteristics), as well as an aggregate measure of vulnerability. Tracts within a state are compared to each other and ranked against others in the state on a scale of 1 to 99, with higher scores indicating increased vulnerability.

The four component themes and the overall SVI Index score for each tract is included in the PHX application. Values are stored in the coi_svi_table.

**Helpful Terms & Facts**

**Social Vulnerability** refers to the resilience of communities (the ability to survive and thrive) when confronted by external stresses on human health, stresses such as natural or human-caused disasters, or disease outbreaks. Reducing social vulnerability can decrease both human suffering and economic loss.

**Socially Vulnerable Populations** include those who have special needs, such as, but not limited to, people without vehicles, people with disabilities, older adults, and people with limited English proficiency.

**Census tracts** are subdivisions of counties for which the Census collects statistical data. The CDC/ATSDR SVI ranks each tract on 16 social factors, including poverty, lack of vehicle access, and crowded housing, and groups them into four related themes. Each tract receives a separate ranking for each of the four themes, as well as an overall ranking.

**Data Sources**: U.S. Census

**CDC/ATSDR SVI Themes & Social Factors:**

- Socioeconomic status (below 150% poverty, unemployed, housing cost burden, no high school diploma, no health insurance)
- Household characteristics (aged 65 or older, aged 17 or younger, civilian with a disability, single-parent households, English language proficiency)
- Racial and ethnic minority status (Hispanic or Latino (of any race); Black and African American, Not Hispanic or Latino; American Indian and Alaska Native, Not Hispanic or Latino; Asian, Not Hispanic or Latino; Native Hawaiian and Other Pacific Islander, Not Hispanic or Latino; Two or More Races, Not Hispanic or Latino; Other Races, Not Hispanic or Latino)
- Housing type & transportation (multi-unit structures, mobile homes, crowding, no vehicle, group quarters)

## Child Opportunity Index

"Neighborhoods matter. Children who live in neighborhoods with quality early childhood education and schools, safe housing, access to healthy food, parks and playgrounds and clean air are more likely to grow into healthy, productive adults than children who don't. The Child Opportunity Index (COI) measures and maps the quality of resources and conditions that matter for children to develop in a healthy way in the neighborhoods where they live."

The Child Opportunity Index, developed and maintained by the Institute for Child, Youth and Family Policy at the Heller School for Social Policy and Management, Brandeis University, can be used to understand opportunity, opportunity gaps, and issues of equity associated with the places where children live.

The COI groups indicators into three themes (Social and Economic, Health and Environment, and Education), and provides an overall opportunity score. Scores are available normed by metro-area, state, or nationally. The COI also makes available numerous component indicators that add additional nuance. For more information, please review the COI technical documentation.

The PHX application includes overall COI grouping, and the three component theme groupings, both metro-normed and state-normed.

Values are stored in the coi_svi_table. Additional values related to COI should be added to this table.

Table 2. COI 2.0 indicators and domains

| Domain | INDICATOR | DESCRIPTION (SOURCE) |
|---|---|---|
| EDUCATION | **Early childhood education (ECE)** | |
| | ECE centers | Number of ECE centers within a 5-mile radius (own data collection from state and federal sources) |
| | High-quality ECE centers | Number of NAEYC accredited centers within a 5-mile radius (own data collection from state and federal sources) |
| | ECE enrollment | Percent 3- and 4-year-olds enrolled in nursery school, preschool or kindergarten (ACS) |
| | **Elementary education** | |
| | Third grade reading proficiency | Percent third graders scoring proficient on standardized reading tests, converted to NAEP scale score points (EDFacts, GS and SEDA) |
| | Third grade math proficiency | Percent third graders scoring proficient on standardized math tests, converted to NAEP scale score points (EDFacts, GS and SEDA) |
| | **Secondary and postsecondary education** | |
| | High school graduation rate | Percent ninth graders graduating from high school on time (EDFacts and GS) |
| | Advanced Placement (AP) course enrollment | Ratio of students enrolled in at least one AP course to the number of 11th and 12th graders (CRDC) |
| | College enrollment in nearby institutions | Percent 18-24 year-olds enrolled in college within 25-mile radius (ACS) |
| | **Educational and social resources** | |
| | School poverty | Percent students in elementary schools eligible for free or reduced-price lunches, reversed (NCES CCD) |
| | Teacher experience | Percent teachers in their first and second year, reversed (CRDC) |
| | Adult educational attainment | Percent adults ages 25 and over with a college degree or higher (ACS) |
| HEALTH & ENVIRONMENT | **Healthy environments** | |
| | Access to healthy food | Percent households without a car located further than a half-mile from the nearest super-market, reversed (USDA) |
| | Access to green space | Percent impenetrable surface areas such as rooftops, roads or parking lots, reversed (CDC) |
| | Walkability | EPA Walkability Index (EPA) |
| | Housing vacancy rate | Percent housing units that are vacant, reversed (ACS) |
| | **Toxic exposures** | |
| | Hazardous waste dump sites | Average number of Superfund sites within a 2-mile radius, reversed (EPA) |
| | Industrial pollutants in air, water or soil | Index of toxic chemicals released by industrial facilities, reversed (EPA) |
| | Airborne microparticles | Mean estimated microparticle (PM2.5) concentration, reversed (CDC) |
| | Ozone concentration | Mean estimated 8-hour average ozone concentration, reversed (EPA) |
| | Extreme heat exposure | Summer days with maximum temperature above 90F, reversed (CDC) |
| | **Health resources** | |
| | Health insurance coverage | Percent individuals ages 0-64 with health insurance coverage (ACS) |
| SOCIAL & ECONOMIC | **Economic opportunities** | |
| | Employment rate | Percent adults ages 25-54 who are employed (ACS) |
| | Commute duration | Percent workers commuting more than one hour one way, reversed (ACS) |
| | **Economic and social resources** | |
| | Poverty rate* | Percent individuals living in households with incomes below 100% of the federal poverty threshold, reversed (ACS) |
| | Public assistance rate* | Percent households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assistance Program, reversed (ACS) |
| | Homeownership rate* | Percent owner-occupied housing units (ACS) |
| | High-skill employment* | Percent individuals ages 16 and over employed in management, business, financial, computer, engineering, science, education, legal, community service, health care practitioner, health technology, arts and media occupations (ACS) |
| | Median household income* | Median income of all households (ACS) |
| | Single-headed households | Percent family households that are single-parent headed, reversed (ACS) |

Notes: We reverse some of the indicators when combining them into the index, e.g., the poverty rate, so that more of that indicator always means more opportunity. *These five indicators are combined into an economic resource index.