2024 Annual Data Release: Analytic Dataset Codebook

The SAS and CSV files provided at https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation contain the measure value, numerator, denominator and confidence intervals, where available.

There are seven descriptive variables included for each row to identify what geographic area the data represent and whether the county was included in the Health Groups. These variables are:

1. statecode = FIPS State Code
2. countycode = FIPS County Code
3. fipscode = 5-digit FIPS Code
4. state = State abbreviation
5. county = County or State name
6. year = Release year
7. county_clustered equals 1 when the county was included in the Health Groups and equals 0 if the county has not grouped. This variable is set to missing for state and national data.

Each measure variable name is coded with a measure ID. These measure IDs can be found in the tables at the end of the documentation. You will find at least five variables for each measure:

Measure value coded as v(measure id)_rawvalue
Numerator coded as v(measure id)_numerator
Denominator coded as v(measure id)_denominator
Lower bound of 95% confidence interval coded as v(measure id)_cilow
Upper bound of 95% confidence interval coded as v(measure id)_cihigh

Non-standard measure variables

Health care provider measures (v004, v088, v062, v131): Four measures have a "ratio" variable included. These measures and their corresponding variables are:

- Primary Care Physicians (variable = v004_other_data_1),
- Dentists (variable = v088_other_data_1),
- Mental Health Providers (variable = v062_other_data_1), and
- Other Primary Care Providers (variable = v131_other_data_1)

For these measures, the value reported for each county is a ratio of population:number of providers. However, in calculating the composite scores used to generate Health Groups, we use the actual *rate* of number of providers/100,000 population. The analytic dataset includes both the rate (as the raw value variable), and the *ratio* (as the "other data 1" variable).

When a county has no reported providers (ratio = [population]:0), the value for the ratio variable is assigned to be the negative population value. For example, v004_other_data_1 stores the ratio variable as a positive number if the ratio is XXXX:1. However, for counties that have zero providers, the value is stored as a negative number to indicate that the ratio is different and is instead XXXX:0. (For example, if the ratio is 2000:0, the value stored in the field would be -2000.)

Drinking Water Violations (v124): is an indicator variable, where a 1 indicates a "Yes", while a 0 indicates a "No".

Severe Housing Problems (v136): In addition to the standard variables (e.g., raw value, confidence intervals), there are three additional variables for this measure:

- Percentage of households with high housing costs (variable = v136_other_data_1),
- Percentage of households with overcrowding (variable = v136_other_data_2),
- Percentage of households with lack of kitchen or plumbing facilities (variable = v136_other_data_3)

Juvenile Arrests (v158): There are two variables that are summed to create the number of juvenile arrests:

- Petitioned cases (variable = v158_other_data_1),
- Non-petitioned cases (variable = v158_other_data_2)

Race-specific data: Several measures are available by race/ethnicity categories for counties with sufficient sub-population sizes. These measures are coded as:

- v(measure ID)_race_aian for the American Indian & Alaska Native population estimate
- v(measure ID)_race_asian for the Asian population estimate
- v(measure ID)_race_black for the Black population estimate
- v(measure ID)_race_hispanic for the Hispanic population estimate
- v(measure ID)_race_nhopi for the Native Hawaiian and Other Pacific Islander population estimate
- v(measure ID)_race_tom for the two or more races population estimate
- v(measure ID)_race_white for the white population estimate

If confidence intervals are available for race-specific data, the variable name is the same as the raw value variable, except with a "_cilow" or "_cihigh" extension. For example, the confidence intervals for American Indian & Alaska Native years of potential life lost values are given by v001_race_aian_cilow and v001_race_aian_cihigh.

For information about the methods used in the 2024 Annual Data Release, and more in-depth descriptions of each of the measures, please visit www.countyhealthrankings.org and review our 2024 Technical Document.