Methods to Obtain County-Level Estimates and Ranks

Data from CDC's Behavioral Risk Factor Surveillance System (BRFSS) and from the US Census Bureau's Population Estimates Program were used to obtain county-level estimates of diagnosed diabetes, newly diagnosed diabetes, obesity, and physical inactivity.[1,2] The BRFSS is an ongoing, monthly, state-based telephone survey of the adult population aged 18 years or older that provides state-specific information on behavioral risk factors and preventive health practices. Respondents were considered,

- To have diagnosed diabetes if they responded "yes" to the question, "Has a doctor ever told you that you have diabetes?" Women who indicated that they only had diabetes during pregnancy were not considered to have diagnosed diabetes. People who reported having diagnosed diabetes were then asked at what age they were diagnosed.
- To have been diagnosed with diabetes in the last year if they reported having diagnosed diabetes and the difference between their age at the time of the survey and the age they provided to the question, "How old were you when you were told you have diabetes?" was less than one. If the difference was between one year and two years, the person was weighted as half a newly diagnosed case.
- To be obese if their body mass index was 30 or greater. Body mass index (weight [kg]/height [m]2) was derived from self-report of height and weight.
- To be physically inactive if they answered "no" to the question, "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?"

You can view trends in county-level data beginning in 2004. Three years of data were used to improve the precision of the year-specific county-level estimates of diagnosed diabetes, newly diagnosed diabetes, obesity, and physical inactivity. For example, 2003, 2004, and 2005 were used for the 2004 estimate and 2004, 2005, and 2006 were used for the 2005 estimate. Estimates were restricted to adults aged 20 years or older to be consistent with the population estimates from the US Census Bureau. The US Census Bureau provides year-specific county population estimates by demographic characteristics - age, sex, race, and Hispanic origin.

The county-level estimates for the over 3,100 counties or county equivalents (e.g., parish, borough, municipality) in the 50 US states, Puerto Rico, and the District of Columbia (DC) were based on indirect model-dependent estimates using Bayesian multilevel modeling techniques.[3-5] This model-dependent approach employs a statistical model that "borrows strength" in making an estimate for one county from BRFSS data collected in other counties and states. Multilevel Binomial regression models with random effects of demographic variables (age 20-44, 45-64, >=65; race/ethnicity; sex) at the county-level were developed. Estimates were age adjusted to the 2000 US standard population using age groups, 20-44, 45-64, and 65 or older.[6]

Ranks of all US counties or county equivalents were based on age-adjusted estimates of diagnosed diabetes, obesity, and physical inactivity described above. Models were fit using a Bayesian simulation method known as Markov chain Monte Carlo.[3-5] As part of the model fitting process we generated and saved 1,000 random samples from the distribution of each county's age-adjusted rate. For each of these random samples we ranked the counties by rate. From the estimated distribution of each county's rank, we used the median for the rank and the 5th and 95th percentiles for a 90% confidence interval. A low rank indicates a low rate. Ranks for Puerto Rico were not generated.

Data Classification

When selecting data filters, the user can choose between two different data classification methods: quartiles and natural breaks. Both methods require 1) sorting the data in ascending order first based on their values and 2) grouping similar observations together into classes.

The quartile classification method divides the data into four classes with equal number of observations in each group irrespective of ties or outliers. Quartile ranges may vary by year for the same indicator. The first and last quartiles may include outliers.

The natural breaks method is also called the Jenks optimization method.[7] Classes are created based on natural groupings inherent in the data that minimize within-class differences and maximize between-class differences. When using this method, 5 classes were pre-specified and data from all years were aggregated so the categories were the same across all years.

References

1. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System www.cdc.gov/brfss. Accessed August 2016.

2. US Census Bureau. Population Estimates Program. Population and Housing Unit Estimates. https://www.census.gov/popest/. Accessed August 2016

3. Rao JNK. *Small Area Estimation*. Hoboken, New Jersey: John Wiley & Sons, Inc; 2003.

4. Cadwell BL, Thompson TJ, Boyle JP, Barker LE. Bayesian small area estimates of diabetes prevalence by U.S. county, 2005. *J Data Sci*. 2010;8(1):173-188.

5. Barker LE, Thompson TJ, Kirtland KA, Boyle JP, Geiss LS, McCauley MM, Albright AL. Bayesian small area estimates of diabetes incidence by United States county, 2009. *J Data Sci*. 2013;11:249-269.

6. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected US population. Healthy People Statistical Notes. 2001;20:1-10.

7. Jenks, George F. 1967. "The Data Model Concept in Statistical Mapping",International Yearbook of Cartography 7: 186-190.