Category: Data Analysis – Work in Progress – Staff

An Accurate Algorithm to Calculate Median Age of Catchment Areas

J. Zhao, K. Tossas

VCU Massey Comprehensive Cancer Center

1. Background

Accurately estimating the median age of populations within specific cancer center catchment areas is crucial for a variety of public health analyses, including cancer epidemiology and health care resource allocation. The aggregation of median age data from lower levels (such as county level) to higher levels (such as catchment areas) requires careful consideration of the appropriate methods to ensure reliability and accuracy.

2. Goals

In this study, we evaluated various methods for aggregating median age estimates from county-level data to the state level, using the actual median age at the state level as the benchmark. Our goal was to identify the most accurate approach for estimating the median age at these higher levels.

3. Solutions and Methods

Four aggregation methods were evaluated to calculate state median age based on county-level median age. The first 3 methods (A, B, and C) use population data from American Community Survey (ACS).

1. Method A: The state's median age is calculated directly as the median of the county-level median ages within the state.
2. Method B: The median age for the state is aggregated by weighting county median ages based on the county populations.
3. Method C: This method refines the calculation by considering the population distribution across age groups, assuming uniform distribution within each age group.
4. Method S: This method uses single-year age population estimates from the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) database, offering a detailed and direct approach for age estimation at the county level.

4. Outcomes

Among the three methods utilizing ACS population data, Method C provided the most accurate results, with a mean percentage difference of just 0.05 percent (SD = 0.11%) when compared to actual state median ages (p-value = 0.003). The largest deviation was seen in Wyoming, where Method C calculated a median age of 38.611 years, slightly higher than the actual median age of 38.5 years. The other two ACS-based methods (A and B) performed less well, with mean percentage differences of 6.7 percent and 0.53 percent (SD = 4.4% and 0.38%, respectively), both showing significant deviations from the actual median age (p-values < 0.0001). However, Method S, which relies on single-year age population estimates from SEER, proved to be the most accurate overall. It yielded a mean percentage difference of only 0.04 percent (SD = 0.18%) compared to the actual median age, with a p-value of 0.15, indicating no significant difference. This superior performance led us to select Method S for further analysis. Using Method S, we estimated the median age for all 63 NCI cancer center catchment areas. The median ages across these areas varied from a low of 33.9 years at the Mays Cancer Center at UT Health San Antonio MD Anderson to a high of 45.9 years at the Dartmouth Cancer Center, with the overall median age for all catchment areas being 38.5 years.

Category: Data Analysis – Work in Progress – Staff

5. Lessons Learned and Future Directions

Method S, which utilizes single-year age estimates from SEER, proved to be the most accurate for aggregating median age data to higher levels. Its precision makes it an excellent choice for estimating the median age across cancer center catchment areas, providing valuable demographic insights for public health planning and cancer research.