An Accurate Algorithm to Calculate Median Age of Catchment Areas

Jinlei Zhao[1], **Katherine Y. Tossas**[1,2]

[1]Massey Comprehensive Cancer Center, Virginia Commonwealth University, Richmond, VA;

[2] Department of Social and Behavioral Sciences, Virginia Commonwealth University School of Public Health, Richmond, VA

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 healthcare 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.
- 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.

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).

Method A: The state's median age is calculated directly as the median of the county-level median ages within the state.

Method B: The median age for the state is aggregated by weighting county median ages based on the county populations.

Method C: This method refines the calculation by considering the population distribution across age groups, assuming uniform distribution within each age group.

Method S: This method uses single-year age population estimates from the SEER database, offering a detailed and direct approach for age estimation at the county level.

Table 1. P-values of Paired t-test between the 4 methods and real value

A	B	C	S
<.0001****	<.0001****	0.0029**	0. 1662

KEY FINDING

Calculation using SEER single year population more accurately estimates median age.

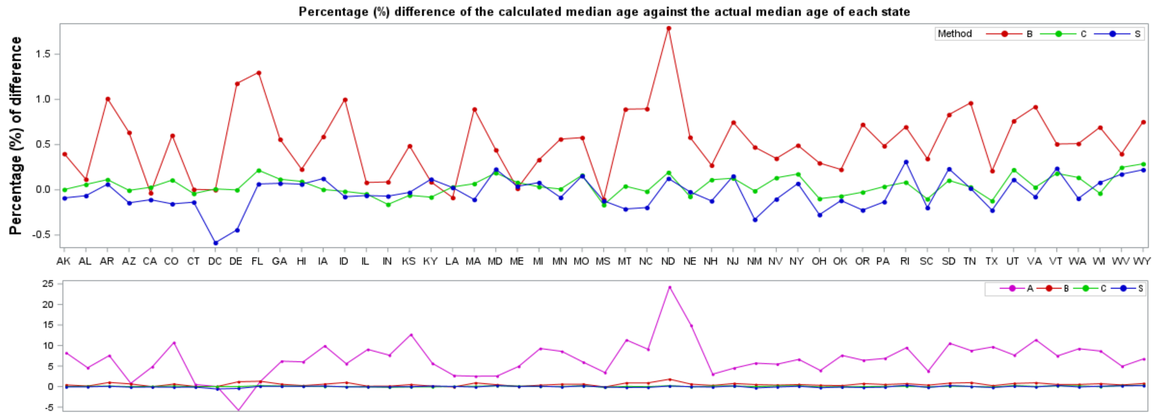

Figure 1. The percentage (%) difference of the calculated median age against the actual median age of each state. Compared to methods B, C and S (top), the median age calculated from method A demonstrates considerable divergence from real median values. Please note the y-axes have different scales. The population from ACS 5-Year Estimates (2022) and SEER (2018-2022) were used in this study.

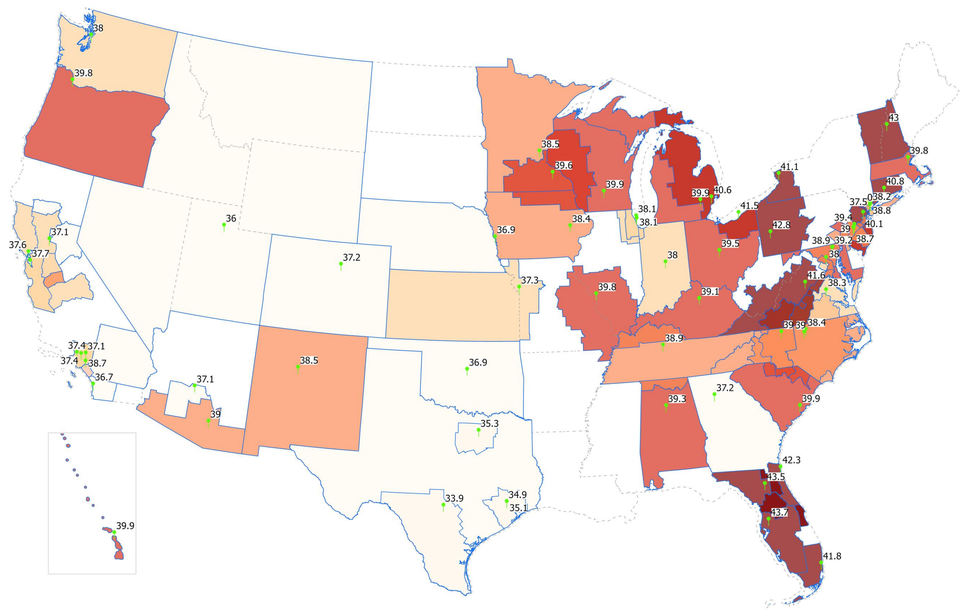

Figure 2. The median ages of NCI cancer center catchment areas were calculated using the method S. The catchment area of Mays Cancer Center at Texas has the lowest median age of 33.9 years and Moffitt Cancer Center at Florida has the highest of 43.7. Due to the irregular shape of the catchment of Memorial Sloan Kettering Cancer Center, it was excluded from the calculation.

RESULTS

- Among the three methods utilizing ACS population data, Method C provided the most accurate results, with a mean percentage difference of just 0.05% (SD = 0.11%) when compared to actual state median ages (p-value = 0.003) (Fig. 1). 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% and 0.53% (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% (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 almost all NCI cancer center catchment areas. The median ages across these areas varied from a low of 33.9 years at the Mays Cancer Center to a high of 43.7 years at the Moffitt Cancer Center, with the overall median age for all catchment areas being 38.7 years (Fig. 2).

CONCLUSIONS

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.

NCI Comprehensive Cancer Center

A Cancer Center Designated by the National Cancer Institute