SSM - Population Health 18 (2022) 101078

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Estimating uncertainty in a socioeconomic index derived from the American community survey

Francis P. Boscoe [a,b,*], Bian Liu [c], Jordana Lafantasie [d], Li Niu [c], Furrina F. Lee [b]

[a] *Pumphandle, LLC, Camden, ME, USA*
[b] *New York State Department of Health, Albany, NY, USA*
[c] *Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
[d] *School of Public Health, University at Albany, Rensselaer, NY, USA*

A B S T R A C T

Socioeconomic indexes are widely used in public health to facilitate neighborhood-scale analyses. Although they are calculated with high levels of precision, they are rarely reported with accompanying measures of uncertainty (e.g., 90% confidence intervals). Here we use the variance replicate tables that accompany the United States Census Bureau's American Community Survey to report confidence intervals around the Yost Index, a socioeconomic index comprising seven variables that is frequently used in cancer surveillance. The Yost Index is reported as a percentile score from 1 (most affluent) to 100 (most deprived). We find that the average uncertainty for a census tract in the United States is plus or minus 8 percentiles, with the uncertainty a function of the value of the index itself. Scores at the extremes of the distribution are more precise and scores near the center are less precise. Less-affluent tracts have greater uncertainty than corresponding more-affluent tracts. Fewer than 50 census tracts of 72,793 nationally have unusual distributions of socioeconomic conditions that render the index uninformative. We demonstrate that the uncertainty in a census-based socioeconomic index is calculable and can be incorporated into any analysis using such an index.

1. Introduction

Socioeconomic indexes are widely used in the social sciences, particularly in the field of public health, and particularly where neighborhood-level distinctions are of interest (Carlson et al., 2021; Deas, Robson, Wong, & Bradford, 2003; Mehaffey et al., 2020; Patel et al., 2016; Lopez-De Fede et al., 2016; Webb et al., 2017). One such index is the Yost Index, developed from a factor analysis of seven variables or combinations of variables from the American Community Survey (ACS) of the United States Census related to household income, poverty, rent, house value, education, employment type and employment status (Yost et al. 2001; Yu, Tatalovich, Gibson, & Cronin, 2014). Most of its applications have been in cancer surveillance (Abdel-Rahman, 2019; Li et al., 2021; Rajeshuni et al., 2020; Ross et al., 2017; Swords, Mulvihill, Brooke, Firpo, & Scaife, 2020), and it is endorsed by the Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute (National Cancer Institute, 2021).

The Yost Index is reported as a percentile score from 1 (most affluent) to 100 (most deprived). This level of precision has been criticized as potentially unwarranted, especially given that associated measures of uncertainty (e.g., 90% confidence interval) are rarely, if ever, reported (Donegan, Chun, & Griffith, 2021; Jung, Thill, & Issel, 2019; Logan et al., 2020; Spielman, Folch & Nagle 2014). The lone article we identified that focused on quantitative measures of uncertainty in derived census measures was by Napierala & Denton (2017), which used the margins of error published by the census to simulate the distribution of measures of racial segregation. Other papers have focused on methods for assessing which variables belong in an index (Spielman et al., 2020), comparisons between indexes (Temam et al., 2017), the degree to which area-based indexes match individual-level data (Bryere et al., 2017), and the choice of geographic scale (Schuurmann, Bell, Dunn, & Oliver, 2007; Cabrera-Barona, Wei, & Hagenlocher, 2016; Cebrecos, Domínguez-Berjón, Duque, Franco, & Escobar, 2018). With respect to determining which variables belong in an index, there is a divergence of opinion between using more objective approaches like factor analysis and principal components analysis and more subjective approaches like expert opinion (Spielman et al., 2020). In general, these papers have found substantially different results between indexes and within indexes when contributing variables and their weights are modified. There have also been several papers critiquing the concept of a socioeconomic index itself and how one can know that it is measuring the conditions it claims to measure (Gordon, 1995; Deas, Robson, Wong, & Bradford, 2003). The present paper is not concerned with these issues but rather simply with the problem of calculating a confidence interval around an established

* Corresponding author. Pumphandle, LLC, Camden, ME, USA.
E-mail address: francis.boscoe@pumphandle.consulting (F.P. Boscoe).

https://doi.org/10.1016/j.ssmph.2022.101078
Received 7 October 2021; Received in revised form 22 December 2021; Accepted 22 March 2022
Available online 17 May 2022

measure using the variance replicate tables that accompany the ACS. Our method is adaptable to any other index derived from the ACS.

2. Theory and calculation

We make use of the variance replicate estimates (VREs) which accompany the ACS (United States Census Bureau, 2020). The VREs consist of 80 replications of the data and are published for the most frequently used ACS tables. The variance and margin of error are obtained by comparing the 80 replicate values to the official estimate using the formulas:

$$\text{Variance} = \frac{4}{80}\sum_{i=1}^{80}(Replicate_i - Estimate)^2 \tag{1}$$

$$\text{Margin of Error (at 90\% confidence)} = \pm 1.645\sqrt{Variance} \tag{2}$$

The above formulas yield the exact margins of error that accompany the ACS data. They can also be used to find margins of error for derived variables, such as the medians and ratio measures used in the construction of the Yost Index, as well as the Yost Index itself. The rationale behind the formulas is given by Fay (1989) and Judkins (1990). In brief, each replicate involves randomly assigning half of the observations within each stratum of the data a weight of 1, one-quarter of the observations a weight of approximately 1.7, and one-quarter of the observations a weight of approximately 0.3. The measure of interest is then calculated, and the process is repeated 80 times. The mean square difference of all the replicate estimates is then directly proportional to the variance by Equation (1). The method is computationally more efficient than jackknife and bootstrap methods, a consideration that was important at the time the method was first developed. A detailed technical explanation is given in Chapter 12 of the ACS Design and Methodology manual (United States Census Bureau, 2014).

3. Material and methods

We obtained the variance replicate estimates at the census tract level for the fifty states, the District of Columbia, and Puerto Rico for 2015–2019 for the seven tables listed in Table 1. For the tables corresponding to the education, employment, and occupation variables we were able to follow the formal definitions published by the National Cancer Institute's SEER Program exactly (National Cancer Institute, 2021). For the poverty table, we used the percentage of households below the poverty level rather than the percentage of households below 150% of the poverty level, as the latter variable was not available in the VRE tables. For the income, rent, and house value variables, we derived the medians from the strata provided as they are not reported directly. This was achieved first by finding the stratum in which the median household fell, then interpolating within the stratum, assuming a uniform distribution. For example, in a calculation of median household income, if there were 4001 people in a particular census tract, we identified the census tract in which the 2000th person fell. Further suppose this was the $50,000 to $59,999 stratum and that this stratum included the 1900th through the 2200th persons. The 2000th person, being at the one-third position within this stratum, is assigned the corresponding income of $53,333.

This approach for calculating the median of stratified data is the same as that used by the Census Bureau, but with the benefit of access to its raw survey data, the Census Bureau uses much finer strata (United States Census Bureau, 2019). As a consequence, our calculated medians differ modestly from published values. For example, the median absolute difference in median household income between our values and the census values is 1% and the mean absolute difference is 3%. Some large differences occur within the uppermost income category, which in the publicly available data is top-coded as $250,000 and above but which the Census Bureau reports with ranges from $250,000 to over $1 million. These large differences are not important for our purposes as once they are converted into ranks they all fall well within the top 1% of median household incomes nationally.

We removed census tracts with fewer than 30 households due to measurement unreliability. This value corresponds roughly to a minimum population size of 100 that has been used previously (Diez Roux et al., 2001) given an average household size of about 3 (United States Census Bureau, 2021). We opted to use a household-based criterion as it also filters out tracts consisting of group quarters such as prisons and college dormitories where people do not live in households and socioeconomic measures are unreliable, if collected at all. A total of 72,793 census tracts met this definition out of 74,001 for which any data are reported. This captures 99.9% of the total United States population using the population for whom poverty status is determined as the denominator.

The factor analysis method we employ requires measurements for each of the component variables. One limitation of this approach is that over 1200 census tracts (about 1.7%) do not contain measures of both housing value and rent, either because all residents surveyed were homeowners or all residents surveyed were renters. To remedy this issue we combined the two measures into a single housing measure by standardizing the values and either taking their average or using the single measure that was available. We also used a single measure when either the number of renters or number of homeowners was below 30, treating that number as unreliable.

Each of the resulting six variables was ranked from 1 to 72,793 such that a rank of 1 corresponded to the highest socioeconomic position. For tied values, the average rank was assigned. A maximum likelihood factor analysis was performed on the ranks using the psych package in R (Revelle. psych, 2021). The ranks serve to standardize the data which varies by orders of magnitude between and within variables (Baxter, 1995; Conover & Iman, 1981). The first principal component, which explained 60% of the variance in the data, was retained for each census tract. The factor weights were similar to those previously reported in similar analyses, where median household income accounts for about half of the total weight and the variables collectively account for the other half (Boscoe et al., 2021; Yu et al., 2014). The resulting factor scores were grouped into percentiles with a value of 1 again corresponding to the highest socioeconomic position and a value of 100 the lowest. 90% confidence intervals around the index values were then calculated using the formulas presented above. We elected to abide by the Census Bureau practice of computing and reporting the 90% confidence intervals rather than the more common 95% confidence intervals. For the sake of readability, we have limited discussion of the technical details here and refer the reader to the published code (Boscoe et al., 2021).

4. Results

The mean margin of error at the 90% confidence level for the 72,793 census tracts was 8.2, meaning that a typical census tract can be thought of as having a Yost index that is precise to about plus or minus 8 points (Fig. 1). The interquartile range was 5.5–10.5. At the extremes, 10 tracts had a margin of error greater than 40, 49 tracts had a difference greater than 30, and 452 tracts had a difference of 0. All but one of the tracts with a difference of 0 had a Yost index of either 1 or 100; the exception had a Yost index of 3.

Table 1
Census tables used in the construction of the yost index.

B15002	Sex by educational attainment for the population aged 25 years and older
B17001	Poverty status by sex by age
B19001	Household income
B23025	Employment status
B25063	Gross rent
B25075	Value (of house)
C24010	Sex by occupation for the civilian employed population 16 years and older

To illustrate these relationships further, we focused on Chatham County, Georgia, which contained a census tract (tract number 101.01) with a 90% confidence interval which spanned nearly the entire range of values, from 2 to 100. Fig. 2 depicts the point estimates of Yost index rank and their 90% confidence intervals for each of 69 census tracts in this county. The results are mapped in Fig. 3, with the left map showing the point estimates and the right map showing the margins of error. (The two tracts labeled as having insufficient data comprise Hutchinson Island, a nearly uninhabited island in the Savannah River that is primarily used for industry and, more recently, recreation. The island was well-populated prior to its conversion to heavy industry during World War II and its original census tract designations have never been altered).

Results for the entire United States are summarized in Fig. 4. Each Yost Index position on the x-axis represents 727 or 728 census tracts. The black horizontal lines correspond to the median margins of error, the gray boxes to the interquartile ranges, and the whiskers to 1.5 times the interquartile ranges. The margins of error are seen to be related to the Yost Index itself, with values as high as 11 near the middle of the distribution and close to 0 at either extreme. The plot is also skewed such that the margins of error are greater among the less-affluent relative to the corresponding position among the more-affluent. The peak margins of error are for a Yost Index of 56.

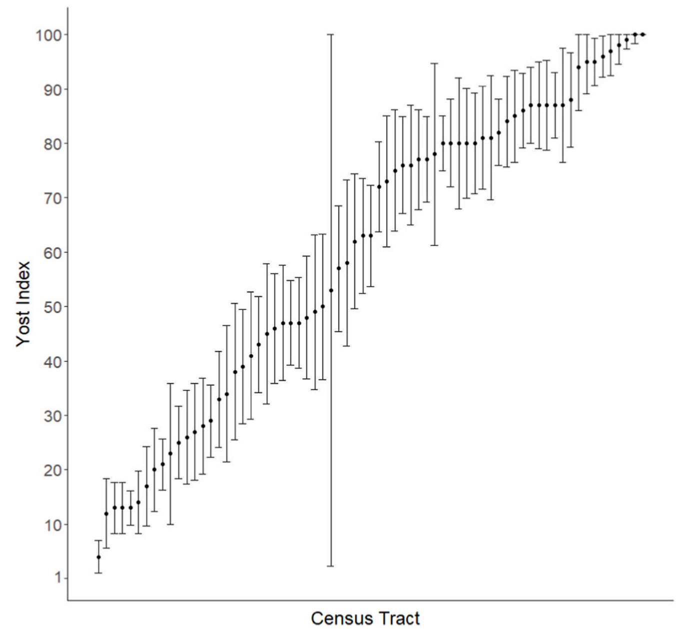

Fig. 2. Distribution of Yost Index values and 90% confidence intervals for 69 census tracts in Chatham County, Georgia. One tract has confidence intervals which span nearly the entire range.

5. Discussion

The Yost Index is precise to within one decile for 71% of census tracts in the United States and precise to within two deciles for 99% (Fig. 1). While these results seem reasonable, they are not ignorable. Past studies that have ignored this error are at risk of having overstated the relationship between socioeconomic status and the health outcomes of interest. Given the magnitude of the margin of error, there is no danger of misclassifying a highly affluent neighborhood as middle-class, a middle-class neighborhood as poor, and so on, as a result of sampling error. But the idea that fine gradations can be made between neighborhoods with similar characteristics through a percentile-based index is not supported by our results.

For a small handful of tracts, as in tract 101.01 in Chatham County, Georgia which stood out in Figs. 2 and 3, the Yost Index is not informative at all. This tract along the Savannah River is dominated by heavy industry. The small resident population is divided between a poor, primarily African-African community on its western edge and a recent upscale, mixed-use development along the river. This development, known as the Eastern Wharf, did not exist when the tract boundaries were last defined in conjunction with the 2010 census. Consequently, roughly half of this tract has low SES values and roughly half has high SES values, with little in between. In the variance replicate estimates, the median person randomly falls in either group, rendering the medians highly unstable, which carries through to the Yost Index. This represents not a shortcoming of the index, but rather a reflection of a highly heterogeneous environment that is atypical of census geographies. While situations like this are rare, with only 0.07% of the census tracts in the entire nation having a margin of error wider than three deciles, without consideration of uncertainty it is not possible to know where they are and have the ability to exclude them from analyses.

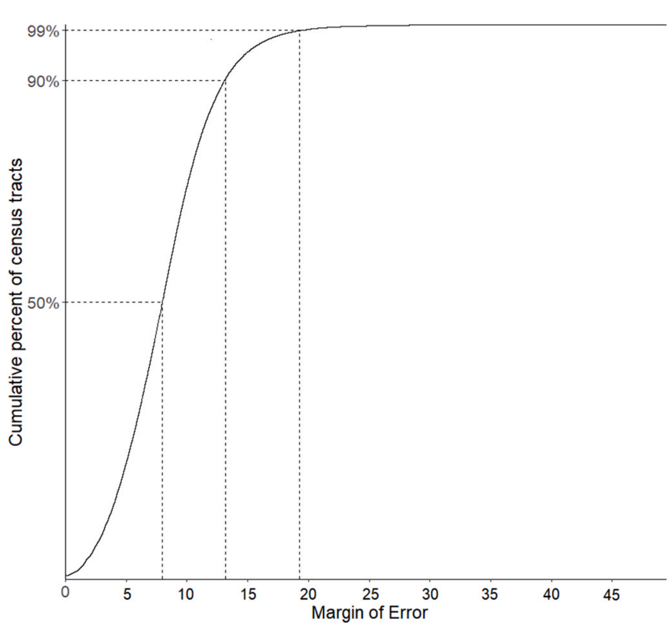

Fig. 1. Cumulative distribution plot of the margin of error for 71,570 United States census tracts.

The uncertainty of the Yost Index was seen to be associated with the index itself, with values near the middle of the distribution being the least precise and values at the extremes being the most precise. This arises for two reasons: first, the definition of a percentile itself is variable. For example, the fifth percentile of median household income ranges from approximately $130,100 to $136,400 while the fiftieth percentile ranges from $59,200 to $59,800 and the ninety-fifth percentile ranges from $26,800 to $28,300. A given amount of sampling error matters more in the middle of the distribution where there are larger numbers of similar values. In addition, there are many more ways for a multi-variable index to come up with a value near the middle: all contributing variables can have ranks near the middle, or a mixture of high, middle, and low ranks can average to a rank in the middle. For values near the extremes of the distribution, in contrast, the ranks of the contributing variables cannot exhibit much variation, by definition. The visible skewness in Fig. 4, indicating that less-affluent tracts have more uncertainty than corresponding more-affluent tracts, is consistent with survey research that routinely finds lower response rates among poorer populations (Spielman et al., 2014).

Some data limitations may have impacted our results. These include the minor adjustments we had to make to the Yost Index to accommodate the variance replicate estimates: the use of the poverty rate rather than 150% of the poverty rate, calculating medians from strata less precise than those used by the Census Bureau, and combining the two

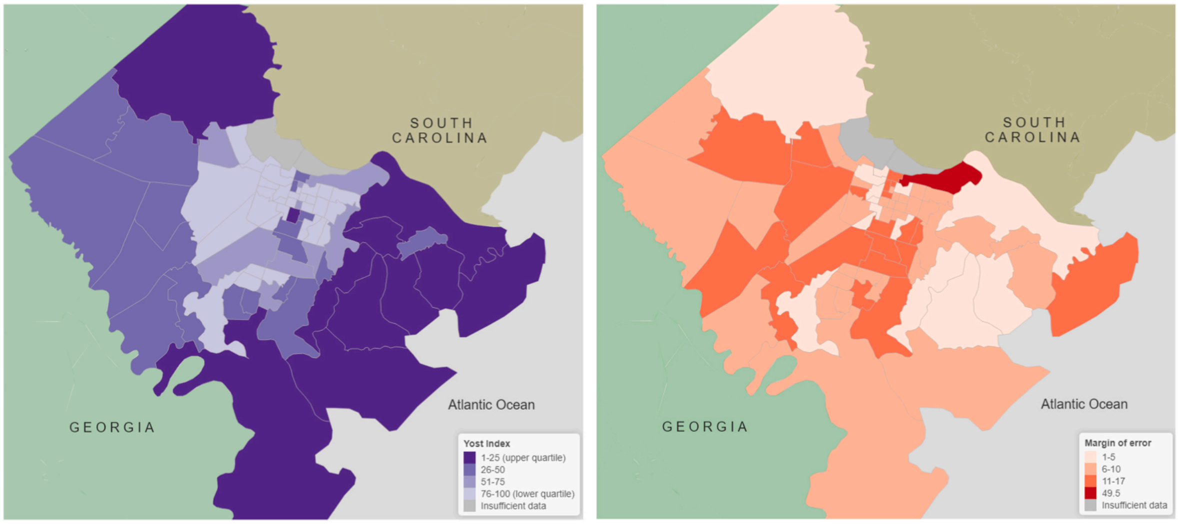

Fig. 3. Maps of census tracts in Chatham County, Georgia. The city of Savannah comprises the smaller tracts at the center of the maps. Left: shaded by Yost Index, based on 2015–2019 American Community Survey data. Right: Same data, shaded by the margin of error.

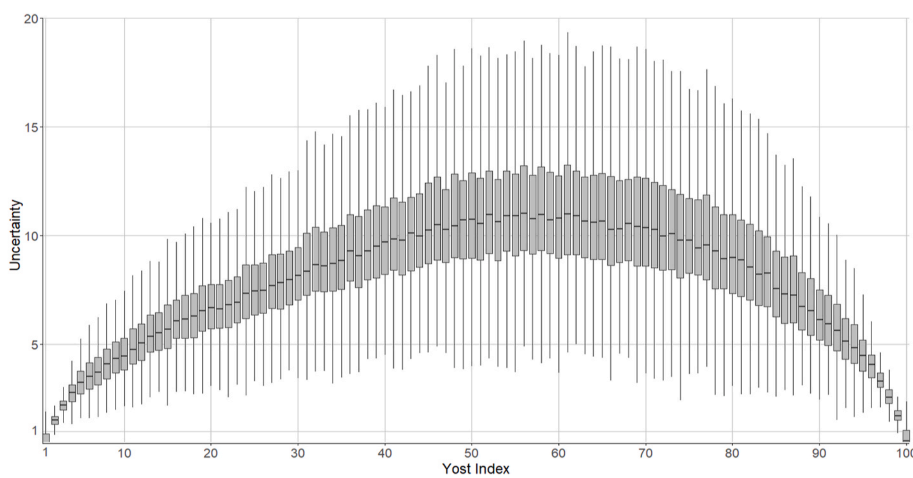

Fig. 4. Box-and-whiskers plot of margin of error as a function of Yost index for 72,793 United States census tracts. Black horizontal lines correspond to median values, gray boxes to the interquartile range, and whiskers to 1.5 times the interquartile range.

housing variables into a single measure prior to the factor analysis. In addition, the variance replicate estimates are intended to be used within a single sample universe, such as population aged 16 and over or households. The seven tables used as inputs to the Yost Index use seven different universes. Consequently, the covariances between these tables are likely lower than what they would be if they were all from the same universe. Collectively these adjustments yield less-precise estimates than would be obtained if we had unfettered access to the raw ACS data, hence our results are likely conservative.

6. Conclusion

As we have shown, the uncertainty in a census-based socioeconomic index is calculable and can be incorporated into any analysis using such an index. While the process is not mathematically complex, the use of the many VRE tables makes the calculation labor-intensive, and so we have made our results available for direct download from a data repository, along with the R code used to generate the index values and all figures (Boscoe et al., 2021).

Funding

This work was supported in part by a grant from the National Cancer Institute (1R21CA235153\05501).

Ethical statement

Research reported in this publication was supported by grant number 1R21CA235153\05501 from the National Institutes of Health, National Cancer Institue (NIH/NCI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the authors' institutions. The authors report no conflicts of interest.

Declaration of competing interest

None.

References

Abdel-Rahman, O. (2019). Outcomes of non-metastatic colon cancer patients in relationship to socioeconomic status: An analysis of SEER census tract-level socioeconomic database. *International Journal of Clinical Oncology, 24*, 1582–1587. https://doi.org/10.1007/s10147-019-01497-9

Baxter, M. J. (1995). Standardization and transformation in principal component analysis, with applications to archaeometry. *Applied Statistics - Journal of the Royal Statistical Society C, 44*(4), 513–527. https://doi.org/10.2307/2986142

Boscoe, F. P., Liu, B., Lafantasie, J., Niu, L., & Lee, F. F. (2021). Yost Index with 90% confidence intervals. *Data file*. https://doi.org/10.6084/m9.figshare.16649773

Bryere, J., Pornet, C., Copin, N., Launay, L., Gusto, G., Grosclaude, P., et al. (2017). Assessment of the ecological bias of seven aggregate social deprivation indices. *BMC Public Health, 17*, 1–7. https://doi.org/10.1186/s12889-016-4007-8

Cabrera-Barona, P., Wei, C., & Hagenlocher, M. (2016). Multiscale evaluation of an urban deprivation index: Implications for quality of life and healthcare accessibility planning. *Applied Geography, 70*, 1–10. https://doi.org/10.1016/j.apgeog.2016.02.009

Carlson, L. C., Kim, J., Samuels-Kalow, M. E., Yun, B. J., Terry, D. F., Weilburg, J. B., et al. (2021). Comparing neighborhood-based indices of socioeconomic risk factors and potentially preventable emergency department utilization. *American Journal of Emergency Medicine, 44*, 213–219. https://doi.org/10.1016/j.ajem.2020.03.035

Cebrecos, A., Domínguez-Berjón, M. F., Duque, I., Franco, M., & Escobar, F. (2018). Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research. *Applied Geography, 95*, 9–18. https://doi.org/10.1016/j.apgeog.2018.04.001

Conover, W. J., & Iman, R. L. (1981). Rank transformation as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*, 124–129. https://doi.org/10.1080/00031305.1981.10479327

Deas, I., Robson, B., Wong, C., & Bradford, M. (2003). Measuring neighbourhood deprivation: A critique of the index of multiple deprivation. *Environment and Planning C: Government and Policy, 21*, 883–903. https://doi.org/10.1068/c0240

Diez Roux, A. V., Kiefe, C. I., Jacobs, D. R., Haan, M., Jackson, S. A., Nieto, F. J., et al. (2001). Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiological studies. *Annals of Epidemiology, 11*, 395–405. https://doi.org/10.1016/s1047-2797(01)00221-6

Donegan, C., Chun, Y., & Griffith, D. A. (2021). Modeling community health with area data: Bayesian inference with survey standard errors and spatial structure. *International Journal of Environmental Research and Public Health, 18*, 6856. https://doi.org/10.3390/ijerph18136856

Fay, R. E. (1989). Theory and application of replicate weighting for variance calculations. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 212–217. Retrieved from www.asasrms.org/Proceedings/papers/1989_033.pdf. (Accessed 16 July 2021).

Gordon, D. (1995). Census based deprivation indices: Their weighting and validation. *Journal of Epidemiology & Community Health, 49*, S39–S44. https://doi.org/10.1136/jech.49.suppl_2.s39

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics, 6*, 223–239.

Jung, P. H., Thill, L. C., & Issel, M. (2019). Spatial autocorrelation and data uncertainty in the American community survey: A critique. *International Journal of Geographical Information Science, 33*, 1155–1175. https://doi.org/10.1080/13658816.2018.1554811

Li, Y., Huang, D., Wang, B. X., Mao, W., Chen, X. W., & Dong, P. (2021). Socioeconomic factors are associated with the prognosis of thyroid cancer. *Journal of Cancer, 12*, 2507–2512. https://doi.org/10.7150/jca.52329

Logan, J. R., Bauer, C., Ke, J., Xu, H. W., & Li, F. (2020). Models for small area estimation for census tracts. *Geographical Analysis, 52*, 325–350. https://doi.org/10.1111/gean.12215

Lopez-De Fede, A., Stewart, J. E., Hardin, J. W., & Mayfield-Smith, K. (2016). Comparison of small-area deprivation measures as predictors of chronic disease burden in a low-income population. *International Journal for Equity in Health, 15*, 8910.1186/s12939-016-0378-9.

Mehaffey, J. H., Hawkins, R. B., Charles, E. J., Turrentine, F. E., Hallowell, P. T., Friel, C., et al. (2020). Socioeconomic "distressed communities index" improves surgical risk adjustment. *Annals of Surgery, 271*, 470–474. https://doi.org/10.1016/j.jamcollsurg.2018.07.243

National Cancer Institute. (2021). *Surveillance, Epidemiology and End results program. Time-dependent county attributes*. Retrieved from https://seer.cancer.gov/seerstat/variables/countyattribs/time-dependent.html. (Accessed 15 July 2021).

Patel, S. A., Sherman, C. G., Khatry, S. K., Leclerq, S. C., Katz, J., Tielsch, J. M., et al. (2016). An index of community-level socioeconomic composition for global health research. *Social Indicators Research, 129*, 639–658. https://doi.org/10.1007/s11205-015-1114-3

Rajeshuni, N., Zubair, T., Ludwig, C. A., Moshfeghi, D. M., & Mruthyunjaya, P. (2020). Evaluation of racial, ethnic and socioeconomic associations with treatment and survival in uveal melanoma, 2004-2014. *JAMA Opthalmology, 138*, 876–884. https://doi.org/10.1001/jamaophthalmol.2020.2254

Revelle psych, W. (2021). *Procedures for psychological, psychometric, and personality research*. Retrieved from version 2.1.6. https://CRAN-R-project.org/package=psych. (Accessed 18 July 2021).

Ross, J., Braswell, K. V., da Silva, L. M., Mujica, F., Stutsman, S., Finan, M. A., et al. (2017). Unraveling the etiology of ovarian cancer racial disparity in the deep south: Is it nature or nurture? *Gynecologic Oncology, 145*, 329–333. https://doi.org/10.1016/j.ygyno.2017.02.025

Schuurmann, N., Bell, N., Dunn, J. R., & Oliver, L. (2007). Deprivation indices, population health and geography: An evaluation of the spatial effectiveness of indices at multiple scales. *Journal of Urban Health, 84*, 591–603. https://doi.org/10.1007/s11524-007-9193-3

Spielman, S. E., Folch, D., & Nagle, N. (2014). Patterns and causes of uncertainty in the American community survey. *Applied Geography, 41*, 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002

Spielman, S. E., Tuccillo, J., Folch, D. C., Schweikert, A., Davies, R., Wood, N. J., et al. (2020). Evaluating social vulnerability indicators: Criteria and their application to the SVI. *Natural Hazards, 100*, 417–436. https://doi.org/10.1007/s11069-019-03820-z

Swords, D. S., Mulvihill, S. J., Brooke, B. S., Firpo, M. A., & Scaife, C. L. (2020). Size and importance of socioeconomic status-based disparities in use of surgery in nonadvanced stage gastrointestinal cancers. *Annals of Surgical Oncology, 27*(2), 333–341. https://doi.org/10.1245/s10434-019-07922-7

Temam, S., Varraso, R., Pornet, C., Sanchez, M., Affret, A., Jacquemin, B., et al. (2017). Ability of ecological deprivation indices to measure social inequalities in a French cohort. *BMC Public Health, 17*, 1–11. https://doi.org/10.1186/s12889-017-4967-3

United States Census Bureau. (2014). *American community survey Design and methodology*. Retrieved from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html. (Accessed 17 July 2021).

United States Census Bureau. (2019). *American community survey and Puerto Rico community survey 2019 subject definitions*. Retrieved from https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2019_ACSSubjectDefinitions.pdf. (Accessed 20 December 2021).

United States Census Bureau. (2020). *Documentation for the 2015-2019 variance replicate estimates tables*. Retrieved from https://www2.census.gov/programs-surveys/acs/replicate_estimates/2019/documentation/5-year/2015-2019_Variance_Replicate_Tables_Documentation.pdf. (Accessed 19 July 2021).

United States Census Bureau. (2021). *Quick facts, United States*. Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045219. (Accessed 20 July 2021).

Webb, S., Janus, M., Duku, E., Raos, R., Brownell, M., Forer, B., et al. (2017). Neighbourhood socioeconomic status and early childhood development. *SSM-Population Health, 3*, 48–56. https://doi.org/10.1016/j.ssmph.2016.11.006

Yost, K., Perkins, C., Cohen, R., Morris, C., & Wright, W. (2001). Socioeconomic status and breast cancer incidence in California for different race/ethnic groups Cancer. *Causes & Control, 12*, 703–711. https://doi.org/10.1023/A:1011240019516

Yu, M. D., Tatalovich, Z., Gibson, J. T., & Cronin, K. A. (2014). Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer Causes & Control, 25*, 81–92. https://doi.org/10.1007/s10552-013-0310-1

Napierala J, N. Denton (2017). Measuring residential segregation with the ACS: How the margin of error affects the dissimilarity index. Demography, 54(1), p.p. 285-309, 10.1007/s13524-016-0545-z.