Original Article

Towards the Use of a Census Tract Poverty Indicator Variable in Cancer Surveillance

Francis P. Boscoe, PhD[a]

Abstract: Incidence rates for many cancer sites are strongly correlated with area measures of socioeconomic conditions such as poverty rate. Analyzing such measures at the county scale produces misleading results by masking enormous within-county variations. The census tract is a more suitable scale for assessing the relationship between cancer and socioeconomics. The North American Association of Central Cancer Registries (NAACCR) developed a census tract-level poverty indicator variable which was included as an optional item in its 2010 Call for Data. This variable does not allow the identification of individual census tracts as long as the county of diagnosis is not known. It is expected that this data item will be made available to researchers in future releases of the CINA Deluxe file.

Key words: *call for data, census tract, disclosure risk, poverty*

Introduction

The North American Association of Central Cancer Registries (NAACCR) included an optional census tract-level poverty indicator variable in its 2010 Call for Data. The purpose of this data item is to provide a measure of local socioeconomic conditions for each cancer case that can be made available to researchers. The socioeconomic environment directly influences cancer rates and can confound other etiologic studies of cancer. This relationship has been well established, though attention has largely been limited to the more common sites of cancer. The monograph by Singh et al,[1] for example, was limited to all cancers combined and 6 individual cancer sites (lung, colorectal, prostate, female breast, cervical, and melanoma). As the relationship between socioeconomic status and cancer is dynamic and can vary by geographic location, it requires ongoing surveillance and study-specific measurement. Lung cancer, for example, was historically associated with higher socioeconomic status but since the 1980s has been associated with lower socioeconomic status,[2] but this relationship varies by race, ethnicity, and geography.

Most cancer epidemiology studies avail themselves of county-level measures of socioeconomic status, as these are relatively easily obtained.[3-5] While well-intentioned, the coarseness of this scale can result in biased findings. One only need consider any large urban county to see the problems inherent in using a county-level variable—assigning identical codes to each of the millions of people living in each of the hundreds or thousands of neighborhoods in Los Angeles County, Manhattan, or Miami-Dade County is obviously flawed. In general, using large and heterogeneous geographic areas for analysis obscures important relationships, sometimes even to the point of reversing the apparent direction of association.[6]

Census tracts, in contrast, are a useful scale at which to identify social gradients in health.[7-10] A census tract is formally defined as a small, relatively permanent statistical subdivision of a county with an optimum size of 4000 people and designed to be relatively homogeneous with respect to population characteristics, economic status, and living conditions.[11] In urban settings, it roughly equates to a neighborhood. As an ecologic unit, census tracts still pose potential inferential problems, but their size and homogeneity make these issues far more manageable.

There are many ways of measuring socioeconomic status, including measures of poverty, education, income, substandard housing, or indexes that combine multiple variables. Of these, poverty rate has been found to be singularly effective, both for its simplicity and ability to capture variations in the health of populations.[1,12] A tract-level poverty rate is properly viewed not as a proxy for an individual's poverty status, but rather as a useful measure of environmental context.

The NAACCR poverty indicator variable assigns each cancer case to 1 of 5 poverty rate categories: less than 5%, 5% to less than 10%, 10% to less than 20%, 20% and above, and undefined. (The latter category applies to rare instances of census tracts with populations but no sampled households, as in some university campuses or prisons, or census tracts with no population at the time of the decennial census but with residents before or after, as with large urban renewal projects. Because this category adds no useful information about local socioeconomic conditions, it would be omitted from any data file made available by NAACCR to researchers.) A SAS program available on the NAACCR Web site allows registrars to assign this code to their own cases.[13] This data element can thus be derived and transmitted without the need to also transmit census tract, which is of concern to some state cancer registries because of potential disclosure risk.

This paper describes how this variable will be useful to researchers and demonstrates how it does not present a disclosure risk, so long as the county of diagnosis is not made available simultaneously.

"Towards the Use of a Census Tract Poverty Indicator Variable in Cancer Surveillance"

[a]New York State Cancer Registry, New York State Department of Health, Albany, NY.

Address correspondence to Francis P. Boscoe, PhD, Research Scientist, New York State Cancer Registry, Riverview Center, 150 Broadway, Suite 361, Menands, NY 12204. Email: fpb01@health.state.ny.us.

Methods and Materials

There were 2 methodological objectives: first, to illustrate how the census tract poverty rate indicator variable highlights substantial differences in cancer risk by cancer site, and second, to assess the potential for disclosure risk. To meet the first objective, the census tract poverty rate indicator variable was assigned to all cancer cases among white non-Hispanics diagnosed between 2003 and 2007 in New York State (n=382,285). White non-Hispanics were selected to minimize confounding by race and ethnicity. Census tracts were available for over 99% of the cases, with the remaining values imputed using a previously published method.[14] Age-adjusted rates standardized to the 2000 US population were calculated by site and poverty category for each site and site grouping listed in the SEER (Surveillance, Epidemiology and End Results Program) ICD-O3 Site Recode table.[15] The rate ratio of living in the highest-poverty category (poverty rate of 20% or higher) to the lowest-poverty category (less than 5%) was calculated for each site. The process was then repeated at the county level. As New York only has a single county with a poverty rate below 5% (Putnam), the cut point for the lowest-poverty category was relaxed to 6% to allow the inclusion of 3 additional counties (Nassau, Suffolk, and Saratoga). There were 3 counties above 20% poverty (Bronx, Brooklyn, and Manhattan).

The potential for disclosure risk was measured by cross-tabulating states (including the District of Columbia), counties, census tracts, and their associated poverty indicator values to determine the number of instances where the census tract of an individual case could be identified. This is well-illustrated through the example of St. Lawrence County, New York, a sparsely populated rural county bordering Canada. St. Lawrence County contains 1 tract with a poverty rate below 5%, 1 that is between 5 and 10%, 1 that is undefined because of an absence of households, and 25 others with poverty rates over 10%. The combination of county and poverty rate can thus potentially identify 3 distinct census tracts, 2 of which would potentially be available to researchers.

Results

Table 1 lists age-adjusted incidence rate ratios and 95% confidence intervals between the highest-poverty and lowest-poverty census tracts for non-Hispanic whites for numerous cancer sites. The table includes all of the most common cancer sites along with several selected subsites and rare sites with unusually high or low values, listed in order by SEER ICD-O3 Site Recode. The table reveals that the number of sites and subsites elevated among residents of the highest-poverty census tracts is twice that of the lowest-poverty census tracts. This is counterbalanced by the fact that several of the most common sites (specifically, prostate, female breast, and melanoma) have higher rates among residents of the lowest-poverty census tracts. For all cancers combined, rates are just 4% higher among residents of the highest-poverty census tracts. When these tract-level results are compared with county-level results, major differences are evident among several the most common sites (Table 2).

Table 1. Age-adjusted cancer incidence rate ratios for the most common cancer sites and other selected sites, highest-poverty census tracts to lowest-poverty census tracts, New York State, white non-Hispanics, 2003–2007.

Site	*Rate Ratio (95% CI)*
All invasive malignant tumors	1.04 (1.02–1.05)
Oral cavity and pharynx	1.41 (1.29–1.52)
Oral cavity	1.20 (1.08–1.33)
Pharynx	1.89 (1.63–2.18)
Esophagus	1.19 (1.06–1.33)
Stomach	1.58 (1.45–1.72)
Colorectal	1.24 (1.19–1.28)
Anus, anal canal and anorectum	2.10 (1.73–2.51)
Liver and IBD	1.62 (1.46–1.80)
Pancreas	1.05 (0.98–1.13)
Larynx	1.77 (1.56–2.01)
Lung and bronchus	1.26 (1.22–1.30)
Melanoma of the skin	0.56 (0.52–0.61)
Female breast	0.90 (0.87–0.93)
Cervix uteri	1.79 (1.55–2.04)
Corpus uterus and NOS	1.17 (1.09–1.25)
Ovary	1.08 (0.98–1.19)
Vagina	1.64 (0.95–2.58)
Prostate	0.78 (0.76–0.81)
Testis	0.94 (0.80–1.08)
Penis	1.73 (1.01–2.65)
Urinary bladder	0.89 (0.84–0.94)
Kidney and renal pelvis	1.09 (1.02–1.16)
Ureter	0.74 (0.51–1.00)
Other urinary organs	0.51 (0.23–0.86)
Brain and other nervous system	1.06 (0.95–1.17)
Cranial nerves/other nervous system	1.55 (1.09–2.13)
Thyroid	0.88 (0.81–0.96)
Hodgkin lymphoma	1.06 (0.91–1.22)
Non-Hodgkin lymphomas	0.95 (0.89–1.00)
Multiple myeloma	1.04 (0.93–1.16)
Leukemia	1.04 (0.97–1.11)
Mesothelioma	0.75 (0.54–0.77)
Kaposi sarcoma	4.18 (2.99–5.88)
Miscellaneous	1.26 (1.19–1.33)

Table 2. Age-adjusted cancer incidence rate ratios for selected sites, census-tract-derived versus county-derived, New York State, white non-Hispanics, 2003–2007.

Site	*Census tract-derived rate ratio (95% CI)*	*County-derived rate ratio (95% CI)*
Oral cavity and pharynx	1.41 (1.29–1.52)	0.88 (0.81–0.95)
Colorectal	1.24 (1.19–1.28)	0.92 (0.89–0.95)
Lung and bronchus	1.26 (1.22–1.30)	0.79 (0.77–0.82)
Melanoma of the skin	0.56 (0.52–0.61)	0.84 (0.80–0.89)
Prostate	0.78 (0.76–0.81)	0.71 (0.69–0.74)

Specifically, the positive associations between poverty rate and oral, colorectal, and lung cancers are all reversed.

The disclosure-risk analysis reveals 1833 census tracts (2.8% of the nationwide total) that would be identifiable in combination with knowledge of county. This includes 205 census tracts which are coterminous with counties; if these are excluded, then the number is 1,628 (2.5%). Given that census tracts are roughly equal in population, this implies that the fraction of cancer cases with an identifiable census tract would also be around 2.5%. However, this subset of census tracts includes many with younger populations at lower risk for cancer, such as universities, Indian reservations, and military bases (2 of the unique tracts in St. Lawrence County describe university campuses). Thus, the total fraction of cancer cases impacted nationally is likely well below 2.5%; in New York State, it is under 0.3%. There are no census tracts that would be identifiable in combination with knowledge of state. Every state has at least 2 census tracts in each of the poverty rate categories.

Discussion

Cancer sites with rates that are elevated among patients residing in census tracts with the highest poverty rates include many associated with smoking (head and neck, stomach, colorectal, liver, lung, female reproductive sites),[16] alcohol consumption (head and neck, colorectal, liver),[17] and sexually transmitted viruses (cervix, head and neck, anogenital, Kaposi's sarcoma).[18] However, not all such sites show statistically significant elevations or even elevations, as with bladder cancer, which is associated with smoking but also with independent occupational and lifestyle factors.[19] Sites that are elevated among residents of the lowest-poverty census are less easy to summarize in terms of shared risk factors, but tend to be characterized by extremely high relative survival, as with breast, thyroid, prostate, and melanoma,[20] suggesting a role of better access to health care for this group. The association between the cranial nerves/other nervous system category and poverty has not been widely identified in the literature, if at all.

But rather than attempt to interpret each of these findings, the main point is simply to illustrate that there are strong associations between socioeconomic status and cancer that exist for many cancer sites, and these are often uncontrolled for or insufficiently controlled for in analyses. When analyzed at the county scale, these relationships can be highly distorted, even reversing the direction of association, as seen for several sites in Table 2. This is a direct consequence of the severe misclassification of poverty that occurs when areas as large and diverse as Manhattan and Brooklyn and the 2 counties comprising Long Island are each classified with a single poverty value. Manhattan, in particular, is counted in the highest-poverty category even though it includes neighborhoods among the wealthiest in the world.

The proposed mechanism for making this data available to researchers is through the CINA (Cancer in North America) Deluxe Analytic File.[21] This file consists of data from 1995 onward from registries which met specific quality standards for each year of data included. To gain access to this file, researchers must submit an application to NAACCR which goes through a review and approval process. Individual registries then grant access to their own data on a project-specific basis. Based on past experience, a large majority of eligible registries consent to most projects. In the case of the census tract poverty indicator, initial participation may be below average because of inadequate geocoding, but a recent analysis by Singh et al finds such states to be in the minority.[22] Geocoding has become dramatically easier and less expensive in recent years, and more and more states are geocoding their cases on a routine basis.[23]

Restricting the simultaneous availability of county and the census tract poverty indicator on this file will minimize disclosure risk by making it impossible to identify the exact census tract for any cancer case. While Howe et al have proposed an acceptable threshold up to 5% record uniqueness in public-use data files,[24] in practice there is little tolerance for any record uniqueness when small geographic units are involved.

Census tract poverty indicator values assigned to 2004–2008 cases will be based on poverty rates from the 2000 census, but in future years will be based on an exact temporal match. Beginning in the winter of 2010–2011, the US Census Bureau's American Community Survey will begin annual releases of poverty rates by census tract averaged over a 5-year period which will correspond with the most recent 5 years of cancer data. This means that analysis of 2005–2009 cancer data will make use of poverty rates for 2005–2009, and so on. This added temporal precision will make this data item even more useful.

In summary, the census tract poverty indicator variable being introduced in the NAACCR's 2010 Call for Data has the promise of becoming a standard item in the cancer epidemiologist's tool kit, promising a better understanding of the relationship between local socioeconomic conditions and cancer incidence and mortality. Moreover, it will permit better control of confounding in etiologic studies generally.

The application provided here using New York State data was intended as a quick and coarse demonstration of its utility. Future researchers will be able to enhance these results through the inclusion of additional registries, race and ethnic groups, confounding variables, and time periods.

Acknowledgments

The author thanks the members of the NAACCR Data Use and Research Committee, particularly Andy Lake, Maria Schymura, and Xiao-Cheng Wu, for their general support.

References

1. Singh GK, Miller BA, Hankey BF, Edwards BK. *Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975–1999*. Bethesda, MD: National Cancer Institute; 2003.
2. Singh GK, Miller BA, Hankey BF. Changing area socioeconomic patterns in US cancer mortality, 1950-1998: Part II—Lung and colorectal cancers. *J Natl Cancer Inst*. 2002;94(12):916–925.
3. Hausauer AK, Keegan THM, Chang ET, Glaser SL, Howe H, Clarke CA. Recent trends in breast cancer incience in US white women by county-level urban/rural and poverty status. *BMC Med*. 2009;7:31.
4. Hao YP, Jemal A, Zhang XY, Ward EM. Trends in colorectal cancer incidence rates by age, race/ethnicity, and indices of access to medical care, 1995-2004 (United States). *Cancer Causes Control*. 2009;20(10):1855–1863.
5. Boscoe FP, Schymura MJ. Solar ultraviolet-B exposure and cancer incidence and mortality in the United States, 1993-2002. *BMC Cancer*. 2006;6:264.
6. Greenland S, Robins J. Ecologic Studies—Biases, Misconceptions, and Counterexamples. *Am J Epidemiol*. 1994;139(8):747–760.
7. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health*. 2003;57(3):186–199.
8. Byers TE, Wolf HJ, Bauer KR, et al. The impact of socioeconomic status on survival after cancer in the United States: findings from the National Program of Cancer Registries Patterns of Care Study. *Cancer*. 2008;113(3):582–591.
9. Abe T, Martin IB, Roche LM. Clusters of census tracts with high proportions of men with distant-stage prostate cancer incidence in New Jersey, 1995 to 1999. *Am J Prev Med*. 2006;30(2):S60–S66.
10. Rehkopf DH, Haughton LT, Chen JT, Waterman PD, Subramanian SV, Krieger N. Monitoring socioeconomic disparities in death: comparing individual-level education and area-based socioeconomic measures. *Am J Public Health*. 2006;96(12):2135–2138.
11. Kreiger N. A century of census tracts: health & the body politic (1906–2006). *J Urban Health*. 2006;83(3):355–361.
12. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *Am J Epidemiol*. 2002;156(5):471–482.
13. North American Association of Central Cancer Registries. *Data Analysis Tools. Poverty and Census Tract Linkage Program*. Available at: **http://www.naaccr.org/Research/DataAnalysisTools.aspx.** Accessed October 28, 2010.
14. Henry KA, Boscoe FP. Estimating the accuracy of geographical imputation. *Int J Health Geographics*. 2008;7:3.
15. Surveillance Epidemiology and End Results (SEER) Program. *SEER Site Recode ICD-O-3 (1/27/2003) Definition*. Available at: **http://seer.cancer.gov/siterecode/icdo3_d01272003/.** Accessed August 1, 2010.
16. Secretan B, Straif K, Baan R, et al. A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol*. 2009;10(11):1033–1034.
17. Baan R, Straif K, Grosse Y, et al. Carcinogenicity of alcoholic beverages. *Lancet Oncol*. 2007;8(4):292–293.
18. zur Hausen H. Papillomaviruses in the causation of human cancers—a brief historical account. *Virology*. 2009;384(2):260–265.
19. Scelo G, Brennan P. The epidemiology of bladder and kidney cancer. *Nature Clin Practice Urol*. 2007;4(4):205–217.
20. Verdecchia A, Francisci S, Brenner H, et al. Recent cancer survival in Europe: a 2000–02 period analysis of EUROCARE-4 data. *Lancet Oncol*. 2007;8(9):784–796.
21. North American Association of Central Cancer Registries. *CINA Deluxe Analytic File*. Available at: **http://www.naaccr.org/Research/CINADeluxe.aspx.** Accessed August 1, 2010.
22. Singh S, Ajani UA, Wilson RJ, Thomas CC. Evaluating quality of census tract data in the NPCR dataset. *J Registry Management*. 2010;36(4):143–147.
23. Goldberg DW. *A Geocoding Best Practices Guide*. Springfield, IL: North American Association of Central Cancer Registries; 2008.
24. Howe HL, Lake AJ, Shen T. Method to assess identifiability in electronic data files. *Am J Epidemiol*. 2007;165(5):597–601.