Neighborhood-level factors contribute to disparities in hepatocellular carcinoma incidence in Texas

Abiodun O. Oluyomi[1,2,3], **Hashem B. El-Serag**[2,4,5,6], **Adegboyega Olayode**[7], **Aaron P. Thrift**[1,2]

[1]Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, Texas

[2]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas

[3]Gulf Coast Center for Precision Environmental Health, Baylor College of Medicine, Houston, Texas

[4]Section of Gastroenterology and Hepatology and Clinical Epidemiology and Comparative Effectiveness Program in the Health Services Research, Michael E. DeBakey VA Medical Center, Baylor College of Medicine, Houston, Texas

[5]Clinical Epidemiology and Comparative Effectiveness Program, Section of Health Services Research, Michael E. DeBakey VA Medical Center, Baylor College of Medicine, Houston, Texas

[6]Texas Medical Center Digestive Diseases Center, Houston, Texas

[7]Division of Hospital Medicine, Emory University School of Medicine, Atlanta, Georgia

Abstract

Background: Texas has the highest hepatocellular carcinoma (HCC) incidence rates in the continental U.S., but these rates vary by race/ethnicity. We examined racial/ethnic disparities through a geospatial analysis of the social determinants of health.

*corresponding author: Abiodun O. Oluyomi, PhD, oluyomi@bcm.edu, Baylor College of Medicine, No. 1 Baylor Plaza, BCM Jewish Institute for Research Bldg., Room 607D, (MS BCM307), Houston, TX 77030, USA, Telephone: 713-798-8462.

Methods: Using data from the Texas Cancer Registry, we assembled 11,547 HCC cases diagnosed between 2011 and 2015 into Texas's census tracts geographic units. Twenty-nine neighborhood measures representing demographics and socioeconomic, and employment domains were retrieved from the US Census Bureau. We performed a series of aspatial and spatially weighted regression models to identify neighborhood-level characteristics associated with HCC risk.

Results: We found positive associations between HCC and proportion of population in census tracts that are black/African American, Hispanic, over 60 years old, in construction industry, and in the service occupation; but an inverse association with the proportion of population employed in the agricultural industry. The magnitude of these associations varied across Texas census tracts.

Conclusion/Impact: We found evidence that neighborhood-level factors are differentially associated with variations in HCC incidence across Texas. Our findings reinforce existing knowledge about HCC risk factors and expose others, including neighborhood-level employment status.

Graphical Abstract
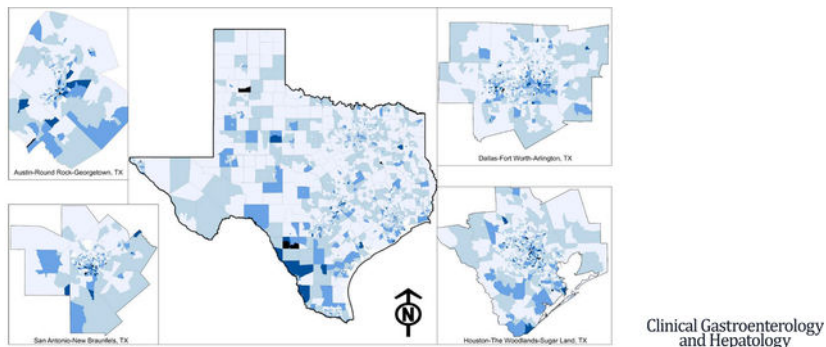
Clinical Gastroenterology and Hepatology

INTRODUCTION

Hepatocellular carcinoma (HCC), which constitutes >90% of primary liver cancer, is an increasingly important health problem in the United States. HCC incidence and mortality rates have increased 3 fold between 1975 and 2014, and these rates continue to increase.[1,2] Texas has the highest HCC incidence rates in the continental U.S.[2] Meanwhile, HCC incidence and mortality rates vary by race/ethnicity with Hispanics now having the highest HCC incidence rates in the U.S.[2–4]

Reasons for the higher incidence of HCC in minority populations in Texas are likely multifactorial, including differences in the prevalence and/or severity of known risk factors for HCC (e.g., chronic hepatitis C and hepatitis B virus infections, alcohol, obesity, diabetes, fatty liver, smoking),[5–7] disparities in prevention and treatment of HCC or its risk factors,[8] and possibly genetic susceptibility.[9,10] In the U.S., renewed attention is directed toward the

significance of the social determinants of health (SDOH) in health disparities.[11,12] HCC, its precursor, cirrhosis and risk factors, and the continuum of care patients undergo can be organized under three major domains of etiologic agents: proximal, intermediate, and distal. Proximal determinants operate at the individual level (e.g., genetic makeup, sex, race/ethnicity, or personal income). Intermediate determinants include the environmental context within which an individual experience their routine/daily activities (e.g., neighborhood-level social and physical environment). Distal determinants include policies that affect the availability, receipt of, and quality of healthcare. Examining geographic neighborhood-level factors provides an important framework for understanding intermediate determinants of SDOH.

Though many studies have examined relationships between demographic and socio-environmental factors and cancers, including HCC,[13–17] only a handful pursued epidemiological studies through spatial analysis methods.[18,19] In the current analysis, we used a suite of aspatial and spatial modeling techniques to correlate HCC incidence rates with neighborhood sociodemographic conditions that are in the intermediate domain of HCC risk factors, hence, identifying neighborhood level characteristics that are associated with higher HCC incidence rates in Texas.

METHODS

Study Setting

Our study population and accompanying data were drawn from the Texas Cancer Registry (TCR) and the U.S. Census Bureau (Census). The TCR provided data on individual patients that were diagnosed with HCC in Texas between 2011 and 2015. We identified HCC cases within the TCR dataset using a combination of International Classification of Diseases for Oncology, Third Edition (ICD-O-3) site code C22.0 and ICD-O-3 histology codes 8170–8175. We used the Census 2011–2015 American Community Survey (ACS) 5-year estimates to compute the neighborhood-level independent variables. The ACS is a nationwide survey that collects and produces information on social, economic, housing, and demographic characteristics about U.S. population. The ACS samples over 3.5 million households annually and produces a rolling 5-year average for each variable it measures.[20] The Census summarizes ACS estimates to specific geographic levels, including the census tract. The census tract, with an optimum population of approximately 4,000 residents or 1,600 housing units, is a small and relatively permanent statistical subdivision of a county designed to be homogeneous in terms of population characteristics, economic status, and living conditions.[21] The census tract was the unit of analysis, and all Texas census tracts (N=5,265) were considered for inclusion in our analysis.

Dependent Variable

The research file from the TCR included location reference data values for each individual HCC patient, including the longitude (X) and latitude (Y) coordinate points that represent a patient's address at the time of HCC diagnosis. We overlaid the X, Y data on top of Texas census tract boundaries and summed the count of HCC cases per census tract. We used the count of HCC cases with an Empirical Bayes smoother that used the census tract population

as the offset term for each tract.[22,23] That is, the model estimates the number of HCC cases per capita given the explanatory variables included in the model.

Explanatory Variables

We assembled US census neighborhood-level variables into five categories that included race/ethnicity, nativity/citizenship, age/sex, socioeconomic status, and employment industry. The variables included for each category are listed in the supplementary materials; Table S1. Except for the area deprivation index (ADI) that was computed, the explanatory variables were retrieved from the U.S. Census ACS 2011–2015 5-year estimates database, entered analysis workflows as percentage values, and were treated as continuous data. The ADI is a composite measure of neighborhood socioeconomic disadvantage that relies on 17 census variables (Table S1) drawn from four categories, including; poverty, housing, employment, and education.[24] Each census tract was assigned a score from 1 to 10 based on a decile classification of the ADI scores.

Data Analysis

After summarizing the HCC cases into census tract boundaries, our analytic dataset was highly skewed. To address the evidence of overdispersion observed in our dataset—the variance of the dependent variable is greater than the mean—we used the Poisson-based modeling with negative binomial regression (NBR) technique.[22,25] For each census tract, the number of HCC cases per capita is assumed to be Poisson distributed and independent. The Poisson–Gamma model was applied to the count of HCC cases in each census tract, while the total population of >20-year-old residents was used as an offset term.[22,23] Given that the dependent variable is treated as a count variable, the Poisson regression models the log of the expected count as a linear function of the explanatory variables. The Poisson regression coefficient is interpreted as follows: for a one-unit change in the explanatory variable, the difference in the logs of expected counts is estimated to change by the respective regression coefficient, while holding the other independent variables constant. The coefficients were expressed in terms of relative risk (RR) by exponentiating the Poisson regression coefficient,[22,23,26] and rescaled per a 10-unit change in the explanatory variable.

Our model selection process involved two stages. **Stage 1.** Explanatory variables that had a bivariate relationship with HCC (at $p \leq 0.10$) were arranged under their respective category and subsequently entered into a series of multivariable NBR models. Eventually, variables that maintained a significant association with HCC (p<0.05) were retained. To minimize multicollinearity among the retained variables, we dropped variables where T < 0.6 (Variance Inflation Factor, VIF > 1.67). As a general practice, the variable with the lowest tolerance was removed first, allowing the model to re-run without them. Candidate variables with tolerances ≥ 0.6 proceeded to Stage 2. **Stage 2.** For our final model estimation, all the variables that passed the multicollinearity test and maintained significance at p<0.05 from Stage 1 were included in a series of multivariable NBR model runs. Variables were dropped from any iteration of the Stage 2 model runs when $p \geq 0.05$ or T < 0.4 (VIF > 2.5). All statistical analyses were done using SPSS 28.0 (SPSS Inc, Chicago, IL, USA) and Stata16.0 (Stata Corp, College Station, TX).

To finalize our analysis, we included the variables in the final model run of Stage 2 in a Geographically Weighted Poisson Regression (GWPR) procedure. GWPR technique extends the conventional regression framework by allowing local variations in rates of change so that the coefficients produced are for specific locations, not global estimates.[27,28] Thus, each census tract has a separate Beta (β) coefficient for the exposure-outcome relationships being modeled. It shows areas where significant relationships are most or least pronounced, or neighborhoods where relationships diverge from what was observed in global models.[29] We ran the GWPR with the MGWR 2.2 software (https://sgsup.asu.edu/sparc/mgwr). The software's functionality and utility were discussed in detail by Oshan and colleagues.[30] We used the adaptive bi-square kernel to remove the effect of observations outside the neighborhood specified for local model fitting, while the "golden section search" function was used to select an optimal bandwidth. The local coefficients that resulted from the GWPR modelling were mapped in ArcGIS Pro (Esri, Redlands, CA).

Using the MGWR 2.2 software, we performed diagnostic tests to assess biases in the local estimates produced by the GWRPR, they were: (1) the Monte Carlo test for spatial variability of parameter estimate surface and (2) the local collinearity diagnostics.[30] The Monte Carlo test is a computationally intensive process that requires the GWPR to be calibrated many times; default number of iterations is 1000. The test constructs pseudo p-values for hypothesis testing; where p-value <0.05 indicates that the spatial variability of a parameter estimate surface is not occurring at random. The local collinearity diagnostics assess local multicollinearity via two methods: the local condition number (local CN) and the local variance inflation factor (local VIF). The local CN provides a single aggregate measure of local multicollinearity while the local VIF is generated for each covariate. Local CN > 30 or VIF >10 suggests there might be issues with local estimates due to multicollinearity.

RESULTS

After removing 44 census tracts with less than 100 residents, 14 where 100% of residents lived in quarters (e.g., correctional facilities), and two with inexplicably high HCC rates, 5,205 census tracts remained for data analysis. A total of 11,547 HCC diagnoses occurred between 2011–2015. Basic demographic data on the HCC cases are shown in Table 1 and the distribution of the average annual age-standardized incidence rates across Texas is shown in Figure 1. The incidence rates ranged from 0 to 208 per 100,000 adult population.

For context, descriptive data on the full list of explanatory variables are shown in Table S2. From the full list, three variables were not included in Stage 1 since they had no bivariate association with HCC at p   0.10, they were: % male, % Non-Hispanic Native Hawaiian, Pacific Islander, and % over 65 years old with no health insurance (Table S3). Stage 1 modeling produced 14 significant variables that were entered into the Stage 2 modeling (see Table 2 for summarized results and Tables S4 and S5 for more details). Seven of the 14 variables that entered Stage 2 remained statistically significantly associated with HCC. A 10-point increase in the percentage of the following characteristics is associated with increased risk for HCC incidence by the corresponding factors: % Non-Hispanic Black RR = 1.10 (95% CI, 1.08–1.12), % Hispanic RR = 1.10 (95% CI, 1.08–1.11), % 60 years and

above RR = 1.32 (95% CI, 1.28–1.34), % employed in the construction industry RR = 1.12 (95% CI, 1.07–1.17), % employed in the other services, except public admin. RR = 1.08 (95% CI, 1.00–1.17). Also, compared to census tracts in the first decile of the ADI, those in decile 10 had RR = 1.71 (95% CI, 1.52–1.91). Conversely, 10-point increase in % employed in the agricultural industry was associated with decreased risk (RR = 0.92; 95% CI, 0.87–0.96) (Table 2; see Table S5 for more details).

The GWPR established evidence of spatial heterogeneity, i.e., the relationships between the HCC incidence and the explanatory variables that remained in the final NBR model were not constant across Texas (Figure 2). In general, the positive association between % NH Black and HCC incidence was pronounced in the southernmost part of the Texas-Mexico border, along the center of the state, and in the north-western part of the state. For % Hispanics, the positive association with HCC incidence was also pronounced in the southernmost part of the Texas-Mexico border, and generally in the western half of the state. Percent ≥ 60 y.o. and the ADI maintained positive associations with HCC incidence across much of the state, except for patches of census tracts on the eastern side of the state for % ≥ 60 y.o., and on some on the northern part for the ADI. For % employed in construction and % employed in the other services (not administration) industries, the local relationships were generally either positive or not significant. Only agricultural/forestry industry had more negative local relationships with HCC incidence than positive relationships, although the relationship was not significant in most parts of the state. The Monte Carlo test suggests that only one explanatory variable (i.e., % in other services, except public administration) was threatened by multiple testing (Monte Carlo pseudo p-value = 0.107). For the remaining six variables the pseudo p-value was <0.05. In terms of local collinearity, the local CN was less than 30 in 5,205 census tracts (91.4%) and none of the explanatory variables had local VIF>10; suggesting that the local estimates were likely not degraded by the presence of collinearity.

DISCUSSION

In this population-based study of HCC in Texas, we found that measures that represent minority population, socioeconomic disadvantage, or blue-collar employment were independently associated with higher risk of HCC incidence while employment in the agricultural industry was associated with a lower risk. However, these relationships are not uniform across the state. The local modeling approach showed significant and spatially varying patterns of relationships between the explanatory variables and HCC risk. Spatial analysis approaches are powerful methods for investigating disease patterns, and they are increasingly used to better understand cancer epidemiology.[19,31,32] By allowing regression coefficients to vary over space, the GWPR allows for the visualization of spatially heterogeneous relationship between each explanatory variable and HCC. In general, the relationships between HCC risk and the demographic/SES variables were more spatially heterogenous than the relationships with the employment industry variables. The relationship with % black/AA appeared to show the most dramatic changes across the state; it is more pronounced in the southernmost and north-western parts.

Our study joins recent and growing research examining various aspects of the relationships between neighborhood level demographic and socioeconomic factors and HCC outcomes

in general,[15,16] and the handful doing so through spatial analytical approaches.[18,19] In a prospective cohort study of seniors living in six US states (California, Florida, Louisiana, New Jersey, North Carolina, and Pennsylvania), area socioeconomic deprivation was associated with increased risk of HCC and chronic liver disease (CLD) mortality after accounting for participants' age, sex, and race, but after accounting for educational attainment and health-risk factors, the socioeconomic deprivation relationship became non-significant for HCC but remained so for CLD mortality.[18] Also, our study adds to the literature in this space by including employment measures in our analysis. Few studies have investigated the relationships between employment and HCC.[33,34] Previous studies have suggested an association between HCC and several blue-collar jobs,[33–35] and a protective association between managerial jobs and HCC.[33] These studies assessed employment measures at the individual level, ours may be the first to do so at the neighborhood level, in terms of relationships with HCC. Indeed, the ecological nature of our study precludes the interpretation of a causal link between working in the industries that remained in our analysis (i.e., construction, other services except public administration, and agriculture) and HCC.

Our study has some limitations. Though we found associations between HCC incidence from the TCR and specific area level measures of demographic, socioeconomic, and employment measures, the measures used were census estimates that match the time of HCC diagnosis. Hence, they may not reflect earlier neighborhood-level exposures. However, because census tracts are stable census standard geographies, and our analysis examined all the census tracts in Texas, we expect our neighborhood-level estimates are fair representations of the historical neighborhood milieu across Texas. Also, because the ACS produces population-level estimates (here, census tract-level) based on information from a sample of the US population, the estimates have standard errors indicating their degrees of uncertainty.[36,37] The implicit nonuniformity of these standard errors over space may bias the results of spatial analyses employing the ACS measures. However, the Census ACS remains a premier source for areal data in the US. Additionally, the measures considered in this analysis may not represent a comprehensive list of all the factors that could influence HCC incidence in Texas neighborhoods. Although we started with many neighborhood measures (N=29), we cannot rule out other unexamined factors. In the current analysis, the lack of individual-level factors (e.g., hepatitis C infection) in our analysis workflow may be a limitation. However, we designed this work to focus exclusively on the contributions of neighborhood-level characteristics to community incidence of HCC. To a large extent, our approach relies on previous assertions of the independent contributions of neighborhood-level social, economic, and environmental factors to health outcomes.[11,38,39] Understanding HCC relationships with factors that operate at the community level may offer new HCC intervention approaches or help strengthen existing ones.

In summary, we found preliminary evidence that neighborhood-level factors may partly explain spatial and racial/ethnic variations in HCC incidence. Future research, including longitudinal exposure assessment studies, are needed to clarify the specific roles of occupations and industries in HCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research. 2014;74(11):2913–2921. [PubMed: 24840647]

2. White DL, Thrift AP, Kanwal F, Davila J, El-Serag HB. Incidence of hepatocellular carcinoma in all 50 United States, from 2000 through 2012. Gastroenterology. 2017;152(4):812–820. e815. [PubMed: 27889576]

3. Zhang X, El-Serag HB, Thrift AP. Sex and Race Disparities in the Incidence of Hepatocellular Carcinoma in the United States Examined through Age-Period-Cohort Analysis. Cancer Epidemiol Biomarkers Prev. 2020;29(1):88–94. [PubMed: 31712271]

4. El-Serag HB, Sardell R, Thrift AP, Kanwal F, Miller P. Texas Has the Highest Hepatocellular Carcinoma Incidence Rates in the USA. Dig Dis Sci. 2020.

5. Caetano R, Ramisetty-Mikler S, Rodriguez LA. The Hispanic Americans Baseline Alcohol Survey (HABLAS): the association between birthplace, acculturation and alcohol abuse and dependence across Hispanic national groups. Drug and alcohol dependence. 2009;99(1–3):215–221. [PubMed: 18945554]

6. Lee DJ, Markides KS, Ray LA. Epidemiology of self-reported past heavy drinking in Hispanic adults. Ethnicity & Health. 1997;2(1–2):77–88. [PubMed: 9395591]

7. Mandayam S, Jamal MM, Morgan TR. Epidemiology of alcoholic liver disease. Paper presented at: Seminars in liver disease2004.

8. Singal AG, Li X, Tiro J, et al. Racial, social, and clinical determinants of hepatocellular carcinoma surveillance. The American journal of medicine. 2015;128(1):90. e91–90. e97.

9. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461(7262):399–401. [PubMed: 19684573]

10. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. Nature genetics. 2010;42(1):21. [PubMed: 19946271]

11. Adler NE, Rehkopf DH. US disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health. 2008;29:235–252. [PubMed: 18031225]

12. Haskins J. Healthy People 2030 to create objectives for health of nation: Process underway for next 10-year plan. American Public Health Association; 2017.

13. Zahnd WE, McLafferty SL. Contextual effects and cancer outcomes in the United States: a systematic review of characteristics in multilevel analyses. Annals of Epidemiology. 2017;27(11):739–748. e733. [PubMed: 29173579]

14. Singh GK, Siahpush M, Altekruse SF. Time trends in liver cancer mortality, incidence, and risk factors by unemployment level and race/ethnicity, United States, 1969–2011. Journal of community health. 2013;38(5):926–940. [PubMed: 23689953]

15. Danos D, Leonardi C, Gilliland A, et al. Increased risk of hepatocellular carcinoma associated with neighborhood concentrated disadvantage. Frontiers in oncology. 2018;8:375. [PubMed: 30254987]

16. Shebl FM, Capo-Ramos DE, Graubard BI, McGlynn KA, Altekruse SF. Socioeconomic status and hepatocellular carcinoma in the United States. Cancer Epidemiol Biomarkers Prev. 2012;21(8):1330–1335. [PubMed: 22669949]

17. Singh GK, Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: over six decades of changing patterns and widening inequalities. Journal of environmental and public health. 2017;2017.

18. Major JM, Sargent JD, Graubard BI, et al. Local geographic variation in chronic liver disease and hepatocellular carcinoma: contributions of socioeconomic deprivation, alcohol retail outlets, and lifestyle. Annals of epidemiology. 2014;24(2):104–110. [PubMed: 24332863]

19. Ford MM, Ivanina E, Desai P, et al. Geographic epidemiology of hepatocellular carcinoma, viral hepatitis, and socioeconomic position in New York City. Cancer Causes & Control. 2017;28(7):779–789. [PubMed: 28573469]

20. Census. American Community Survey Information Guide. 2017; https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf. Accessed April 20, 2019.

21. Census. Geographic Areas Reference Manual, Chapter 10: Census Tracts and Block Numbering Areas. 1994: https://www2.census.gov/geo/pdfs/reference/GARM/Ch10GARM.pdf. Accessed October 3, 2020.

22. Cameron AC, Trivedi PK. Regression analysis of count data. Vol 53: Cambridge university press; 2013.

23. Hilbe JM. Modeling count data. Cambridge University Press; 2014.

24. Singh GK. Area deprivation and widening inequalities in US mortality, 1969–1998. American journal of public health. 2003;93(7):1137–1143. [PubMed: 12835199]

25. Lord D. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention. 2006;38(4):751–766. [PubMed: 16545328]

26. Yang T-C, Shoff C, Matthews SA. Examining the spatially non-stationary associations between the second demographic transition and infant mortality: A Poisson GWR approach. Spatial demography. 2013;1(1):17–40. [PubMed: 25383259]

27. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis. 1996;28(4):281–298.

28. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons; 2003.

29. Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted Poisson regression for disease association mapping. Statistics in medicine. 2005;24(17):2695–2717. [PubMed: 16118814]

30. Oshan TM, Li Z, Kang W, Wolf LJ, Fotheringham AS. MGWR: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS International Journal of Geo-Information. 2019;8(6):269.

31. Cheng EM, Atkinson PM, Shahani AK. Elucidating the spatially varying relation between cervical cancer and socio-economic conditions in England. International journal of health geographics. 2011;10(1):51. [PubMed: 21943079]

32. Mansori K, Solaymani-Dodaran M, Mosavi-Jarrahi A, et al. Determination of effective factors on geographic distribution of the incidence of colorectal cancer in Tehran using geographically weighted Poisson regression model. Medical journal of the Islamic Republic of Iran. 2019;33:23. [PubMed: 31380313]

33. Abaza Y, Abdel-Wahab R, Li D, et al. Association between the job types and the risk of hepatocellular carcinoma in the United States. J Epidemiol Res. 2016;3:1.

34. Ferrand J-F, Cénée S, Laurent-Puig P, et al. Hepatocellular carcinoma and occupation in men: a case-control study. Journal of occupational and environmental medicine. 2008;50(2):212–220. [PubMed: 18301178]

35. Porru S, Placidi D, Carta A, et al. Primary liver cancer and occupation in men: A case-control study in a high-incidence area in northern Italy. International journal of cancer. 2001;94(6):878–883. [PubMed: 11745492]

36. Spielman SE, Folch DC. Reducing uncertainty in the American Community Survey through data-driven regionalization. PloS one. 2015;10(2):e0115626. [PubMed: 25723176]

37. Koo H, Wong DW, Chun Y. Measuring global spatial autocorrelation with data reliability information. The Professional Geographer. 2019;71(3):551–565. [PubMed: 31787781]

38. Macintyre S, Ellaway A. Neighborhoods and health: an overview. Neighborhoods and health. 2003:20–42.

39. Robert SA. Socioeconomic position and health: the independent contribution of community socioeconomic context. Annual review of sociology. 1999;25(1):489–516.

What You Need to Know

BACKGROUND:

Racial/ethnic minorities in the United States have disproportionately high rates of hepatocellular carcinoma (HCC); however, the reasons for this high HCC burden remains unclear.

FINDINGS:

We identified novel neighborhood-level factors that are associated with variations in HCC incidence across Texas, including proportion of population in census tracts that are black/African American and Hispanic.

IMPLICATIONS FOR PATIENT CARE:

Intermediate social determinants of health, including neighborhood-level social and physical environment, may influence HCC risk and may be a target for primary prevention of HCC among minority populations.
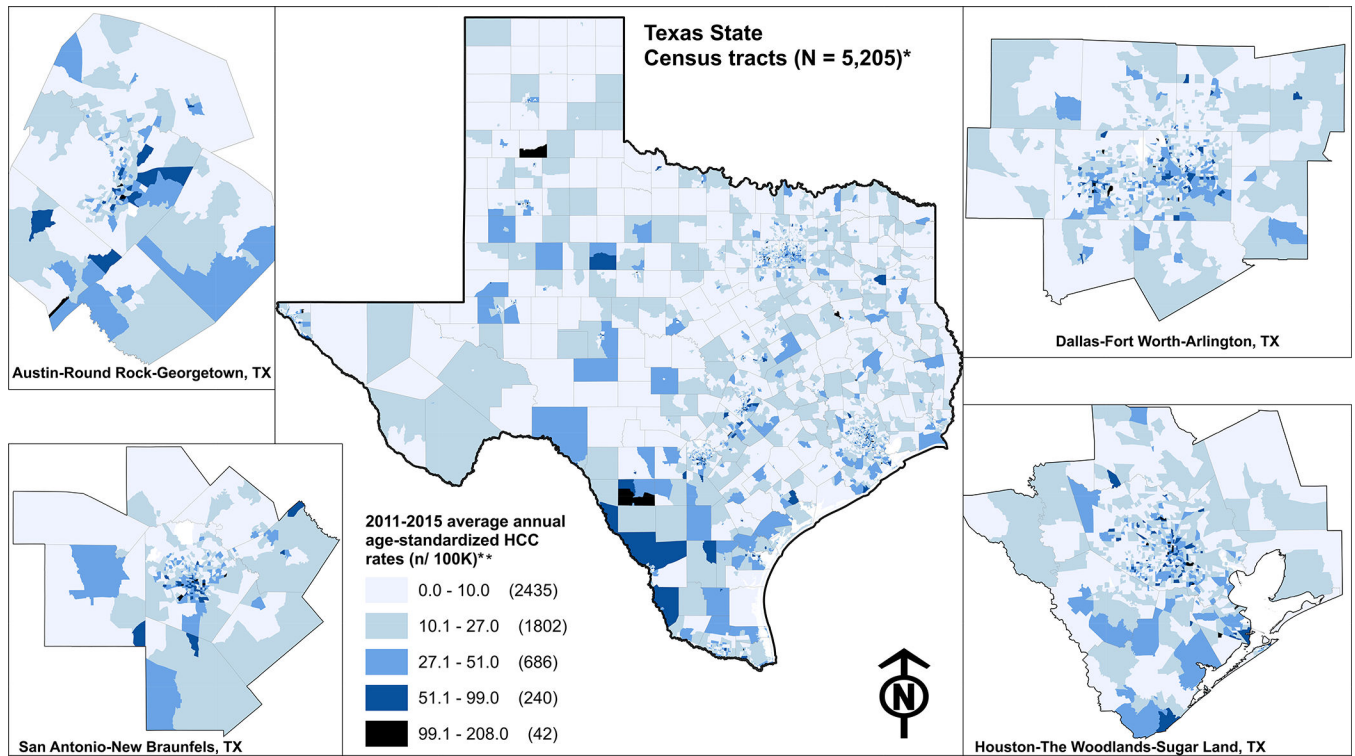
Figure 1.
The 2011–2015 average annual age-adjusted hepatocellular carcinoma (HCC) incidence rates shown for Texas' census tracts. * 5,205 of the 5,265 Texas census tracts were included for analysis. ** The natural breaks technique (i.e., Jenks optimization) was used to classify the data; Jenks seeks to minimize intra-class variance and simultaneously maximize inter-class.
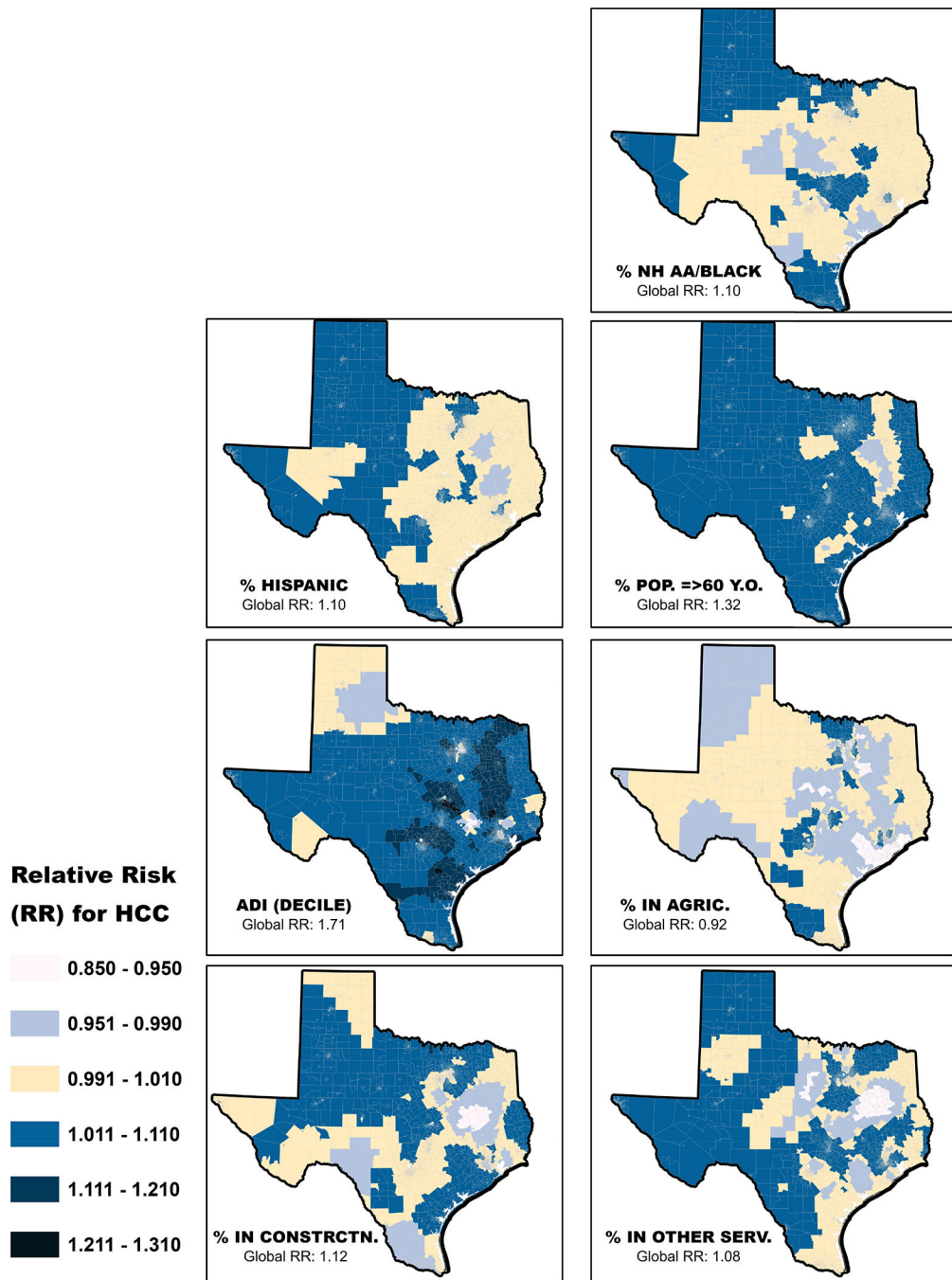
Figure 2.

The local beta coefficients were exponentiated (RR) to show the sensitivity of HCC incidence to a change of 10-unit difference in each of the neighborhood characteristics shown above, specific to each census tract. The class labeled 0.991–1.010 crosses 1.0. To simplify interpretation, colors before yellow shade (from top to bottom) suggest census tracts where an increase in the proportion of a given independent variable was associated with decreased RR for HCC, while colors after the yellow shade suggest census tracts where

an increase in the proportion of the explanatory variables was associated with increased RR for HCC.

Table 1:

Basic demographic data on the HCC cases diagnosed in Texas between 2011–2015.

Characteristics	N = 11,547 (%)
Sex	
Male	8736 (75.7)
Female	2808 (24.3)
Other	3 (0.03)
Age at diagnosis	
Mean (SD)	63.1 (10.8)
<50	808 (7.0)
50–69	7789 (67.5)
70+	2950 (25.5)
Race/ethnicity	
NHW	5033 (43.6)
NHB	1615 (14.0)
Hispanic	4301 (37.2)
Other	598 (5.2)
Year of diagnosis	
2011	2030 (17.6)
2012	2140 (18.5)
2013	2282 (19.8)
2014	2522 (21.8)
2015	2573 (22.3)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Categories of neighborhood-level SDOH and incidence of HCC in Texas (N census tracts = 5,205).

Neighborhood Characteristics	Stage 1 IRR (95% CI) [a,b]	Stage 2 [c]	Final Stage 2 IRR (95% CI) [a,d]
Race / Ethnicity			
% Non-Hispanic White [e]			
% Non-Hispanic Black or African American	1.11 (1.10–1.13)	X	1.10 (1.08–1.12)
% Non-Hispanic American Indian and Alaska Native [f]	1.34 (0.86–2.08)		
% Non-Hispanic Asian	0.77 (0.74–0.81)	X	
% Non-Hispanic Native Hawaiian, Pacific Islander [e]			
% Non-Hispanic Other + 2 or more races	0.86 (0.79–0.93)	X	
% Hispanic or Latino	1.10 (1.09–1.11)	X	1.10 (1.08–1.11)
Citizenship			
% Population; not a US Citizen	0.91 (0.84–0.99)	X	
% Population; foreign-born [b]	1.06 (0.99–1.13)		
% of Foreign-born; born in Latin America	1.13 (1.12–1.14)	X	
Age and Sex			
% of Population; 50 to 59 y.o. [f]	0.96 (0.90–1.04)		
% of Population; 60 y.o. and above	1.12 (1.08–1.16)	X	1.32 (1.28–1.34)
% of Population; Male [e]			
Socioeconomic Status			
Area Deprivation Index (decile)	2.78 (2.46–3.13)	X	1.71 (1.52–1.91)
% of 18 to 34 years with no health insurance [g]	1.08 (1.06–1.10)		
% of 35 to 64 years with no health insurance [g]	0.97 (0.94–1.00)		
% of 65 years and over with no health insurance [e]			
Industry			
% in Agriculture, forestry industry	0.90 (0.86–0.95)	X	0.92 (0.87–0.96)
% in Construction industry	1.43 (1.37–1.51)	X	1.12 (1.07–1.17)

Neighborhood Characteristics	Stage 1 IRR (95% CI) [a,b]	Stage 2 [c]	Final Stage 2 IRR (95% CI) [a,d]
% in Manufacturing category	0.95 (0.90–1.00)	X	
% in Transportation and warehousing etc.	1.20 (1.11–1.29)	X	
% in Professional, scientific etc.	0.87 (0.82–0.92)	X	
% in Educational services and health care etc. [g]	1.10 (1.06–1.15)		
% in Other services, except public admin.	1.20 (1.11–1.29	X	1.08 (1.00–1.17)

[a] The coefficient values were exponentiated; expressed in terms of relative risk (RR). Explanatory variables were rescaled to interpret results as an increase or a decrease in the risk of HCC incidence associated with a 10-unit change in predictor variable.

[b] Effect estimates shown were from running a series of multivariable models where variables that represent each category were entered into the model together. Variables across categories were not included in any single model. Only variables significant at p 0.10 during the bivariate analysis were used.

[c] Variables that were allowed to enter Stage 2.

[d] Effect estimates shown were from running a series of multivariable models where variables were added across categories. Only variables significant at $p < 0.05$ during Stage 1 and those not affected by multicollinearity were used.

[e] Variables that were examined in bivariate analysis but not included in the category-specific multivariable models (i.e., not included in Stage 1) because P value > 0.10. Of note NH White was excluded to avoid model overfitting in the race/ethnicity category.

[f] Variables that were not significant at $p < 0.05$ during Stage 1 model runs.

[g] Variables that were dropped due to multicollinearity issues during Stage 1 model runs (Tolerance < 0.6; VIF > 1.67).