



## Towards precision public health: Geospatial analytics and sensitivity/specificity assessments to inform liver cancer prevention

Shannon M. Lynch<sup>a,\*</sup>, Daniel Wiese<sup>b</sup>, Angel Ortiz<sup>a</sup>, Kristen A. Sorice<sup>a</sup>, Minhhuyen Nguyen<sup>a</sup>, Evelyn T. González<sup>a</sup>, Kevin A. Henry<sup>a,b</sup>

<sup>a</sup> Cancer Prevention and Control, Fox Chase Cancer Center, Philadelphia, PA, USA

<sup>b</sup> Geography and Urban Studies, Temple University, Philadelphia, PA, USA

### ARTICLE INFO

**Keywords:**  
Geospatial  
Liver cancer  
Sensitivity  
Specificity  
Neighborhood  
Disparities  
Precision public health

### ABSTRACT

**Objectives:** Liver cancer (LC) continues to rise, partially due to limited resources for prevention. To test the precision public health (PPH) hypothesis that fewer areas in need of LC prevention could be identified by combining existing surveillance data, we compared the sensitivity/specificity of standard recommendations to target geographic areas using U.S. Census demographic data only (percent (%) Hispanic, Black, and those born 1950–1959) to an alternative approach that couples additional geospatial data, including neighborhood socioeconomic status (nSES), with LC disease statistics.

**Methods:** Pennsylvania Cancer Registry data from 2007–2014 were linked to 2010 U.S. Census data at the Census tract (CT) level. CTs in the top 80th percentile for 3 standard demographic variables, %Hispanic, %Black, %born 1950–1959, were identified. Spatial scan statistics (SatScan) identified CTs with significantly elevated incident LC rates ( $p$ -value < 0.05), adjusting for age, gender, diagnosis year. Sensitivity, specificity, and positive predictive value (PPV) of a CT being located in an elevated risk cluster and/or testing positive/negative for at least one standard variable were calculated. nSES variables (deprivation, stability, segregation) significantly associated with LC in regression models ( $p < 0.05$ ) were systematically evaluated for improvements in sensitivity/specificity.

**Results:** 9,460 LC cases were diagnosed across 3,217 CTs. 1,596 CTs were positive for at least one of 3 standard variables. 5 significant elevated risk clusters (CTs = 402) were identified. 324 CTs were positive for a high risk cluster AND standard variable (sensitivity = 92%; specificity = 37%; PPV = 17.4%). Incorporation of 3 new nSES variables with one standard variable (%Black) further improved sensitivity (93%), specificity (62.9%), and PPV (26.3%).

**Conclusions:** We introduce a quantitative assessment of PPH by applying established sensitivity/specificity assessments to geospatial data. Coupling existing disease cluster and nSES data can more precisely identify intervention targets with a liver cancer burden than standard demographic variables. Thus, this approach may inform prioritization of limited resources for liver cancer prevention.

### 1. Introduction

Incidence and mortality rates in liver cancer are on the rise in the U.S., increasing by close to 3% per year since 2000 (Ryerson et al., 2016). In the United States, 42,030 new cases of liver cancer will be diagnosed and about 31,780 people will die of liver cancer. By 2030, liver cancer is expected to exceed breast cancer as the second leading cause of cancer death in the U.S. (Altekruse, Henley, Cucinell, McGlynn, 2014).

Compared to non-Hispanic Whites (NHW-6.3/100,000), incidence rates are higher in Blacks (10.2/100,000), Hispanics (13/100,000) and Asians (13.5/100,000) (Wang et al., 2016). In Pennsylvania, these racial trends in liver cancer are similar to those in the U.S., with Blacks having triple the rate of liver cancer incidence compared to NHWs (American Cancer Society, 2018).

Compared to other cancer sites, pathways to liver cancer are largely known and potentially modifiable (Singal, Pillai, Tiro, 2014). Up to 30%

\* Corresponding author. Cancer Prevention and Control, Fox Chase Cancer Center, Young Pavilion P4159, 333 Cottman Avenue, Philadelphia, PA, 19111, USA.  
E-mail addresses: [Shannon.Lynch@fcc.edu](mailto:Shannon.Lynch@fcc.edu) (S.M. Lynch), [tug30358@temple.edu](mailto:tug30358@temple.edu) (D. Wiese), [Angel.Ortiz@fcc.edu](mailto:Angel.Ortiz@fcc.edu) (A. Ortiz), [Kristen.Sorice@fcc.edu](mailto:Kristen.Sorice@fcc.edu) (K.A. Sorice), [Minhhuyen.Nguyen@fcc.edu](mailto:Minhhuyen.Nguyen@fcc.edu) (M. Nguyen), [Evelyn.Gonzalez@fcc.edu](mailto:Evelyn.Gonzalez@fcc.edu) (E.T. González), [khenry1@temple.edu](mailto:khenry1@temple.edu) (K.A. Henry).

<https://doi.org/10.1016/j.ssmph.2020.100640>

Received 5 February 2020; Received in revised form 29 June 2020; Accepted 1 August 2020

Available online 7 August 2020

2352-8273/© 2020 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of liver cancer cases are attributed to Hepatitis B (HBV) and Hepatitis C (HCV) viral infection (Makarova-Rusher et al., 2015; Wetzel et al., 2013). The fraction of liver cancer cases in African American and Asian patients attributed to chronic HBV or HCV is much higher, closer to 40%–50% (Wetzel et al., 2013). HCV and HBV infection are often contracted through modifiable risk behaviors, including sexual activity, drug use, and unsanitary tattoo and nail salon practices (El-Serag & Rudolph, 2007). Additionally, treatments with high cure rates exist for HCV, and vaccination can help prevent HBV infection (NCCN, 2017). Alcohol consumption and metabolic disorders, including diabetes and obesity, are also associated with liver cancer. NHWs are more likely to develop liver cancer through metabolic disorders; Hispanics through HCV infection and increased alcohol consumption (Makarova-Rusher et al., 2015). Similarly, metabolic syndrome and alcohol consumption are associated with diet and lifestyle behaviors that could be modified through educational interventions and policies. Despite this, disproportionate rates of liver cancer and related risk behaviors persist across race/ethnic groups, suggesting that evidence-based interventions are not reaching vulnerable, high risk populations and that health disparities and health equity issues are major contributors to the growing burden of liver cancer in this country.

As detailed by a number of multilevel conceptual frameworks (Lynch & Rebbeck, 2013; Warnecke et al., 2008), beyond a person's race/ethnicity, social environmental factors, particularly the neighborhood in which a person lives, also inform cancer disparities (Lynch & Rebbeck, 2013). The neighborhood social and economic environment or status (nSES) is often defined in cancer studies by U.S. Census variables related to the economic (e.g., employment, income), physical (e.g., housing/transportation structure), and social (e.g., poverty, education), characteristics of a geographic area (Diez Roux & Mair, 2010). Previous studies have demonstrated that nSES independently effects liver cancer incidence and cancer mortality more broadly, even after adjustments for an individual's race/ethnicity and socioeconomic status (SES) (e.g., a person's education, income, and poverty level) (Chang et al., 2010; Makarova-Rusher et al., 2015; Wetzel et al., 2013). However, rarely are nSES factors considered when identifying vulnerable, high risk populations for cancer prevention.

While screening guidelines exist for risk factors for liver cancer, including HBV and HCV (NCCN, 2017), screening guidelines for liver cancer for the general population are lacking. Thus, current recommendations for liver cancer prevention focus on targeting high risk, minority populations including: Hispanics, Blacks, and those born between 1950 and 1959 who are at risk for HCV infection (Petrick, Kelly, Altekruze, McGlynn, & Rosenberg, 2016). As a result, U.S. cancer centers, who are often tasked with implementing cost-effective educational and behavioral interventions for liver cancer prevention, commonly utilize publically-available neighborhood demographic data from the U.S. Census to identify these vulnerable communities in their catchment—defined as the neighborhoods where their patients reside (Blake, Ciolino, & Croyle, 2019). Neighborhood data is used because studies suggest an individual's demographics are similar to the neighborhood in which they live, particularly at smaller geographic areas (Tunstall, 2005). However, there are a few problems with this approach. First, the geographic unit of analysis used to define (and subsequently prioritize) neighborhoods in need of cancer prevention is typically quite large, at a regional, state, or county level. Using Pennsylvania as an example, Philadelphia County has the largest population of Blacks and Hispanics; however, there are approximately 1.5 million people living in Philadelphia. Beyond demographic data, cancer rates are also traditionally reported at State and county levels. However, geospatial methods allow for small area estimations of disease risk and can be used to identify neighborhood clusters that have higher than expected rates of cancer at smaller geographic units than county (Sahar et al., 2019; Sherman et al., 2014). Thus, to maximize often limited resources available for liver cancer prevention at the local level, narrowing down geographic areas, from counties to Census tracts, for instance, which contain on average

about 4,000 residents, would prove useful. Second, combining existing demographic data with cancer incidence and mortality data, as well as nSES measures, could also further narrow down neighborhoods for cancer prevention. However, traditional neighborhood health rankings typically report prevalence rates of single behavioral risk factors or cancer mortality separately (Erwin, Myers, Myers, & Daugherty, 2011; Kanarek, Tsai, & Stanley, 2011; Oliver, 2010), and often without consideration of health disparity measures, like nSES (Thornton-Wells, Moore, & Haines, 2004). Coupling multiple sources of surveillance data to guide interventions that can benefit populations more efficiently is a strategy referred to as *precision public health*. Precision public health is being applied to infectious diseases and in developing countries to narrow down geographic areas most in need of interventions, but it has yet to be applied in a cancer prevention setting (Dowell, Blazes, & Desmond-Hellmann, 2016).

In this study, we merge liver cancer surveillance data from the Pennsylvania (PA) State Cancer registry with U.S. Census data in order to identify geographic areas at the Census tract level that contain (alone or in combination): a) a high burden of liver cancer incidence; b) a high proportion of Blacks, Hispanics, or those born 1950–1959 (standard demographic variables); c) unfavorable nSES conditions found to be associated with liver cancer incidence in PA. Introducing a sensitivity/specificity assessment that we derived from patient-level clinical tests and applied to area-level surveillance data, we then compare the number of Census tracts identified for cancer prevention using only standard recommendations to combined approaches that link liver cancer disease rates with often underutilized nSES measures. Our goal was to test the precision public health hypothesis that a smaller number (i.e., fewer) Census tracts in need of intervention could be identified by combining existing surveillance data, and to evaluate which combinations (or number) of nSES and demographic variables were needed to improve sensitivity/specificity assessments. Thus, this study serves as a quantitative assessment of the precision public health framework.

## 2. Methods

### 2.1. Study sample

Incident liver cancer cases diagnosed between 2007 and 2014 (n = 9466) were ascertained from the Pennsylvania (PA) Cancer Registry ([dataset] Pennsylvania Cancer Registry), which is a state-wide North American Association of Central Cancer Registries (NAACCR) gold certified data system that collects basic demographics, including age (0–102), gender (male/female), race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Hispanic), address at diagnosis, as well as clinical data, including diagnosis data, stage (In-Situ, Local, Regional, Distant), and treatment information. Cases without address data, or only P.O. Box data were removed from the dataset (n = 6). The PA registry does not typically release prisoner data. A total of 9,460 cases of liver cancer were included in this analysis.

Using the ESRI ArcGIS geocoder with StreetMap Premium streets NAACCR standards (Goldberg, 2008), we were able to match and geocode patient addresses at time of diagnosis and link the data to the Census tract for over 98% of patients. Thus, the geographic boundary used to define neighborhood in this study is the administrative Census tract (CT) in which the case lived at time of diagnosis, which was derived from the 2010 Census tract boundaries from the U.S. Census Bureau data. In the State of PA, there are a total of 3,217 CTs (average of 3,973 residents). Studies show that Census tracts can serve as useful units of analysis to study associations between cancer outcomes and related disease determinants (Boscoe et al., 2014; Krieger et al., 2002).

### 2.2. Statistical analysis

#### 2.2.1. Disease Outcome: Identification of liver cancer disease clusters

For spatial analyses that calculated adjusted liver cancer incidence

rates, we grouped single-year Census tract level residential population estimates by race/ethnicity, sex and 19 age-groups (5-year ranges) from the American Community Survey 2007–2011 (diagnosis years 2007–2011) and 2011–2015 (diagnosis years 2011–2015) to generate denominator data. For spatial cluster detection, we applied spatial scan statistics using SaTScan software, version 9.6 (<https://www.satscan.org/>). The spatial scan statistic provides evidence whether a disease is clustered or randomly distributed throughout the study area. This cluster analysis was applied at the Census tract level using a Poisson model and an elliptical spatial window with the maximum cluster size set up to 50% of the population at risk (Kulldorff, Huang, Pickle, & Duczmal, 2006). Using Monte Carlo 9,999 simulations testing statistical significance, clusters of Census tracts with significantly higher than expected rates of liver cancer are reported using P values < 0.05, adjusted for multiple testing (Kulldorff, Huang, & Konty, 2009). The tested clusters were adjusted for year at the diagnosis, sex, race/ethnicity, and age at diagnosis (categorized into 19 age groups).

### 2.2.2. Neighborhood measures

To characterize the socioeconomic status of a neighborhood or Census tract area-based measures of disparity, we selected variables from the American Community Survey (ACS) 2007–2011 and 2011–2015 that have been previously investigated in other cancer studies (Gomez et al., 2015). Variables of interest include standard demographic variables derived from U.S. Census data only (Petrick et al., 2016): 1) race/ethnicity (% Non-Hispanic Black (NHB); % Hispanic); 2) age (born in the 1950–1959 birth cohort-yes/no); as well as additional nSES variables commonly assessed in neighborhood and cancer studies: 3) poverty (% population 18 and older living below the federal poverty level (CT-Poverty); 4) immigration (% foreign born population; % English language proficient); 5) migration/stability (% of households still living in same house as one year ago); 6) racial segregation or concentration, where we used Massey's (2001) formula (Massey, Booth, & Crouter, 2001) and instructions for the integration of neighborhood income and race/ethnicity data from Krieger et al. (2016) (Krieger et al., 2016) to calculate the index of concentration at the extremes (ICE) that compares the most privileged race/ethnic group (White, Non-Hispanics) to Blacks or Hispanics across income levels (Krieger et al., 2016); 7) neighborhood deprivation indices, which are composite or summary scores of the education, employment, housing, and access (defined in terms of transportation) of a neighborhood or Census tract. Specifically, we evaluated the Townsend Deprivation Score (TDS) (Rice et al., 2014) which is a summary score of the following z-transformed variables: % with no access to a car, % of crowded households, % of rented households, % unemployed, as well as a deprivation index we previously created using a principal component analysis of indicator variables related to poverty (CT-Poverty), education (% No-High-School) and income (Median household income) (Supplementary File 1A2).

In order to reduce the number of explanatory variables ( $n = 14$ ), we applied a logistic regression model using SAS 9.1 where the outcome of interest was whether a patient was located in a high-risk liver cancer cluster (from disease outcome statistical analysis section above; 1 = located in an elevated disease cluster; 0 = not located in a high risk cluster-See Supplementary File 1A2). For neighborhood variables where quartile summary estimates included zero observations, binary variables were created using the percentage of Census tracts above the State average as a cut-point (% foreign-born). The number of Census tracts located in the most unfavorable category for each neighborhood measure were then plotted and visualized geospatially (Supplementary File 1A2/3) and a frequency analysis (Supplementary File 1B), along with area under the curve estimates (AUC; Supplementary File 1D) were conducted to further optimize and compare sensitivity/specificity assessments (described below).

### 2.2.3. Sensitivity/specificity assessments

We first compared the number of Census tracts identified as having a

higher than expected rate of liver cancer in the State of PA ( $n = 402$ ) to the number of Census tracts identified as having at least one ( $n = 1596$ ) or all of the standard recommendation variables ( $n = 9$ ). We then combined the disease measures AND the standard demographic measures from the U.S. Census in order to: a) further reduce the number of Census tracts by identifying areas with both a liver cancer burden and higher proportion of Blacks, Hispanics and those in the birth cohort; b) quantify and compare the number of Census tracts that might have been targeted for prevention efforts based on standard demographic variables alone, but did not have an actual disease burden. We did these comparisons by adapting sensitivity/specificity clinical assessments, often used to evaluate patient-level diagnostic tests, to our geospatial data (See Table 1).

Next, we determined if sensitivity/specificity assessments could be improved with the addition of nSES variables. We developed a systematic analytic pipeline (See Supplementary File-1B-D) that evaluated changes in sensitivity and specificity for each addition of a single nSES variable, as well as all possible combinations of these variables. The assessment with the best (highest percent) sensitivity/specificity is reported here.

## 3. Results

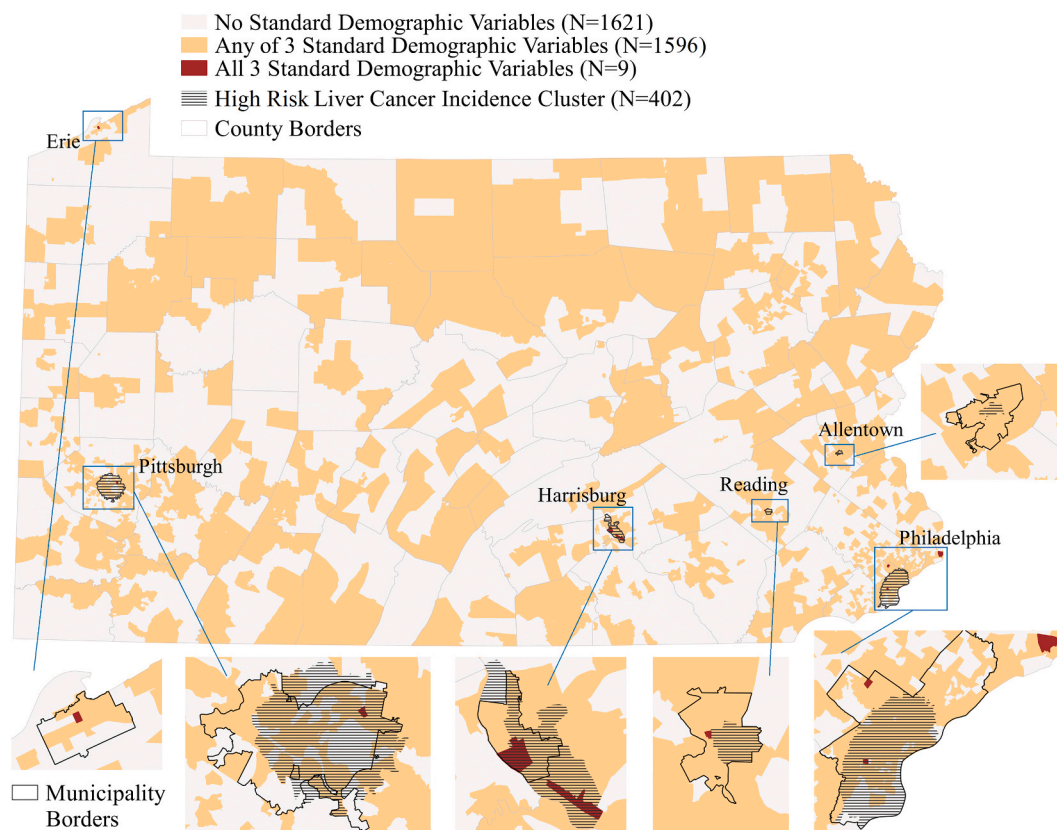
Referring to Fig. 1, using standard demographic variables to target geographic areas with higher percentages of Blacks, Hispanics, or the birth cohort at risk for Hepatitis C, we identified 1,596 Census tracts that would meet these criteria (light-orange), while only 9 Census tracts met the criteria for all 3 demographic variables (dark-red; e.g., Erie). Using spatial scan statistics, we identified five clusters ( $n = 402$  Census tracts) near Philadelphia, Pittsburgh, Allentown, Harrisburg, and Reading with higher than expected liver cancer incidence rates (hashed). The Allentown cluster had the highest relative risk of 3.69 ( $p < 0.01$ ), followed by Philadelphia 2.87 ( $p < 0.01$ ). Table 2 summarizes the basic demographics of areas located in a high risk cluster compared to other areas of the State of PA. A higher proportion of cases located in the high risk clusters were males between the ages of 45 and 65 years old, which corresponds to those born in the 1950-59 birth cohort, compared to the rest of the State of PA, which had higher proportions of those over the age of 65. The majority of cases in the Philadelphia and Harrisburg clusters were non-Hispanic Black. Pittsburgh, Allentown, and Reading clusters contained majority non-Hispanic White cases, but Allentown and Reading had a higher proportion of Hispanic cases compared to liver cancer cases in the rest of the State of PA. Areas identified as having higher than expected rates of liver cancer also tended to have higher poverty, lower nSES, and higher % of foreign-born residents compared to the rest of the State, suggesting the potentially important role of nSES in helping to identify high risk cluster areas.

In Fig. 2, we first assess the sensitivity/specificity of utilizing

**Table 1**  
Combining liver cancer disease clusters and neighborhood measures for sensitivity/specificity assessments to evaluate precision public health approaches.

	Census Tracts in Statistically Significant Elevated Disease Cluster (Disease)	Census Tracts Outside a Significantly Elevated Disease Cluster (NonDisease)	Total
Positive (has at least one standard demographic variable)	A (True Positive)	B (False Positive)	Total Positive
Negative (has no standard demographic variables)	C (False Negative)	D (True Negative)	Total Negative
	Total Elevated Risk	Total Non-Risk	TOTAL





**Fig. 1.** Location and Number (N) of Census Tracts (CT) in Pennsylvania by Standard Demographic Variables (any 1 out of 3 standard demographic variables (light-orange) or all 3 standard demographic variables (dark-red)) and Liver Cancer Cluster Analysis (hashed). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

standard demographic variables with liver cancer disease cluster data. In assessments with any of the three standard demographic variables (% non-Hispanic Black; % Hispanic, % birth cohort), sensitivity was 80.6%, specificity 54.8% and positive predictive value (PPV) 20.3%. Overlapping areas (dark-orange) indicate Census tracts "at-risk" or identified as being located in a disease cluster and containing the highest quartile of at least one of 3 standard demographic variables (i.e., true positives;  $n = 324$  Census tracts). Yellow-shaded areas were not detected in a high risk cluster, but were identified to contain at least 1 standard demographic variable (i.e., false positives;  $n = 1272$  Census tracts). These findings demonstrate that using standard approaches, areas without a liver cancer burden could be targeted (yellow Census tracts). Further, combining disease data with standard demographic variables from the U.S. Census would reduce targets ( $n = 324$  Census tracts from true positives) more than disease cluster ( $n = 402$  Census tracts) or standard demographic variables ( $n = 1596$  Census tracts) alone.

Next, we determined if the addition of nSES variables could improve sensitivity/specificity assessments. First, nSES measures that were significantly related to being located in a high-risk cluster were identified (Supplementary File 1B), and systematically evaluated using frequency analysis in order to reduce the number of explanatory variables to optimize sensitivity/specificity assessments (Supplementary File 1B/C). After these assessments, 4 nSES variables remained that were significantly associated with liver cancer incidence and occurred in high frequency within high risk liver cancer clusters: % Non-Hispanic Blacks, the Hispanic-ICE, TDS, and neighborhood instability. Comparisons of spatial patterns and changes in sensitivity/specificity assessments using different combinations of the 4 nSES variables alone and in combination with the 3 standard demographic variables were conducted to identify the assessment with the highest sensitivity/specificity, and to determine if the addition of more variables (i.e., all 7 versus 3 variables, etc.) would

result in the best sensitivity/specificity assessment (Supplementary File 1C/D). The final (and best) assessment included Census tracts with highest percentage (i.e., positive for the highest quartile) of the following 4 nSES measures: % Non-Hispanic Black, Hispanic-ICE, TDS, and neighborhood instability. This assessment had a sensitivity of 92.8%, specificity of 67.3% and PPV 28.8%. This was chosen as the final model given that spatial patterns indicated that fewer Census tracts were classified as false positive (Fig. 3-yellow areas) in comparison to the model when 3 standard demographic variables were used (Fig. 2-yellow areas); i.e., more Census tracts in actual high risk clusters were identified ( $n = 374$ ), and AUC estimates were most improved using this approach (0.80 vs 0.87) (Supplementary File 1D). Additionally, when applying this model at the case level instead of the Census tract level, this model also had similar sensitivity/specificity (95.9%/59.2%), meaning a high proportion of current liver cancer cases would be identified for liver cancer interventions. Using this final, 4 variable assessment as an example, we further determined if sensitivity/specificity assessments could be improved if we limited these calculations to areas that contained all 4 nSES variables vs. 3 or more, 2 or more variables, etc. We found that the PPV improved up to 50% if areas only positive for all 4 nSES variables were identified, but this was at the expense of sensitivity (which reduced down to 25%) (Supplementary File 1D2).

### 3.1. Application of precision public health to liver cancer prevention in Philadelphia

Utilizing findings from the best model of the sensitivity/specificity assessments (Fig. 3), we apply this knowledge to outline priority areas in Philadelphia to target for liver cancer prevention (Fig. 4). Philadelphia County is located in Southeast Pennsylvania. It is the most populated city/county in Pennsylvania (1.5 Million residents, Census 2010), and it

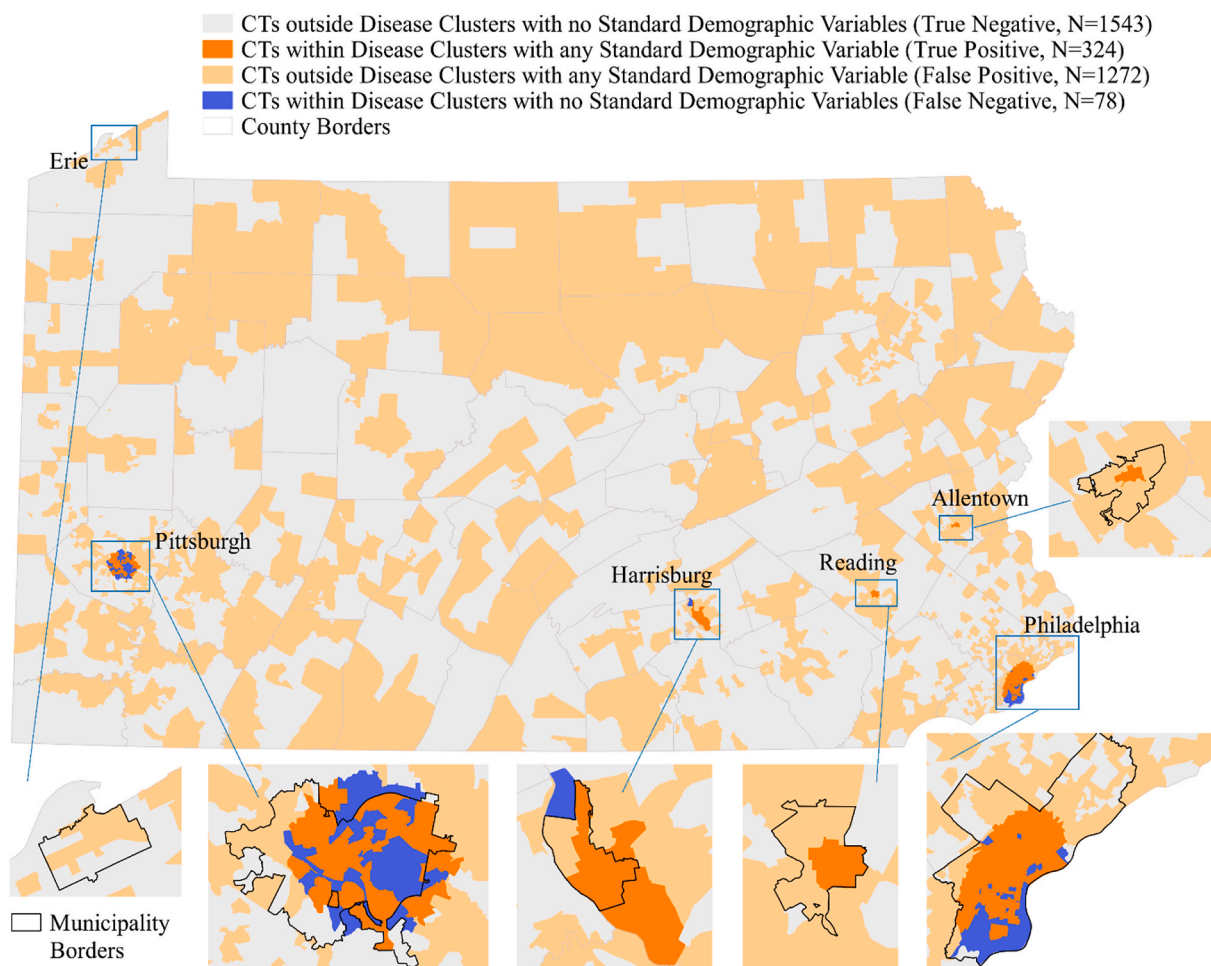
**Table 2**  
Baseline demographics of cases and census tracts located inside and outside of liver cancer disease clusters in Pennsylvania (PA).

Disease Rates	State of PA		Cluster Areas with Higher than Expected Rates of Liver Cancer Incidence										Rest of PA (outside of clusters)	
			Philadelphia		Pittsburgh		Allentown		Harrisburg		Reading			
Census Tracts (n)	3217		231		132		8		19		12		2815	
Cases (n)	9460		1240		339		42		87		47		5658	
Mean Relative Risk (p-Value)	1.0 (Reference)		<0.01		<0.01		<0.01		<0.01		<0.01		N/A	
<b>Patient Characteristics</b>	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<b>Age at Diagnosis (years)</b>														
0-45	313	3.3	55	4.4	7	2.06	0	0.0	2	2.3	2	4.3	176	3.1
46-65	4818	50.9	780	62.9	199	58.7	30	71.4	56	64.4	31	65.9	2812	49.7
>66	4329	45.8	405	32.7	133	39.2	12	28.6	29	33.3	14	29.8	2670	47.2
<b>Sex</b>														
Male	6810	72.0	929	74.9	257	75.8	32	76.2	74	85.1	39	82.9	4008	70.9
Female	2650	28.0	311	25.1	82	24.2	10	23.8	13	15.0	8	17.0	1650	29.2
<b>Race/Ethnicity</b>														
White Non-Hispanic	7217	76.3	414	33.4	177	52.2	27	64.3	29	33.3	29	61.7	4653	82.2
Black Non-Hispanic	1655	17.5	664	53.6	145	42.8	6	14.3	48	55.2	7	14.9	713	12.6
Hispanic	116	1.2	26	2.1	3	0.9	3	7.1	1	1.2	9	19.2	63	1.1
Asian/Pacific Island	324	3.4	93	7.5	8	2.4	1	2.4	7	8.1	0	0.0	161	2.9
Other	148	1.6	43	3.5	6	1.8	5	11.9	2	2.3	2	4.3	68	1.2
<b>Select Census Tract Characteristics</b>														
Total Population	12779559		922469		289547		26849		65495		38913		7344275	
<b>Age (%)</b>														
0-45	55.6		67.5		63.3		73.5		64.5		71.0		54.5	
46-65	28.0		22.2		22.6		19.6		24.3		20.5		28.5	
>66	16.3		10.3		14.1		6.9		11.2		8.5		17.0	
<b>Race/Ethnicity (%)</b>														
White Non-Hispanic	78.1		29.2		62.9		19.7		32.4		22.7		79.2	
Black Non-Hispanic	10.5		43.7		26.3		15.5		42.7		9.0		9.9	
Hispanic	6.4		17.1		2.7		61.2		17.0		64.5		6.3	
Asian/Pacific Island	3.1		7.5		5.2		1.4		3.9		0.7		2.7	
Other	1.8		2.2		2.7		2.0		3.8		3.1		1.9	
<b>Neighborhood Instability (%Population Living at the Same Place as 1 Year Ago)</b>	87.6		83.2		80.0		65.9		79.9		76.3		88.2	
<b>Neighborhood Poverty level</b>														
Q1 < 5.6% (LOW)	28.5		1.8		5.1		0.0		2.3		0.0		29.0	
Q2 < 10.16	25.9		8.3		12.9		0.0		16.0		0.0		26.5	
Q3 < 17.6	24.1		14.4		20.8		0.0		17.1		0.0		25.4	
Q4 > 17.6 (HIGH)	21.5		75.5		61.1		100		64.6		100		19.1	
<b>Townsend Deprivation Score</b>														
Q1 -Very Low Deprivation Level	26.4		0.0		5.8		0.0		3.2		0.0		27.6	
Q2	26.8		0.3		6.3		0.0		2.3		0.0		26.1	
Q3	24.1		9.1		32.6		0.0		29.9		7.1		26.7	
Q4-Very High Deprivation Level	22.7		90.6		55.3		100		64.6		92.9		19.6	
<b>ICE (Hispanic Households)</b>														
Q1 - Very Low Concentration of Hispanic Households	24.7		0.4		4.4		0.0		0.0		0.0		22.0	
Q2	25.8		2.9		8.9		0.0		0.0		0.0		26.7	
Q3	25.3		9.5		39.2		0.0		3.2		0.0		27.4	
Q4- Very High Concentration of Hispanic Households	24.1		87.3		47.5		100		96.8		100		23.9	

Q1 = quartile 1.  
Q2 = quartile 2.  
Q3 = quartile 3.  
Q4 = quartile 4.

contains 384 Census tracts. Of the 384 Census tracts, 231 Census tracts were identified as being located in a significant cluster of elevated relative risk, demonstrating the high burden of liver cancer in the city. Using the 4 selected nSES measures from the final model, we plan to maximize our limited resources, and focus on Census tracts that also contain a high burden of disparity (i.e., that contain the highest percentage of all 4 (Category 1) or at least 3 nSES variables (Category 2)). This approach allows us to reduce intervention targets identified with a disease burden down to 179 Census tracts with the highest local rates of liver cancer (Category 1 relative risk (RR) = 2.96; Category 2 RR = 2.95). However, in the absence of more sophisticated geospatial analyses that identify clusters of higher than expected rates of liver cancer, if we were to only use the 4 nSES variables that are available by downloading Census tract-level data and identify those Census tracts with the highest burden of all 4 or at least 3 of these variables, we would be

targeting 66 Census tracts that do not have a statistically significant elevated risk of liver cancer compared to the rest of the State of PA (i.e., not in a liver cancer cluster), but that do have a significantly elevated local risk of liver cancer (Category 1 outside of the disease cluster RR = 1.59). Further, this number is much lower than if we were to use the 3 standard demographic variables for liver cancer prevention in Philadelphia, where 324 Census tracts and additional 126 Census tracts would be unnecessarily targeted. This suggests that utilizing nSES variables (with or without disease data) when identifying intervention targets for liver cancer could help to maximize limited resources by more precisely pinpointing areas that are likely to have a disease burden.



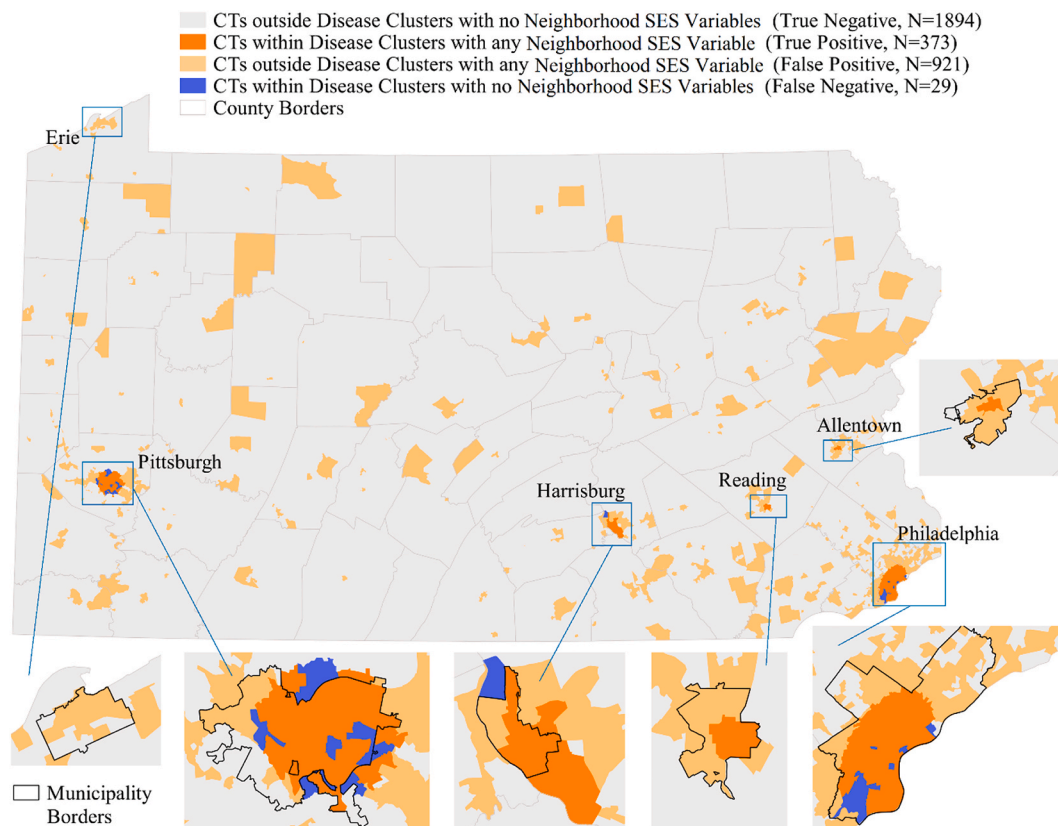
**Fig. 2.** Spatial application of the Sensitivity/Specificity approach to Census Tracts (CT) in Pennsylvania using Liver Cancer Disease Cluster Data and the 3 Standard Demographic Variables (%Black, %Hispanic, %born 1950–1959) from the U.S. Census. Note: Sensitivity = True Positive/(True Positive + False Negative)\*100 – 80.6% Specificity = True Negative/(False Positive + True Negative)\*100 – 54.8% Positive Predictive Value = True Positive/(True Positive + False Negative)\*100 – 20.3% Negative Predictive Value = True Negative/(True Negative + False Negative)\*100 – 95.2%.

#### 4. Discussion

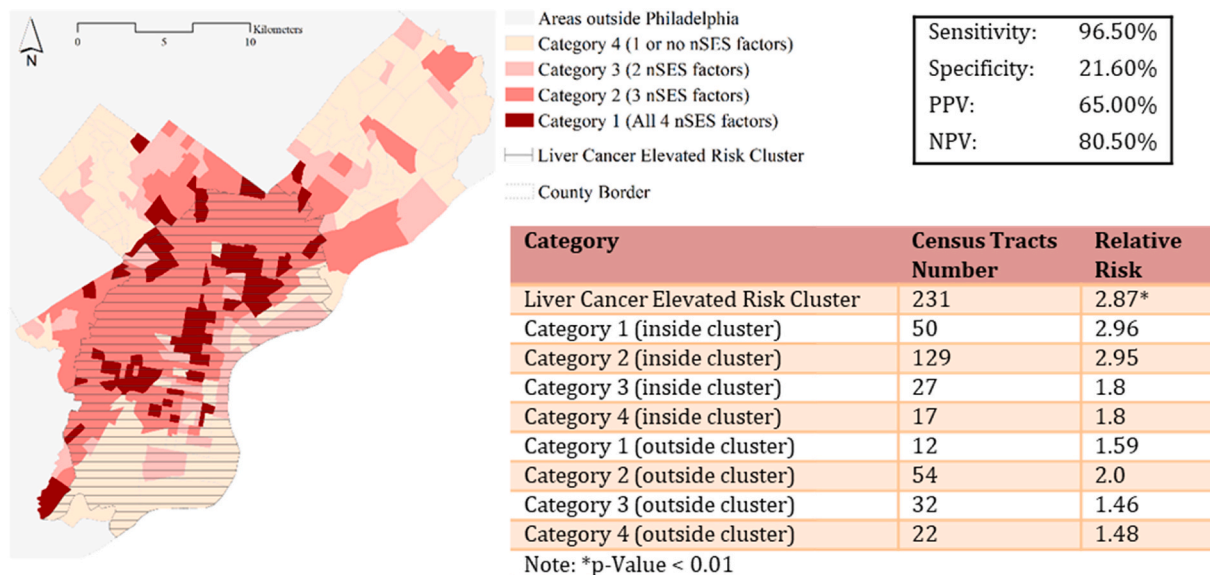
Precision public health requires the linkage of multiple primary surveillance data resources, and the rapid application of sophisticated analytics to track the geospatial distribution of disease in order to reduce geographic targets and act on this information in the form of interventions (Dowell et al., 2016). In this study, we applied precision public health approaches to inform liver cancer prevention efforts in Pennsylvania. We found that combinations of surveillance data, including neighborhood measures from the U.S. Census together with liver cancer disease rates generated from Pennsylvania State cancer registry data, can narrow down Census tracts to target for liver cancer prevention more than standard approaches that use demographic data (race/ethnicity and age) from the U.S. Census only. Using this approach, we are also able to account for or target 4,825 (51%) of the total number of 9,460 liver cancer cases in PA. To our knowledge, we are one of the first studies to quantitatively evaluate precision public health approaches by applying sensitivity/specificity assessments to linked surveillance resources. Utilizing sensitivity/specificity assessments, we were able to evaluate the utility of precision public health by quantifying the number of Census tracts without a known liver cancer burden that might have been targeted using standard recommendations (i.e., identify false positives;  $n = 1272$ ). Given the high false positive rate, we then sought to determine if nSES factors could improve sensitivity/specificity assessments. This is because in the disease cluster analysis,

nSES factors related to income, deprivation, stability, and immigration status were found in higher proportions in high-risk cluster areas compared to the rest of the State of PA. Further, previous population-based studies have found that these nSES measures contribute to both liver cancer incidence, race/ethnic disparities (Nguyen & Thuluvath, 2008), and are also correlated with access to care measures, such as screening utilization (Diez Roux & Mair, 2010). Our model with the highest sensitivity (92.8%) and specificity (67.3%) included one standard demographic variable (% Non-Hispanic Black) and 3 additional nSES variables related to segregation (Hispanic-ICE), deprivation (Townsend Index), and neighborhood instability (% of households still living in same house as one year ago). These findings suggest that nSES could serve as an additional informative marker for high-risk populations in need of liver cancer prevention, particularly in the absence of available disease cluster data, as demonstrated by the application of precision public health approaches to the city of Philadelphia. Thus, moving forward, the incorporation of nSES to prioritize neighborhoods for future community-based liver cancer prevention efforts appears warranted.

While the incorporation of nSES factors improved sensitivity/specificity assessments, the specificity and PPV estimates were still low. In a clinical setting, the goal is to achieve measures above 90%. It is possible that other data resources that include additional risk factor information at the neighborhood level, such as Hepatitis C or B rates, could further improve specificity. The inclusion of additional liver cancer-related



**Fig. 3.** Spatial application of the Sensitivity/Specificity approach to Census Tracts (CT) in Pennsylvania using Liver Cancer Disease Cluster Data and the final assessment, which included one standard demographic variable (%Black), and 3 neighborhood socioeconomic (nSES) variables (Hispanic-Index of Concentration at the Extremes (ICE); Townsend Index, neighborhood instability). Note: Sensitivity = True Positive/(True Positive + False Negative)\*100 - 93.0% Specificity = True Negative/(False Positive + True Negative)\*100 - 62.9% Positive Predictive Value = True Positive/(True Positive + False Negative)\*100 - 26.3% Negative Predictive Value = True Negative/(True Negative + False Negative)\*100 - 98.4%.



**Fig. 4.** Application of Precision Public Health to Liver Cancer Prevention in the City of Philadelphia: Combination of selected neighborhood socioeconomic (nSES) Measures and Disease cluster data from the final sensitivity/specificity assessment helps to prioritize Census tracts for intervention. Note: Sensitivity = True Positive/(True Positive + False Negative)\*100 Specificity = True Negative/(False Positive + True Negative)\*100 Positive Predictive Value (PPV) = True Positive/(True Positive + False Negative)\*100 Negative Predictive Value (NPV) = True Negative/(True Negative + False Negative)\*100.



disease (i.e., Hepatitis B and C surveillance data) and behavior-related risk factors (i.e., obesity, alcohol drinking) could improve not only sensitivity/specificity assessments, but could also lead to the generation of neighborhood profiles that would tell us not only “where” to target liver cancer prevention, but “what type” of intervention would be most useful. For instance, we would target Hepatitis B vaccination in areas with high Hepatitis B rates, but not in areas with low Hepatitis B rates. This targeted approach would further support the application of precision public health for liver cancer prevention. However, ongoing preventive programming and monitoring of the targeted regions will likely be needed to monitor the potential for precision public health approaches to truly reduce regional LC burdens over time.

There are a number of limitations in this study to note. Sensitivity/specificity assessments in a clinic setting rely on a “gold standard” for disease identification. There are no “gold standards” for disease cluster identification. SatScan software is one of the most reliable and commonly-used methods to define spatial clusters of high risk, but it is possible that areas with a liver cancer burden might not have been detected or that Census tracts might have been included in a high risk cluster due to aggregation assumptions within scanning windows (Ozdenrol, Williams, Kang, & Magsumbol, 2005). In the present study, we used the GINI method (Han et al., 2016), and 50% scanning window size was found to be most suitable. Although not reported, we did compare cluster results from SatScan to another software package, BayesX and results were similar. Additionally, our evaluation of the effect of nSES measures on liver cancer incidence was not comprehensive, and it’s possible other nSES measures may be better suited for this type of analysis (Krieger et al., 2002; Wiese, Stroup, Crosbie, Lynch, & Henry, 2019). Additionally, given that the frequency of nSES variables likely changes across State and geographic scale, it’s possible that findings from this study might not be generalizable to other States. Measures of race/ethnic concentration that were found to be important in this study may be related to the fact that the majority of the LC cases are clustered in urban areas, which tend to be racially segregated (Massey, 1990), and therefore have high concentration of a single race/ethnicity in certain neighborhoods. In rural areas with less racial/ethnic segregation and concentration, racial/ethnic ICE measures might not be as effective. Further, it is possible that cases from mental health/treatment facilities could have been included in this analysis and impacted cluster results; however, there were 9,286 unique addresses out of the 9,460 cases, suggesting this effect would be minimal. Additionally, the application of similar methodology to other diseases may require adjustments in scanning window size selection and alternative nSES variables. Finally, utilization of administrative Census tract boundaries may not reflect the true neighborhood utilized or perceived by the population. It’s possible that residents within a Census tract may also be influenced by neighboring Census tracts (Sperling, 2012). Future studies may consider using Census-derived measures that are estimated using surrounding areas to ensure inclusion of neighborhoods with similar conditions. In this way, contiguous geographic areas with similar profiles may be considered as target intervention sites.

## 5. Conclusion

The methods and subsequent findings in this study are particularly informative, given that public health and community outreach organizations from U.S. cancer centers are increasingly tasked with implementing cost-effective educational, behavioral, and screening related interventions that have the broadest reach in their cancer center catchment areas (i.e., areas where their patient populations reside). Due to limited resources, the majority of these centers implement community-based interventions and select priority neighborhoods for intervention, not based on disease outcomes, but on demographic data from the U.S. Census that is publically available and easily accessible. In this study, utilizing our novel strategy of combining established sensitivity/specificity assessments with geospatial cluster analysis, we found

that using only standard approaches would lead to targeting lower risk areas, and not using limited prevention resources efficiently. Analyses that couple disease and Census data, should be standard moving forward; however, in the absence of having geospatial expertise or disease data, coupling demographic and nSES data could help reduce targets for intervention. Further exploration of the present methodology for liver cancer and other diseases across different States, as well as the integration of additional neighborhood SES factors are needed, but findings do support the utilization of precision public health approaches for cancer prevention.

## Author contributions

Shannon Lynch: conceptualization, formal analysis, funding acquisition, methodology, writing – original draft, review, and editing; Daniel Wiese: data curation, formal analysis, writing – review and editing; Angel Ortiz: formal analysis, writing – review and editing; Kristen Sorice: data curation, project administration, resources, writing – review and editing; Minhhuyen Nguyen: supervision, writing – review and editing; Evelyn González: project administration, supervision, writing – review and editing; Kevin Henry: data curation, formal analysis, writing – review and editing.

## Ethics approval

These data were collected following approval from the Pennsylvania Department of Health’s Bureau of Health Statistics & Registries. This research was approved by the Fox Chase Cancer Center Institutional Review Board (protocol #17–9031).

## Declaration of competing interest

None.

## Acknowledgments

These data were supplied by the Bureau of Health Statistics & Registries, Pennsylvania Department of Health, Harrisburg, Pennsylvania. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions. This work was supported by pilot funding from the National Cancer Institute of the National Institutes of Health under Cancer Center Support Grant P30 CA006927 and Mentored Research Scholar Grant in Applied and Clinical Research, MRS-18-098-01-CPHPS, from the American Cancer Society.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssmph.2020.100640>.

## References

- Altekruse, S. F., Henley, S. J., Cucinell, J. E., & McGlynn, K. A. (2014). Changing hepatocellular carcinoma incidence and liver cancer mortality rates in the United States. *American Journal of Gastroenterology*, 109(4), 542–553.
- American Cancer Society. (2018). *Pennsylvania at a glance*. Retrieved from <https://cancerstatisticscenter.cancer.org/#/state/Pennsylvania>.
- Blake, K. D., Ciolino, H. P., & Croyle, R. T. (2019). Population health assessment in NCI-designated cancer center catchment areas. *Cancer Epidemiology Biomarkers & Prevention*, 28(3), 428–430. <https://doi.org/10.1158/1055-9965.epi-18-0811>.
- Boscoe, F. P., Johnson, C. J., Sherman, R. L., Stinchcomb, D. G., Lin, G., & Henry, K. A. (2014). The relationship between area poverty rate and site-specific cancer incidence in the United States. *Cancer*, 120(14), 2191–2198. <https://doi.org/10.1002/cncr.28632>.
- Chang, E. T., Yang, J., Alfaro-Velcamp, T., So, S. K., Glaser, S. L., & Gomez, S. L. (2010). Disparities in liver cancer incidence by nativity, acculturation, and socioeconomic status in California Hispanics and Asians. *Cancer Epidemiology Biomarkers & Prevention*, 19(12), 3106–3118. <https://doi.org/10.1158/1055-9965.Epi-10-0863>.



- Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1), 125–145. <https://doi.org/10.1111/j.1749-6632.2009.05333.x>.
- Dowell, S. F., Blazes, D., & Desmond-Hellmann, S. (2016). Four steps to precision public health. *Nature*, 540(7632), 189–191. <https://doi.org/10.1038/540189a>.
- El-Serag, H. B., & Rudolph, K. L. (2007). Hepatocellular carcinoma: Epidemiology and molecular carcinogenesis. *Gastroenterology*, 132(7), 2557–2576. <https://doi.org/10.1053/j.gastro.2007.04.061>.
- Erwin, P. C., Myers, C. R., Myers, G. M., & Daugherty, L. M. (2011). State responses to America's health rankings: The search for meaning, utility, and value. *Journal of Public Health Management and Practice*, 17(5), 406–412. <https://doi.org/10.1097/PHH.0b013e318211b49f>.
- Goldberg, D. W. (2008). *A geocoding best practices guide*. Retrieved from Springfield, IL.
- Gomez, S. L., Shariff-Marco, S., DeRouen, M., Keegan, T. H., Yen, I. H., Mujahid, M., et al. (2015). The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. *Cancer*, 121(14), 2314–2330. <https://doi.org/10.1002/cncr.29345>.
- Han, J., Zhu, L., Kulldorff, M., Hostovich, S., Stinchcomb, D. G., Tatalovich, Z., et al. (2016). Using Gini coefficient to determine optimal cluster reporting sizes for spatial scan statistics. *International Journal of Health Geographics*, 15(1), 27. <https://doi.org/10.1186/s12942-016-0056-6>.
- Kanarek, N., Tsai, H. L., & Stanley, J. (2011). Health ranking of the largest US counties using the Community Health Status Indicators peer strata and database. *Journal of Public Health Management and Practice*, 17(5), 401–405. <https://doi.org/10.1097/PHH.0b013e318205413c>.
- Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M. J., Subramanian, S. V., & Carson, R. (2002). 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. *American Journal of Epidemiology*, 156(5), 471–482. <https://doi.org/10.1093/aje/kwf068>.
- Krieger, N., Waterman, P. D., Spasojevic, J., Li, W., Maduro, G., & Van Wye, G. (2016). Public health monitoring of privilege and deprivation with the index of concentration at the extremes. *American Journal of Public Health*, 106(2), 256–263. <https://doi.org/10.2105/AJPH.2015.302955>.
- Kulldorff, M., Huang, L., & Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics*, 8, 58. <https://doi.org/10.1186/1476-072x-8-58>.
- Kulldorff, M., Huang, L., Pickle, L., & Duczmal, L. (2006). An elliptic spatial scan statistic. *Statistics in Medicine*, 25(22), 3929–3943. <https://doi.org/10.1002/sim.2490>.
- Lynch, S. M., & Rebbeck, T. R. (2013). Bridging the gap between biologic, individual, and macroenvironmental factors in cancer: a multilevel approach. *Cancer Epidemiology Biomarkers & Prevention*, 22(4), 485–495. <https://doi.org/10.1158/1055-9965.EPI-13-0010>.
- Makarova-Rusher, O., Altekruse, S., McNeel, T., Graubard, B. I., Greten, T., & McGlynn, K. A. (2015). Population attributable fractions of risk factors for hepatocellular carcinoma in the United States. *Cancer*, 122(11), 1757–1765.
- Massey, D. S. (1990). American apartheid: Segregation and the making of the underclass. *American Journal of Sociology*, 96(2), 329–357. Retrieved from [www.jstor.org/stable/2781105](http://www.jstor.org/stable/2781105).
- Massey, D. S., Booth, A., & Crouter, A. C. (2001). The prodigal paradigm returns: Ecology comes back to sociology. *Does it take a village? Community Effects on Children, Adolescents, and Families*, 41–48.
- NCCN, N. C. C. N. (2017). *Hepatocellular carcinoma clinical practice guidelines*. Retrieved from [www.nccn.org](http://www.nccn.org).
- Nguyen, G. C., & Thuluvath, P. J. (2008). Racial disparity in liver disease: Biological, cultural, or socioeconomic factors. *Hepatology*, 47(3), 1058–1066. <https://doi.org/10.1002/hep.22223>.
- Oliver, T. R. (2010). Population health rankings as policy indicators and performance measures. *Preventing Chronic Disease*, 7(5), A101.
- Ozdenerol, E., Williams, B. L., Kang, S. Y., & Magsumbol, M. S. (2005). Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *International Journal of Health Geographics*, 4(1), 19. <https://doi.org/10.1186/1476-072x-4-19>.
- Pennsylvania Cancer Registry. Retrieved from: <https://www.health.pa.gov/topics/Reporting-Registries/Cancer-Registry/Pages/Cancer%20Registry.aspx>.
- Petrick, J. L., Kelly, S. P., Altekruse, S. F., McGlynn, K. A., & Rosenberg, P. S. (2016). Future of hepatocellular carcinoma incidence in the United States forecast through 2030. *Journal of Clinical Oncology*, 34(15), 1787–1794. <https://doi.org/10.1200/JCO.2015.64.7412>.
- Rice, L. J., Jiang, C., Wilson, S. M., Burwell-Naney, K., Samantapudi, A., & Zhang, H. (2014). Use of segregation indices, Townsend index, and air toxics data to assess lifetime cancer risk disparities in metropolitan Charleston, South Carolina, USA. *International Journal of Environmental Research and Public Health*, 11(5), 5510–5526. <https://doi.org/10.3390/ijerph110505510>.
- Ryerson, A. B., Ehemann, C. R., Altekruse, S. F., Ward, J. W., Jemal, A., Sherman, R. L., ... Kohler, B. A. (2016). Annual Report to the Nation on the Status of Cancer, 1975–2012, featuring the increasing incidence of liver cancer. *Cancer*, 122(9), 1312–1337. <https://doi.org/10.1002/cncr.29936>.
- Sahar, L., Foster, S. L., Sherman, R. L., Henry, K. A., Goldberg, D. W., Stinchcomb, D. G., et al. (2019). GIScience and cancer: State of the art and trends for cancer surveillance and epidemiology. *Cancer*, 125(15), 2544–2560. <https://doi.org/10.1002/cncr.32052>.
- Sherman, R. L., Henry, K. A., Tannenbaum, S. L., Feaster, D. J., Kobetz, E., & Lee, D. J. (2014). Applying spatial analysis tools in public health: an example using SaTScan to detect geographic targets for colorectal cancer screening interventions. *Preventing Chronic Disease*, 11, E41. <https://doi.org/10.5888/pcd11.130264>.
- Singal, A. G., Pillai, A., & Tiro, J. (2014). Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: A meta-analysis. *PLoS Medicine*, 11(4), Article e1001624. <https://doi.org/10.1371/journal.pmed.1001624>.
- Sperling, J. (2012). The tyranny of census geography: Small-area data and neighborhood statistics. *Cityscape*, 14(2), 219–223. Retrieved from [www.jstor.org/stable/41581107](http://www.jstor.org/stable/41581107).
- Thornton-Wells, T. A., Moore, J. H., & Haines, J. L. (2004). Genetics, statistics and human disease: Analytical retooling for complexity. *Trends in Genetics*, 20(12), 640–647. <https://doi.org/10.1016/j.tig.2004.09.007>.
- Tunstall, H. (2005). *Neighbourhoods and health*. Kawachi I and Berkman LF (eds). New York: Oxford University press Inc, USA, 2003, pp. 320, £39.50. ISBN 0195138384. *International Journal of Epidemiology*, 34(1), 231–232. <https://doi.org/10.1093/ije/dyh387>.
- Wang, S., Sun, H., Xie, Z., Li, J., Hong, G., Li, D., ... Ma, H. (2016). Improved survival of patients with hepatocellular carcinoma and disparities by age, race, and socioeconomic status by decade, 1983–2012. *Oncotarget*, 7(37), 59820–59833. <https://doi.org/10.18632/oncotarget.10930>.
- Warnecke, R. B., Oh, A., Breen, N., Gehlert, S., Paskett, E., Tucker, K. L., ... Hiatt, R. A. (2008). Approaching health disparities from a population perspective: The National Institutes of Health centers for population health and health disparities. *American Journal of Public Health*, 98(9), 1608–1615. <https://doi.org/10.2105/AJPH.2006.102525>.
- Wetzel, T. M., Graubard, B. I., Quraishi, S., Zeuzem, S., Davila, J. A., El-Serag, H. B., et al. (2013). Population-attributable fractions of risk factors for hepatocellular carcinoma in the United States. *American Journal of Gastroenterology*, 108(8), 1314–1321.
- Wiese, D., Stroup, A. M., Crosbie, A., Lynch, S. M., & Henry, K. A. (2019). The impact of neighborhood economic and racial inequalities on the spatial variation of breast cancer survival in New Jersey. *Cancer Epidemiology Biomarkers & Prevention*, 28(12), 1958–1967. <https://doi.org/10.1158/1055-9965.EPI-19-0416>.