Novel Models to Identify Census Tracts for Hepatitis C Screening Interventions

Thomas Ludden, PhD, Lindsay Shade, PA-C, Jeremy Thomas, MSW, Brisa Urquieta de Hernandez, BUS, Sveta Mohanan, MD, Mark W. Russo, MD, Michael Leonard, MD, Philippe J. Zamor, MD, Charity G. Patterson, PhD, and Hazel Tapp, PhD

Background: Increased screening efforts and the development of effective antiviral treatments have led to marked improvement in hepatitis C (HCV) patient outcomes. However, many people in the United States are still believed to have undiagnosed HCV. Geospatial modeling using variables representing at-risk populations in need of screening for HCV and social determinants of health (SDOH) provide opportunities to identify populations at risk of HCV.

Methods: A literature review was conducted to identify variables associated with patients at risk for HCV infection. Two sets of variables were collected: HCV Transmission Risk and SDOH Level of Need. The variables were combined into indices for each group and then mapped at the census tract level (n = 233). Multiple linear regression analysis and the Pearson correlation coefficient were used to validate the models.

Results: A total of 4 HCV Transmission Risk variables and 12 SDOH Level of Need variables were identified. Between the 2 indexes, 21 high-risk census tracts were identified that scored at least 2 standard deviations above the mean. The regression analysis showed a significant relationship with HCV infection rate and prevalence of drug use ($B = 0.78, P < .001$). A significant relationship also existed with the HCV infection rate for households with no/limited English use ($B = -0.24, P = .001$), no car use ($B = 0.036, P < .001$), living below the poverty line ($B = 0.014, P = .009$), and median household income ($B = -0.00, P = .009$).

Conclusions: Geospatial models identified high-priority census tracts that can be used to map high-risk HCV populations that may otherwise be unrecognized. This will allow future targeted screening and linkage-to-care interventions for patients at high risk of HCV. (J Am Board Fam Med 2020;33:407–416.)

Keywords: Hepatitis C, Regression Analysis, Risk Factors, Social Determinants of Health
Screening efforts and the development of effective antiviral treatments have led to improvements in HCV patient outcomes.10 Observational studies demonstrate a 40% reduction in hepatocellular carcinoma and liver failure in patients who achieve virologic cure, along with a reduction in mortality rates.11–14 Patient experience a 5% to 20% increased risk for liver cirrhosis due to delayed treatment, putting patients at high risk for additional complications, such as hepatocellular carcinoma.10 In 2016, over 18,000 people had HCV as an underlying cause of death.15 Moreover, nearly 42% of primary care physicians reported being unfamiliar with the Centers for Disease Control and Prevention guidelines for HCV screening in a survey of community-based physicians.16

Many factors cause delays in the screening and treatment of patients with HCV. Because social determinants of health (SDOH) play a role in overall health outcomes and, therefore, may lead to delays in screening and treatment, developing solutions to increase HCV screening requires an understanding of the underlying SDOH variables that are correlated with HCV infection rates and affect decision making around screening and treatment of HCV.17–19 Known SDOH variables can be used to identify patients who would benefit from interventions to reduce the delay in screening and treatment. Although some electronic health records (EHRs) are building the capacity to capture SDOH variables to provide service options to patients, this information is not yet consistently captured.

One method to efficiently and effectively identify patients with SDOH is by geographically mapping known SDOH variables and indexing the data at the census tract level.17 Previous work used Geographic Information System mapping combined with participatory research techniques to look at areas in a community most in need of access to primary care services.20 Maps were created using SDOH and community-level attributes to identify geographic regions most in need of access to health care services. Results of SDOH mapping were used to develop and implement interventions around patient populations impacted by SDOH to increase access to primary care services with the intent to improve screenings and reduce time to treatment.21–23

Here, we show how patient HCV characteristics and SDOH variables can be mapped both individually and in combination to identify patient populations most in need of interventions to screen and treat for HCV. Patients diagnosed with HCV and SDOH variables can be mapped out at the census tract level to identify priority patient populations that overlap with high index values of SDOH variables pertinent to HCV.

The objectives of this article are to (1) identify the variables that correlate with the HCV transmission risk, (2) identify the SDOHs that are barriers to screening and treating patients with HCV, (3) identify high-priority census tracts to identify patients and practices for screening interventions for HCV, and (4) validate the results of the of the HCV and SDOH models. We hypothesized that overlap would occur between the census tracks with high risk for HCV transmission and high index values of SDOH variables, identifying priority areas associated with at-risk patient populations.

Methods

The literature was reviewed to identify the most common variables associated with HCV transmission and SDOH that are related to the screening and treatment of HCV patients.5,24–27 The literature review was conducted until we reached saturation and did not find any additional variables. HCV and SDOH variables identified by the literature review were presented to the Social and Economic Team (SET) committee, a project advisory board composed of representatives from a large health care system that includes representation from behavioral health, primary care, social work, case management, human resources, virtual care, quality, and research.

In establishing the working relationship with the SET, we leveraged completed work that was used for a report evaluating SDOH needs pertinent to access health care and a website, including SDOH variables in the following indicator categories: general population, economic, housing and transportation, social resources and education, and food access (Table 1). Although these variables were not specific to HCV, they represented the collective input by the committee to systematically identify the SDOH variables that were representative of the patients seeking health care in the Charlotte, NC, region.28

From the list of identified variables, data were collected from publicly available sources and from the EHR from Atrium Health (formerly Carolinas
HealthCare System), an integrated health care system based in Charlotte, with over 12 million patient encounters per year. The percentage of the population for each census tract that were patients with Atrium Health ranged from 13% to 96% with an average of 56%. Across the system, approximately 57% of patients seen identify as female; 70% are white and 19% African American; 15% are 70 years old, 26% are between the ages of 50 and 69 years, 34% are between 20 and 49 years of age, and 25% are 19 years old or younger; payer mix is 16% Medicare, 15% Medicaid, 49% commercial, and 20% self pay or other.

Data were coded to correspond to census tracts in Mecklenburg County, NC. The variables were then mapped out and reviewed by the research team and SET. The mapped values were standardized and then combined into 2 groups: HCV Transmission Risk and SDOH Level of Need. Using these 2 index maps, we identified priority areas using risk variables mapped at the census tract level in Mecklenburg County, NC, to identify census tracts associated with at-risk patient populations for HCV and SDOH. ArcGIS 10.5.1 (Environmental Research Institute, Redlands, CA) was used to create maps to identify patients in priority census tracts.

The SET reviewed the results of the geospatial models based on the best evidence from the literature and qualitative assessments. The SET viewed the census tracts selected and agreed that, in their opinion and via previous impressions, these areas were likely to benefit from increased screening. The SET agreed that the identified areas were likely to have high rates of undiagnosed HCV based on known risk factors and should be used to target community outreach to increase HCV screening.20,22,23,29–31 The research team identified priority tracts, and draft maps were vetted again by the SET and providers within the health care system through key informant interviews.

### Statistical Methods

HCV Transmission Risk and SDOH Level of Need indices were created using standardized values by calculating the mean and standard deviation for a variable. For each variable value for each census tract, the mean was subtracted and divided by the standard deviation to obtain the standardized score. The standardized scores for all variables for each census tract were then added together and divided by the number of variables to get an average standardized variable. The summarized index score was then again standardized. Multiple linear regression analysis was used to validate the results of the HCV Transmission Risk Index and SDOH Level of Need Index. A Pearson correlation coefficient was computed to assess the relationship between HCV Transmission Risk and SDOH Level of Need. A P value of <0.05 was considered significant. All analyses were conducted using R v3.43 (R Foundation for Statistical Computing, Vienna, Austria).

### Results

The literature review identified 8 variables, and the SET included 10 additional variables related to SDOH based on previous projects evaluating variables in the environment associated with a wide range of health and quality of life outcomes.28

Two of the variables, incarcerated and homeless populations, were not included because they could not be mapped at the census tract level. Injection drug use was modified to drug use, as mode of delivery could not always be verified in the EHR. For the racial minority variable, we chose percent

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Literature Review</th>
<th>Social and Economic Team Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income</td>
<td>Individuals with less than a high school education</td>
<td></td>
</tr>
<tr>
<td>Racial minorities</td>
<td>Households with no/limited English</td>
<td></td>
</tr>
<tr>
<td>Poor access to healthcare services</td>
<td>Single parent households</td>
<td></td>
</tr>
<tr>
<td>Injection drug use</td>
<td>Low access to food sources</td>
<td></td>
</tr>
<tr>
<td>Baby Boomer birth cohort</td>
<td>Median household income</td>
<td></td>
</tr>
<tr>
<td>Uninsured/under-insured/publicly insured</td>
<td>Unemployed individuals</td>
<td></td>
</tr>
<tr>
<td>Homelessness</td>
<td>Households living in rental housing</td>
<td></td>
</tr>
<tr>
<td>Current or history of incarceration</td>
<td>Households paying &gt;30% of income for rent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Households without transportation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crowded households (&gt;1 person per room)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Variables Identified by a Literature Review and the Social and Economic Team Committee Associated with Hepatitis C Transmission Risk and Social Determinants of Health Level of Need
white population and interpreted the results inversely. Of the remaining 16 variables collected at the census tract level, the variables were divided into SDOH variables (n = 12) (Table 2) and HCV Transmission Risk variables (n = 4) (Table 3). Each set of variables was mapped and converted into an index score (Figure 1 and 2). Between the 2 indexes, 21 high-risk census tracts were identified that scored at least 2 standard deviations above the mean of 0. Two census tracts overlapped between the 2 indexes; 11 were uniquely identified as priority tracts for HCV Transmission Risk by virtue of their scores for HCV prevalence, known percentage of drug use, Baby Boomer age cohort, and percentage of white persons in the census tracts. Eight were uniquely identified as priority tracts for SDOH Level of Need (Figure 3). The location of the high priority tracts for the HCV Transmission Risk index mainly cluster around the center city, whereas the high-priority tracts for the SDOH Level of Need index are more dispersed around the center city.

**Table 2. Social Determinants of Health Variables Collected for Mecklenburg County, NC, Census Tracts Used to Identify Populations at Risk for Hepatitis C Infection**

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition and Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals with less than a high school education</td>
<td>Percentage of individuals over the age of 25 who have not completed a high school education*</td>
</tr>
<tr>
<td>Households with no/limited English</td>
<td>Percentage of households over the age of 5 who speak no or limited English at home*</td>
</tr>
<tr>
<td>Single-parent households</td>
<td>Percentage of households who have only 1 parent present in household*</td>
</tr>
<tr>
<td>Low access to food sources</td>
<td>Percentage of the population living more than ½ mile (urban areas) or more than 10 miles (rural areas) from the nearest supermarket, supercenter, or large grocery store†</td>
</tr>
<tr>
<td>Median household income</td>
<td>Median household income*</td>
</tr>
<tr>
<td>Individuals living below the federal poverty line</td>
<td>Percentage of households living below the federal poverty line*</td>
</tr>
<tr>
<td>Unemployed individuals</td>
<td>Percentage of individuals who do not have any form of health insurance*</td>
</tr>
<tr>
<td>Uninsured individuals</td>
<td>Percentage of individuals ages 16 to 64 in the civilian labor force and not employed divided by the number of individuals ages 16 to 64 in the civilian labor force*</td>
</tr>
<tr>
<td>Households living in rental housing</td>
<td>Percentage of households who are paying greater than 30% of their income in housing*</td>
</tr>
<tr>
<td>Households paying &gt;30% of income for rent</td>
<td>Percentage of households without access to a vehicle*</td>
</tr>
<tr>
<td>Households without transportation</td>
<td>Percentage of households who have more than 1 person per room*</td>
</tr>
<tr>
<td>Crowded households (&gt;1 person per room)</td>
<td>Percentage of households who have more than 1 person per room*</td>
</tr>
</tbody>
</table>

*US Census.  
†US Department of Agriculture.

**Table 3. Hepatitis C Transmission Risk Variables Collected for Mecklenburg County, NC, Census Tracts Used to Identify Populations at Risk for Hepatitis C Infection**

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition and Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Caucasian population</td>
<td>Caucasian population divided by total population*</td>
</tr>
<tr>
<td>Percent of population born 1945–1965</td>
<td>Population born between 1945 and 1965 divided by total population*</td>
</tr>
<tr>
<td>Prevalence of drug use</td>
<td>Number of patients who indicate they currently or previously used drugs divided by population*</td>
</tr>
<tr>
<td>Prevalence of hepatitis C infection</td>
<td>Number of patients who have been diagnosed with hepatitis C divided by the total number of patients†</td>
</tr>
</tbody>
</table>

*US Census.  
†Atrium Health, US Census.
available ($B = 0.036, P < .001$), living below the poverty line ($B = 0.014, P = .009$), and median household income ($B = -0.00, P = .009$).

The Pearson correlation between the HCV Transmission Risk Index and the SDOH Level of Need Index showed a positive correlation, ($r = 0.62, n = 233, P < .001$). Although the high-priority tracts between the 2 models did not overlap except for 2 census tracts, the correlation between the 2 indexes were relatively high. The results suggest that the SDOH Level of Need index provided a viable alternative to HCV Transmission Risk mapping for identifying neighborhoods for an intervention.

**Discussion**

Since many patients who are infected with HCV are unaware of their infection, screening patients in primary care provides an opportunity to identify these patients for linkage-to-care and treatment. In addition to the existing known variables, such as patients born between 1945 to 1965 and patients with current or past histories of drug use, other SDOH factors can also be included to further identify high need areas to improve patient screening. Although the SDOH variables overlapped with the existing poverty and low-income variables used in the model, it was important to keep the broad range of variables originally selected by the SET. The results of this study show that the correlation between the HCV Transmission Risk and SDOH Level of Need is high. However, more research is needed to determine if additional screening criteria will enhance the number of patients diagnosed with HCV.4,34–38

With the increase in opioid drug use and drug injection users in the younger, suburban populations as well economically depressed rural areas, the epidemiology of HCV infection is changing. With the development of highly effective, well-tolerated therapy, primary care providers may elect to diagnose and treat HCV. An understanding of how these patients were infected is important for determining who to screen and treat outside the existing Baby Boomer cohort.39 Although using SDOH variables will not
necessarily pick up suburban drug use, our results around the data validation of the HCV infection rate suggest that geographic areas outside the traditional center city that are challenged with both risks for HCV transmission and poor outcomes due to SDOH variables should not be excluded from future interventions.

Epidemiologic research has used geospatial models in many ways to identify at-risk patients with chronic diseases. A study in the United Kingdom used hot spot mapping for diabetic patients to identify potential patients in need of care.\textsuperscript{40} Other research has predicted the prevalence of certain diseases by using spatial regression models that accounted for additional environmental factors and longitudinal frequencies along with socioeconomics characteristics.\textsuperscript{40–45}

**Future Research**

A future study objective is to identify patient populations that correspond with the high-risk census tracts identified here. By identifying these patient populations while acknowledging the sensitivity and cultural aspects of this type of research, the practices with the highest number of patients receiving care can receive educational interventions around screening and treatment of HCV.\textsuperscript{44–46} These interventions may also result in shorter duration of diagnosis to treatment of HCV. In addition to this proposed work, the next steps to establish index utility would be to screen in areas identified as low risk by using the same index and compare the results.

Because the United States Prevention Task Force now recommends screening for all adults aged 18 to 79, future research should include the 18
to 79 age range as the variable in the HCV Transmission risk model. To distinguish the difference in age groups, subgroups could be analyzed, such as 18 to 34, 35 to 49, and 50 to 79, capturing the nuances of each age group and their current and past behaviors.

**Limitations**

This study used publicly available data with data obtained from 1 health care system and, therefore, does not represent the entire population of the study area. Because we are only using data for 1 health care system, we did not have full information for HCV prevalence. Drug use information included in the EHR did not always specify intravenous as the mode of delivery. Since data for formerly incarcerated and homeless populations are not available at the census tract level, the applicability of the outcomes of this model are limited. Another potential SDOH variable indicating a barrier to HCV treatment is the requirement that patients with a history of substance use in the last year in North Carolina must be enrolled in a treatment program and agree to abstinence during treatment as well as toxicology screening.

**Table 4. Validation of the Hepatitis C Transmission Risk Index Using Multiple Linear Regression**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Prevalence of Hepatitis C Infection</th>
<th>B (SE B)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>0.0004 (0.00069)</td>
<td>0.49</td>
</tr>
<tr>
<td>Percent Caucasian Population</td>
<td></td>
<td>0.0014 (0.00094)</td>
<td>0.13</td>
</tr>
<tr>
<td>Percent of Population Born 1945-1965</td>
<td></td>
<td>0.0043 (0.0033)</td>
<td>0.20</td>
</tr>
<tr>
<td>Prevalence of Drug Use</td>
<td></td>
<td>0.78 (0.042)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

*P values in bold are significant.
B, Coefficient Estimate; SE B, Coefficient – Standard Error; R², Multiple R-Squared.
In addition, the data analyzed and presented are at the census tract level. Due to the modifiable area unit problem, where the summary statistics of a census tract do not apply to an individual person, the results of our analysis could not be applied to each individual household of the census tract. The data only encompass the central county of a large metropolitan area in the southeast, and therefore, the results may not be applicable to other types of municipalities in the south or other regions of the United States.

**Conclusions**

Geospatial models identified a selection of census tracts that can be used to identify populations at risk of HCV infection. Mapping will allow future targeted screening and linkage-to-care interventions for patients at a high risk of HCV. This mapping method is applicable to other regions able to assess regional SDOH need and HCV risk variables.

We acknowledge the Social and Economic Team of Atrium Health for their contribution on this project.

To see this article online, please go to: http://jabfm.org/content/33/3/407.full.

### References


