

# County-level Vulnerability Assessment for Rapid Dissemination of HIV or HCV Infection Among Persons who Inject Drugs, United States – Supplemental Appendix

## Contents

<b>Supplemental Methods</b> .....	2
Regression Modeling Analyses .....	2
Modeling Procedure .....	3
Continuous Indicators Linearity Assessment .....	3
Collinearity Assessment of Indicators.....	4
Standardized Regression Coefficients.....	5
Composite Index (Vulnerability) Score and Rank.....	5
<b>Supplemental Results</b> .....	6
Model Fit Results .....	6
Composite Index (Vulnerability) Score and Rank.....	6
Counties Identified as Vulnerable .....	7
<b>References</b> .....	7
<b>Tables</b> .....	7
Table S1. Counties identified in the top 5% of vulnerability ranks by state and rank .....	8
Table S2. States with at least one county identified in the top 5% of highest vulnerability ranks by number of vulnerable counties and percentage of all state counties identified as vulnerable.....	9
<b>Figures</b> .....	9
Figure S1. County-level indicators investigated for association with acute HCV infection. ....	9
Figure S2. Acute HCV infection rate by county. Reported rate of acute HCV infection by county, NNDSS 2012-2013 and model-estimated rate of acute HCV infection by county .....	18
Figure S3. A. Histogram showing vulnerability scores by number of counties in each score group (eg, 0 to 0.25), B. Sigmoid curve showing vulnerability scores by county rank and identifying the scores bordering the top 5% cut-off, and C. Caterpillar curve of vulnerability scores by mean rank 90% confidence intervals bordering the top 5% rank cut-off (red dash line). ....	19

## Supplemental Methods

We identified indicators associated with acute hepatitis C virus (HCV) infection to develop a composite index score (vulnerability score) ranking each county's vulnerability to rapid dissemination of IDU-associated HIV if introduced, and new or continuing high numbers of HCV infections among persons who inject drugs (PWID). We chose acute HCV infection as the outcome that best serves our purpose because it is collected at the county-level for almost all states.

### Regression Modeling Analyses

We modeled the number of acute HCV infections by county using a multilevel Poisson model with the county population set as the offset.<sup>1</sup> Our data have a multilevel structure with the  $i^{th}$  year (2012, 2013) nested in the  $j^{th}$  county and  $j^{th}$  county nested in the  $k^{th}$  state. The Poisson distribution is defined as:

$$Pr(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where  $y = 0, 1, 2, \dots$ , and  $\lambda$  is the expected rate. The Poisson model uses the  $\log_e$  function that relates the expected value of the response variable to the linear predictor. Hence, the expected rate,  $\lambda$ , is modeled using the link function  $\log_e$  as:

$$\log_e(\lambda) = \eta = \log_e(\text{Population}) + \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

where  $X$  is the  $i^{th}$  indicator and  $\beta$  is the associated model parameter, and  $\beta_0$  the intercept (i.e., overall mean). The offset,  $\log_e(\text{Population})$ , is the county population. We have multilevel data and we model the levels (i.e., state and county nested within a state) as random effects to account for spatial heterogeneity (i.e., overdispersion). We modified our model so the  $\log_e$  link function relates the conditional mean (i.e., conditional on the random effects) of the response variable (i.e., acute HCV rate) to the linear indicator of the fixed and random effects. Our model including random effects for the county and state is given by:

$$\log_e(\lambda|b) = \eta = \log_e(\text{Population}) + \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + b_0 Z_{jk} + b_1 Z_k$$

Where the random effects,  $b_0$  and  $b_1$ , are assumed to be  $N(0, \sigma^2_{jk})$  and  $N(0, \sigma^2_k)$ , respectively. We used SAS GLIMMIX<sup>2</sup> and the residual subject-specific pseudo-likelihood (RSPL) model estimation method.

## Modeling Procedure

We fit a univariable Poisson random-effects models for each of the 15 considered indicators. Figure S1 depicts county-level data by class for each of the 15 considered indicators. Per capita income and population density were modeled on  $\log_{10}$  scale. Aside from urgent care and highway exit, which were coded as yes/no, the other indicators were treated as continuous variables. Our goal was to develop a parsimonious model that is significantly associated with acute HCV infection rate. We entered all 15 indicators in the multivariable model and removed the indicators with the highest p-value. We removed and added indicators in a backwards stepwise procedure until all remaining indicators had a p-value < 0.05.

## Continuous Indicators Linearity Assessment

We assessed linearity for the 13 continuous indicators. Our assumption was that these indicators were linear on the  $\log_e(\text{acute HCV rate})$  scale. To assess the assumption of linearity of the rate on the log-scale we used the following procedure.

1. Calculate the quintiles for the indicator
2. Calculate the acute HCV rate by quintile
3. Plot the  $\log_e(\text{acute HCV rate})$  versus the quintile for the indicator

4. Estimate the slope and intercept of the  $\log_e(\text{acute HCV rate})$  versus quintile
5. Visually assess the assumption of linearity of the indicator

### Collinearity Assessment of Indicators

If an indicator is nearly a linear combination of other indicators in the model, the affected estimates may be unstable and have high standard errors. This situation is usually referred to as collinearity or multicollinearity. We used a generalized linear model (GLM) with counts as the outcome, which required a different procedure to assess collinearity than for a linear regression model. To assess collinearity we relied on three calculated statistics: eigenvalue, condition index, and principal component proportion of variation. An eigenvalue is a computed value that characterizes the essential properties and numerical relationships within a matrix. Eigenvalues that are close to zero may be indicative of a matrix that is close to singular, which indicates collinearity. Eigenvalues  $<0.01$  are usually thought to be close to zero. The condition index is defined as the square root of the ratio of the largest eigenvalue to each individual eigenvalue. The largest condition index (i.e., the square root of the ratio of the largest to the smallest eigenvalue) is the condition number of the scaled X matrix which, as noted by Belsley et al (1990), suggest that when this number is nearing 10, weak dependencies might start to affect the regression estimates.<sup>3</sup> When this number is larger than 100, the estimates likely include significant numerical error.

To calculate the eigenvalues, condition indices, and proportions of variation we used a two-step process in SAS.<sup>2</sup> First, we fit the Poisson model using PROC GENMOD and output the Hessian weights. Secondly, we fit a linear model, using PROC REG, with the Hessian weights defined in the weight statement to obtain the collinearity diagnostics.

## Standardized Regression Coefficients

Standardized regression coefficients for our final multivariable model were calculated to determine the relative importance of each indicator. We calculated the standardized regression coefficients using:

$$\beta_p^s = \frac{Std(X_p)}{Std(y^{pseudo})} \beta_p$$

Where  $\beta_p$  is the estimated regression coefficient from the final multivariable model,  $Std$  is the standard deviation of the  $X^{th}$  indicator and  $pseudo$  outcome  $y$ . The  $pseudo$  outcome is the outcome estimated on the  $log_e$  scale from the estimated regression model.

## Composite Index (Vulnerability) Score and Rank

Our primary goal was to develop a composite index score for ranking the county vulnerability to rapid dissemination of IDU-associated HIV if introduced, and new or continuing high numbers of acute HCV infection among PWID. We developed a vulnerability score using data from the indicators identified in the final multivariable model and the following method to rank counties from lowest to highest vulnerability. We used regression coefficients and observed values to compute the index score for each county. The score for the  $j^{th}$  county was calculated using the regression coefficients ( $\beta$ ) and indicators ( $X$ ) as given by:

$$S_j^u = \beta_1 X_1 + \dots + \beta_p X_p$$

The intercept,  $\beta_0$ , is not used because it is a constant and has no impact on the ranking of counties based on the scores. Once the vulnerability score was calculated for each county, including those not used in

fitting the model, they were ranked from 1 - 3143 with higher scores interpreted as being more vulnerable. Ranks using regression coefficients include uncertainty. To account for uncertainty in the ranks we used simulation to estimate the 90% confidence interval (CI) for each county's rank. We drew 10,000 samples from a normal distribution for each regression coefficient using their estimate and standard error of the estimate. For each of the 10,000 samples we calculated the county's vulnerability score and rank and then obtained a CI for each county's rank.

The threshold for classifying the most vulnerable counties was set at the 95th percentile (top 5%). The 95<sup>th</sup> percentile threshold of the ranks was calculated using all 3,143 counties as  $0.95 * 3,143 = 2985.85$ . We used the upper bound of the 90% CI to determine if a county's rank was within the 95<sup>th</sup> percentile. Once we determined the counties that were within the threshold we ranked them using the inverse of their mean estimated rank (1=highest vulnerability).

## **Supplemental Results**

### **Model Fit Results**

The final multivariable model with the 6 indicators closely aligned the reported HCV rates with the model-estimated HCV rates. To illustrate the model fit we mapped the reported and model-estimated rates of acute HCV infection per 10,000 population (Figure S2). Fewer than 15% (469 of 3,143) of counties varied by more than 1 class when comparing the reported and model-estimated rate of acute HCV infection. The average absolute difference in the model-estimated rates was 0.011 per 10,000 population lower than the actual rates.

### **Composite Index (Vulnerability) Score and Rank**

Figure S3a shows a histogram of the vulnerability scores by number of counties in each score group (e.g., 0 to 0.25). Figure S3b shows a sigmoid curve of the vulnerability scores by county rank.

Using the mean average rank, 157 counties were ranked above the inclusion threshold. The red circle identifies the counties that border the top 5% cut-off. Figure S3c shows a caterpillar curve of the 90% confidence intervals (CIs) bordering the top 5% cut-off. An additional 63 counties were identified above the threshold based on their 90% CI for a total of 220 vulnerable counties.

### Counties Identified as Vulnerable

Table S1 lists the counties identified within the top 5% threshold of vulnerability ranks by state and rank. Table S2 summarizes information on the 220 counties by state; including information on the number of counties identified. Seven states had 10 or more vulnerable counties: Indiana, Kentucky, Michigan, Missouri, Ohio, Tennessee, and West Virginia.

### References

1. Gelman, A and Hill, J. Data Analysis Using Regression and Multilevel/Hierarchical Models. (2007). Cambridge University Press. 625 p.
2. SAS Institute, Inc. SAS<sup>®</sup>: Version 9.3 for Windows. Cary (NC): SAS Institute, Inc.; 2012.
3. Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. (1990). John Wiley & Sons Inc.

### Tables

Table S1. Counties identified in the top 5% of vulnerability ranks by state and rank

FIPS	County	Rank	FIPS	County	Rank	FIPS	County	Rank	FIPS	County	Rank
<b>Alabama</b>			<b>Kentucky (cont.)</b>			<b>Missouri (cont.)</b>			<b>Tennessee (cont.)</b>		
01127	Walker	37	21133	Letcher	50	29153	Ozark	185	47063	Hamblen	138
01093	Marion	100	21115	Johnson	53	29229	Wright	194	47007	Bledsoe	139
01133	Winston	109	21207	Russell	54	<b>Montana</b>			47159	Smith	140
01059	Franklin	206	21063	Elliott	56	30061	Mineral	161	47109	McNairy	141
<b>Arizona</b>			21125	Laurel	65	30103	Treasure	211	47139	Polk	142
04015	Mohave	208	21041	Carroll	67	<b>Nevada</b>			47089	Jefferson	149
<b>Arkansas</b>			21217	Taylor	75	32029	Storey	52	47163	Sullivan	151
05135	Sharp	157	21081	Grant	77	32009	Esmeralda	118	47181	Wayne	160
05075	Lawrence	201	21001	Adair	93	<b>North Carolina</b>			47101	Lewis	168
<b>California</b>			21137	Lincoln	97	37043	Clay	63	47091	Johnson	169
06063	Plumas	152	21231	Wayne	99	37193	Wilkes	104	47099	Lawrence	172
06033	Lake	199	21057	Cumberland	101	37075	Graham	124	47179	Washington	198
<b>Colorado</b>			21077	Gallatin	108	37023	Burke	176	47177	Warren	203
08025	Crowley	220	21011	Bath	125	37039	Cherokee	189	47095	Lake	216
<b>Georgia</b>			21085	Grayson	126	<b>Ohio</b>			<b>Texas</b>		
13111	Fannin	82	21089	Greenup	129	39001	Adams	51	48155	Foard	204
13281	Towns	120	21087	Green	132	39131	Pike	72	<b>Utah</b>		
13213	Murray	159	21045	Casey	153	39079	Jackson	111	49007	Carbon	84
13143	Haralson	200	21043	Carter	154	39105	Meigs	123	49001	Beaver	114
<b>Illinois</b>			21171	Monroe	163	39015	Brown	127	49015	Emery	186
17069	Hardin	68	21079	Garrard	167	39145	Scioto	136	<b>Vermont</b>		
<b>Indiana</b>			21201	Robertson	175	39163	Vinton	146	50009	Essex	143
18143	Scott	32	21135	Lewis	178	39053	Gallia	155	50025	Windham	219
18175	Washington	57	21061	Edmonson	179	39009	Athens	173	<b>Virginia</b>		
18149	Starke	70	21003	Allen	180	39027	Clinton	190	51027	Buchanan	28
18041	Fayette	81	21019	Boyd	187	39071	Highland	196	51051	Dickenson	29
18155	Switzerland	94	21105	Hickman	191	<b>Oklahoma</b>			51167	Russell	61
18025	Crawford	112	21027	Breckinridge	202	40067	Jefferson	89	51105	Lee	73
18065	Henry	128	21037	Campbell	212	40025	Cimarron	217	51195	Wise	78
18079	Jennings	158	21167	Mercer	214	<b>Pennsylvania</b>			51185	Tazewell	96
18137	Ripley	195	<b>Maine</b>			42079	Luzerne	38	51141	Patrick	166
18029	Dearborn	213	23027	Waldo	135	42021	Cambria	131	51197	Wythe	210
<b>Kansas</b>			23025	Somerset	145	42039	Crawford	188	<b>West Virginia</b>		
20207	Woodson	144	23029	Washington	170	<b>Tennessee</b>			54047	McDowell	2
20001	Allen	171	23011	Kennebec	193	47067	Hancock	13	54059	Mingo	7
20205	Wilson	181	<b>Michigan</b>			47087	Jackson	19	54109	Wyoming	16
20153	Rawlins	218	26129	Ogemaw	86	47005	Benton	24	54081	Raleigh	18
<b>Kentucky</b>			26035	Clare	87	47151	Scott	26	54045	Logan	20
21237	Wolfe	1	26135	Oscoda	88	47135	Perry	33	54005	Boone	22
21025	Breathitt	3	26119	Montmorency	91	47071	Hardin	36	54019	Fayette	27
21193	Perry	4	26085	Lake	137	47029	Cocke	41	54065	Morgan	44
21051	Clay	5	26141	Presque Isle	174	47015	Cannon	42	54063	Monroe	47
21013	Bell	6	26001	Alcona	184	47137	Pickett	43	54029	Hancock	49
21131	Leslie	8	26143	Roscommon	192	47013	Campbell	46	54015	Clay	60
21121	Knox	9	26039	Crawford	197	47019	Carter	59	54099	Wayne	62
21071	Floyd	10	26079	Kalkaska	207	47027	Clay	64	54009	Brooke	76
21053	Clinton	11	26031	Cheboygan	215	47057	Grainger	66	54053	Mason	85
21189	Owsley	12	<b>Mississippi</b>			47073	Hawkins	71	54013	Calhoun	90
21235	Whitley	14	28141	Tishomingo	164	47173	Union	74	54067	Nicholas	98
21197	Powell	15	<b>Missouri</b>			47059	Greene	79	54089	Summers	110
21119	Knott	17	29179	Reynolds	55	47025	Claiborne	80	54101	Webster	113
21195	Pike	21	29123	Madison	58	47085	Humphreys	83	54043	Lincoln	121
21153	Magoffin	23	29187	St. Francois	69	47145	Roane	92	54011	Cabell	122
21065	Estill	25	29039	Cedar	107	47133	Overton	95	54091	Taylor	133
21129	Lee	30	29093	Iron	117	47041	DeKalb	102	54055	Mercer	147
21165	Menifee	31	29223	Wayne	119	47143	Rhea	103	54007	Braxton	150
21159	Martin	34	29221	Washington	130	47121	Meigs	105	54095	Tyler	162
21021	Boyle	35	29055	Crawford	148	47129	Morgan	106	54087	Roane	165
21127	Lawrence	39	29085	Hickory	156	47049	Fentress	115	54051	Marshall	182
21203	Rockcastle	40	29013	Bates	177	47111	Macon	116	54003	Berkeley	205
21095	Harlan	45	29181	Ripley	183	47185	White	134	54039	Kanawha	209
21147	McCreary	48									



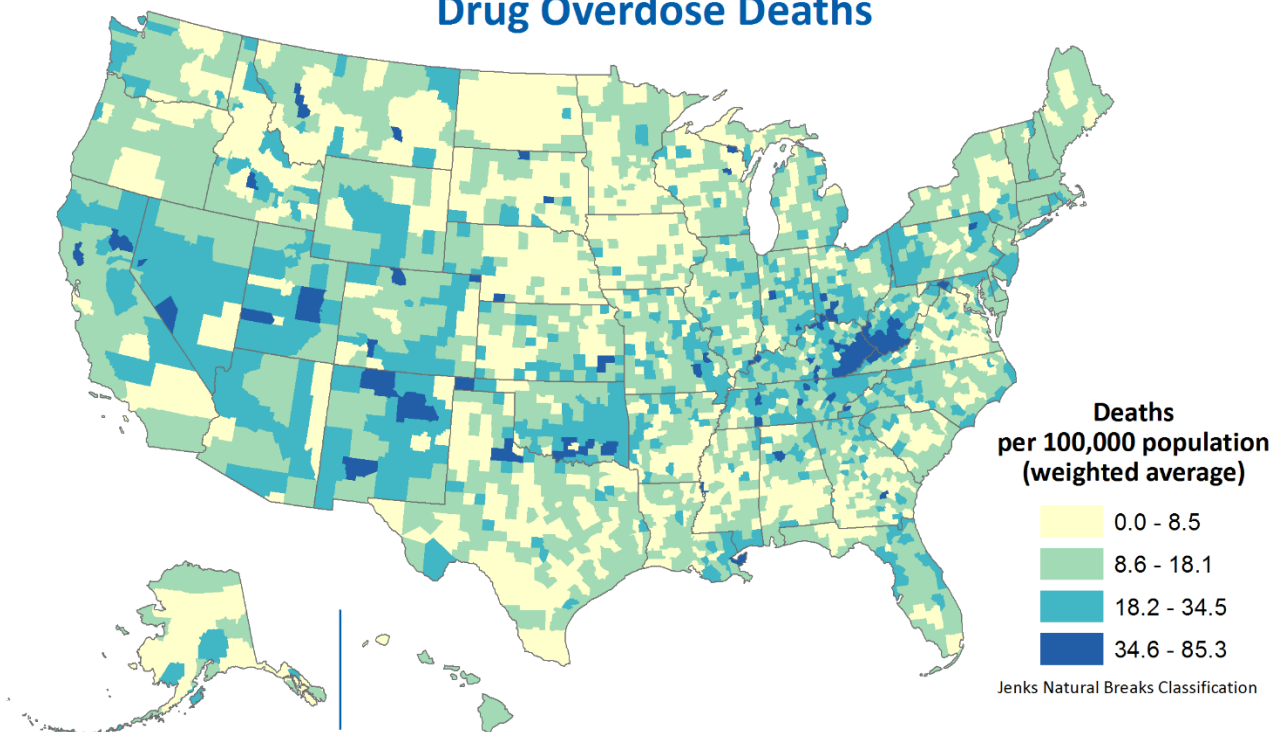
Table S2. States with at least one county identified in the top 5% of highest vulnerability ranks by number of vulnerable counties and percentage of all state counties identified as vulnerable.

<b>State</b>	<b>Vulnerable #</b>	<b>Total #</b>	<b>Identified Vulnerable (%)</b>
Alabama	4	67	6.0
Arizona	1	15	6.7
Arkansas	2	75	2.7
California	2	58	3.5
Colorado	1	64	1.6
Georgia	4	159	2.5
Illinois	1	102	1.0
Indiana	10	92	10.9
Kansas	4	105	3.8
Kentucky	54	120	45.0
Maine	4	16	25.0
Michigan	11	83	13.3
Mississippi	1	82	1.2
Missouri	13	115	11.3
Montana	2	56	3.6
Nevada	2	17	11.8
North Carolina	5	100	5.0
Ohio	11	88	12.5
Oklahoma	2	77	2.6
Pennsylvania	3	67	4.5
Tennessee	41	95	43.2
Texas	1	254	0.4
Utah	3	29	10.3
Vermont	2	14	14.3
Virginia	8	134	6.0
West Virginia	28	55	50.9
<b>Total</b>	<b>220</b>	<b>2139</b>	<b>10.3%</b>

## Figures

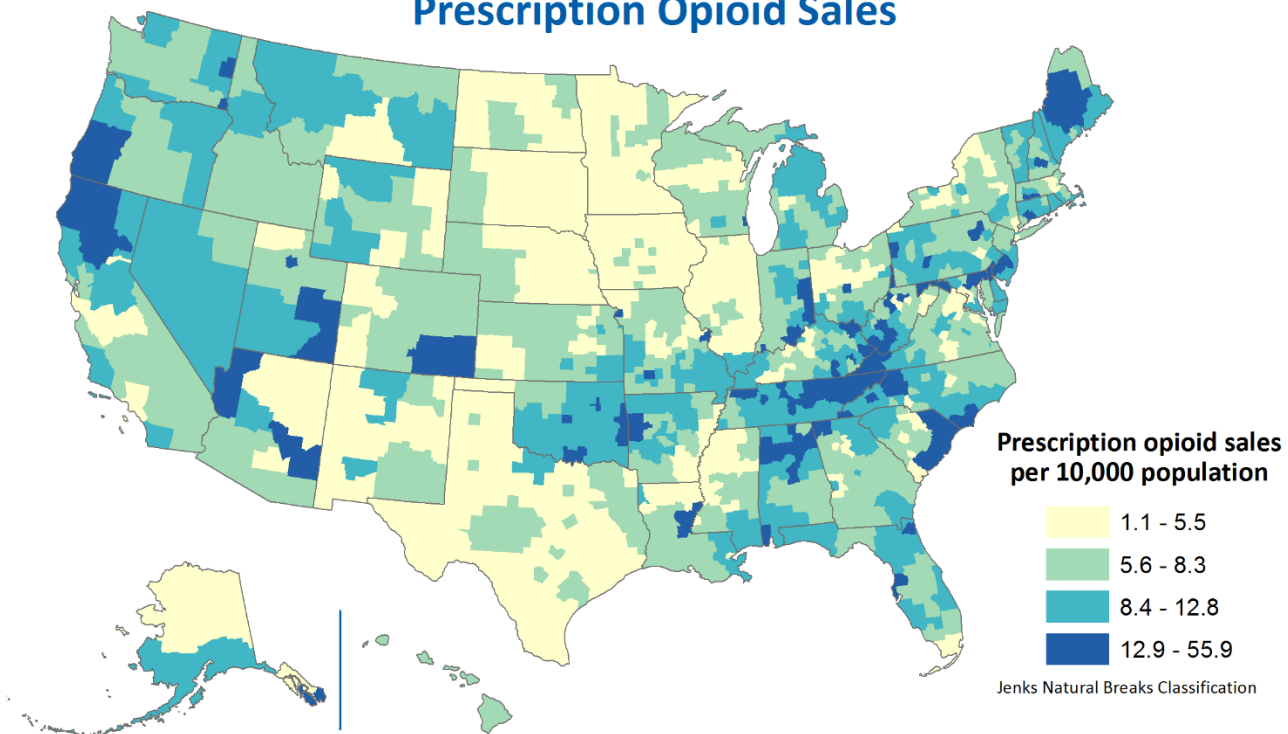
Figure S1. County-level indicators investigated for association with acute HCV infection.

## Drug Overdose Deaths



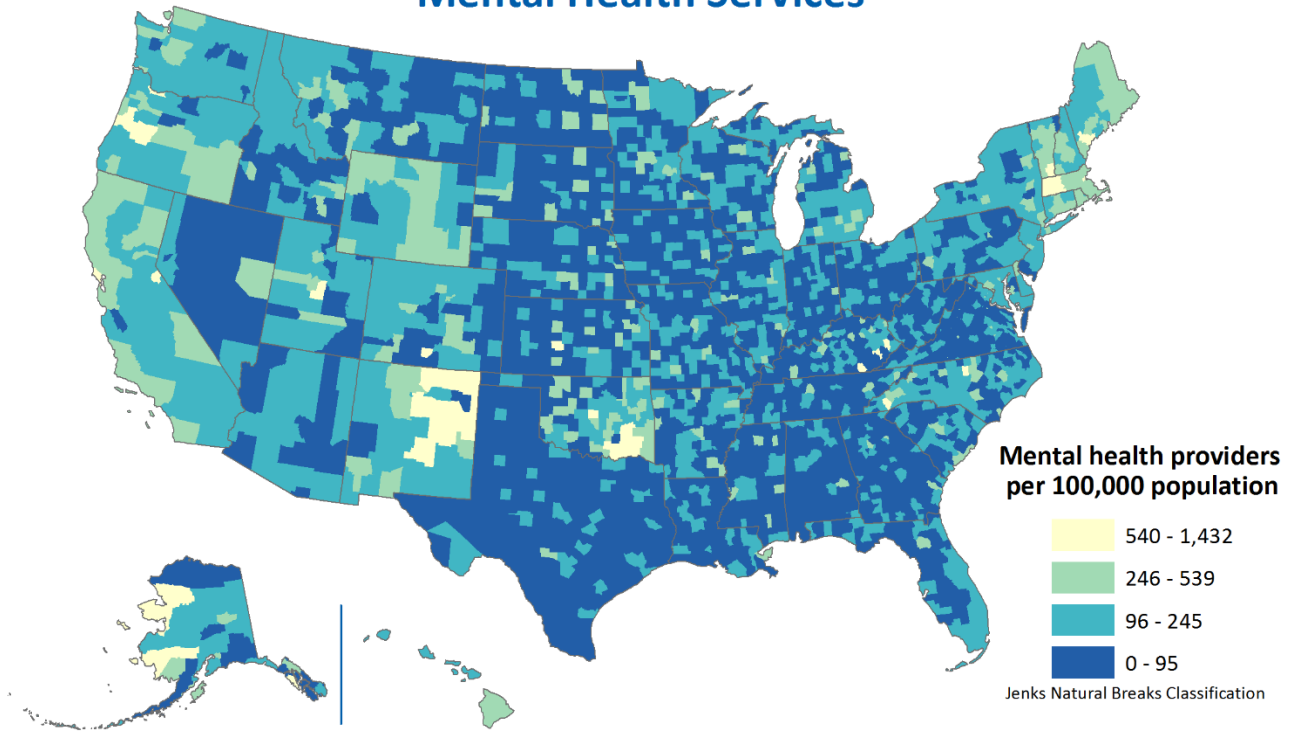
Source: National Center for Health Statistics (NCHS)/National Vital Statistics, 2012 and 2013

## Prescription Opioid Sales



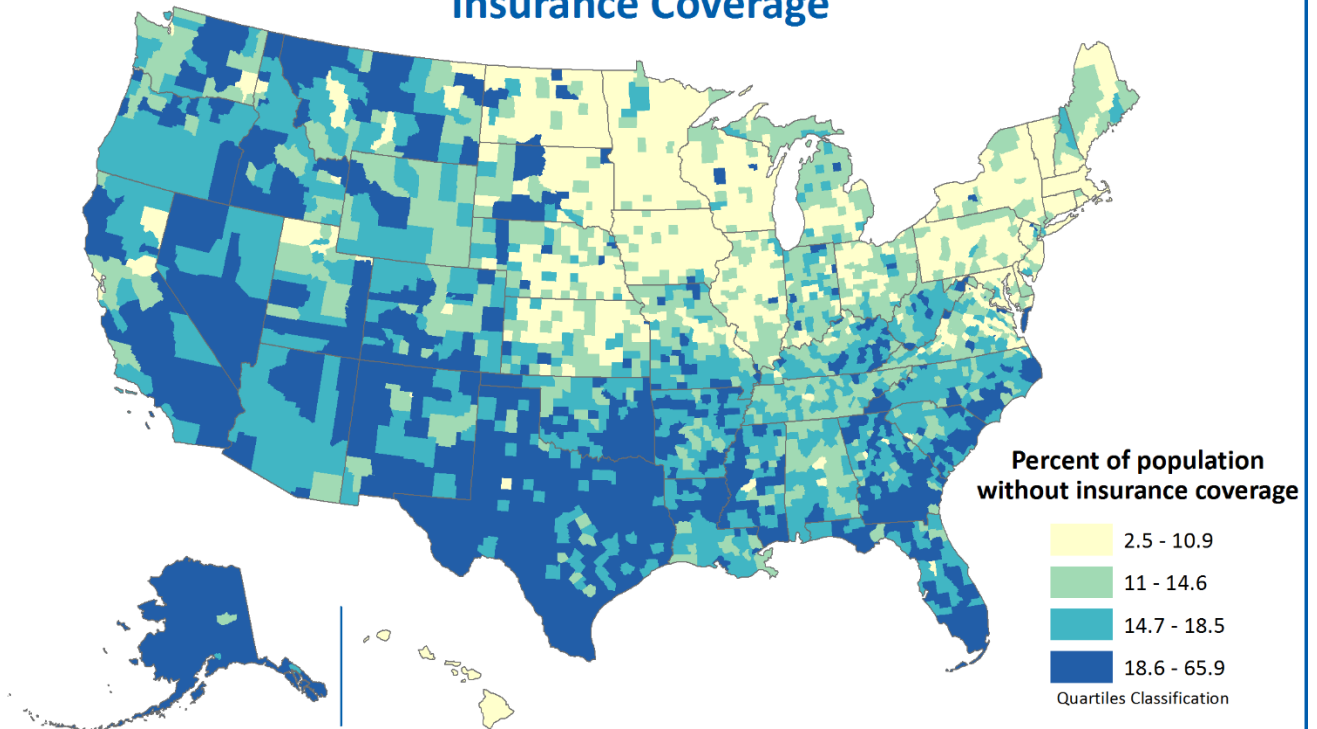
Source: Drug Enforcement Administration (DEA) – Automation of Reports and Consolidated Orders System (ARCOS), 2013. DEA provided 3-digit ZIP data and authors transformed data to 5-digit county FIPS.

## Mental Health Services



Source: CMS, National Provider Identification, 2014

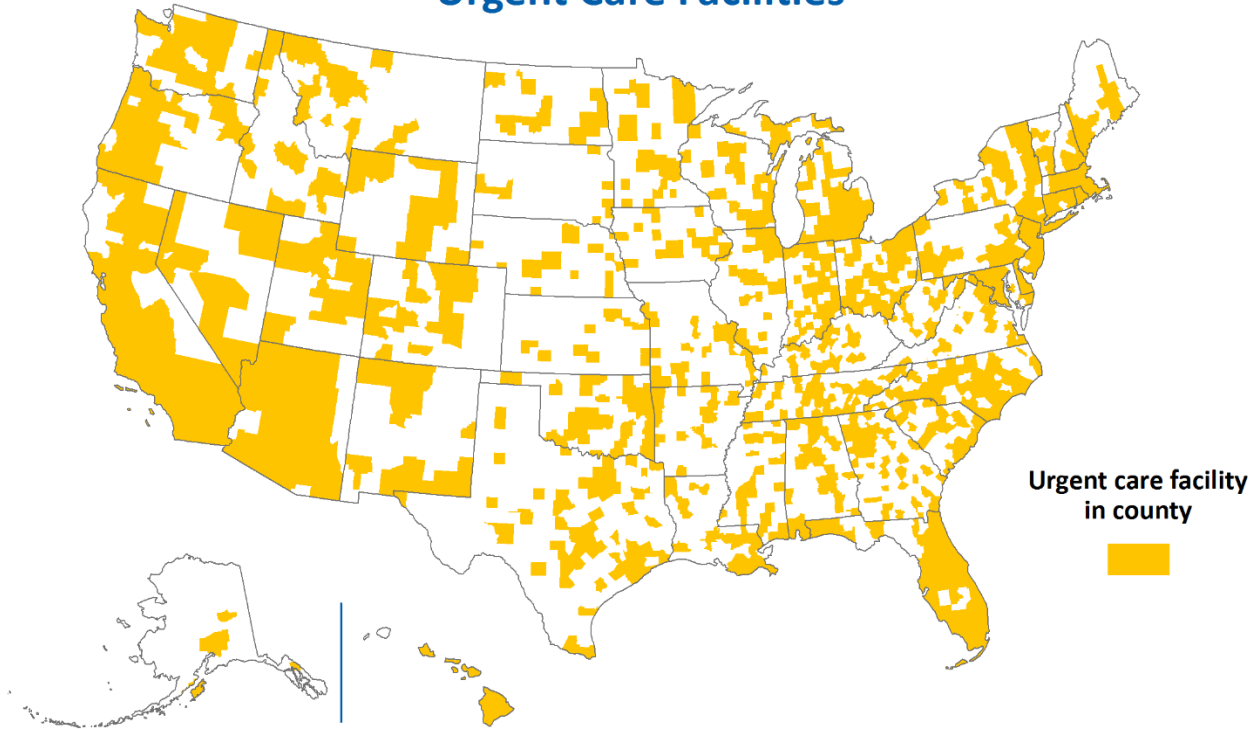
## Insurance Coverage



Source: American Community Survey, 2012 and 2013 5-year estimates

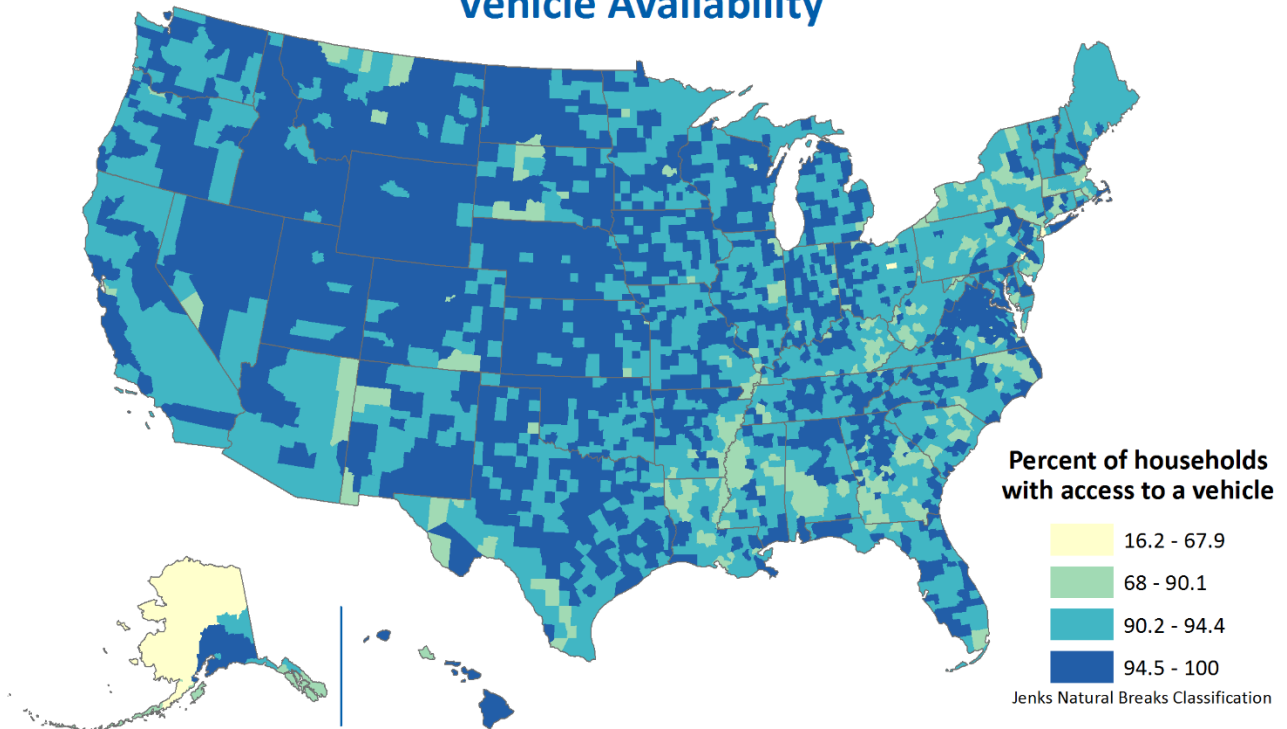
Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).

## Urgent Care Facilities



Source: HSIP 2012 Gold; last updated 2009

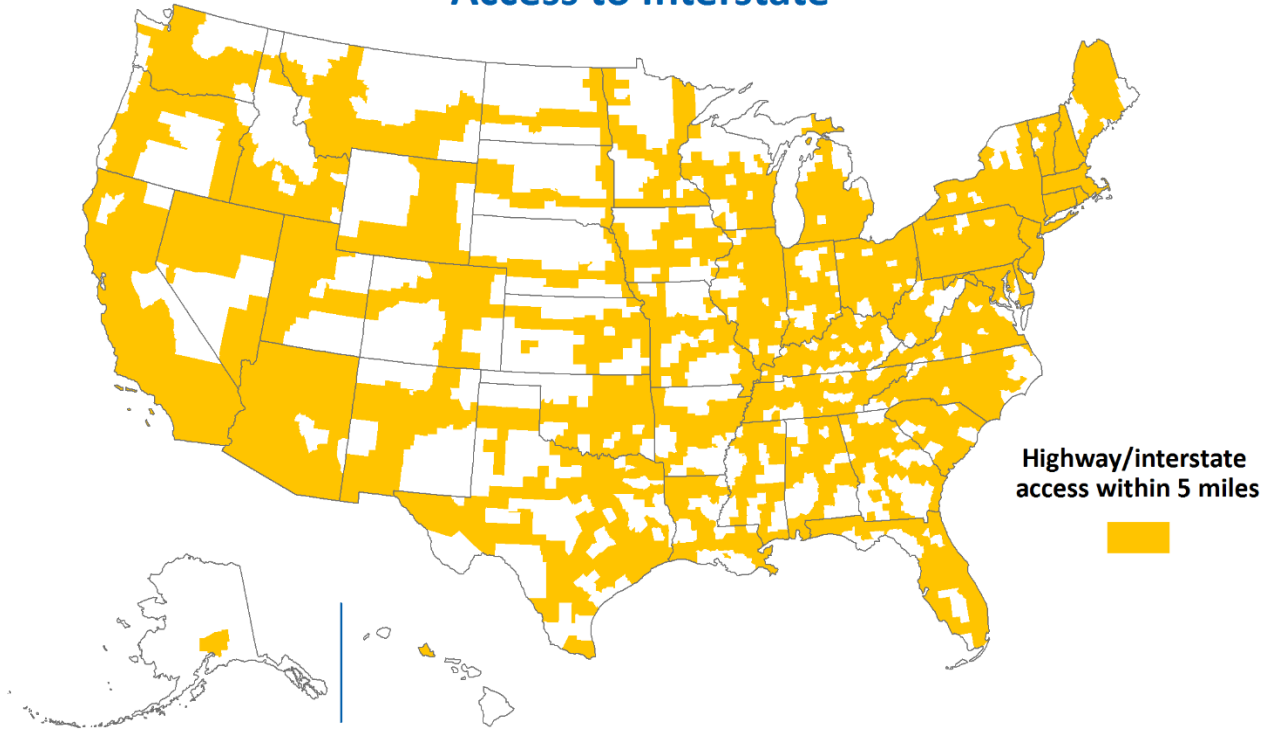
## Vehicle Availability



Source: American Community Survey, 2012 and 2013 5-year estimates

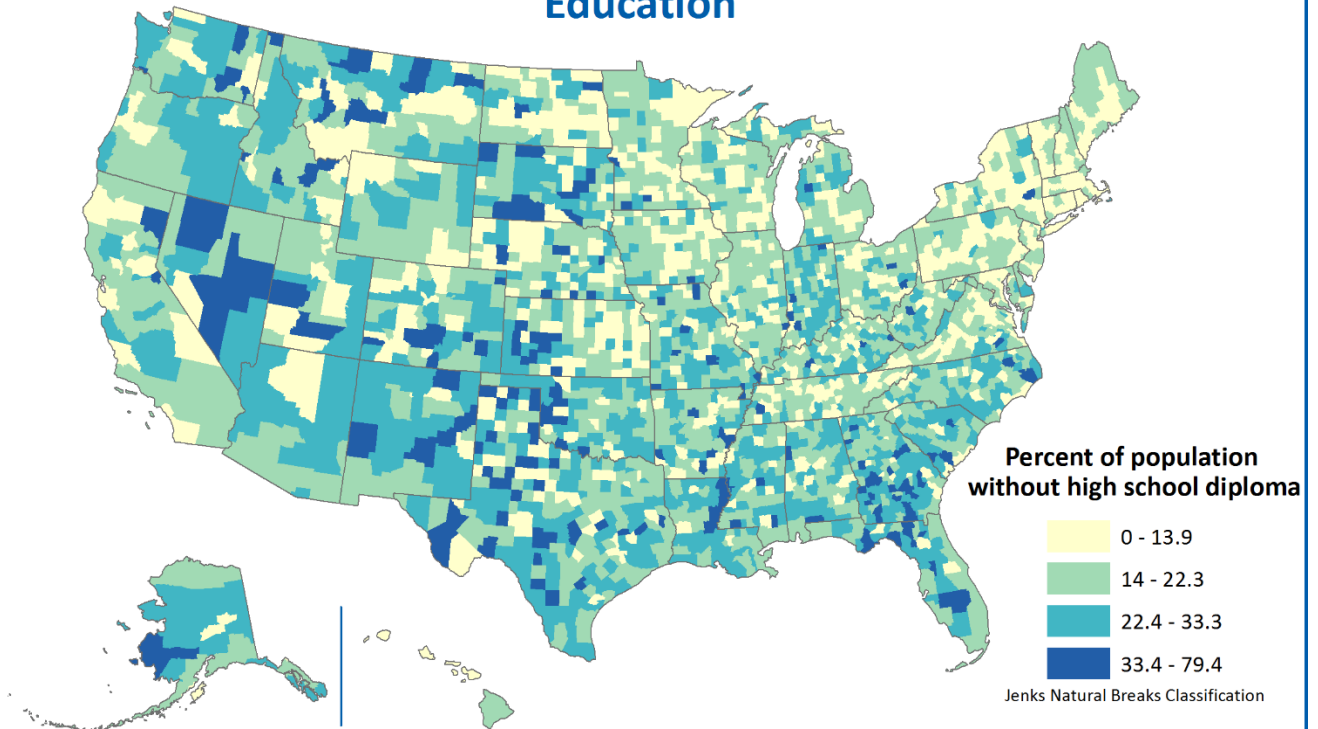
Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).

## Access to Interstate



Source: ESRI maps and data 2014; last updated 2010

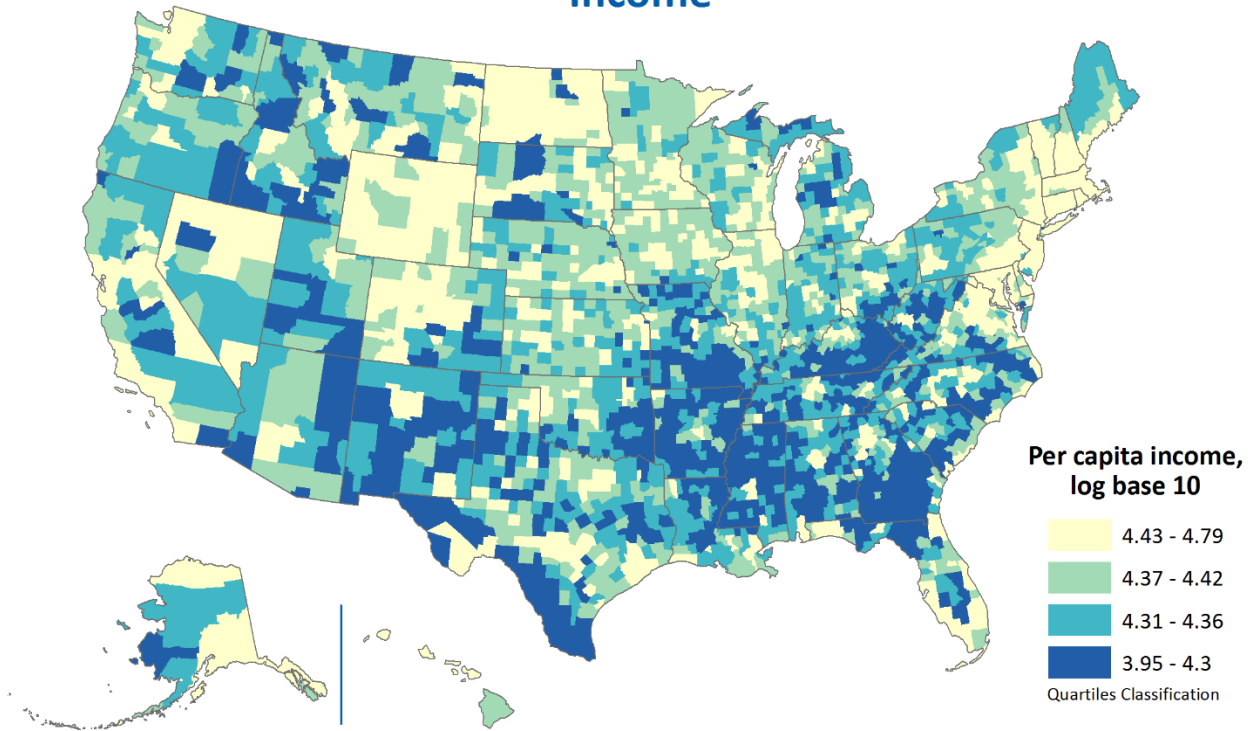
## Education



Source: American Community Survey, 2012 and 2013 5-year estimates

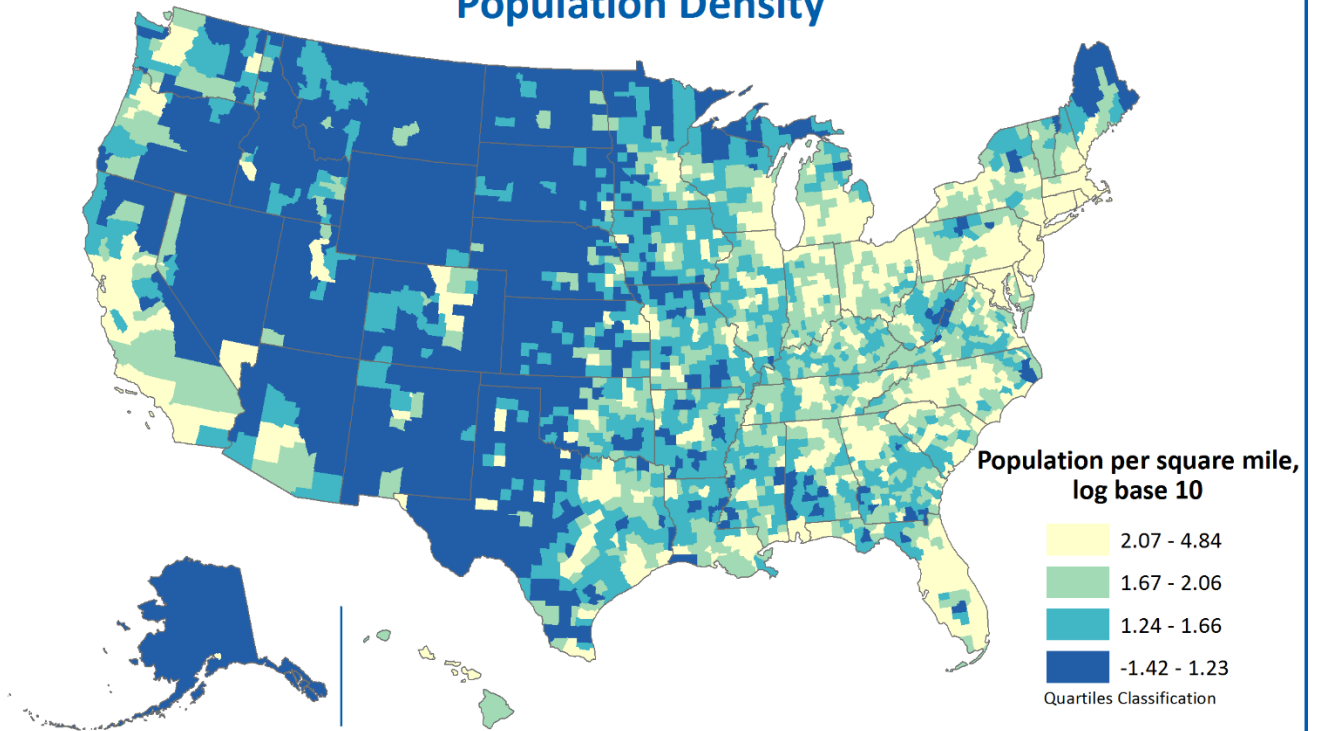
Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).

## Income



Source: American Community Survey, 2012 and 2013 5-year estimates

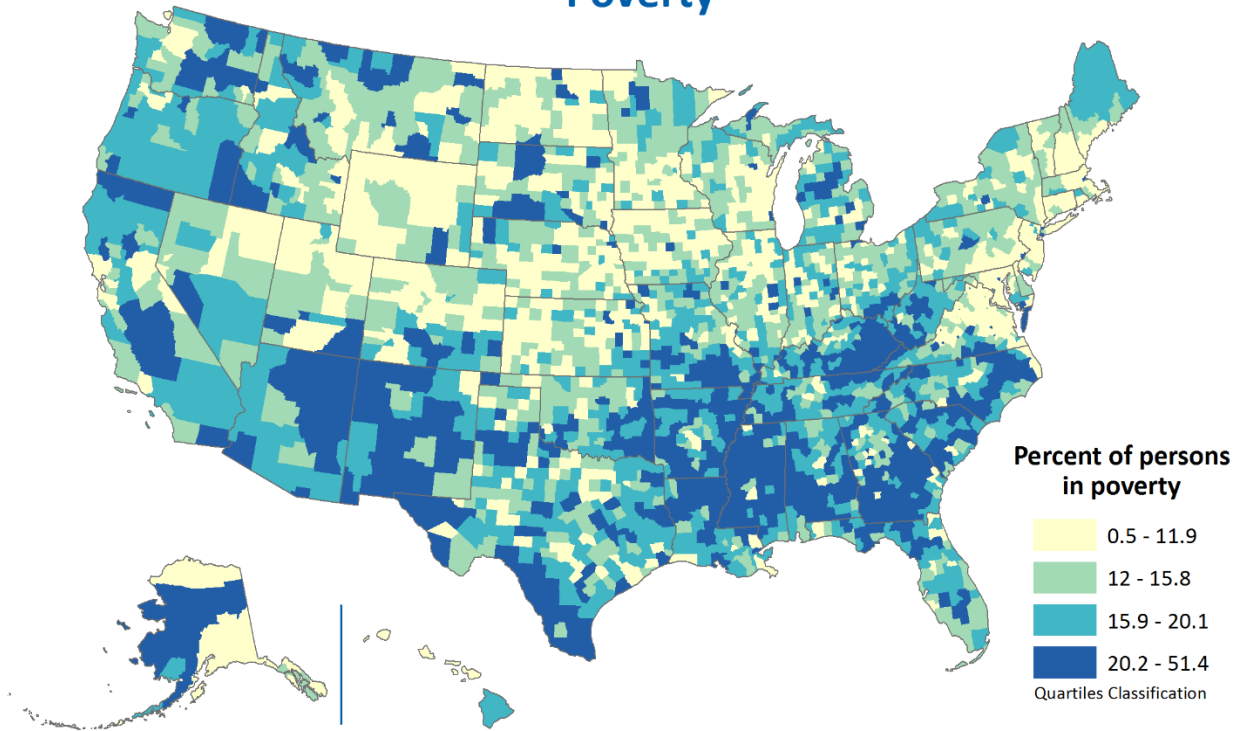
## Population Density



Source: U.S. Census

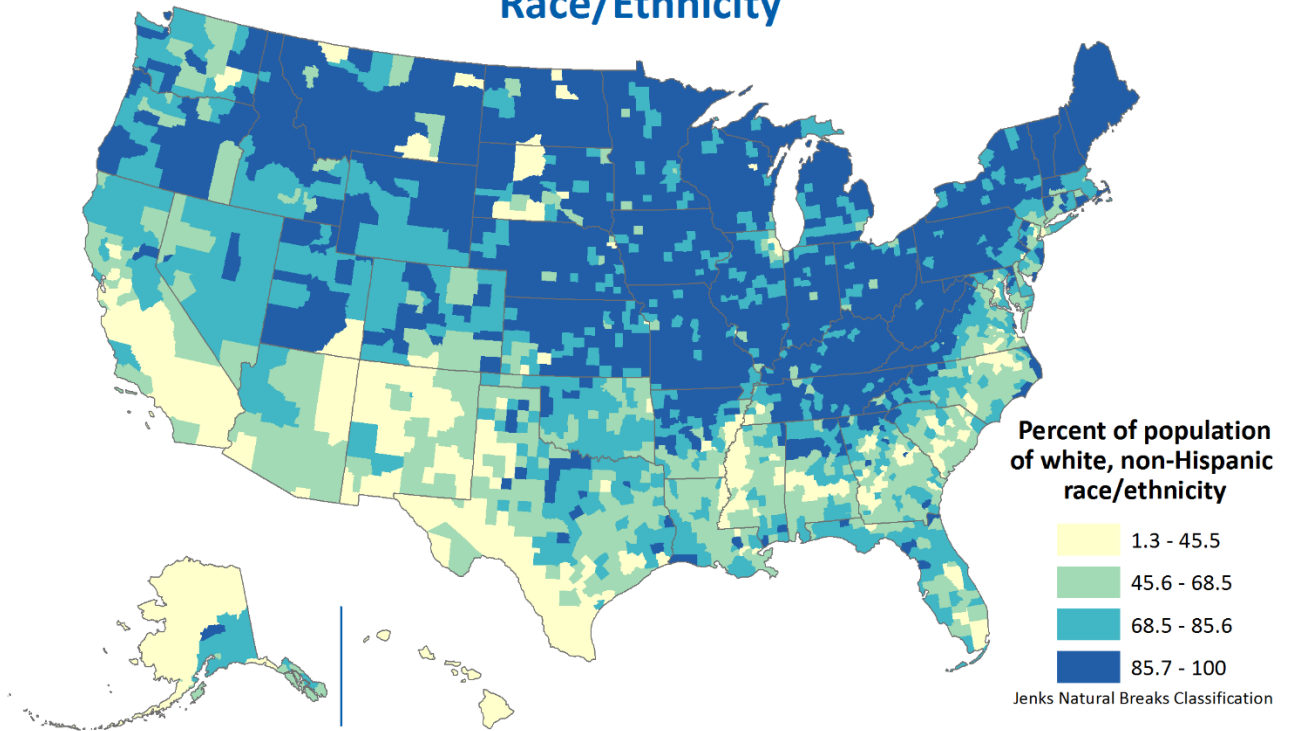
Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).

## Poverty



Source: American Community Survey, 2012 and 2013 5-year estimates

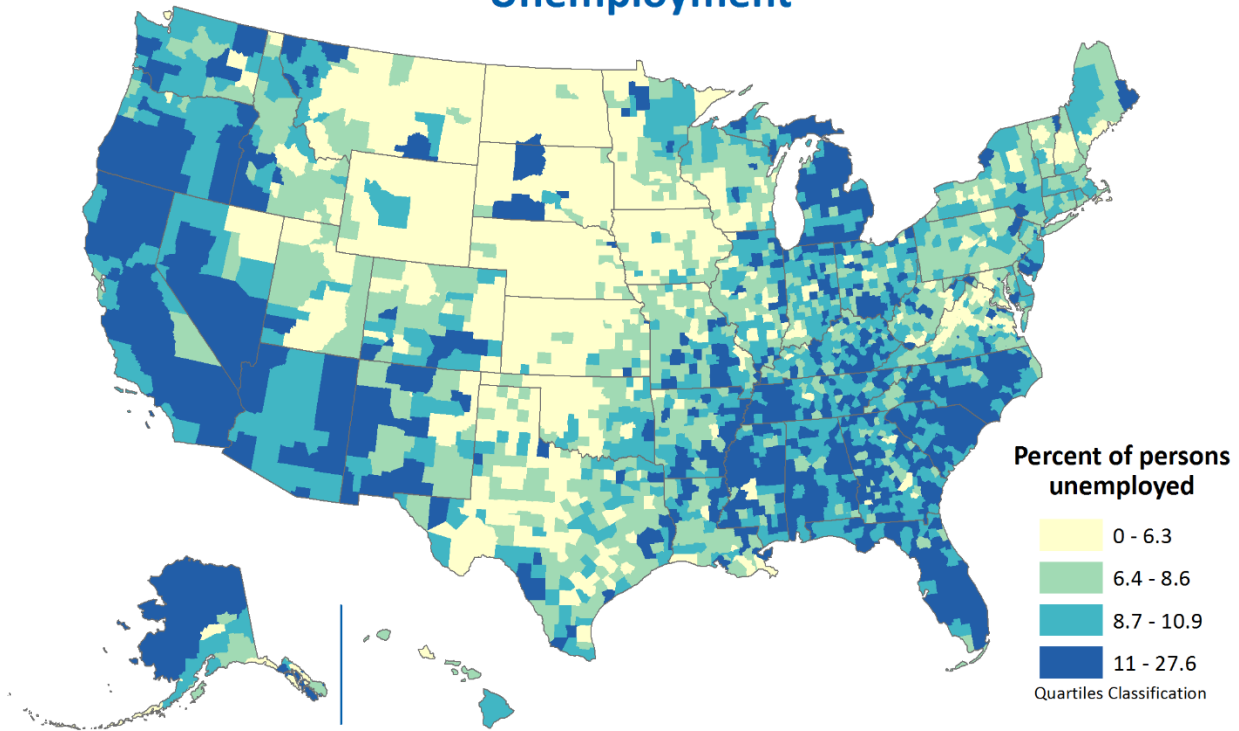
## Race/Ethnicity



Source: American Community Survey, 2012 and 2013 5-year estimates

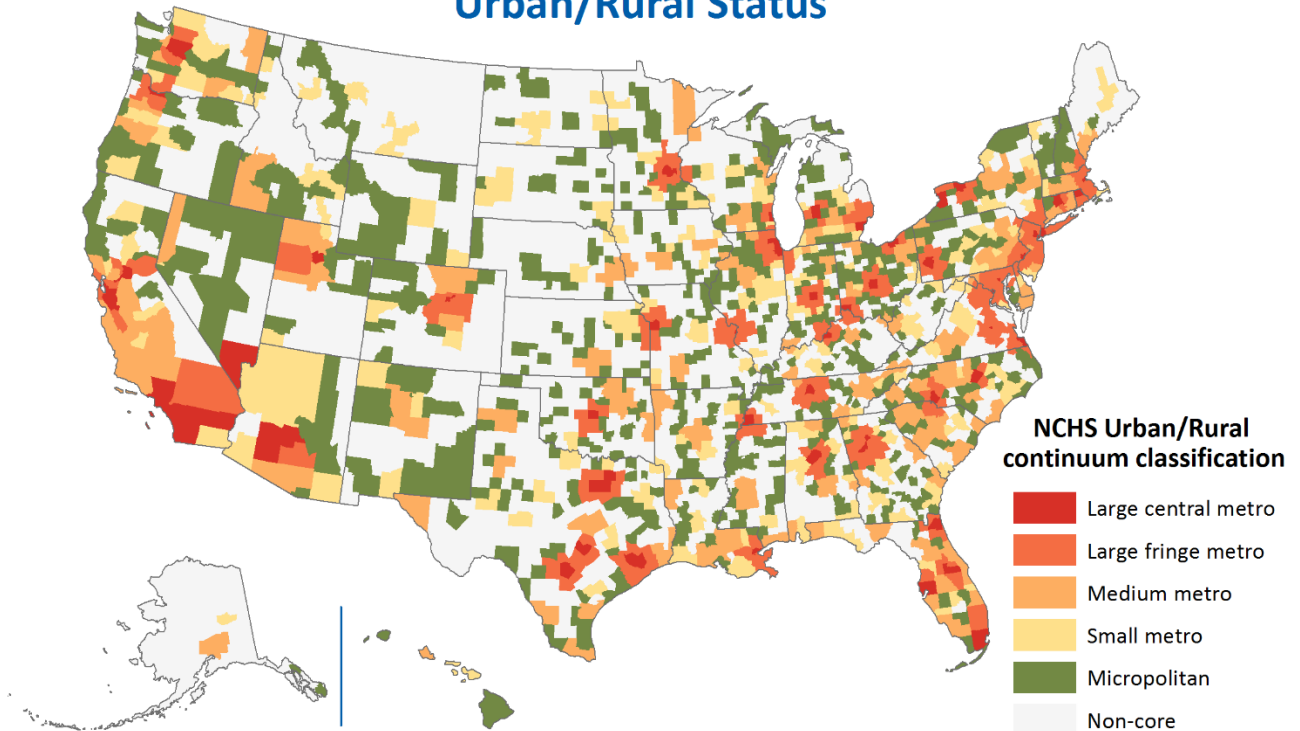
Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).

## Unemployment



Source: American Community Survey, 2012 and 2013 5-year estimates

## Urban/Rural Status

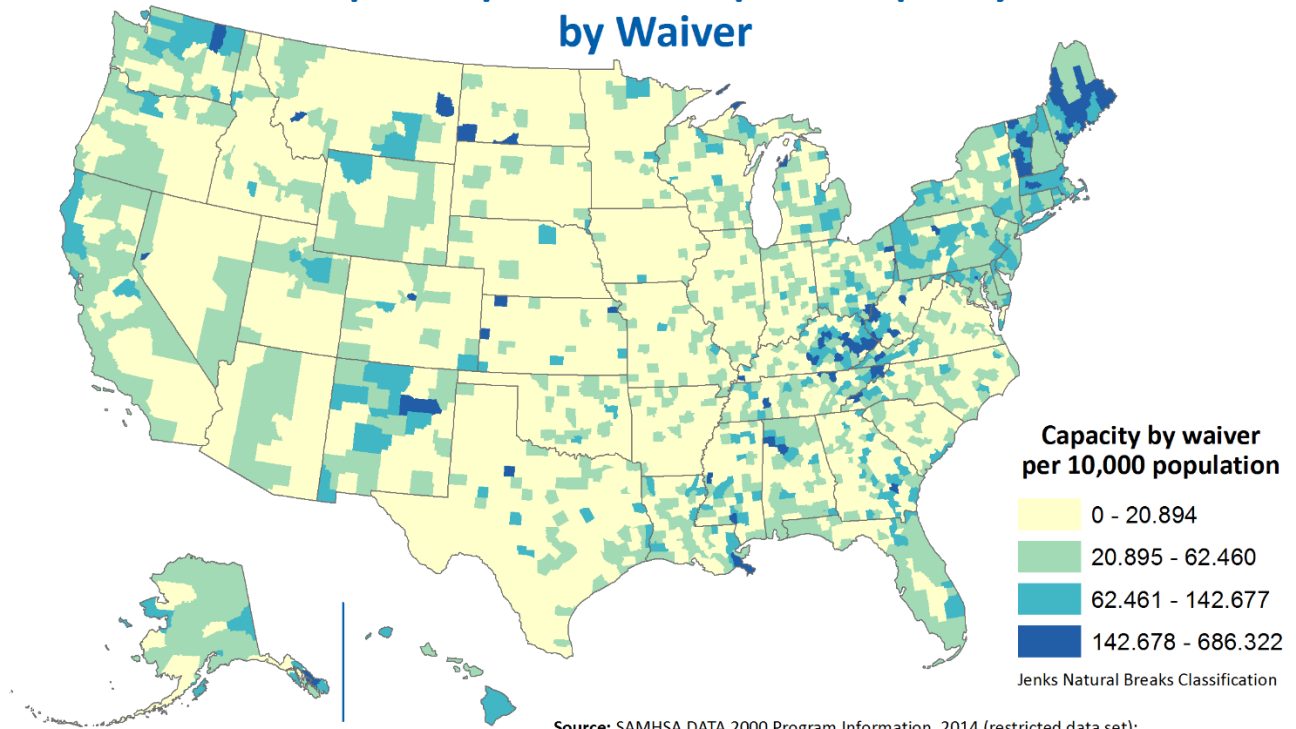


Source: NCHS, 2013

Notes: Map and Analysis provided by the Geospatial Research, Analysis, and Services Program (GRASP), Div of Toxicology and Human Health Sciences, ATSDR (2015).



## Buprenorphine Prescription Capacity by Waiver



Source: SAMHSA DATA 2000 Program Information, 2014 (restricted data set);  
rates calculated using ACS 2012 and 2013 5-year population estimates

Figure S2. Acute HCV infection rate by county. Reported rate of acute HCV infection by county, NNDSS 2012-2013 and model-estimated rate of acute HCV infection by county

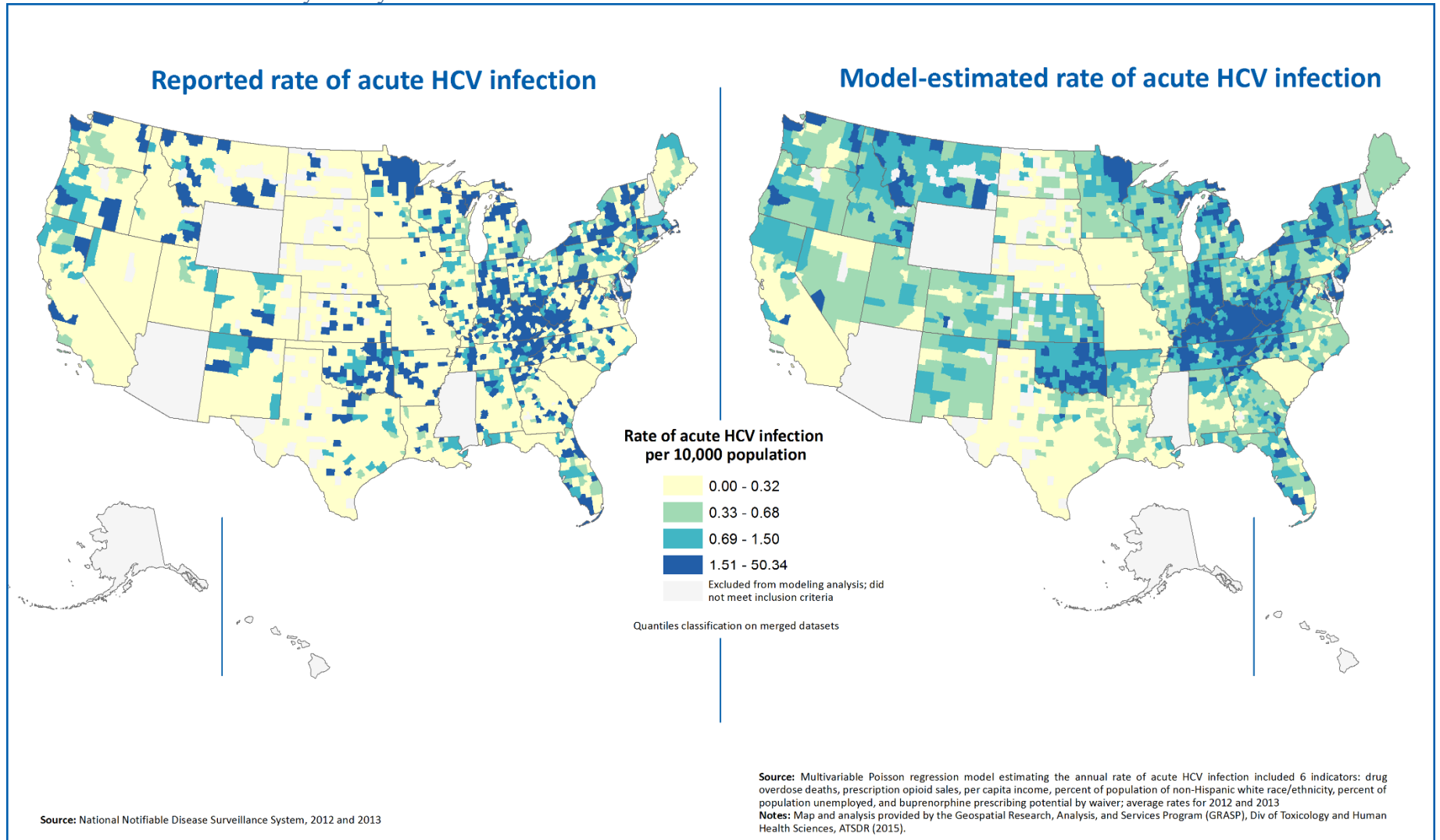


Figure S3. A. Histogram showing vulnerability scores by number of counties in each score group (eg, 0 to 0.25), B. Sigmoid curve showing vulnerability scores by county rank and identifying the scores bordering the top 5% cut-off, and C. Caterpillar curve of vulnerability scores by mean rank 90% confidence intervals bordering the top 5% rank cut-off (red dash line).

