
Eli Rosenberg, Emory University
Eric W. Hall, Emory University
Patrick Sullivan, Emory University
Travis Sanchez, Emory University
Kimberly Workowski, Emory University
John Ward, Emory University
Deborah Holtzman, Centers for Disease Control and Prevention

Journal Title: Clinical Infectious Diseases
Volume: Volume 64, Number 11
Publisher: Oxford University Press (OUP) | 2017-06-01, Pages 1573-1581
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1093/cid/cix202
Permanent URL: https://pid.emory.edu/ark:/25593/s2wn7

Final published version: http://dx.doi.org/10.1093/cid/cix202

Copyright information:
Copyright © The Author 2017. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed March 11, 2019 1:26 AM EDT

Eli S. Rosenberg,1 Eric W. Hall,1 Patrick S. Sullivan,1 Travis H. Sanchez,1 Kimberly A. Workowski,2 John W. Ward,3 and Deborah Holtzman3

1Department of Epidemiology, Emory University Rollins School of Public Health; 2Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine; and 3Division of Viral Hepatitis, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia

**Background.** Hepatitis C virus (HCV) infection is the most common chronic blood-borne infection in the United States and a leading cause of morbidity and mortality. Previous analyses of the US National Health and Nutrition Examination Survey (NHANES) indicated approximately 3.6 million noninstitutionalized persons with antibody to HCV (anti-HCV). However, state-level prevalence remains less understood and cannot be estimated reliably from NHANES alone.

**Methods.** We used 3 publicly available government data sources to estimate anti-HCV prevalence in each US state among noninstitutionalized persons aged ≥18 years. A small-area estimation model combined indirect standardization of NHANES-based prevalence with logistic regression modeling of mortality data, listing acute or chronic HCV infection as a cause of death, from the National Vital Statistics System during 1999–2012. Model results were combined with US Census population sizes to estimate total number and prevalence of persons with antibody to HCV in 2010.

**Results.** National anti-HCV prevalence was 1.67% (95% confidence interval [CI], 1.53–1.90), or 3 911 800 (95% CI, 3 589 400–4 447 500) adults in 2010. State-specific prevalence ranged from 0.71% (Illinois) to 3.34% (Oklahoma). The West census region had the highest region-specific prevalence (2.14% [95% CI, 1.96–2.48]); 10 of 13 states had rates above the national average. The South had the highest number of persons with anti-HCV (n = 1 561 600 [95% CI, 1 427 700–1 768 900]). The Midwest had the lowest region-specific prevalence (1.14% [95% CI, 1.04–1.30%]).

**Conclusions.** States in the US West and South have been most impacted by hepatitis C. Estimates of HCV infection burden are essential to guide policy and programs to optimally prevent, detect, and cure infection.

**Keywords.** hepatitis C; surveillance; prevalence; National Health and Nutrition Examination Survey; National Vital Statistics System.

Hepatitis C virus (HCV) infection is the most common blood-borne infection in the United States [1]; untreated infection is a leading cause of cirrhosis and hepatocellular carcinoma, is the most frequent cause of liver failure requiring transplantation, and causes more deaths annually than human immunodeficiency virus (HIV) [2–5]. About 70% of persons with untreated HCV infection remain infected for life. HCV infection may be asymptomatic and approximately half of persons are unaware of their infection [6]. Within 30 years, 41% of infected persons progress to cirrhosis, leading to liver failure, hepatocellular carcinoma, and death from liver-related causes [2].

The annual number of new HCV infections in the United States was highest before HCV was discovered to be the cause of “non-A non-B hepatitis” in 1989; incidence declined after prevention guidelines and blood donor screening were implemented in the 1990s. After years of level incidence, new HCV infections began to rise nationally in 2010, increasing >2-fold by 2014. Incidence of acute HCV infections has risen among males, persons aged 20–29 years, and American Indians/Alaska Natives [7]. The prevalence of HCV infection can be measured by a positive test for HCV antibody. Chronic HCV infection is associated with substantial morbidity and mortality, is highest among males and those born during 1945–1965, and has been associated with rising HCV-related mortality in recent years [1, 4, 7–9].

Current US surveillance programs provide incomplete estimates of HCV infection prevalence. In 2014, the Centers For Disease Control and Prevention (CDC) National Notifiable Diseases Surveillance System (NNDSS) included passively collected laboratory reports from 40 states of acute HCV infection, and from 34 states reporting chronic HCV infection. US national and state-specific estimates of HCV infection prevalence cannot be ascertained from NNDSS data [7, 10].
National estimates of HCV prevalence have been produced using the US National Health and Nutrition Examination Survey (NHANES), which includes interviews and physical examinations of noninstitutionalized persons aged ≥6 years [11]. NHANES data for 2003–2010 estimated that 3.6 million persons had antibodies to HCV (anti-HCV), indicative of past or current HCV infection, corresponding to a national prevalence of 1.3% [8]. Subsequent systematic reviews estimate an additional 1 million HCV infections from populations excluded and underrepresented in NHANES, primarily homeless or incarcerated persons [12, 13].

State-level estimates of the prevalence of HCV infection are essential for guiding intervention programs, research, and federal assistance funding priorities among US states. In 2016, CDC funded 7 jurisdictions to conduct active case surveillance for HCV infection, and case counts of HCV diagnoses have been produced [7]. However, these states have been conducting active surveillance for few years, and most states lack local resources to provide similar information regarding the number of persons diagnosed with HCV; surveillance data are limited by variability in testing of persons at risk for infection and reporting to public health authorities.

Some jurisdictions have created their own estimates of prevalence using various methods, most commonly by applying NHANES-derived HCV infection prevalence rates to state populations [14]. The assumption that all states have the national NHANES prevalence of HCV infection might produce inaccurate state-specific estimates, because risk of HCV infection likely varies by state. Some researchers have attempted to improve upon this simple approach by standardizing the national NHANES estimate to local demographic profiles or creating models using case surveillance data (Supplementary Table 1). Despite these informative efforts, there is no complete set of state-specific estimates of HCV infection prevalence for all US states that is based on accurate and consistent methods.

Small-area estimation methods are statistical approaches that can be used for determining state-level HCV infection prevalence by allocating national prevalence estimates into state-specific components using data sources indicating state-level markers of HCV infection prevalence [15]. State-level data sources include electronic medical records (EMRs), insurance claims, and laboratory and mortality data. Individual commercially available EMR, claims, and laboratory datasets are inconsistent in their geographic and demographic representation [16]. Alternatively, mortality is systemically recorded for all decedents in the United States and is publicly available through the population-based National Vital Statistics System (NVSS) [4]. NVSS mortality data provide detailed information regarding decedent age, sex, and race/ethnicity, which are covariates of state-specific HCV infection prevalence.

Here, we describe an application of small-area methodology to deconstruct national NHANES-estimated HCV infection prevalence into state-specific components using HCV-related mortality as a state-level data source. Sensitivity

Table 1. Data Sources and Purposes for Primary and Sensitivity Analyses

<table>
<thead>
<tr>
<th>Data source</th>
<th>Years Represented</th>
<th>Purpose</th>
<th>No. of Individuals Represented</th>
<th>No. of Cases</th>
<th>Data Extraction Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Vital Statistics System</td>
<td>1999–2012</td>
<td>Hepatitis C-related mortality, for primary analysis</td>
<td>33,540,118 decedents aged ≥18 y who resided in the 50 states or Washington, DC</td>
<td>185,285 with HCV as underlying or multiple cause of death</td>
<td>ICD-10 codes included acute viral hepatitis C (B17.1), chronic viral hepatitis C (B18.2)</td>
</tr>
<tr>
<td>National Vital Statistics System</td>
<td>1999–2012</td>
<td>Cirrhosis-related mortality, for bias analysis</td>
<td>33,540,118 decedents aged ≥18 y who resided in the 50 states or Washington, DC</td>
<td>427,404 with cirrhosis as underlying or multiple cause of death</td>
<td>ICD-10 codes included hepatic fibrosis (K74.0); hepatic sclerosis (K74.1); hepatic fibrosis with hepatic sclerosis (K74.12); or other and unspecified cirrhosis of liver (K74.8)</td>
</tr>
<tr>
<td>National Vital Statistics System</td>
<td>1999–2012</td>
<td>Hepatocellular carcinoma-related mortality, for bias analysis</td>
<td>33,540,118 decedents aged ≥18 y who resided in the 50 states or Washington, DC</td>
<td>197,976 with hepatocellular carcinoma as underlying or multiple cause of death</td>
<td>ICD-10 codes included liver cell carcinoma (C22.0); primary malignant neoplasm of liver, type unspecified (C22.8); or malignant neoplasm of liver, not specified as primary or secondary (C22.9)</td>
</tr>
</tbody>
</table>

Abbreviations: HCV, hepatitis C virus; ICD-10, International Classification of Diseases, Tenth Revision; NHANES, National Health and Nutrition Examination Survey.
analyses consider the uncertainty of the estimates associated with state-specific variation in diagnosing and treating HCV infection and in recording HCV-related mortality. We present state-level estimates of anti-HCV prevalence and compare these to previously published state estimates.

METHODS

We employed a synthetic small-area estimation approach that combined indirect standardization of NHANES data with regression-based estimates of state-level HCV-related mortality. The data sources and approach are described below and in Table 1 and Figure 1.

Data Sources

**NHANES 1999–2012**

Annually, NHANES uses a complex, multistage sampling design to select approximately 5,000 persons in 15 counties, representing the noninstitutionalized population of persons aged ≥6 years residing in the 50 states and Washington, District of Columbia [11].

---

**Equation 1.** Estimated total persons with anti-HCV, in state $i$

$$
\hat{T}_i = \left(\hat{\beta}_i \times N_{ij}\right) \times \left(\frac{\hat{b}_{ij}}{\sum_{j=1}^{k} \left(\hat{\beta}_j \times N_{ij}\right)}\right)
$$

**Equation 2.** Estimated prevalence rate of persons with anti-HCV, in state $i$

$$
\hat{\lambda}_i = \frac{\hat{T}_i}{N_i}
$$

**Equation 3.** Model for estimating $\hat{b}_{ij}$, HCV-related mortality, in stratum $j$ of state $i$

$$
\logit(P[H]) = a_1 + \sum_{j=1}^{k} \beta_{1j}X_{1j} + \sum_{j=1}^{k} \beta_{2j}X_{2j} + \sum_{j=1}^{k} \beta_{3j}X_{3j} + \sum_{j=1}^{k} \beta_{4j}X_{4j} + \sum_{j=1}^{k} \beta_{5j}X_{5j} + \sum_{j=1}^{k} \beta_{6j}X_{6j} + \sum_{j=1}^{k} \beta_{7j}X_{7j} + \sum_{j=1}^{k} \beta_{8j}X_{8j} + \sum_{j=1}^{k} \beta_{9j}X_{9j} + \sum_{j=1}^{k} \beta_{10j}X_{10j}
$$

**Equation 4.** Adjusted HCV-related death counts in stratum $j$ of state $i$ used in bias analyses

$$
\text{Adjusted } D_{h_{ij}} = D_{h_{ij}} + (PAF_{h_{ij}} \times D_{h_{ij}}) - D_{sh_{ij}}
$$

Where:

- $i = $ states 1 to 51 ($i$ = 51)
- $j = $ stratum 1 to 24 ($j$ = 24), formed by combination of sex $s$, birth cohort $b$, race $r$
- $\hat{T}_i = $ Estimated total persons with anti-HCV, in state $i$
- $\hat{\beta}_i = $ Estimated weighted HCV prevalence rate, in stratum $j$
- $\hat{b}_{ij} = $ Estimated probability of HCV-related mortality, in stratum $j$ of state $i$
- $N_{ij} = $ Adult population in stratum $j$ of state $i$
- $N_{ij}^* = $ Adult population-years in stratum $j$ of state $i$
- $\hat{\lambda}_i = $ Estimated prevalence rate of persons with anti-HCV, in state $i$
- $H = $ HCV-related death among NVSS decedents
- $x_{are} = $ indicator variables for state $i$, sex level $s$, race level $r$, birth cohort level $b$
- $x = 1$ for cirrhosis, 2 for hepatocellular carcinoma
- $PAF = $ Population attributable fraction for HCV infection among individuals with disease $x$
- $D_{h_{ij}} = $ Deaths with acute/chronic viral hepatitis C listed as underlying or multiple cause of death, in stratum $j$ of state $i$
- $D_{s_{ij}} = $ Deaths with disease $x$ listed as underlying or multiple cause of death, in stratum $j$ of state $i$
- $D_{h_{ij}} = $ Deaths with both disease $x$ and acute/chronic viral hepatitis C listed as underlying or multiple cause of death, in stratum $j$ of state $i$

Figure 1. Modeling equations. Abbreviations: anti-HCV, hepatitis C virus antibody; NVSS, National Vital Statistics System.
District of Columbia and were aged ≥18 years at death. Records also contained International Classification of Diseases, Tenth Revision (ICD-10) codes for multiple underlying causes of death (N = 33 540 118). Demographic covariates included sex, race/ethnicity, and birth cohort. In the primary analysis, any records including the ICD-10 code for acute viral hepatitis C (B17.1) or chronic viral hepatitis C (B18.2) as an underlying cause of death were considered to signal HCV-related mortality (n = 185 285).

US Census Inter-censal Data (1999–2012) and 2010 US Census Data
Intercensal population estimates from the US Vintage 2000, 2009, and 2014 datasets provided denominators for HCV-related mortality rates during 1999–2012 [21]. Data were grouped into the same sex, race/ethnicity, and birth categories as the NHANES and NVSS data. State-by-demographic group population totals from the 2010 US Census were used to calculate 2010 state HCV prevalence [21].

Analysis
The number of persons in each state with anti-HCV was computed using the standardization-based estimator in Figure 1, equation 1. First, we calculated direct weighted estimates of national HCV-antibody prevalence \( \hat{\lambda} \) for 24 strata (sex × race/ethnicity × birth cohort), using standard methodology (Figure 1, equation 1) [15]. We multiplied weighted estimates by state-by-demographic stratum 2010 population counts to generate crude state-level estimates. These were adjusted by the ratio of state-by-demographic stratum effects, based on the average HCV-related death rate \( \hat{\theta}_j \) in the 14-year period. We fit a high-order logistic regression model that approximated full stratification (several of 1224 strata had zero cells), permitting detection of heterogeneity among strata (Figure 1, equation 3). We assessed collinearity and model fit by comparing observed state-level HCV-related mortality totals to model predictions [15, 22]. Mortality-adjusted HCV infection prevalence totals were summarized to yield estimated state-level totals \( T_i \) (rounded to nearest hundred persons), with prevalence rates \( \hat{\lambda}_i \) (Figure 1, equations 1 and 2). Supplementary analyses estimated state-level chronic HCV infection, defined as a positive or indeterminate anti-HCV test and a positive HCV RNA test, using the above approach but with a 12-stratum model (race/ethnicity considered white non-Hispanic or not), due to more sparse NHANES data.

Sensitivity Analyses
To account for the joint statistical uncertainty in the stratified NHANES estimates and model-based HCV-related mortality estimates, we conducted a Monte Carlo simulation that respectively sampled from logit-normal and normal distributions (\( k = 10000 \) runs), using the standard errors for the original estimates, to produce 95% confidence intervals (CIs) for state-level estimates.

There might be state-level variability in HCV diagnosis and treatment that produces state-level variability in HCV-related deaths or proper attribution (specific ICD-10 codes) of deaths to an HCV cause, although likely limited [5]. We repeated all analyses with a broader definition of HCV-related deaths that used a combination of the HCV-specific ICD-10 codes and less-specific, more sensitive, ICD-10 codes representing cirrhosis-related and hepatocellular carcinoma–related (HCC) causes of death (Table 1) [5]. To increase specificity of these additional codes, we applied available estimates of the population attributable fraction (PAF) due to HCV infection (cirrhosis: 42%, HCC: 48%; Figure 1, equation 4) [23]. We secondarily considered other PAF estimates in less representative populations [24, 25].

We descriptively compared model findings to other publicly available reports of state estimates. Reports were excluded if HCV infection estimates solely involved applying national NHANES HCV infection prevalence estimates to total state population or if estimation was exclusively based on partial case surveillance data for HCV infection. Reports not describing the methodology used for prevalence estimation were included to facilitate more comparisons. Where possible, abstracted prevalence estimates were restricted to comparable noninstitutionalized populations.

RESULTS
The estimated national prevalence of anti-HCV in 2010 was 1.67% (95% CI, 1.53%–1.90%), corresponding to 3 911 800 (95% CI, 3 589 400–4 447 500) US adults with past or current HCV infection. Demographic stratum-specific estimates ranged from 0.26% for other race/ethnicity females born after 1965 to 8.02% for black males born 1945–1965 (Supplementary Table 2).

The prevalence of anti-HCV varied by state (Table 2 and Figure 2). State-specific prevalence rates ranged from 0.71% in Illinois to 3.34% in Oklahoma, with >2.5% additionally in District of Columbia, New Mexico, Oregon, and Tennessee. By census region and division, the West contained the most high-prevalence states: 10 of 13 states were above the national average, and the region-specific prevalence was 2.14%. The Pacific West division was most affected (2.30% prevalence; 865 400 persons who had been infected with HCV). The South had the second-highest anti-HCV prevalence (1.80%) and the largest number of persons with anti-HCV (n = 1 561 600). Within the South, the West South Central had the highest prevalence (2.19%), while the South Atlantic division had the most persons with anti-HCV (n = 712 900). The Midwest (1.14%) and Northeast (1.43%) had relatively lower anti-HCV prevalence. Of the 21 states in these regions, only Rhode Island had prevalence above the national average (2.12%). Although having prevalence rates lower than the national average, Michigan, New York, Pennsylvania, and Ohio each had >100 000 persons with anti-HCV.
<table>
<thead>
<tr>
<th>Region/Division/State</th>
<th>2010 Census Population</th>
<th>Total Persons With Anti-HCV</th>
<th>Anti-HCV Prevalence Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. (95% CI)</td>
<td>No. (95% CI)</td>
<td>Rate per 100 (95% CI)</td>
</tr>
<tr>
<td>NORTHEAST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New England</td>
<td>42 984 048</td>
<td>614 300 (565 700–693 900)</td>
<td>1.43 (1.22–1.62)</td>
</tr>
<tr>
<td>Connecticut</td>
<td>11 293 971</td>
<td>157 800 (143 400–182 100)</td>
<td>1.40 (1.27–1.61)</td>
</tr>
<tr>
<td>Maine</td>
<td>2 757 082</td>
<td>36 800 (33 500–42 400)</td>
<td>1.33 (1.21–1.54)</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>105 382 8</td>
<td>11 200 (960 0–13 900)</td>
<td>1.06 (0.91–1.32)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>5 128 706</td>
<td>74 100 (66 900–85 600)</td>
<td>1.44 (1.30–1.67)</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>1 0 29 236</td>
<td>11 000 (93 000–13 700)</td>
<td>1.07 (0.91–1.33)</td>
</tr>
<tr>
<td>Vermont</td>
<td>82 861 1</td>
<td>17 600 (15 600–21 000)</td>
<td>2.12 (1.89–2.53)</td>
</tr>
<tr>
<td>Middle Atlantic</td>
<td>496 508</td>
<td>72 000 (61 000–94 000)</td>
<td>1.45 (1.23–1.89)</td>
</tr>
<tr>
<td>New Jersey</td>
<td>31 690 077</td>
<td>456 600 (421 300–513 200)</td>
<td>1.44 (1.33–1.62)</td>
</tr>
<tr>
<td>New York</td>
<td>6 726 680</td>
<td>90 700 (83 000–103 000)</td>
<td>1.35 (1.23–1.53)</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>15 053 173</td>
<td>223 700 (207 000–252 700)</td>
<td>1.49 (1.38–1.68)</td>
</tr>
<tr>
<td>MINNESOTA</td>
<td>9 910 224</td>
<td>142 100 (129 200–169 900)</td>
<td>1.43 (1.30–1.62)</td>
</tr>
<tr>
<td>East North Central</td>
<td>50 798 893</td>
<td>578 600 (527 700–659 200)</td>
<td>1.14 (1.04–1.30)</td>
</tr>
<tr>
<td>Iowa</td>
<td>35 269 776</td>
<td>378 800 (345 200–429 200)</td>
<td>1.07 (0.98–1.22)</td>
</tr>
<tr>
<td>Kansas</td>
<td>2 126 179</td>
<td>29 900 (26 600–35 200)</td>
<td>1.41 (1.25–1.66)</td>
</tr>
<tr>
<td>Nebraska</td>
<td>4 0 19 862</td>
<td>41 500 (37 600–48 500)</td>
<td>1.03 (0.94–1.21)</td>
</tr>
<tr>
<td>North Dakota</td>
<td>52 720</td>
<td>44 000 (37 000–52 000)</td>
<td>0.83 (0.71–1.19)</td>
</tr>
<tr>
<td>South Dakota</td>
<td>611 383</td>
<td>6300 (5400–7800)</td>
<td>1.03 (0.88–1.42)</td>
</tr>
<tr>
<td>SOUTH</td>
<td>86 766 987</td>
<td>1 561 600 (1 427 700–1 768 900)</td>
<td>1.80 (1.65–2.04)</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>46 020 846</td>
<td>712 900 (652 100–805 700)</td>
<td>1.55 (1.42–1.76)</td>
</tr>
<tr>
<td>Delaware</td>
<td>692 169</td>
<td>13 600 (12 200–16 100)</td>
<td>1.97 (1.76–2.33)</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>500 908</td>
<td>16 400 (14 400–19 500)</td>
<td>3.27 (2.87–3.90)</td>
</tr>
<tr>
<td>Florida</td>
<td>14 799 219</td>
<td>245 600 (221 700–287 700)</td>
<td>1.66 (1.50–1.90)</td>
</tr>
<tr>
<td>Georgia</td>
<td>7 196 101</td>
<td>84 500 (76 800–96 100)</td>
<td>1.17 (1.07–1.34)</td>
</tr>
<tr>
<td>Maryland</td>
<td>4 4 2 0 5 8 8</td>
<td>82 000 (74 900–93 100)</td>
<td>1.86 (1.69–2.11)</td>
</tr>
<tr>
<td>North Carolina</td>
<td>7 253 348</td>
<td>117 300 (106 600–133 100)</td>
<td>1.62 (1.47–1.83)</td>
</tr>
<tr>
<td>South Carolina</td>
<td>3 544 890</td>
<td>62 300 (56 200–71 400)</td>
<td>1.76 (1.58–2.01)</td>
</tr>
<tr>
<td>Virginia</td>
<td>6 147 347</td>
<td>66 700 (60 800–76 000)</td>
<td>1.09 (0.99–1.24)</td>
</tr>
<tr>
<td>East South Central</td>
<td>14 025 119</td>
<td>264 300 (236 500–303 300)</td>
<td>1.88 (1.69–2.17)</td>
</tr>
<tr>
<td>Arkansas</td>
<td>3 647 277</td>
<td>52 400 (47 000–60 400)</td>
<td>1.44 (1.29–1.66)</td>
</tr>
<tr>
<td>Kentucky</td>
<td>3 315 996</td>
<td>54 200 (47 200–63 800)</td>
<td>1.63 (1.42–1.92)</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>2 821 685</td>
<td>94 200 (83 800–112 900)</td>
<td>3.34 (2.97–4.00)</td>
</tr>
<tr>
<td>Texas</td>
<td>18 279 737</td>
<td>37 600 (34 500–42 800)</td>
<td>2.06 (1.89–2.34)</td>
</tr>
<tr>
<td>WEST</td>
<td>54 014 143</td>
<td>1 157 400 (1 060 100–1 341 100)</td>
<td>2.14 (1.96–2.48)</td>
</tr>
<tr>
<td>Mountain</td>
<td>16 368 084</td>
<td>291 900 (266 100–339 700)</td>
<td>1.78 (1.63–2.07)</td>
</tr>
<tr>
<td>Arizona</td>
<td>4 763 003</td>
<td>90 000 (81 400–104 600)</td>
<td>1.89 (1.71–2.20)</td>
</tr>
<tr>
<td>Colorado</td>
<td>3 803 587</td>
<td>66 100 (60 000–76 500)</td>
<td>1.74 (1.58–2.01)</td>
</tr>
<tr>
<td>Idaho</td>
<td>1 138 510</td>
<td>16 400 (14 200–20 200)</td>
<td>1.44 (1.25–1.77)</td>
</tr>
<tr>
<td>New Mexico</td>
<td>1 540 507</td>
<td>42 600 (37 900–51 400)</td>
<td>2.76 (2.46–3.34)</td>
</tr>
</tbody>
</table>
In sensitivity analyses that considered the addition of 2 potential indicators of HCV-related deaths, the above patterns were preserved, generally yielding state-level estimates closer to the national average (Figure 3 and Supplementary Tables 3–4). Inclusion of cirrhosis-related mortality resulted in 39 of 51 state estimates within the 95% CI bounds for the primary estimates that used only HCV-specific mortality. The median absolute deviation of state estimates between these 2 estimation approaches was 8% (interquartile range [IQR], 5%–16%, multiplicative scale). Inclusion of HCC-related mortality resulted in 50 of 51 state estimates that were within the 95% CI bounds for the primary estimates using HCV-specific mortality. The median absolute deviation of state estimates between these 2 estimation approaches was 3% (IQR, 2%–6%). Results were similar when considering alternative, but less representative PAF estimates (Supplementary Tables 5–6). There was high agreement between the primary anti-HCV prevalence estimates and 11 external estimates of state-level HCV infection prevalence (Supplementary Table 1). The 2 external estimates that utilized state-level HCV testing data (Arkansas, Oregon) were closest to our primary estimates (5.5% and 4.8% higher, respectively). To facilitate comparisons, Supplementary Tables 7–8 display model-based chronic HCV infection and NHANES standardization–based anti-HCV estimates.

DISCUSSION
We used a small-area estimation approach to synthesize data from 3 publicly available systematic, population-based data systems and provide the first comprehensive state estimates of the prevalence of anti-HCV in the United States. The estimates were robust to multiple approaches based on additional mortality data, and were comparable to independent state estimates. The method and its results can be applied by state and local health officials to guide program planning, set priorities for resource allocation, and evaluate interventions.

These state-specific estimates bring a new level of understanding to our prior knowledge about the epidemiology of HCV infection in the United States. States with high estimated anti-HCV prevalence also contain highly urban populations (eg, Rhode Island, District of Columbia) and high proportions of the population who inject drugs (eg, Vermont, Tennessee, West Virginia) or who are Native Americans (eg, Oklahoma, New Mexico; the model did not explicitly include Native American race) [26, 27]. The current anti-HCV prevalence estimates highlight states which, by their demographic structure alone, might be predicted to have lower prevalence. For example, simple NHANES standardization yields 44,000 estimated cases in Oklahoma; incorporating additional data on HCV-related deaths in Oklahoma suggests an estimated prevalence more than twice that high.

Our approach has significant methodologic strengths, as the data sources are population-based and representative of the underlying populations. However, our analysis has important limitations. We aggregated data across multiple years of NHANES to achieve a sufficient sample for stable estimates. It is possible that there were secular trends over the NHANES period. The NHANES population samples noninstitutionalized adults and therefore excludes incarcerated persons, homeless persons, those in active military service, and persons on tribal lands. These populations are critical constituents of the US HCV epidemic, comprising up to a fifth of prevalent antibody-positive persons in the United States [12]. Thus, our estimates apply only to noninstitutionalized US populations. Mortality data used to allocate the NHANES–based anti-HCV total included deaths from institutionalized populations. Complete laboratory data in NHANES were nearly universally available for the anti-HCV test, but were less complete for RNA testing.
Thus, our primary estimates of persons with anti-HCV includes the 15%–25% of people infected by HCV who may have cleared their HCV infections but retained detectable antibodies [28]. Antibody-based prevalence of individuals previously or currently infected with HCV provides a conservative overestimate of persons indicated for clinical evaluation and HCV-related medical services [8]. The HCV RNA test more accurately indicates active infection needing treatment. Importantly, the proportion of anti-HCV positive persons with detectable HCV by polymerase chain reaction is also influenced by the number of persons successfully treated and cured of their HCV infection. Over time, this proportion should decline as more persons are identified, treated, and cured; however, the proportion will vary among populations with differential access to HCV-related care and treatment.

Finally, NHANES is likely less representative of populations of persons infected in recent years. Incidence of HCV infection has been increasing since 2010, and those infected in more recent epidemics are younger, more likely to live in rural areas, and likely to acquire HCV through needle-sharing behaviors associated with opioid use [29]. Nationally, the number of recent new infections is small relative to the prevalent population, although some states with modest or low prevalence of infection have had larger increases than reflected in national
Methods that incorporate recent trends in incident HCV infection and in treatment can improve future iterations of this model.

There is some mismatch between the NHANES and mortality data: mortality data represent older trends in the epidemic. Also, HCV infection may not be consistently diagnosed or recorded on death certificates by state. However, this low sensitivity of identified HCV infection in death records should not cause bias in our estimates unless the likelihood of diagnosis or recording on death certificates varies across states. We posit that deficiencies in diagnosis or recording likely occur, but are at a facility level and do not vary systematically by state. The impact of varying sensitivities of our case definition in mortality data were explored in extensive sensitivity analyses, and did not result in substantial changes to our estimates.

CONCLUSIONS

Although national recommendations for HCV prevention, testing, and clinical management are developed by CDC and other authorities [30], decisions regarding the capacity to deliver these services are made at the state level. State-level estimates can inform these decisions in multiple ways. First, these data can prompt reconsiderations of HCV disease burden. These modeled prevalence estimates are based on publicly available data, allowing local authorities to assess biases in data sources used in the model (ie, mortality data) and the representativeness of the modeled data in comparisons with other local data sources (ie, HCV surveillance). Second, these results, supported by local public health authorities, provide new information to engage stakeholders, resulting in agreed-upon state/local estimates of HCV infection prevalence. State-level prevalence estimates allow state and local health officials to consider more investments in surveillance and collection of other strategic data for refining our estimates. These data can also be used to revise HCV-related prevention plans and guide prevention initiatives. For example, these data can help state Medicaid programs to budget funds for HCV testing and treatment. More broadly, having state-level estimates calculated consistently across states will allow states to assess their standing in relation to other states and to the nation as a whole, and to adapt their
prevention and control efforts to national or other state programs that have been shown to be effective.

The approach reported here can be routinely updated as new data become available. The same approach to estimation can also be extended to yield estimates for demographic strata, areas smaller than states, and excluded populations in each state. Modeling approaches can serve as a useful method to quantify the HCV epidemic in the absence of national case surveillance data. However, this should not minimize the imperative of continuing to enhance local surveillance efforts, especially in the context of curative HCV therapies. Surveillance data may provide more accurate and reliable estimates of the burden of HCV infection and serve as the basis for public health programs to diagnose persons living with hepatitis C and link them to appropriate clinical services and treatment.

**Supplementary Data**

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

**Notes**

**Acknowledgments.** We acknowledge Kevin Weiss for his assistance in compiling external state prevalence estimates for Supplementary Table 1 and Dr Jeremy Grey for his input on estimation methodology.

**Financial support.** We acknowledge funding from the Centers for Disease Control and Prevention (CDC) National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Epidemic and Economic Modeling Agreement (U38 PS004646).

**Potential conflicts of interest.** E. S. R. has received grants from CDC and personal fees from Medidata Inc and Cengage Learning. P. S. S. has received grant support through the CDC, National Institutes of Health, MAC AIDS Fund, and Gilead Sciences. T. H. S's institution has a project received grant support through the CDC, National Institutes of Health, and personal fees from Medidata Inc and Cengage Learning. P. S. S. has received personal fees from Medidata Inc.

**Supplementary Table 1**

Consisting of data provided by the authors to benefit the reader, the posted data are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

**References**