The Predictive Value of International Classification of Disease Codes for Chronic Hepatitis C Virus Infection Surveillance: The Utility and Limitations of Electronic Health Records

Winston E. Abara, MD, PhD[1], **Anne C. Moorman, MPH**[1], **Yuna Zhong, MD, MS**[1], **Melissa G. Collier, MD, MPH**[1], **Lora B. Rupp, MPH**[2], **Stuart C. Gordon, MD**[2], **Joseph A. Boscarino, PhD**[3], **Mark A. Schmidt, PhD**[4], **Connie Mah Trinacty, PhD**[5], **Scott D. Holmberg, MD**[1], **CHeCS Investigators**

[1]Division of Viral Hepatitis, CDC, Atlanta, Georgia.

[2]Henry Ford Hospital, Detroit, Michigan.

[3]Geisinger Health System, Danville, Pennsylvania.

[4]Kaiser Permanente Northwest, Portland, Oregon.

[5]Kaiser Permanente Hawaii, Honolulu, Hawaii.

Abstract

Surveillance of chronic hepatitis C virus (HCV) cases faces limitations that result in delays and under-reporting. With increasing use of electronic health records (EHRs), the authors evaluated the predictive value of using *International Classification of Diseases, Ninth Revision* (ICD-9) codes to identify chronic HCV cases from EHR data. Longitudinal EHR data from 4 health care systems during 2006–2012 were evaluated. Using chart abstraction and review to confirm chronic HCV cases ("gold standard" definition), the authors calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of 2 case definitions: (1) 2 ICD-9 codes separated by 6 months and (2) 1 positive HCV RNA (ribonucleic acid) test. Among 2,718,995 patients, 20,779 (0.8%) with ICD-9 codes indicating a likely diagnosis of chronic HCV infection were identified; 13,595 (65.4%) of these were randomly selected for review. Case definition 1 (2 ICD-9 codes separated by 6 months) had 70.3% sensitivity, 91.9% PPV, 99.9% specificity, and 99.9% NPV while case definition 2 (1 positive HCV RNA test) had 74.1% sensitivity, 97.4% PPV, 99.9% specificity, and 99.9% NPV. The predictive values of these alternate EHR-derived ICD-9 code-based case definitions suggest that these measures may be useful in capturing the burden of diagnosed chronic HCV infections. Their use can augment current chronic HCV case surveillance efforts; however, their accuracy may vary by length of observation and completeness of EHR data.

Address correspondence to: Winston E. Abara, MD, PhD, Division of Viral Hepatitis, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G-37, Atlanta, GA 30333, wabara@cdc.gov.

Keywords

Hepatitis C; public health surveillance; electronic health records; International Classification of Diseases

Background

THERE ARE APPROXIMATELY 3 million persons living with the hepatitis C virus (HCV) in the United States and at least half are not aware of their infection.[1–3] An estimated 30,000 new HCV infections occurred in the United States in 2013.[4,5] Between 75% and 85% of acute HCV infections become chronic HCV infections.[6] Persons who are infected with chronic HCV are at increased risk of chronic liver disease and hepatocellular carcinoma.[7,8] Chronic HCV infection is also the leading indication for liver transplants in the United States.[9] Deaths from chronic HCV infection are on the rise, with most deaths related to liver cirrhosis and hepatocellular cancer. Mortality attributed to HCV infection has surpassed mortality attributed to all other nationally notifiable infectious conditions.[10]

Viral hepatitis surveillance is critical to accurately estimating the true burden of HCV infection in the United States.[4] However, the true prevalence of chronic HCV infection cannot be determined from national surveillance data[11,12] because current surveillance is based on a passive reporting system, (ie, a system wherein state and local health departments have variable ability to detect and report viral hepatitis A, B, and C cases through the National Notifiable Diseases Surveillance System [NNDS]).[4,5] The surveillance infrastructure to monitor chronic HCV infections at the local and state level is inadequate.[13] Chronic HCV infection surveillance has been limited by the lack of resources and personnel at the local and state levels to assess and follow-up HCV case reports to determine chronicity, resulting in considerable delay, under-ascertainment, and underreporting of cases. Such efforts also are labor intensive for public health department staff, because, for example, an average of 4 documents per case may need to be evaluated to distinguish between acute, resolved, or chronic HCV case status.[14] Thus, in 2013, chronic HCV surveillance cases were reported by only 37 of the 50 states.[4] Accordingly, the National Health and Nutrition Examination Survey (NHANES) has mainly been used to estimate the burden of chronic HCV in the United States.[1,15] NHANES is a random survey of US residents and obtains representative data on the health and nutritional status of the population. Though it has been used to estimate national HCV infection prevalence,[1] the NHANES sample is a low-risk, low-prevalence population, consequently limiting its generalizability.[1,16] NHANES also includes surveillance data from persons who are currently infected and those ever infected with HCV, and therefore may not accurately capture cases of chronic HCV infection.[1]

Adoption of electronic health records (EHRs) has increased significantly over the past decade in the United States.[17] With the increasing availability of EHRs, the use of the *International Classification of Diseases, Ninth Revision* (ICD-9) codes to complement public health surveillance has been proposed.[18] The ICD-9 coding system is a system that assigns specific codes to clinical diagnoses and procedures that are associated with health care utilization in the United States. Prior work has demonstrated that ICD-9 codes have good

sensitivity and specificity for public health surveillance of some disease conditions.[18–21] For example, one study showed that the use of EHR-derived ICD-9 codes for acute hepatitis B had good sensitivity and specificity when compared to chart abstractor confirmed diagnoses and can be used for disease surveillance[18] Other studies have shown that EHR-derived ICD-9 codes can be effective for the public health surveillance of influenza-like illnesses, *Clostridium difficile* infections, community-acquired pneumonia, and other acute communicable diseases.[19,20,22,23] There are limited data on the sensitivity and specificity of ICD-9 or ICD-10 codes for chronic HCV infection[24] and their utility in the surveillance of chronic HCV infections despite the aforementioned barriers to current surveillance strategies.

With increasing use of EHRs in the United States, the predictive values of ICD-9 and ICD-10 codes for identifying true cases of chronic HCV infection warrants evaluation. The use of EHR-derived ICD-9 and ICD-10 codes to identify true cases of chronic HCV infection can further complement surveillance data obtained from NNDS and NHANES. In this study, data from the Chronic Hepatitis Cohort Study (CHeCS) were used to examine the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of using 2 alternate EHR-derived ICD-9 code-based case definitions, (1) 2 hepatitis C ICD-9 codes separated by 6 months and (2) 1 positive HCV RNA (ribonucleic acid) test, when compared to true cases of chronic HCV infection, ascertained by abstractor chart review ("gold standard" case definition).

Methods

CHeCS is a cohort study drawing patients from the EHR and claims databases of 4 large integrated health systems with data on more than 2.7 million patients. This analysis includes patients aged 18 years or older who used any health service between January 1, 2006, and December 31, 2012.[25] The 4 integrated health systems include Geisinger Health System (GHS), Danville, PA; Henry Ford Health System (HFHS), Detroit, MI; Kaiser Permanente-Hawaii, (KPHI), Honolulu, HI; and Kaiser Permanente-Northwest (KPNW), Portland, OR. GHS serves approximately 2.6 million Pennsylvania residents. HFHS provides health care services to more than 1 million southeast Michigan residents. KPHI provides health care services to about 220,000 health plan members in Hawaii while KPNW provides health care services to about 500,000 health plan members in Oregon.

Eligible patients for this study were selected from the CHeCS cohort between January 1, 2006, and December 31, 2012 because chronic HCV infections could be confirmed by abstractor chart review (gold standard) during this time period. The study team obtained retrospective and prospective longitudinal EHR data related to patients' demographic characteristics, medical diagnoses, and laboratory results starting from the time of the patients' first visit to the health system when EHRs were available in order to identify chronic HCV infections. Data were evaluated both retrospectively and prospectively through the end of 2013. Given that CHeCS patients' records have received ICD-9 coding to date, patients with at least 1 of the following were considered eligible for further evaluation: any acute or chronic HCV ICD-9 code (070.44, 070.54, 070.70, 070.71, 070.41, 070.51), a positive HCV antibody test, or a positive HCV RNA test (nucleic acid test [NAT];

qualitative, quantitative, or genotype test). Each year, patients' records were evaluated to determine whether the cohort inclusion criteria were met, and patients who did not meet the cohort inclusion criteria in one year were reevaluated for cohort inclusion in subsequent years. Trained medical abstractors reviewed the EHRs of patients meeting cohort inclusion criteria to collect summary data and confirm chronic HCV infection status. Because they might in fact be true chronic HCV cases, charts lacking documentation that a specialist had diagnosed chronic HCV or with documentation of acute HCV were flagged and reviewed under the supervision of a physician using hepatologist-developed criteria.[25] Flagged cases for which chronic viral hepatitis infection could not be confirmed were excluded from the study.

The demographics of patients with chronic HCV infection meeting cohort inclusion criteria and confirmed by trained medical abstractors (the gold standard), were compared to alternate "chronically infected cases" identified solely by EHR-derived ICD-9 codes (2 hepatitis C ICD-9 codes separated by 6 months) or laboratory results (1 positive HCV RNA test). The sensitivity, specificity, PPV, and NPV of the 2 alternate case definitions were compared to the gold standard. All analyses were conducted using SAS v.9.3 (SAS Institute, Inc., Cary, NC).

Results

Figure 1 is a flowchart that shows CHeCS cohort eligibility and chronic HCV case identification. Of 2,718,995 eligible health system patients, the study team identified 20,779 (0.8% of the total health system population) unique individuals with a likely diagnosis of chronic HCV infection (1 positive anti-HCV or NAT, 2 HCV ICD-9 codes, or 1 HCV ICD-9 code plus a positive anti-HCV or NAT). Of these likely cases, ~65% were randomly selected for review by trained medical abstractors because of funding restrictions. Numbers of confirmed chronic HCV infections and chronic HCV infections ruled out after medical abstractor chart review are shown in Figure 1. Less than 10% of patients had insufficient evidence–primarily lacking an HCV RNA test–to determine chronic HCV infection status.

The study team compared the demographic characteristics of the 11,273 abstractor-confirmed (gold standard) chronic HCV cases with the 7928 confirmed chronic HCV cases identified using the alternate EHR-derived ICD-9 case definition, 2 HCV ICD-9 codes separated by 6 months, and 8349 confirmed chronic HCV cases identified using the EHR-derived laboratory-based case definition, 1 positive HCV RNA test. The age and sex composition of each of the 3 case definitions (gold standard case definition and the 2 alternate case definitions) are similar (Table 1). There are slight differences in the composition of White, non-Hispanic and Black non-Hispanic participants by the case definitions. White, non-Hispanic participants accounted for between 58.8% and 68.4% of respondents by case definition while Black, non-Hispanic participants accounted for between 19.4% and 29.7% of respondents. The composition of Asian, non-Hispanic, Native American, Hawaiian/Pacific Islander, and unknown race respondents was similar across all 3 case definitions. HFHS accounted for the greatest proportion of respondents identified using the gold standard case definition (abstractor confirmed chronic HCV infection cases) and the EHR-derived laboratory-based case definition, 1 positive HCV RNA, while KPNP

accounted for the greatest proportion of respondents identified used alternate EHR-derived ICD-9 case definition, 2 HCV ICD-9 codes separated by 6 months (Table 1).

Chronic HCV case definition: 2 HCV ICD-9 codes separated by 6 months

The case definition, 2 HCV ICD-9 codes separated by 6 months, identified 8622 chronic HCV infection cases (Fig. 1A). However, this definition did not identify 38.8% of true chronic HCV cases (cases identified by the gold standard of abstractor chart review). Numbers of confirmed chronic HCV infections (true positives) and those ruled out as chronic HCV infections (false positives) are shown in Figure 1A.

Table 2 shows the sensitivity, specificity, PPV, and NPV of this case definition. Using retrospective longitudinal data, the sensitivity, the ability to identify true chronic HCV infections, of this case definition was 70.3%, while its specificity, the ability to correctly identify non–chronic HCV infections, was almost 100%. The PPV describes the proportion of chronic HCV infections identified by this case definition that are confirmed chronic HCV infections while the NPV describes the proportion of non–chronic HCV infections identified by this case definition that were ruled out as chronic HCV infections. The PPV and NPV of this case definition were over 91% (Table 2).

Chronic HCV case definition: 1 positive HCV RNA test

The case definition, 1 positive HCV RNA test, identified 8574 chronic HCV cases (Fig. 1B). This definition did not identify 34.1% true chronic HCV infections (cases identified by the gold standard of abstractor chart review). Numbers of confirmed chronic HCV infections (true positives) and those ruled out as chronic HCV infections (false positives) are shown in Figure 1B. The reasons for ruling out chronic HCV infection among these cases included errors in lab data and spontaneous clearance of infection.

Table 2 shows the sensitivity, specificity, PPV, and NPV of this case definition. The sensitivity of this case definition was 74.1% and the specificity was almost 100%. The PPV and NPV of this case definition were over 97% (Table 2).

Discussion

Currently, passive national surveillance conducted through the NNDS only captures a small percentage of all chronic HCV cases[5]; hence, this study explored the utility of EHRs and found some evidence that its use could supplement current HCV infection surveillance efforts. The findings of this analysis suggest that using the 2 alternate EHR-derived case definitions in the longitudinal data set (ie, 2 HCV ICD-9 codes separated by 6 months, or 1 positive HCV RNA test) to identify chronic HCV infections yielded moderate but varying sensitivity and PPV. The sensitivity and PPV of using 1 positive HCV RNA test to identify chronic HCV infection was greater than using 2 HCV ICD-9 codes separated by 6 months. However, the 2 alternate case definitions had the same specificity and NPV. The greater sensitivity and PPV associated with using 1 positive HCV RNA test as a case definition may be related to the fact that patients who have an HCV RNA test are those who may have received a positive antibody test to screen for HCV infection previously, may have elicited a greater index of clinical suspicion from their physicians, or may be

more clinically symptomatic. Hence, these patients may be more likely to have true chronic HCV infections than those who only have HCV ICD-9 codes. Among persons with ≥2 HCV ICD-9 codes separated by ≥6 months, the first HCV ICD-9 code might be a "rule out" diagnosis code and may not reflect a true diagnosis, consequently affecting the sensitivity of this case definition. The lower sensitivity associated with using ≥2 HCV ICD-9 codes separated by ≥6 months also may be affected by time duration (≥6 months) associated with this case definition compared with a case definition that requires ≥1 positive HCV RNA test that is not associated with any time duration. Additionally, because chronic HCV infection is often an underlying disease condition, physicians may fail to include this diagnosis at every clinic visit, further decreasing the sensitivity of using the alternate case definition, ≥2 HCV ICD-9 separated by ≥6 months.

Despite the increased sensitivity and PPV of using ≥1 positive HCV RNA test, this case definition is limited by the fact that it cannot completely distinguish between acute and chronic HCV infections. HCV RNA can be detected in blood as early as 2 to 3 weeks after infection and laboratory records alone do not provide a comprehensive clinical picture of chronicity.[26] The increased PPV of using ≥1 positive HCV RNA may be attributed to the difference in prevalence of confirmed chronic HCV cases identified by both case definitions. The case definition, ≥1 positive HCV RNA, identified more confirmed chronic HCV infections than the other case definition, ≥2 HCV ICD-9 codes separated by ≥6 months.

This analysis is subject to several limitations. These findings may not be generalizable to other populations because the prevalence of chronic HCV infection in this sample may be higher than the general population[1] and data used were obtained from the EHRs of 4 health systems that primarily provided health care to certain catchment areas in 4 US states. Using EHR laboratory data to distinguish true from questionable or non–cases of HCV infection required very extensive data processing and selected chart review, primarily because laboratory test codes and reporting formats are not standardized between laboratories, and these codes are subject to change over time. This analysis did not evaluate whether the time for data capture affects the sensitivity and specificity estimates of the case definitions. Especially for chronic conditions such as chronic HCV infection, a longer time for data capture may increase the number of true cases identified.

Another limitation of this study is that its findings were based on ICD-9 codes that have recently been replaced by ICD-10 codes. Despite the improved specificity of ICD-10 codes, the study team expects the conclusions to remain the same because the ICD-9 codes the EHR-derived case definition was based on are consistent with the new ICD-10 codes. It is also possible that some true cases were missed among the patients excluded from chart review who had only a single ICD-9 code for HCV but no positive anti-HCV or RNA tests. This analysis was limited to patients who had access to health care. Undiagnosed cases also could not be ascertained.[2] The study cohort was geographically and demographically diverse and as such may not reflect the demographic composition of patients in other geographic regions, hence limiting its generalizability. However, this patient population reflects real-life clinical care at 4 sites across the United States with broad catchment areas. There also are human and economic limitations regarding the feasibility of using ICD codes for public

health surveillance that can be attributed to the extensive data processing and chart review that are required to ensure data completeness and accuracy.

In conclusion, the predictive values of these alternate EHR-derived ICD-9 code-based case definitions suggest that these measures may be useful in complementing current strategies to capture the burden of diagnosed chronic HCV infections. Although results should be interpreted cautiously as reliability may vary greatly according to the completeness of EHR data and EHR data observation period, their use offers a timely chronic HCV surveillance approach that can augment, but not replace, current chronic HCV infection surveillance efforts that are based on data from NNDS and NHANES.

Author Disclosure Statement

References

1. Denniston MM, Jiles RB, Drobeniuc J, et al. Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. Ann Intern Med 2014;160:293–300. [PubMed: 24737271]

2. Spradling PR, Rupp L, Moorman AC, et al. Hepatitis B and C virus infection among 1.2 million persons with access to care: factors associated with testing and infection prevalence. Clin Infect Dis 2012;55:1047–1055. [PubMed: 22875876]

3. Roblin DW, Smith BD, Weinbaum CM, Sabin ME. Hepatitis C virus screening practices and prevalence in an MCO, 2000–2007. Am J Managed Care 2011;17:548–55.

4. Centers for Disease Control and Prevention. Viral hepatitis surveillance United States, 2013. www.cdc.gov/hepatitis/statistics/2013surveillance/pdfs/2013hepsurveillancerpt.pdf Accessed March 21, 2016.

5. Centers for Disease Control and Prevention. Viral hepatitis surveillance United States, 2014. www.cdc.gov/hepatitis/statistics/2014surveillance/pdfs/2014hepsurveillancerpt.pdf Accessed September 13, 2016.

6. Ghany MG, Strader DB, Thomas DL, Seeff LB; American Association for the Study of Liver Diseases. Diagnosis, management, and treatment of hepatitis C: an update. Hepatology 2009;49:1335–1374. [PubMed: 19330875]

7. Velazquez RF, Rodriguez M, Navascues CA, et al. Prospective analysis of risk factors for hepatocellular carcinoma in patients with liver cirrhosis. Hepatology 2003;37: 520–527. [PubMed: 12601348]

8. Ly KN, Xing J, Klevens M, Jiles R, Ward JW, Holmberg SD. The increasing burden of mortality from viral hepatitis in the United States between 1999 and 2007. Ann Intern Med 2012;156:271–278. [PubMed: 22351712]

9. Kim WR, Terrault NA, Pedersen RA, et al. Trends in waiting list registration for liver transplantation for viral hepatitis in the United States. Gastroenterology 2009;137: 1680–1686. [PubMed: 19632234]

10. Ly KN, Hughes EM, Jiles RB, Holmberg SD. Rising mortality associated with hepatitis C virus in the United States, 2003–2013. Clin Infect Dis 2016;62:1287–1288. [PubMed: 26936668]

11. Edlin BR, Eckhardt BJ, Shu MA, Holmberg SD, Swan T. Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology 2015;62: 1353–1363. [PubMed: 26171595]

12. Smith BD, Morgan RL, Beckett GA, et al. Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945–1965. MMWR Recomm Rep 2012;61(RR-4):1–32.

13. Institute of Medicine. Hepatitis and liver cancer: a national strategy for prevention and control of hepatitis B and C. Washington, DC: National Academies Press, 2010.

14. Klevens RM, Miller J, Vonderwahl C, et al. Population-based surveillance for hepatitis C virus, United States, 2006–2007. Emerg Infect Dis 2009;15:1499–1502. [PubMed: 19788825]

15. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 2006;144:705–714. [PubMed: 16702586]

16. National Institutes of Health. National Instutues of Health consensus development conference statement: management of hepatitic C: 2002–June 10–12, 2002. Hepatology 2002; 36(5 suppl 1):S3–S20. [PubMed: 12407572]

17. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. NCHS Data Brief, no 143. Hyattsville, MD: National Center for Health Statistics, 2014.

18. Klompas M, Haney G, Church D, Lazarus R, Hou X, Platt R. Automated identification of acute hepatitis B using electronic medical record data to facilitate public health surveillance. PLoS One 2008;3:e2626. [PubMed: 18612462]

19. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenzalike illness by International Classification of Diseases, Ninth Revision. Emerg Infect Dis 2007;13:207. [PubMed: 17479881]

20. Sickbert-Bennett EE, Weber DJ, Poole C, MacDonald PD, Maillard JM. Utility of International Classification of Diseases, Ninth Revision, Clinical Modification codes for communicable disease surveillance. Am J of Epidemiol 2010; 172:1299–1305. [PubMed: 20876668]

21. Byrne DD, Newcomb CW, Carbonari DM, et al. Prevalence of diagnosed chronic hepatitis B infection among US Medicaid enrollees, 2000–2007. Ann Epidemiol 2014;24:418–423. [PubMed: 24703196]

22. Dubberke ER, Reske K, McDonald LC, Fraser VJ. ICD-9 codes and surveillance for Clostridium difficile–associated disease. Emerg Infect Dis 2006;12:1576–1579. [PubMed: 17176576]

23. van de Garde EM, Oosterheert JJ, Bonten M, Kaplan RC, Leufkens HG. International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. J Clin Epidemiol 2007;60:834–838. [PubMed: 17606180]

24. Mahajan R, Moorman AC, Liu SJ, Rupp L, Klevens RM; Chronic Hepatitis Cohort Study (CHeCS) Investigators. Use of the International Classification of Diseases, 9th revision, coding in identifying chronic hepatitis B virus infection in health system data: implications for national surveillance. J Am Med Inform Assoc 2013;20:441–445. [PubMed: 23462875]

25. Moorman AC, Gordon SC, Rupp LB, et al. Baseline characteristics and mortality among people in care for chronic viral hepatitis: the chronic hepatitis cohort study. Clin Infect Dis 2013;56:40–50. [PubMed: 22990852]

26. Pawlotsky JM. Use and interpretation of virological tests for hepatitis C. Hepatology 2002;36(5 suppl 1):S65–S73. [PubMed: 12407578]
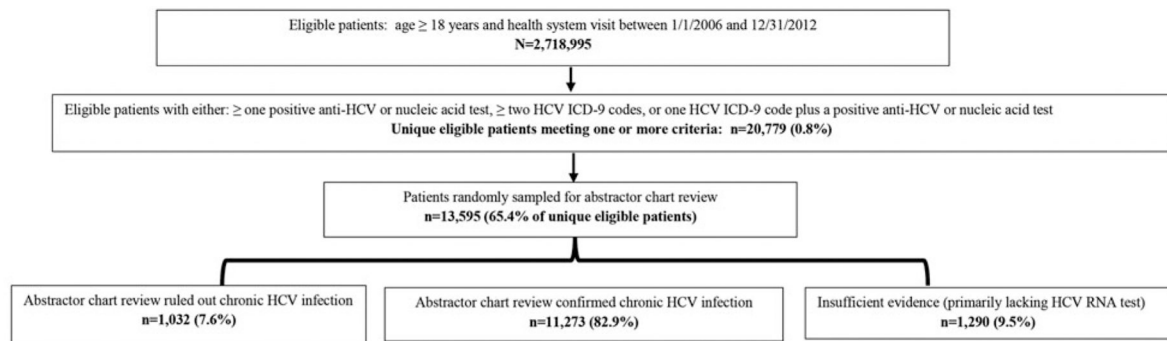
Eligible patients: age ≥ 18 years and health system visit between 1/1/2006 and 12/31/2012
N=2,718,995

Eligible patients with either: ≥ one positive anti-HCV or nucleic acid test, ≥ two HCV ICD-9 codes, or one HCV ICD-9 code plus a positive anti-HCV or nucleic acid test
Unique eligible patients meeting one or more criteria: n=20,779 (0.8%)

Patients randomly sampled for abstractor chart review
n=13,595 (65.4% of unique eligible patients)

Abstractor chart review ruled out chronic HCV infection
n=1,032 (7.6%)

Abstractor chart review confirmed chronic HCV infection
n=11,273 (82.9%)

Insufficient evidence (primarily lacking HCV RNA test)
n=1,290 (9.5%)

Figure 1A. Identification of chronic HCV cases using an alternate case definition, among 11,967 (88%) patients sampled for abstractor chart review and meeting this definition: ≥ two HCV ICD-9 codes separated by ≥ six months (070.44, 070.54, 070.70, 070.71, 070.41, 070.51) separated by > six months
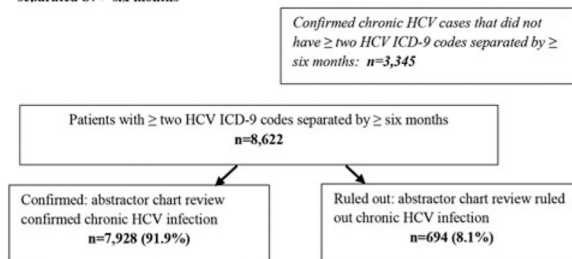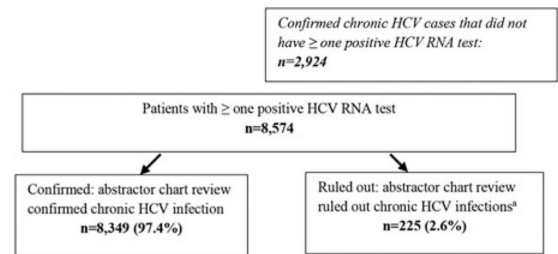
Confirmed chronic HCV cases that did not have ≥ two HCV ICD-9 codes separated by ≥ six months: **n=3,345**

Patients with ≥ two HCV ICD-9 codes separated by ≥ six months
n=8,622

Confirmed: abstractor chart review confirmed chronic HCV infection
n=7,928 (91.9%)

Ruled out: abstractor chart review ruled out chronic HCV infection
n=694 (8.1%)

Figure 1B. Identification of chronic HCV cases using an alternate case definition, among 11,498 (84.6%) patients sampled for abstractor chart review and meeting this definition: ≥ one positive HCV RNA test

Confirmed chronic HCV cases that did not have ≥ one positive HCV RNA test: **n=2,924**

Patients with ≥ one positive HCV RNA test
n=8,574

Confirmed: abstractor chart review confirmed chronic HCV infection
n=8,349 (97.4%)

Ruled out: abstractor chart review ruled out chronic HCV infections[a]
n=225 (2.6%)

[a]reasons cited for ruling out chronic HCV infection included errors in lab data for true negative patients and spontaneous clearance of infection

FIG. 1.
Flow chart for cohort eligibility and chronic HCV case identification. HCV, hepatitis C virus; ICD-9, International Classification of Diseases, Ninth Revision; RNA, ribonucleic acid.

Table 1.

Comparison of Demographic and Site Characteristics of Confirmed Chronic Hepatitis C Virus Infections by Case Definition, Chronic Hepatitis Cohort Study, 2006–2012

Variable	Abstractor chart review confirmed chronic HCV cases (gold standard), N = 11,273, n (%)	Confirmed chronic HCV cases with ≥2 HCV ICD-9 codes separated by ≥6 months, N = 7928, n (%)	Confirmed chronic HCV cases with ≥1 positive HCV RNA test, N = 8349, n (%)
Age group (years)			
18–29	607 (5.4)	354 (4.5)	458 (5.5)
30–39	586 (5.2)	381 (4.8)	410 (4.9)
40–49	1964 (17.4)	1422 (17.9)	1302 (15.6)
50–59	5604 (49.7)	4073 (51.4)	4156 (49.8)
60–69	1914 (16.9)	1316 (16.6)	1562 (18.7)
70–79	449 (3.9)	298 (3.8)	349 (4.2)
≥80	149 (1.3)	84 (1.1)	112 (1.3)
Sex			
Male	6741 (59.8)	4706 (59.4)	5042 (60.4)
Female	4532 (40.2)	3222 (40.6)	3307 (39.6)
Race			
White, non-Hispanic	7246 (64.3)	5425 (68.4)	4907 (58.8)
Black, non-Hispanic	2725 (24.2)	1537 (19.4)	2478 (29.7)
Asian, non-Hispanic	390 (3.5)	298 (3.8)	308 (3.7)
Native American	185 (1.6)	139 (1.8)	128 (1.5)
Hawaiian/Pacific Islander	160 (1.4)	139 (1.8)	139 (1.7)
Unknown	567 (5.0)	390 (4.9)	389 (4.7)
Site			
Kaiser Permanente-Northwest, Portland, OR	3259 (28.9)	2712 (34.2)	1576 (18.9)
Kaiser Permanente-Honolulu, HI	1006 (8.9)	806 (10.2)	875 (10.5)
Henry Ford Health System, Detroit, MI	4617 (41.0)	2624 (33.1)	4377 (52.4)
Geisinger Health System, Danville, PA	2391 (21.2)	1786 (22.5)	1521 (18.2)

HCV, hepatitis C virus; RNA, ribonucleic acid.

Table 2.

Sensitivity, Specificity, Positive and Negative Predictive Values of Two Alternate Electronic Health Record-Derived ICD-9 Code-Based Chronic HCV Case Definitions, Chronic Hepatitis Cohort Study, 2006–2012

	Confirmed chronic HCV cases, *n*	Ruled out chronic HCV cases, *n*	Total, *N*	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Sensitivity, specificity, positive and negative predictive values of using 2 HCV ICD-9 codes[a] separated by 6 months							
2 HCV ICD-9 codes separated by 6 months							
Yes	7928	694	8622	70.3		91.9	
No	3345	2,707,028	2,710,373		99.9		99.9
Total	11,273	2,707,722	2,718,995				
Sensitivity, specificity, positive and negative predictive values of using 1 positive HCV RNA test							
1 positive HCV RNA test							
Yes	8349	225	8574	74.1		97.4	
No	2924	2,707,497	2,710,421		99.9		99.9
Total	11,273	2,707,722	2,718,995				

[a]070.44, 070.54, 070.70, 070.71, 070.41, 070.51.

EHR, electronic health record; HCV, hepatitis C virus; ICD-9, International Classification of Diseases, Ninth Revision; RNA, ribonucleic acid.