Validation of diagnoses codes for identifying patients with hepatitis C virus in EHR in a US, real-world population

RWD7

Ellen Stein[1], ellen.stein@trinetx.com
[1]TriNetX, LLC., Cambridge, MA, USA.

OBJECTIVES

There is a dearth of research validating diagnostic codes to identify patients with hepatitis C virus (HCV) in the US,[1,2] yet these codes are used in studies globally to identify HCV patients.[3,4] There are inconsistencies in existing studies using laboratory results versus laboratory results with diagnostic codes to identify chronic versus active HCV.[5-8] Given these shortcomings, this study aims to validate using diagnoses codes to identify chronic and active HCV infection as compared with the gold standard of lab-confirmed infection among the general US, healthcare-seeking population.

METHODS

This study leverages TriNetX's Linked EHR + Closed Claims network on the TriNetX Live platform, a network of de-identified EHR data within the US linked to closed claims, to identify patients with least two documented encounters at least 365 days apart since 2016, who are at least 18 years of age at second encounter. Among those patients:

- **True positives** were defined as those with a diagnosis of HCV and a positive laboratory test for HCV (see Table 1 for gold standard designation by code; used LOINC codes with a positive/negative result with at least 100 observations in the network).
- **True negatives** have neither documented diagnosis nor positive laboratory test for HCV.
- **False positives** were defined by a diagnosis of HCV but no positive laboratory test,.
- **False negatives** were defined has having no documented diagnosis of HCV but a positive laboratory test.
- **Sensitivity**, **specificity**, and **positive predictive value (PPV)** were calculated.

1. Sheu, M.J., et al., *Validity of ICD-10-CM Codes Used to Identify Patients with Chronic Hepatitis B and C Virus Infection in Administrative Claims Data from the Taiwan National Health Insurance Outpatient Claims Dataset.* Clin Epidemiol, 2020. **12**: p. 185-192.
2. Niu, B., K.A. Forde, and D.S. Goldberg, *Coding algorithms for identifying patients with cirrhosis and hepatitis B or C virus using administrative data.* Pharmacoepidemiol Drug Saf, 2015. **24**(1): p. 107-11.
3. Ashimkhanova, A., et al., *Epidemiological Characteristics of Chronic Viral Hepatitis in Kazakhstan: Data from Unified Nationwide Electronic Healthcare System 2014-2019.* Infect Drug Resist, 2022. **15**: p. 3333-3346.
4. Nakazawa, S., et al., *Occupational class and risk of hepatitis B and C viral infections: A case-control study-based data from a nationwide hospital group in Japan.* J Infect Public Health, 2022. **15**(12): p. 1415-1426.
5. Kapadia, S.N., et al., *Hepatitis C Treatment Initiation Among US Medicaid Enrollees.* JAMA Netw Open, 2023. **6**(8): p. e2327326.
6. Scott, J., et al., *A Population-Based Intervention to Improve Care Cascades of Patients With Hepatitis C Virus Infection.* Hepatol Commun, 2021. **5**(3): p. 387-399.
7. Duffell, E., et al., *Estimating the attributable fraction of cirrhosis and hepatocellular carcinoma due to hepatitis B and C.* J Viral Hepat, 2021. **28**(8): p. 1177-1189.
8. Wu, H.C., et al., *Incidence of hepatocellular carcinoma in a community-based Taiwanese population without chronic HBV/HCV infection.* JHEP Rep, 2022. **4**(2): p. 100410.

ISPOR 2026 | Philadelphia, PA, USA
Design services were provided by Kelly Marsocci and were funded by TriNetX, LLC.

Table 1. Diagnoses codes included for validation

ICD-CM-10 code*	Used to classify	Gold standard*
B17.1	New, active infection	RNA +, Ab-
B18.2	Chronic infection	RNA+, Ab+

*ICD-10 is the classification system used by TriNetX to represent diagnoses. ICD-9 and other extensions/modifications of ICD are mapped to ICD-10. As such, when this document refers to an ICD-10 code, it implicitly encompasses the corresponding ICD-9 codes as well.

Table 2. Designation of true positives, true negatives, false positives, and false negatives.

		Lab-confirmed HCV: Yes	Lab-confirmed HCV: No
Diagnosis of HCV	**Yes**	True positive	False positive
	No	False negative	True negative

Table 3. Results from Linked Closed Claims + EHR network.*

	TOTAL	True positive	True negative	False positive	False negative
B17.1	15,262,512	10	15,226,308	36,167	27
B18.2	16,261,817	919	16,026,259	234,521	118

*Data as of 4/21/2026, please note that TriNetX Live platform is dynamic and thus totals may change over time.

Table 4. Sensitivity, specificity and PPV by ICD-CM-10 code.

	Sensitivity	Specificity	PPV
B17.1	0.2703	0.9976	0.0003
B18.2	0.8862	0.9856	0.0039

RESULTS

Among the 22,025,599 patients in the Linked EHR + Closed Claims network, 16,309,104 patients met the eligibility criteria. For B17.1 (Acute hepatitis C), there were 10 true positive patients, 15,226,308 true negatives, 36,167 false positives, and 27 false negatives.

RESULTS (cont'd)

Sensitivity was 0.2703. Specificity was 0.9976. PPV was 0.0003. For B18.2 (chronic viral hepatitis C), there were 919 true positives, 16,026, 259 true negatives, 234,521 false positives, and 118 false negatives. Sensitivity was 0.8862. Specificity was 0.9856. PPV was 0.0039.

CONCLUSIONS

Within the TriNetX Linked EHR + Closed Claims network, using diagnosis codes to identify HCV as compared with the gold standard of lab-confirmed HCV, the diagnosis code for acute HCV was not very sensitive, highly specific, and had a low PPV; the diagnosis code for chronic HCV was moderately sensitive, highly specific, and had a low PPV.

Low PPV for both codes likely reflects low prevalence in the overall population, which was crudely calculated as 0.000% of total patients for acute HCV (B17.1), and 0.006% for chronic HCV (B18.2). This aligns with other published literature, though PPV is lower than among other cohorts, such as cirrhotic patients.[2]

Like any analysis, this one is subject to limitations. First, the limitations of RWD are well established; patients may receive care at multiple institutions, which means care outside of the institution under investigation may not be captured. This analysis leveraged the Linked EHR + Closed Claims network, which helps mitigate this issue; however, this may be why we see such a low prevalence within this population. Second, the temporal features of the TriNetX Live platform was unable to capture the complex relationship between RNA and antibody testing that plays out after viral infection, thus tests at any time during the study period were used. Future analyses using a downloaded dataset from TriNetX would allow for more nuance. Third, the TriNetX Live platform rounds counts under 10 to 10, meaning the true positive value for B17.1 may be an over-estimate.

Download Poster PDF