Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2015 January ; 24(1): 107–111. doi:10.1002/pds.3721.

Coding algorithms for identifying patients with cirrhosis and hepatitis B or C virus using administrative data

Bolin Niu, MD¹, Kimberly A Forde, MD, MHS^{2,3}, and David S. Goldberg, MD, MSCE^{2,3,4}

¹Department of Medicine, University of Pennsylvania

²Division of Gastroenterology, Department of Medicine, University of Pennsylvania

³Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania

⁴Leonard Davis Institute of Health Economics, University of Pennsylvania

Abstract

Background & Aims—Despite the use of administrative data to perform epidemiological and cost-effectiveness research on patients with hepatitis B or C virus (HBV, HCV), there are no data outside of the Veterans Health Administration validating whether International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) codes can accurately identify cirrhotic patients with HBV or HCV. The validation of such algorithms is necessary for future epidemiological studies.

Methods—We evaluated the positive predictive value (PPV) of ICD-9-CM codes for identifying chronic HBV or HCV among cirrhotic patients within the University of Pennsylvania Health System, a large network that includes a tertiary care referral center, a community-based hospital, and multiple outpatient practices across southeastern Pennsylvania and southern New Jersey. We reviewed a random sample of 200 cirrhotic patients with ICD-9-CM codes for HCV and 150 cirrhotic patients with ICD-9-CM codes for HBV.

Results—The PPV of 1 inpatient or 2 outpatient HCV codes was 88.0% (168/191, 95% CI: 82.5–92.2%), while the PPV of 1 inpatient or 2 outpatient HBV codes was 81.3% (113/139, 95% CI: 73.8–87.4%). Several variations of the primary coding algorithm were evaluated to determine if different combinations of inpatient and/or outpatient ICD-9-CM codes could increase the PPV of the coding algorithm.

Conclusions—ICD-9-CM codes can identify chronic HBV or HCV in cirrhotic patients with a high PPV, and can be used in future epidemiologic studies to examine disease burden and the proper allocation of resources.

Corresponding author: David Goldberg, MD, MSCE, Blockley Hall, 423 Guardian Drive, Room 703, Philadelphia, PA 19104, david.goldberg@uphs.upenn.edu, Phone: 215-746-8598, Fax: 215-349-5915.

Conflict of interest: None of the authors have any relevant conflicts of interest.

Statement about prior postings and presentations: No prior postings or presentations have been performed at this time.

Introduction

Cirrhosis and chronic liver disease account for over 200,000 hospitalizations and 30,000 deaths annually in the United States.¹ Two of the most common etiologies of cirrhosis, hepatitis B virus (HBV) and hepatitis C virus (HCV), account for over 1 million outpatient visits, and approximately 40% of all liver transplants in the US.^{2,3} Over the next 10–20 years, the burden of hepatitis-related liver disease is projected to rise, making this a continued public health problem.⁴

Administrative databases are valuable reservoirs of information for epidemiological and outcomes research using International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes. Yet research using such databases has inherent limitations as it is unknown if ICD-9-CM billing codes accurately identify patients with the clinical conditions associated with such codes. Algorithms to identify patients with HCV and HBV have been developed and validated in the Veterans Health Administration (VHA) system. However, the VHA is unique in that: 1) ICD-9 codes associated with visits are not used for billing; 2) the prevalence of HBV and HCV in the VHA is greater than the general commercially-insured population; and 3) the VHA population is composed of > 90% males. Thus, it is necessary to validate an algorithm to identify patients with HBV and HCV outside of the VHA as given its limited generalizability outside the VHA system. Furthermore, while a recent publication in the Medicaid population validated an algorithm to identify chronic HBV, it did not specifically focus on patients with cirrhosis.⁵

Several recent high-impact studies examining the epidemiology and costs associated with chronic hepatitis care have been conducted in administrative datasets.^{6–8} Yet these studies used ICD-9-CM codes to identify the cohorts with HBV or HCV, without validating such codes. Despite emerging treatments for HCV, administrative databases will continue to be the primary mechanism to perform large-scale retrospective epidemiological and cost-effectiveness research. Therefore, the aim of this study was to evaluate the performance of ICD-9-CM codes for identifying chronic HBV and HCV infection among a cohort of patients with cirrhosis in an administrative database.

Methods

Study design and data source

We conducted a retrospective study using administrative data from the University of Pennsylvania Health System (UPHS). UPHS encompasses a tertiary-care academic hospital, a community medical center, and outpatient practices across southeastern Pennsylvania and New Jersey. Data were obtained from the Penn Data Store (PDS), a collection of all administrative data in UPHS, which includes ICD-9-CM codes, laboratory test results, and ambulatory electronic health records. PDS is a clinical data warehouse containing over 3 billion records of data, and integrates data from multiple systems into a consolidated dataset. While the Penn Data Store contains data beyond just ICD-9-CM codes, the ICD-9-CM codes used for this study to identify patients with HCV and HBV are the same billing codes submitted to insurers for reimbursement, and thus are the same codes that would be contained in an insurance administrative database.

Selection of subjects and coding algorithm derivation

Subjects with cirrhosis were first identified using a validated ICD-9-CM algorithm: a) at least one inpatient cirrhosis ICD-9-CM code (571.2 or 571.5) or b) at least two outpatient ICD-9-CM cirrhosis codes based on physician face-to-face encounters between January 1, 2002 and October 20, 2013. This algorithm for cirrhosis was validated within the UPHS database.⁹

A random sample of 200 cirrhotic patients 18 years of age with either at least one inpatient or at least two outpatient ICD-9-CM codes for HCV was selected (Table 1; outpatient codes based on physician face-to-face encounters), while a random sample of cirrhotic patients with HBV ICD-9-CM codes (Table 1) was chosen similarly. A sample size of 150 HBV patients was used given the smaller pool of patients with HBV cirrhosis, and estimates that 150 patients yielded >90% confidence to determine a positive predictive value (PPV) with a 95% CI of $\pm 10\%$ assuming a PPV of 80%. ICD-9-CM codes were required to occur within 365 days of the index date of cirrhosis diagnosis (either the date of the first inpatient ICD-9-CM code or second outpatient ICD-9-CM code).

Confirmation of HCV

HCV infection was confirmed by laboratory values or physician documentation. HCV was confirmed if any one of the following was present: a) detectable HCV RNA using available data from the PDS; b) detectable HCV RNA based on a report from an outside laboratory result not captured in the PDS; c) detectable HCV RNA documented in a physician note; or d) HCV physician documentation of HCV infection or sustained virologic response, using individual medical records (inpatient discharge summaries, inpatient progress notes, outpatient encounter notes). HCV medications alone were not used as a criterion given the poor accuracy of coding such medications, especially if treated by an outside provider.

Confirmation of HBV

Chronic HBV was confirmed if either of these was present in Penn Data Store data: a) detectable HBV surface antigen or b) detectable HBV DNA in a patient with cirrhosis without another identifiable etiology of cirrhosis. Among cirrhotic patients with negative laboratory data or no available laboratory data, HBV positivity was defined based on medical record review if any of the following was present: 1) physician documentation of HBV based on a detectable HBV DNA; 2) HBV therapy with lamivudine, entecavir, tenofovir, adefovir, interferon, and/or telbivudine with physician documentation of prior detectable HBV DNA, or 3) physician documentation of successful clearance of HBV with anti-viral therapy in the setting of HBV cirrhosis.

Data analysis

We determined the PPVs of our algorithm to identify clinically confirmed HCV or HBV infection, because if this parameter is sufficiently high, we can have confidence that the algorithm will identify future patient samples with high probability of having HBV or HCV and cirrhosis.

All data were analyzed using Stata 13.0 (Stata Corp, College Station, TX, USA).

Results

HCV

Among the 200 randomly selected cirrhotic subjects meeting the HCV coding algorithm, 191 (95.5%) had medical records available for review. Three-quarters of the patients were male, 118 (61.8%) were white and 60 (31.4%) were black, with a median age of 55.4 years (IQR: 52.0–61.7).

Of these 191, 168 had documented HCV (PPV: 88.0%, 95% CI: 82.5–92.2%). These 168 true positives were documented based on: a) detectable quantitative or qualitative HCV RNA identified during the PDS screen (n=91, 54.2%); b) detectable quantitative or qualitative HCV RNA based on reports from an outside laboratory or UPHS laboratory not detected through the PDS screen (n=27, 16.1%); c) inpatient notes or discharge summaries (n=23, 13.7%); d) clinician note documenting detectable HCV RNA without an available laboratory report in the record (n=21, 12.5%); e) documentation of prior treatment with SVR (n=4, 2.4%); or f) documentation of HCV positivity in clinical note without an RNA value (n=2, 1.2%). Of note, the 91 patients with detectable quantitative or qualitative HCV RNA represented 92.9% (n=98) of the total population with available HCV PCR test results in the electronic medical record.

Several variations in the coding algorithm increased the PPV to identify HCV at the expense of decreasing sensitivity (Table 1).

HBV

Among the 150 randomly selected cirrhotic subjects meeting the HBV ICD-9-CM coding algorithm, 139 (92.7%) had available electronic laboratory data or medical records available for review (Table 1). The median age patients meeting the HBV ICD-9-CM algorithm was 56.1 (IQR: 46.9–62.1), with 110 (79.3%) males. There was a broad racial/ethnic distribution: 49 (35.3%) patients were black, 42 (30.2%) white, and 32 (23.0%) were Asian.

Of the 139 cirrhotic subjects with available records, 113 were confirmed to have chronic HBV based on laboratory and/or medication data (PPV: 81.3%, 95% CI: 73.8–87.4%). Among the 26 patients negative for HBV by the above laboratory data, 20 had chronic HCV and were isolated HBV core IgG positive. When restricting the algorithm to at least 1 inpatient and 1 outpatient ICD-9-CM code for HBV, the PPV was 100.0 % (95% CI: 91.8–100.0%), although the number of detected cases was lowered (Table 1).

Discussion

The study demonstrates the ability of ICD-9-CM codes to accurately identify subjects with HCV or HBV among patients with diagnostic codes for cirrhosis. In this administrative dataset, among identified subjects with cirrhosis, 88% who had either at least one inpatient or at least two outpatient ICD-9-CM codes for HCV in fact had the disorder. Furthermore, variations on this algorithm can yield a PPV of >95%, albeit at the expense of sensitivity. ICD-9-CM-based algorithms for cirrhotic patients coded for HBV also performed well, with PPVs of greater than 80%. Results derived from this database have been shown to perform

similarly in other administrative databases,^{9,10} thus we believe this algorithm may perform similarly well in other claims-based databases, although external validation is needed.

Although HCV ICD-9-CM codes have been validated in the VA healthcare system, this is the first study to validate HCV ICD-9-CM codes in a cohort of subjects with diagnostic codes for cirrhosis outside of the VA and includes not only ICD-9-CM codes for HCV, but HBV as well. Furthermore, the UPHS cohort is demographically diverse, and is enriched with a population with Medicaid, Medicare, and commercial insurance. However, given our algorithms have not been externally validated, they may not be completely generalizable to other databases. Interestingly, 20 out of 26 cases that were incorrectly coded as HBV were in fact patients with chronic HCV and positive only for HBV core IgG. These were likely patients tested for HBV in light of their diagnosis of HCV.

This work is important for future epidemiologic research on cirrhosis and hepatitis, the diagnosis and treatment of which are projected to rise considerably in the next decade.^{11,12} Prior large-scale studies have used ICD-9-CM codes to determine study populations with HBV or HCV without such validation.⁶⁻⁸ As increasingly convenient treatment modalities become available for the treatment of HCV, determination of disease burden continues to rely on accurate assessment in administrative databases.¹³ Large population studies with valid hepatitis diagnoses are necessary especially at this point in time as they determine proper allocation of resources.

Nevertheless, this study has limitations. The UPHS has a tertiary care center offering liver transplantation, with a large referral base from Pennsylvania, New Jersey, and Delaware. It is expected that there would be a large cohort of patients with cirrhosis at this site, which could introduce spectrum bias. However, the PPV we found is similar to the VA population.¹⁴ Furthermore, our algorithms are based on HBV and HCV codes in patients with diagnostic codes for cirrhosis only, and may not apply to non-cirrhotic patients. We could not assess the performance of ICD-10 codes, as the UPHS database currently uses only ICD-9 codes. However, all US administrative databases prior to October 1, 2015 will still use ICD-9 codes.

In conclusion, a coding algorithm including at least one inpatient or two outpatient ICD-9-CM codes for HBV or HCV respectively in those with cirrhosis had a high PPV for identifying patients with one of these forms of viral hepatitis. This algorithm can be used in future epidemiological studies to identify these patients to monitor long-term outcomes and healthcare utilization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support

1. NIH grant 1K08DK098272-01A1 (DG)
2. NIH grant K23DK090209 (KF)

References

1. [Accessed February 26, 2014] FASTSTATS - Chronic Liver Disease or Cirrhosis. Available at: <http://www.cdc.gov/nchs/fastats/liverdis.htm>.
2. Myers RP, Tandon P, Ney M, et al. Validation of the five-variable Model for End-stage Liver Disease (5vMELD) for prediction of mortality on the liver transplant waiting list. *Liver Int.* 2013
3. OPTN: Organ Procurement and Transplantation Network. [Accessed February 26, 2014] Available at: <http://optn.transplant.hrsa.gov/latestData/rptData.asp>.
4. Hepatitis C Task Force, Boldt MD, Brill JV, et al. Hepatitis C screening: summary of recommendations from the clinical decision tool. *Gastroenterology.* 2013; 145(5):1146–1149. [PubMed: 24139703]
5. Byrne DD, Newcomb CW, Carbonari DM, et al. Prevalence of diagnosed chronic hepatitis B infection among U.S. Medicaid enrollees, 2000–2007. *Annals of Epidemiology.* 2014
6. Gordon SC, Hamzeh FM, Pockros PJ, et al. Hepatitis C virus therapy is associated with lower health care costs not only in noncirrhotic patients but also in patients with end-stage liver disease. *Aliment Pharmacol Ther.* 2013; 38(7):784–793. [PubMed: 23981040]
7. Gordon SC, Pockros PJ, Terrault NA, et al. Impact of disease severity on healthcare costs in patients with chronic hepatitis C (CHC) virus infection. *Hepatology.* 2012; 56(5):1651–1660. [PubMed: 22610658]
8. Lo Re V 3rd, Volk J, Newcomb CW, et al. Risk of hip fracture associated with hepatitis C virus infection and hepatitis C/human immunodeficiency virus coinfection. *Hepatology.* 2012; 56(5): 1688–1698. [PubMed: 22619086]
9. Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V 3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. *Pharmacoepidemiol Drug Saf.* 2012; 21(7):765–769. [PubMed: 22674685]
10. Nehra MS, Ma Y, Clark C, Amarasingham R, Rockey DC, Singal AG. Use of administrative claims data for identifying patients with cirrhosis. *J Clin Gastroenterol.* 2013; 47(5):e50–e54. [PubMed: 23090041]
11. Razavi H, Elkhoury AC, Elbasha E, et al. Chronic hepatitis C virus (HCV) disease burden and cost in the United States. *Hepatology.* 2013; 57(6):2164–2170. [PubMed: 23280550]
12. Zalesak M, Francis K, Gedeon A, et al. Current and future disease progression of the chronic HCV population in the United States. *PLoS ONE.* 2013; 8(5):e63959. [PubMed: 23704962]
13. Lawitz E, Mangia A, Wyles D, et al. Sofosbuvir for previously untreated chronic hepatitis C infection. *N Engl J Med.* 2013; 368(20):1878–1887. [PubMed: 23607594]
14. Kramer JR, Davila JA, Miller ED, Richardson P, Giordano TP, El-Serag HB. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther.* 2008; 27(3):274–282. [PubMed: 17996017]

Key points

1. ICD-9-CM administrative data can identify chronic hepatitis B or hepatitis C virus in cirrhotic patients with a high positive predictive value (PPV), and can be used in future epidemiologic studies to examine disease burden and the proper allocation of resources.
2. PPV of 1 inpatient or 2 outpatient hepatitis C virus codes was 88.0% (168/191, 95% CI: 82.5–92.2%).
3. PPV of 1 inpatient or 2 outpatient hepatitis B virus codes was 81.3% (113/139, 95% CI: 73.8–87.4%).
4. Variations of the primary coding algorithm could increase the PPV at the expense of decreasing sensitivity by identifying fewer subjects.

Table 1

Results of coding algorithms for HCV and HBV for patients with cirrhosis

Algorithm coding requirements	No. inpatient HCV ICD-9-CM code(s)	No. outpatient HCV ICD-9-CM code(s)	No. meeting algorithm	No. with confirmed HCV	PPV (95% CI):
Hepatitis C					
Inpatient <u>or</u> outpatient codes for HCV*					
Algorithm 1 [†]	1	2	191	168	88.0 (82.5–92.2)
Algorithm 2	2	2	141	134	95.0 (90.0–98.0)
Inpatient <u>and</u> outpatient codes					
Algorithm 3	1	1	65	63	96.9 (89.3–99.7)
Inpatient codes <u>only</u>					
Algorithm 4 [‡]	1	0	89	72	80.9 (71.2–88.5)
Algorithm 5	2	0	39	38	97.4 (86.6–99.9)
Outpatient codes <u>only</u>					
Algorithm 6 [‡]	0	2	37	33	89.2 (74.6–97.0)
Algorithm 7	0	3	28	24	85.7 (67.3–95.9)
Hepatitis B					
Inpatient <u>or</u> outpatient codes for HBV**					
Algorithm 1 ^{††}	1	2	139	113 [‡]	81.3 (73.8–87.4)
Algorithm 2	2	2	107	89	83.2 (74.7–89.7)
Inpatient <u>and</u> outpatient codes					
Algorithm 3	1	1	43	43	100.0 (91.8–100.0)

Abbreviations: HCV=HCV virus; HBV=hepatitis B virus; ICD-9-CM=International Classification of Diseases, Ninth Revision, Clinical Modification; PPV=positive predictive value

* ICD-9-CM code for HCV within 365 days of index date of electronic diagnosis of cirrhosis (1 inpatient or 2 outpatient ICD-9-CM codes for cirrhosis: ICD-9-CM-571.2, 571.5). ICD-9-CM codes for HCV: 070.41, 070.44, 070.51, 070.54, 070.7, 070.70, 070.71.

[†] Combination of 1 inpatient or 2 outpatient codes was primary analytic method to identify patients with HCV.

[‡]Number of patients meeting algorithms 4 and 6 for HCV do not add up to equal Algorithm 1 because Algorithm 4 requires 1 inpatient code in the absence of any outpatient codes, while Algorithm 6 requires 2 outpatient codes in the absence of any inpatient codes.

^{**} ICD-9-CM code for HBV within 365 days of index date of electronic diagnosis of cirrhosis (1 inpatient or 2 outpatient ICD-9-CM codes for cirrhosis: ICD-9-CM-571.2, 571.5), ICD-9-CM codes for HBV: 070.2, 070.20, 070.21, 070.22, 070.23, 070.30, 070.31, 070.32, 070.33.

^{††} Combination of 1 inpatient or 2 outpatient codes was primary analytic method to identify patients with HBV.