Published in final edited form as:
Pharmacoepidemiol Drug Saf. 2012 July ; 21(7): 765–769. doi:10.1002/pds.3290.

Validation of a coding algorithm to identify patients with end-stage liver disease in an administrative database

D Goldberg, MD, MSCE[1,2], **JD Lewis, MD, MSCE**[1,2,3], **SD Halpern, MD, PhD**[2,3,4], **Mark Weiner, MD**[2,3,5], and **Vincent Lo Re III, MD, MSCE**[2,6]

[1]Department of Medicine, Division of Gastroenterology

[2]Clinical Center for Biostatistics and Epidemiology, University of Pennsylvania

[3]Leonard Davis Institute of Health Economics, University of Pennsylvania

[4]Department of Medicine, Division of Pulmonary, Allergy, and Critical Care

[5]Department of Medicine, Division of General Internal Medicine

[6]Department of Medicine, Division of Infectious Diseases

Abstract

Purpose—Use of administrative or population-based databases for post-marketing pharmacoepidemiology research in patients with end-stage liver disease (ESLD) has been limited by the difficulty of accurately identifying such patients. Algorithms to identify patients with ESLD using ICD-9-CM codes have not been developed outside of the Veterans Affairs healthcare setting.

Methods—We queried electronic medical records at two tertiary care hospitals to identify patients with ICD-9-CM codes indicative of ESLD. Coding algorithms were developed to identify patients with confirmed ESLD, and these were tested to determine their positive predictive value (PPV).

Results—The presence of one inpatient or outpatient ICD-9-CM code for: a) cirrhosis, b) chronic liver disease, and c) a hepatic decompensation event yielded a PPV of 85.2% (167/196; 95% CI: 79.4%–89.9%). The PPV increased to 89.3% (150/168; 95% CI: 83.6%–93.5%) when the algorithm required 2 or more ICD-9-CM codes for a hepatic decompensation. However, an algorithm requiring only one ICD-9-CM code for a) cirrhosis and b) a hepatic decompensation event, in the absence of a chronic liver disease code, yielded a PPV of 85.7% (30/35; 95% CI: 69.7%–95.2%).

Conclusions—A coding algorithm that includes at least one ICD-9-CM code for cirrhosis plus one ICD-9-CM code for a hepatic decompensation event has a high PPV for identifying patients with ESLD. The inclusion of at least 2 codes indicative of chronic liver disease increased the PPV.

Corresponding Author: David Goldberg, Hospital of the University of Pennsylvania, 3400 Spruce Street, 9 Penn Tower, Philadelphia, PA 19104, Phone: 646-242-6349, Fax: 215-349-5915, david.goldberg@uphs.upenn.edu.

Disclosures

The authors of this manuscript have no conflicts of interest to disclose as described by Pharmacoepidemiology and Drug Safety. There have been no prior postings or presentations of this research.

This algorithm can be used in future epidemiologic studies to examine the outcomes of a variety of long-term medical therapies in patients with ESLD.

Keywords

Validation; hepatic decompensation; cirrhosis; ICD-9-CM code; end-stage liver disease

Introduction

Most medications are not studied in patients with end-stage liver disease (ESLD) prior to marketing. Use of administrative or population-based databases for post-marketing pharmacoepidemiology research in patients with ESLD has been limited by the ability to accurately identify such patients. While International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes have been validated for hepatic decompensation events, viral hepatitis, and cirrhosis in the Veterans Affairs (VA) administrative databases, (1, 2) algorithms to identify ESLD have not been developed and validated in other settings.

The ability to identify patients with ESLD would allow for a better understanding of patterns of medication use among these patients and permit evaluations of the safety of marketed therapies in this high-risk population. We therefore determined the ability of diagnostic codes to identify patients with ESLD in two tertiary care hospitals.

Methods

Study design and data source

We conducted a cross-sectional study among patients cared for at two hospitals in the University of Pennsylvania Health System (UPHS): the Hospital of the University of Pennsylvania and Penn Presbyterian Medical Center. The UPHS maintains the Pennsylvania Integrated Clinical and Administrative Research Database (PICARD), a warehouse of patient data that includes ICD-9-CM codes, laboratory test results, and ambulatory electronic health records for patients receiving care within the UPHS. The study was approved by the Institutional Review Board of the University of Pennsylvania.

Study Subjects and Coding Algorithm Derivation

We queried the PICARD database to identify a random sample of 300 adult patients ≥18 years of age with ICD-9-CM codes potentially indicative of ESLD, including chronic liver disease, cirrhosis, hepatic decompensation, or conditions associated with ESLD (e.g., jaundice, coagulopathy), recorded between January 1, 1997 and December 31, 2011 (see list of ICD-9-CM codes in Table 1). We included diagnosis codes for bleeding and non-bleeding esophageal varices in the ESLD code list because patients with non-bleeding varices represent a cohort of patients with significant portal hypertension and an increased risk of liver-related morality.(3, 4) Patients may have received inpatient or outpatient care at one or both of the participating hospitals.

The initial query was overly inclusive to allow for testing of multiple potential coding algorithms to best identify patients with ESLD, with the goal to identify an algorithm(s) with a positive predictive value (PPV) of at least 80%.

Primary Outcomes

The main outcome of this study was clinically confirmed ESLD (also referred to as decompensated liver disease). ESLD was defined by the presence of a definitive diagnosis of cirrhosis and a definitive diagnosis of a hepatic decompensation event (ascites, hepatorenal syndrome [HRS], spontaneous bacterial peritonitis [SBP], hepatic encephalopathy [HE), or gastric/esophageal variceal bleeding) in the medical record.

Cirrhosis was confirmed if a patient had a: a) liver biopsy demonstrating cirrhosis, b) radiographic imaging study (abdominal ultrasound, CT scan, or MRI) reporting cirrhosis, or c) physician's note documenting cirrhosis based on one of the two prior criteria.(5)

Definitions for hepatic decompensation events were based on guidelines published by the American Association for the Study of Liver Diseases (AASLD).(6) Confirmation of an event required physical exam (ascites, encephalopathy), imaging (ascites), or endoscopic (variceal bleeding) findings confirming a decompensation event, or documentation of an event by a treating physician within the medical record. Patients were required to have had a decompensation event within 365 days of receiving an ICD-9-CM code for that event.

Medical records were reviewed by a hepatologist (D.G.) to confirm outcomes. Chart review also determined if patients would be potential liver transplant candidates based on AASLD criteria.(7)

Data Analysis

We determined the PPVs of three ICD-9-CM-based coding algorithms to identify clinically confirmed ESLD (Table 1). Our focus was on PPV because if this parameter is sufficiently high, we and other researchers will have confidence that the algorithm identified ESLD with minimal misclassification. Algorithm 1 required a diagnosis of a hepatic decompensation event plus a chronic liver disease diagnosis. Algorithm 2 required a diagnosis of a hepatic decompensation event plus a cirrhosis diagnosis. Algorithm 3 required a diagnosis of a hepatic decompensation event, chronic liver disease, and cirrhosis. The rationale for these algorithms was based on observations that some patients with ESLD may only be coded for a chronic liver disease and not cirrhosis, or vice versa. The large sample size of algorithm 3 permitted determination of a variety of PPVs using different cut-off points for the minimum number of hepatic decompensation codes. Finally, to estimate the likelihood of missing ESLD events with the specified algorithms, we evaluated for the presence of ESLD in a sample of patients who did not meet the algorithms but who had non-specific diagnoses that typically accompany ESLD (e.g. coagulopathy, jaundice, hyponatremia, and portal hypertension).

All data were analyzed using State 12.0 (Stata Corp, College Station, TX, USA).

Results

Of the 266 patients with at least one ICD-9-CM code for a hepatic decompensation event, 244 (91.8%) had medical records available for review. Table 1 reports the PPVs of the coding algorithms.

Algorithm 1 yielded a PPV of only 7.7% (1/13), as only 1 patient had cirrhosis. Algorithm 2 had a PPV of 85.7% (30/35; 95% CI: 69.7%–95.2%). Algorithm 3 yielded a PPV of 85.2% (167/196; 95% CI: 79.4%–89.9%). For this algorithm, as the minimum number of required ICD-9-CM codes for a hepatic decompensation event increased, the PPV to identify patients with ESLD increased—from 85.2% for only 1 ICD-9-CM code to 95.7% (88/92) when at least 8 were required. However, increasing the minimum number of ICD-9-CM codes resulted in detecting 79 (47.3%) fewer cases of ESLD.

In addition to the high PPV to identify patients with ESLD, 80.0% and 83.2% of patients identified with algorithms 2 and 3, respectively, met criteria for listing for liver transplantation.

Finally, among 34 patients with an ICD-9-CM code for coagulopathy, jaundice, hyponatremia, and/or portal hypertension, in the absence of an ICD-9-CM code for a discrete hepatic decompensation event, only 3/34 (8.8%) were confirmed to have had ESLD.

Discussion

This study examined the ability of ICD-9-CM diagnostic codes to identify patients with ESLD. The presence of one inpatient or outpatient ICD-9-CM code for cirrhosis and one inpatient or outpatient ICD-9-CM code for a hepatic decompensation event (ascites, SBP, HRS, HE, or variceal bleeding), with or without an additional ICD-9-CM code for a chronic liver disease, had a PPV of 85% in identifying patients with confirmed ESLD. The PPV increased to greater than 90% when the algorithm required 3 or more ICD-9-CM codes for a diagnosis of a hepatic decompensation event, but this requirement decreased the number of cases identified.

Administrative claims data is an important data source for large-scale epidemiologic studies. However, the potential for misclassification exists by relying only on billing codes without proper validation. Within the field of hepatology, there have been few validations of ICD-9-CM codes, and they have focused on coding within the VA system. Given the different billing structure of the VA healthcare system, such validations are not generalizable to the non-VA setting. This validation is the first to identify patients with ELSD based on ICD-9-CM codes in the non-VA setting.

This work is important for future epidemiologic research. Many medications (e.g. antibiotics for SBP prophylaxis, beta-blockers for prevention of variceal hemorrhage, and statins in patients with concomitant dyslipidemias) are prescribed in this group of patients without large-scale, long-term follow-up data detailing the potential risks and side effects of this high-risk population of patients. Given the potential adverse hepatic consequences in these patients (8–12) for whom the pharmacokinetics and pharmacodynamics are not fully

understood, coinciding with the rising prevalence of ESLD, detailing the potential side effects of long-term medication therapy in this group is critical. However, prior work has been hampered by the inability to accurately identify this cohort of patients. We have now demonstrated that such studies are possible.

Our study has limitations. First, since the Hospital of the University of Pennsylvania is a tertiary care center and offers liver transplantation with a referral base from across Pennsylvania, Delaware, and New Jersey, it is expected that there would be a large cohort of patients with ESLD cared for at this site, which could introduce spectrum bias. However, only approximately 50% of the patients meeting transplant criteria were actually listed for transplantation. Second, it is possible that the coding algorithm we evaluated might miss ESLD events. However, given that fewer than 1% of patients have ESLD,(2) the negative predictive value of this algorithm is expected to be extremely high when utilized in a population-based database. In addition, we evaluated ESLD events in a limited sample of patients who did not meet the algorithm but who had non-specific diagnoses that typically accompany ESLD (e.g. coagulopathy, jaundice, and hyponatremia) and identified very few events.

In conclusion, a coding algorithm that includes at least one ICD-9-CM code for cirrhosis plus one ICD-9-CM code for a hepatic decompensation event, with or without concomitant codes indicative of chronic liver disease, had a high PPV for identifying patients with ESLD. This algorithm can be used in future epidemiologic studies to examine the outcomes of a variety of long-term medical therapies in patients with ESLD.

Acknowledgments

We thank Dr. Daniel Mines for his help in designing the coding algorithms to be used for this manuscript.

Financial Support

1. NIH/NIDDK F32 1-F32-DK-089694-01 Grant (DG)
2. NIH/NIAID K01 AI-070001 (VLR)
3. NIH K24-DK078228 (JL)

References

1. 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):274–82. Epub 2007/11/13. [PubMed: 17996017]
2. Lo Re V 3rd, Lim JK, Goetz MB, Tate J, Bathulapalli H, Klein MB, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. Pharmacoepidemiol Drug Saf. 2011; 20(7):689–99. Epub 2011/06/01. [PubMed: 21626605]
3. Arvaniti V, D'Amico G, Fede G, Manousou P, Tsochatzis E, Pleguezuelo M, et al. Infections in patients with cirrhosis increase mortality four-fold and should be used in determining prognosis. Gastroenterology. 2010; 139(4):1246–56. 56 e1–5. Epub 2010/06/19. [PubMed: 20558165]
4. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: In search of a pathophysiological classification of cirrhosis. Hepatology. 2010; 51(4): 1445–9. Epub 2010/01/16. [PubMed: 20077563]

5. Kudo M, Zheng RQ, Kim SR, Okabe Y, Osaki Y, Iijima H, et al. Diagnostic accuracy of imaging for liver cirrhosis compared to histologically proven liver cirrhosis. A multicenter collaborative study. Intervirology. 2008; 51(Suppl 1):17–26. Epub 2008/06/27. [PubMed: 18544944]

6. Mathurin P, Moreno C, Samuel D, Dumortier J, Salleron J, Durand F, et al. Early liver transplantation for severe alcoholic hepatitis. N Engl J Med. 2011; 365(19):1790–800. Epub 2011/11/11. [PubMed: 22070476]

7. Murray KF, Carithers RL Jr. AASLD practice guidelines: Evaluation of the patient for liver transplantation. Hepatology. 2005; 41(6):1407–32. Epub 2005/05/10. [PubMed: 15880505]

8. 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(10): 705–14. [PubMed: 16702586]

9. Goldstein ST, Zhou F, Hadler SC, Bell BP, Mast EE, Margolis HS. A mathematical model to estimate global hepatitis B disease burden and vaccination impact. Int J Epidemiol. 2005; 34(6): 1329–39. Epub 2005/10/27. [PubMed: 16249217]

10. Kim WR, Brown RS Jr, Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. Hepatology. 2002; 36(1):227–42. Epub 2002/06/27. [PubMed: 12085369]

11. Sorrell MF, Belongia EA, Costa J, Gareen IF, Grem JL, Inadomi JM, et al. National Institutes of Health Consensus Development Conference Statement: management of hepatitis B. Ann Intern Med. 2009; 150(2):104–10. Epub 2009/01/07. [PubMed: 19124811]

12. WHO. Hepatitis C—global prevalence (update). Wkly Epidemiol Rec. 2000; 75:18–9. [PubMed: 10686829]

Take-home messages

1. Post-marketing pharmacoepidemiology research in patients with ESLD has been limited by the difficulty of accurately identifying such patients.
2. ICD-9-CM codes for coagulopathy, hyponatremia, and jaundice do not accurately identify patients with end-stage liver disease
3. An ICD-9-CM code for cirrhosis, in the absence of a code for a chronic liver disease, has a high positive predictive value for identifying cirrhotic patients in an administrative database
4. A coding algorithm combining ICD-9-CM codes for cirrhosis and hepatic decompensation events, with or without ICD-9-CM codes for a chronic liver disease, can identify patients with end-stage liver disease
5. Future pharmacoepidemiology research on patients with end-stage liver disease can be conducted using this coding algorithm

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2014 December 16.

Table 1

Algorithm Number	Chronic liver Disease ICD-9-CM code*	Cirrhosis ICD-9-CM code†	Hepatic decompensation ICD-9-CM code‡	N	PPV for Cirrhosis, % (N)	PPV for hepatic decompensation, % (N)	PPV for ESLD, % (N)	PPV for orthotopic liver transplant % (N)**
Algorithm 1	+	−	+	13	7.7 (1)	69.2 (9)	7.7 (1)	7.7 (1)
Algorithm 2	−	+	+	35	94.3 (33)	91.4 (32)	85.7 (30)	80.0 (28)
Algorithm 3	+	+	≥1	196	93.9 (184)	89.8 (176)	85.2 (167)	83.2 (163)
	+	+	≥2	168	93.5 (157)	95.2 (160)	89.3 (150)	85.7 (144)
	+	+	≥3	151	92.7 (140)	97.4 (147)	90.8 (137)	86.1 (130)
	+	+	≥4	134	93.2 (125)	98.5 (132)	92.5 (124)	88.8 (119)
	+	+	≥6	114	93.8 (107)	100.0 (114)	94.7 (108)	90.4 (103)
	+	+	≥8	92	94.6 (87)	100.0 (92)	95.7 (88)	91.3 (84)
	+	+	≥10	72	93.1 (67)	100.0 (72)	94.4 (68)	89.9 (64)

*ICD-9-CM codes: 070.20, 070.21, 020.22, 070.23, 070.3, 070.31, 070.32, 070.33 (hepatitis B); 070.40, 070.49, 070.59, 070.60, 070.70, 070.71, 070.90, 573.1 (viral hepatitis NOS); 070.41, 070.44, 070.51, 070.54, 070.70, 070.71 (hepatitis C); 070.42, 070.52 (hepatitis D with hepatitis B); 273.4 (alpha-1-antitrypsin deficiency); 275.0 (disorders iron metabolism); 275.1 (disorders copper metabolism); 453.0 (Budd-Chiari syndrome); 571.0, 571.1, 571.1 (alcoholic liver disease); 571.40, 571.41, 571.8, 571.9, 573.0, 573.3, 573.8, 573.9 (hepatitis, liver disorders NOS); 517.42 (autoimmune hepatitis); 576.1 (cholangitis). + = ICD-9-CM code present; − = ICD-9-CM code absent

†ICD-9-CM codes: 571.2 (alcoholic cirrhosis), 571.5 (cirrhosis without mention of alcohol, 571.6 (biliary cirrhosis). + = ICD-9-CM code present; − = ICD-9-CM code absent

‡ICD-9-CM codes: 456.0, 456.20 (esophageal varices with bleeding); 456.1, 456.21 (esophageal varices without bleeding); 789.5, 789.59 (ascites); 572.2 (hepatic coma)l 567.0, 567.2, 567.21, 567.21, 567.29, 567.8, 567.89, 567.9 (peritonitis); 572.4 (hepatorenal syndrome). + = ICD-9-CM code present; − = ICD-9-CM code absent

**Based on AASLD criteria for listing a patient for liver transplantation (7)