Published in final edited form as:

J Clin Gastroenterol. 2013 ; 47(5): e50–e54. doi:10.1097/MCG.0b013e3182688d2f.

Use of Administrative Claims Data for Identifying Patients with Cirrhosis

Mahendra S. Nehra, MD MPH¹, Ying Ma, PhD², Christopher Clark, BS², Ruben Amarasingham, MD MBA^{2,3}, Don C. Rockey, MD¹, and Amit G. Singal, MD MS^{1,3}

¹Department of Internal Medicine, UT Southwestern Medical Center and Parkland Health and Hospital System, Dallas, TX

²Center of Clinical Innovation, Parkland Health and Hospital System, Dallas, TX

³Department of Clinical Sciences, University of Texas Southwestern, Dallas, TX

Abstract

Background—Administrative data are used in clinical research, but the validity of ICD-9 codes to identify cirrhotic patients has not been well established.

Goals—Determine the diagnostic accuracy of ICD-9 codes for cirrhosis in clinical practice

Study—We conducted a retrospective cohort study of patients from a safety-net hospital between 2008-2011. Patients were initially identified using ICD-9 codes for cirrhosis or a resultant complication. The gold-standard diagnosis of cirrhosis was histology and/or imaging based on medical record review. Sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) for each ICD-9 code were calculated. Diagnostic accuracy was assessed by the c-statistic using receiver operator characteristic curve analysis.

Results—We identified 2,893 patients with an ICD-9 code for cirrhosis, of whom 50.2% had one ICD-9 code, 20.3% had two different codes, and 29.5% had three or more codes. Cirrhosis was confirmed in 44.0% of patients with one ICD-9 code, 82.6% with two codes, and 95.7% of those with at least three codes. Ascites had a significantly lower PPV for cirrhosis than other ICD-9 codes ($p < 0.001$). The optimal combination of ICD-9 codes to identify cirrhotic patients included all codes except that of ascites, with a c-statistic of 0.71 in our derivation cohort. The sensitivity of this combination was confirmed to be 98% in a validation cohort of 285 patients with known cirrhosis.

Correspondence: Amit G. Singal, M.D., M.S. Dedman Scholar of Clinical Care Division of Gastroenterology University of Texas Southwestern 5959 Harry Hines Blvd, POB 1, Suite 420 Dallas TX 75390-8887 Tel: 214-645-6029 Fax: 214-645-6294 amit.singal@utsouthwestern.edu.

Conflicts of Interest: The authors have no other financial disclosures. No conflicts of interest exist.

Author Contributions: Mahendra Nehra - acquisition of data, analysis and interpretation of the data, and drafting of the manuscript. Ying Ma - acquisition of data and critical revision for important intellectual content. Christopher Clark - acquisition of data and critical revision for important intellectual content.

Ruben Amarasingham - acquisition of data and critical revision for important intellectual content.

Don Rockey - acquisition of data and/or analysis and interpretation of the data and critical revision for important intellectual content.

Amit Singal - conception and design of the study, acquisition of data, analysis and interpretation of the data, drafting of the manuscript, critical revision for important intellectual content, obtained funding, and study supervision.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—Administrative data can identify patients with cirrhosis with high accuracy, although ascites has a significantly lower positive predictive value than other ICD-9 codes.

Keywords

ICD-9 codes; accuracy; cirrhosis

BACKGROUND

Cirrhosis affects more than 5.5 million patients and costs nearly \$4 billion annually ¹. Further, the burden of cirrhosis is anticipated to increase over the next 10 years, beyond the 1 million hospitalizations per year currently attributed to liver disease ^{2, 3}. Given that cirrhosis is a common condition with high morbidity and high resource utilization, it is the subject of a growing number of outcome studies ^{1, 2}.

Accurate Identification of patients with cirrhosis from administrative data is an area of particular importance. Administrative databases, which are traditionally used for reimbursement, budgetary planning, and monitoring of clinical activities, are valuable sources of information for health services and outcomes research. These databases contain demographic, pharmacy, and diagnostic data using codes, such as the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9 CM). However, use of administrative data is subject to inherent limitations, including incompleteness and variable accuracy of coding.

Validation of administrative data is best accomplished by comparing elements of administrative data with review of medical records. Validation studies in other disease states have yielded mixed results. These variable results have been related to several factors, including the specific codes used for case identification, the source of the data, and whether the data was taken from inpatient or outpatient databases.⁴⁻⁶ However, the number of studies in the area of cirrhosis is limited and the accuracy of ICD-9 CM codes for the diagnosis of cirrhosis has not been extensively studied ^{7, 8}. We hypothesized that a carefully selected panel of ICD-9 codes would be able to predict cirrhosis with a high degree of accuracy. Therefore, the aim of our study was to identify and validate a set of ICD-9 codes for cirrhosis among a large cohort of patients.

METHODS

Study Settings

This retrospective cohort study was conducted at Parkland Memorial Health and Hospital System (PHHS), the safety-net hospital system for Dallas County and teaching hospital of UT Southwestern Medical Center. With eleven primary care clinics in low-income neighborhoods, Parkland serves a large population of patients with cirrhosis in Dallas County. This study was approved by the Institutional Review Board of the University of Texas Southwestern Medical Center.

Data Sources

We tested the validity of ICD-9 codes in an administrative database from Parkland Health and Hospital System, as described above. Parkland is one of the few safety-net hospitals in the country with an integrated electronic medical record for both the hospital and outpatient clinics, including primary care clinics.

Parkland hospital uses ICD-9 coding to record inpatient and outpatient diagnoses. The codes for both inpatients and outpatients are entered by professional coders, who base them on

findings in the clinicians' progress notes and laboratory test results of the electronic medical record. We estimated the validity of select ICD-9 codes by comparing these codes with diagnoses derived from a comprehensive review of the patients' medical records. PHHS uses an integrated electronic medical record system, EPIC, which is available in the hospital as well as outpatient clinics. It is a full electronic medical record system, containing a problem list, pharmacy data, orders, laboratory results, progress notes, vital signs, radiology results, transcribed documents and reports from various studies such as endoscopy. EPIC was implemented for inpatient documentation starting in April 2009 and outpatient documentation starting in September 2009. All inpatient and outpatient medical records are available in this single electronic medical record for patients treated at PHHS.

Derivation Study Population

We identified 3,277 patients at PHHS with at least one ICD-9 code for cirrhosis or one of its resultant complications (Table 1). We did not include the ICD-9 code for biliary cirrhosis (571.6) as this code was often associated with early stage PBC or PSC instead of cirrhosis. We included inpatients from January 2008 to December 2010 and outpatients from January 2010 to July 2011. The gold standard for cirrhosis, based on medical chart review, was defined as either stage 4 cirrhosis on liver biopsy or a cirrhotic appearing liver on abdominal imaging (nodular, shrunken liver on ultrasound, computerized tomography or magnetic resonance imaging) in combination with evidence of portal hypertension (i.e., varices, ascites, splenomegaly). Laboratory data alone were not used to define cirrhosis. Patients with a cirrhotic appearing liver without evidence of portal hypertension were excluded given the non-specific nature of superficial liver nodularity in isolation^{9, 10}. One trained clinician (AS), blinded to the ICD-9 codes, abstracted the medical record using a standardized abstraction form to determine the presence or absence of cirrhosis.

Validation Study Populations

To validate the sensitivity of the ICD-9 codes studied here, we examined a cohort of 285 patients who had been prospectively determined to have cirrhosis¹¹. These patients were an independent convenience sample of those admitted to an Internal Medicine service at PHHS between January 2008 and December 2010. Cirrhosis had been diagnosed by independent, prospective review of each patient's medical record at the time of admission. Cirrhosis was verified by: a) the presence of histological stage 4 fibrosis or b) a cirrhotic appearing liver in combination with signs of portal hypertension on abdominal imaging, clinical evidence of portal hypertension on admission (e.g. ascites, esophageal or gastric varices) or a complication of cirrhosis (e.g. hepatic encephalopathy or hepatocellular cancer).

To validate the negative predictive value of the ICD-9 codes, we identified 116 patients with liver disease but without any ICD-9 codes for cirrhosis. Patients with liver disease were identified using ICD-9 codes for hepatitis C, hepatitis B, and alcohol abuse. Medical charts were reviewed to determine if patients had underlying cirrhosis, as described above.

Statistical Analysis

We calculated the sensitivity, specificity, positive predictive value, and negative predictive value of each ICD-9 code, individually as well as in combination. Sensitivity indicates the probability of an ICD-9 code being present in those with cirrhosis, whereas specificity indicates the probability that an ICD-9 code would be absent in those without cirrhosis. Sensitivity and specificity are not affected by disease prevalence. Positive predictive value indicates the probability that a patient with the ICD-9 code actually has cirrhosis, whereas negative predictive value indicates the probability that a patient without a code actually does not have cirrhosis.

Sensitivity= # true positives/ (# true positives + # false negatives)

Specificity= # true negatives/ (# true negatives + # false positives)

Positive predictive value= # true positives/ (# true positives + # false positives)

Negative predictive value= # true negatives/ (# true negatives+ # false negatives)

The overall diagnostic accuracy of the ICD-9 codes, indicating the degree of correct classification for patients having or not having cirrhosis, was assessed by the c-statistic using ROC analysis¹². The c-statistic ranges from 0 to 1, with 1 indicating perfect prediction and 0.5 indicating prediction due to chance alone. A c-statistic between 0.7 and 0.8 is generally considered acceptable. The c-statistics for ICD-9 code combinations were compared using the Delong method¹³. All statistical analysis was performed using SAS 9.2 (SAS Institute, Cary, NC, USA) and Stata 11.2 (College Station, TX).

RESULTS

Study Population

From January 2008 to July 2011 we identified 3,277 subjects with at least ICD-9 code for cirrhosis or one of its complications. We excluded 384 patients given inadequate data on chart review to accurately determine the presence or absence of cirrhosis, leaving 2893 patients eligible for further analysis. There were 1128 inpatients, 365 outpatients, and 1300 patients who were seen as both inpatients and outpatients over the study period. Patients had a median age of 54 years, and the majority (63%) was male. Our population was racially diverse, with 29% African American, 32% non-Hispanic Caucasians, and 35% Hispanic Caucasians. Over one-third of patients were uninsured, 22% had Medicaid, 17% had Medicare, and only 7% had private health insurance. The majority of patients (50.2%) only had 1 ICD-9 code, whereas 20.3% of patients had 2 different codes and 29.5% had 3 or more different codes. Cirrhosis was confirmed in 44.0% of patients with 1 ICD-9 code, 82.6% with 2 codes, and 95.7% of those with 3 or more codes.

Performance of Individual ICD-9 Codes

Table 2 highlights the diagnostic test characteristics of each individual ICD-9 code. Most ICD-9 codes had sufficient specificity except cirrhosis without alcohol and ascites, which both had significantly lower specificities ($p < .001$ for both comparisons). Similarly, each of the codes had a high positive predictive value above 75%, except ascites, which had a significantly lower positive predictive value ($p < 0.001$). The performance characteristics for individual ICD-9 codes did not significantly differ between inpatients and outpatients. However, one notable difference is a higher sensitivity of non-alcoholic cirrhosis (571.5) among outpatients than inpatient for detecting patients with cirrhosis (86% vs. 71%, $p < 0.001$).

Performance of ICD-9 Code Combinations

Table 3 shows performance characteristics of ICD-9 code combinations. The optimal combination of ICD-9 codes included all codes except for that of ascites (789.59). This set of ICD-9 codes was able to identify 98% of patients with cirrhosis in our cohort, with a positive predictive value of 78% and negative predictive value of 91%. This combination of ICD-9 codes had a c-statistic of 0.71 (Figure 1), which was significantly higher than the combination with all codes ($p < 0.001$). Although the combination without non-alcohol cirrhosis (571.5) and ascites had a higher c-statistic (0.72 vs. 0.71), this difference was not statistically significant ($p = 0.36$). Our set of ICD-9 codes was able to identify over 98% of both inpatients and outpatients, and was able to identify cirrhotic patients with fair accuracy in both populations (C-statistics of 0.72 and 0.68 respectively, $p = 0.02$). The addition of

ascites significantly reduced the accuracy of identifying patients with cirrhosis in both subgroups of patients ($p < 0.001$ for both comparisons).

Independent Validation of Negative Predictive Value of ICD-9 Codes

Of the 116 patients with liver disease but without ICD-9 codes for cirrhosis, 93 had one code for hepatitis C, hepatitis B or alcohol abuse and 23 had more than one of these ICD-9 codes. Forty-two patients had underlying hepatitis C, and 13 patients had hepatitis B infection. The negative predictive value of the ICD-9 codes in this population was confirmed to be sufficiently high at 94.0%.

Independent Validation of Sensitivity of ICD-9 Codes

To further validate the sensitivity of the identified ICD-9 codes, we examined our set of ICD-9 codes in a set of 285 patients with cirrhosis that was prospectively identified by tracking all patients admitted to PHHS over a two-year period; our set of ICD-9 codes identified 280 (98.2%) of these patients. Only two patients in this sample had an ICD-9 code for ascites, without any co-existing ICD-9 codes for cirrhosis or other cirrhotic complications.

DISCUSSION

To the best of our knowledge, this is the largest study to validate ICD-9 codes for cirrhosis. We have demonstrated that most individual ICD-9 codes, except for that of ascites, have a high positive predictive value for identifying cirrhosis. A combination of codes is required to maximize the sensitivity for identifying cirrhosis and can identify over 95% of patients accurately. Our set of ICD-9 codes, including those for cirrhosis and its resultant complications except for that of ascites, demonstrated a positive predictive value of 78%, a negative predictive value of 91% for cirrhosis, and a c-statistic of 0.71.

A study of Veteran Administration (VA) data with 331 patients reported a positive predictive value of 90% and a negative predictive value of 87% when using any one of three ICD-9 cirrhosis codes (571.2, 571.5, and 571.6)⁷. However, this study did not report the value of including additional codes for cirrhosis complications. Furthermore, prior studies have demonstrated the accuracy of ICD-9 codes vary between VA and non-VA settings so external validation is crucial. A study conducted at a tertiary care non-VA institution also recently reported a high positive predictive value of 88% when using one of several ICD-9 codes for cirrhosis or one of its resultant complications, but this study was limited by a small sample size of only 100 patients⁸.

In contrast to a previous report⁷, our study suggests that using only codes for cirrhosis (571.2 and 571.5) is not likely to identify patients with cirrhosis with a high degree of sensitivity. In this previous study, very few patients (0.3%) had a cirrhosis complication code without a code for cirrhosis. In our study, nearly 5% of patients with cirrhosis only had ICD-9 codes for a complication of cirrhosis, without a corresponding code for cirrhosis. Our study suggests that including ICD-9 codes for complications of cirrhosis can increase the sensitivity of codes for identifying patients with cirrhosis and are of additional benefit.

Results from our study are particularly important for researchers interested in using administrative databases to examine questions regarding patients with cirrhosis. Results of this study will be useful for clinical researchers who are interested in cirrhosis research using ICD-9 codes but are cautious of using single ICD-9 code for cirrhosis diagnosis due to their poor diagnostic accuracy. Due to findings of poor sensitivity and accuracy of majority of single ICD-9 code in cirrhosis diagnosing, use of single ICD-9 code is discouraged. Our

ICD-9 code combination can be used to predict most of the cirrhosis cases without a need for confirmation by review of patient charts with fair accuracy.

We recognize limitations of this study. For example, our study was conducted in a single safety-net hospital system and our results may not be generalized to coding practices at other practice settings. However, EPIC is a widely available EMR that is being used by several tertiary care academic and safety net institutions nationwide. A second limitation is that the diagnosis of cirrhosis was based on histology and/or imaging, which are both imperfect gold standards. Liver biopsy is prone to sampling error and abdominal imaging is known to be insensitive for diagnosing cirrhosis¹⁴⁻¹⁶. Furthermore, chart validation may be subject to misclassification bias in case ascertainment. However, we minimized this bias by using an experienced hepatologist to conduct primary review of the charts, and we used a standardized, search strategy to abstract the presence or absence of cirrhosis from each chart. Finally, patients in our primary cohort were identified by ICD-9 codes for cirrhosis or one of its complications, which might bias the reported sensitivity upward. However, we assessed for this bias by validating the sensitivity of these codes in an independent cohort. Overall, we believe the limitations of our study are outweighed by its notable strengths including its relatively large sample size, its racially diverse population, and its inclusion of both inpatients and outpatients.

In summary, this study demonstrates that the diagnostic codes for cirrhosis and its complications, except for that of ascites, have a relatively high positive predictive value and overall accuracy, as demonstrated in this single large healthcare system of Dallas County. However, ICD-9 codes for cirrhosis alone may be insufficiently sensitive in isolation and should likely be combined with ICD-9 codes for cirrhosis complications. These ICD-9 codes can be reliably used in both VA and non-VA administrative databases for research for accurately identifying patients with cirrhosis. Similar studies will need to be performed to assess the diagnostic accuracy of ICD-10 codes, which will be implemented soon.

Acknowledgments

Grant Support: This work was conducted with support from UT-STAR, NIH/NCATS Grant Number KL2 RR024983-05 and the ACG Junior Faculty Development Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of UT-STAR, the University of Texas Southwestern Medical Center at Dallas and its affiliated academic and health care centers, the National Center for Advancing Translational Sciences, or the National Institutes of Health.

REFERENCES

1. Kanwal F, Gralnek IM, Hays RD, et al. Health-related quality of life predicts mortality in patients with advanced chronic liver disease. *Clin Gastroenterol Hepatol*. Jul; 2009 7(7):793–799. [PubMed: 19306949]
2. Davis GL, Albright JE, Cook SF, Rosenberg DM. Projecting future complications of chronic hepatitis C in the United States. *Liver Transpl*. Apr; 2003 9(4):331–338. [PubMed: 12682882]
3. Kanwal F, Hoang T, Kramer JR, et al. Increasing prevalence of HCC and cirrhosis in patients with chronic hepatitis C virus infection. *Gastroenterology*. Apr; 2011 140(4):1182–1188. e1181. [PubMed: 21184757]
4. Benesch C, Witter DM Jr, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology*. Sep; 1997 49(3):660–664. [PubMed: 9305319]
5. Marcus P, Braman SS. International classification of disease coding for obstructive lung disease: does it reflect appropriate clinical documentation? *Chest*. Jul; 2010 138(1):188–192. [PubMed: 20605817]

6. Vance GA, Niederhauser A, Chauhan SP, et al. Does the International Classification of Disease (ICD-9) code accurately identify neonates who clinically have hypoxic-ischemic encephalopathy? *Gynecol Obstet Invest.* 2011; 71(3):202–206. [PubMed: 21160147]
7. 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.* Feb 1; 2008 27(3):274–282. [PubMed: 17996017]
8. Rakoski MO, McCammon RJ, Piette JD, et al. Burden of cirrhosis on older Americans and their families: Analysis of the health and retirement study. *Hepatology.* Jan; 2012 55(1):184–191. [PubMed: 21858847]
9. Filly RA, Reddy SG, Nalbandian AB, Lu Y, Callen PW. Sonographic evaluation of liver nodularity: Inspection of deep versus superficial surfaces of the liver. *J Clin Ultrasound.* Sep; 2002 30(7):399–407. [PubMed: 12210457]
10. Vigano M, Visentin S, Aghemo A, Rumi MG, Ronchi G. US features of liver surface nodularity as a predictor of severe fibrosis in chronic hepatitis C. *Radiology.* Feb.2005 234(2):641. author reply 641. [PubMed: 15671013]
11. Patel B, Rockey D. The use of opioids in cirrhotics. *Hepatology.* 2010; 52(S1):1252A.
12. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation.* Feb 6; 2007 115(5):654–657. [PubMed: 17283280]
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* Sep; 1988 44(3): 837–845. [PubMed: 3203132]
14. Colli A, Fraquelli M, Andreoletti M, Marino B, Zuccoli E, Conte D. Severe liver fibrosis or cirrhosis: accuracy of US for detection--analysis of 300 cases. *Radiology.* Apr; 2003 227(1):89–94. [PubMed: 12601199]
15. Poynard T, Munteanu M, Imbert-Bismut F, et al. Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. *Clin Chem.* Aug; 2004 50(8):1344–1355. [PubMed: 15192028]
16. Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol.* Oct; 2002 97(10):2614–2618. [PubMed: 12385448]

Figure 1. Receiver operating characteristic curve for a combination of ICD-9 codes to identify cirrhosis

Table 1

ICD-9 Codes for Cirrhosis

ICD-9 Code	Meaning
456.0	Esophageal varices with bleeding
456.1	Esophageal varices without bleeding
456.2	Varices in diseases classified elsewhere with bleeding
456.21	Varices in diseases classified elsewhere without bleeding
567.23	Spontaneous bacterial peritonitis (SBP)
571.2	Alcoholic cirrhosis
571.5	Cirrhosis without alcohol
572.2	Hepatic encephalopathy
572.3	Portal hypertension
572.4	Hepatorenal syndrome
789.59	Ascites

ICD – International Classification of Diseases

Table 2

Performance Characteristics of Individual ICD-9 Codes (n=2893)

ICD-9 Code	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	C-statistic
456.0	2.5	99.6	92.3	32.8	0.51
456.1	17.8	96.2	90.6	36.5	0.57
456.2	0.6	99.7	78.6	33.0	0.50
456.21	7.7	99.1	94.3	34.5	0.53
567.23	2.7	98.9	84.1	33.3	0.51
571.2	52.3	84.0	87.0	46.3	0.68
571.5	72.1	67.7	79.8	52.5	0.67
572.2	15.6	97.1	91.5	36.1	0.56
572.3	31.0	94.3	91.8	40.1	0.67
572.4	1.8	99.7	92.1	33.2	0.51
789.59	43.7	48.3	63.3	29.6	0.46

ICD – International Classification of Diseases

Table 3

Performance Characteristics of ICD-9 Code Combinations (n=2893)

ICD-9 Code Combinations	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	C-statistic
572.4 and 456.2	2.4	99.4	88.5	33.3	0.51
Above ICD-9 codes plus 456.0	4.6	98.9	90.0	33.7	0.52
Above ICD-9 codes plus 456.21	11.2	98.3	93.2	35.2	0.55
Above ICD-9 codes plus 567.23	13.4	97.3	90.9	35.5	0.55
Above ICD-9 codes plus 572.2	25.2	94.6	90.6	38.3	0.60
Above ICD-9 codes plus 456.1	33.8	91.6	89.1	40.4	0.63
Above ICD-9 codes plus 572.3	49.4	87.2	88.7	45.8	0.68
Above ICD-9 codes plus 571.2	69.4	73.9	84.5	54.3	0.72
Above ICD-9 codes plus 571.5	97.9	43.0	78.0	91.2	0.71
Above ICD-9 codes plus 789.59	100.0	0.0	67.1	N/A	0.50

ICD – International Classification of Diseases