JGH Open

ORIGINAL ARTICLE OPEN ACCESS

Validation of ICD-10 Consensus Code Set for Cirrhosis Detection Using Electronic Health Records in an Asian Population

Jason Pik-Eu Chang[1,2] | Hong-Yi Lin[3] | Pooi-Ling Loi[1] | Jeanette Pei-Xuan Ng[1] | Marianne De Roza[2,4] | Rahul Kumar[2,5] | Hiang-Keat Tan[1,2] | Chanda Kendra Ho[1,2,6] | Wei-Quan Teo[6] | Amber Hwa Hwa Chung[6] | Prema Raj[2,6]

[1]Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore, Singapore | [2]Duke-NUS Medical School, Singapore, Singapore | [3]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore | [4]Department of Gastroenterology and Hepatology, Sengkang General Hospital, Singapore, Singapore | [5]Department of Gastroenterology and Hepatology, Changi General Hospital, Singapore, Singapore | [6]SingHealth Duke-NUS Transplant Centre, Singapore, Singapore

Correspondence: Jason Pik-Eu Chang (jason.chang@singhealth.com.sg)

Received: 10 October 2024 | **Accepted:** 1 April 2025

Funding: The authors received no specific funding for this work.

Keywords: Asian | cirrhosis | electronic health records | ICD-10 | validation

ABSTRACT

Background: Systematic identification of patients with cirrhosis through electronic healthcare records (EHRs) using ICD-10 codes is essential for epidemiological research but is prone to discrepancies. We aim to validate and improve a recent consensus code set of nine ICD-10 codes to identify cirrhosis in a multi-ethnic Asian population.

Methods: We applied an initial broad algorithm of 25 ICD-10 codes related to cirrhosis and its complications to identify patients potentially with cirrhosis admitted to Singapore General Hospital in 2018 and confirmed true cirrhosis cases via manual EHR review. We evaluated the consensus code set's sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) in identifying cirrhosis cases. We examined alternative code sets to improve cirrhosis identification and validated them in another local hospital.

Results: One thousand, seven hundred thirty-three patients potentially with cirrhosis were identified, with 937 (54.1%) confirmed. The median age at diagnosis was 71 years (IQR: 64–78), with 65.6% males, 75.2%/8.8%/9.3%/6.7% Chinese/Indians/Malays/Others, and 56.7% Child-Pugh A. The main etiologies were chronic hepatitis B (29.5%) and metabolic dysfunction–associated steatotic liver disease (25.5%). The consensus code set demonstrated sensitivity/specificity/PPV/NPV of 76.1%/82.0%/83.3%/74.5%, respectively. We identified a set of 10 ICD-10 codes (SingHealth Chronic Liver Disease Registry [SoLiDaRity]-10) with sensitivity/specificity/PPV/NPV of 76.5%/84.8%/85.6%/75.4%, respectively, demonstrating an improved specificity versus the consensus code set ($p = 0.001$). External validation in another local hospital with 578 patients potentially with cirrhosis demonstrated improved sensitivity of the SoLiDaRity-10 code set versus the consensus code set ($p = 0.033$) (sensitivity/specificity/PPV/NPV: 78.0%/93.6%/94.1%/76.4% vs. 76.2%/93.6%/94.0%/75.0%, respectively).

Conclusions: While the consensus code set performs well in identifying patients with cirrhosis in a multi-ethnic Asian population, we propose the improved SoLiDaRity-10 code set.

Abbreviations: EHR, electronic health records; HCC, hepatocellular carcinoma; ICD, International Classification of Diseases; IQR, interquartile range; MASLD, metabolic dysfunction–associated steatotic liver disease; NPV, negative predictive value; PPV, positive predictive value; SGH, Singapore General Hospital.

1 | Introduction

Liver cirrhosis is the terminal stage of chronic liver disease. It is the leading cause of liver-related death globally, accounting for nearly 4% of total deaths worldwide in 2023 [1]. Furthermore, complications from decompensated cirrhosis result in high rates of 90-day and 1-year morbidity at 13.2% and 26.9%, respectively [2], representing major global challenges in healthcare utilization. Accurate identification of patients with cirrhosis can facilitate the longitudinal study of the disease, allowing timely monitoring of disease progression, audit of outcomes, and ultimately better care.

Electronic health records (EHR) are increasingly adopted worldwide to facilitate disease identification, data storage, and accurate retrieval of patient information [3]. The International Classification of Diseases (ICD) coding system is the international standard used for coding patient diseases, with its 10th Revision (ICD-10) released in 2015 [4]. It is widely used to identify patients with specific diagnoses from the EHR for administrative purposes, billing, research, and patient care [5]. ICD-10 codes have been reported to be useful in identifying cirrhosis and cirrhosis-related complications [6], but their widespread use is limited by variations in coding practices and coding inaccuracies. This is due to the large number of disease codes related to liver cirrhosis and its complications, which results in great variability when individual ICD-10 codes are used to define the presence of cirrhosis [7, 8]. As such, well-defined combinations of related ICD-10 codes are required to reliably identify patients with cirrhosis.

Several studies in North America and Europe have explored the optimal ICD-10 code sets for identifying cirrhosis patients, summarized by Shearer et al. [9] in a recent review. Analyzing 18 studies with 1626 records, a set of nine ICD-10 codes was selected to identify cirrhosis. This "consensus code set" was proposed as an international standard to unify EHR-based cirrhosis studies, fostering collaboration and enabling meaningful comparisons across populations. In external validation studies, the consensus code set had a sensitivity of 61% in a United Kingdom (UK) cohort of 300 patients and 100% in a United States (US) cohort of 113 patients [9]. Its positive predictive value (PPV) was 83% in a UK cohort of 335 patients and 89% in a US cohort of 241 patients [9]. However, there are no validation studies for this code set in Asian populations. Given the variability in ICD-10 code set performance and differing cirrhosis patterns across populations [10, 11], validating and improving this algorithm is an important unmet need to improve collaborative research and patient care in Asian healthcare systems.

The primary aim of this study is to externally validate the performance of Shearer et al.'s consensus code set to identify cirrhosis patients in a diverse, multi-ethnic Asian population. The secondary aim is to improve case-finding of cirrhosis patients using alternative code sets relevant to the multi-ethnic Asian healthcare context.

2 | Methods

2.1 | Study Design

We conducted a retrospective study on the EHR database of patients admitted to Singapore General Hospital (SGH) from January 1, 2018 to December 31, 2018. SGH is the largest tertiary care hospital in Singapore and is part of the SingHealth cluster, which forms the largest group of healthcare institutions comprising four public hospitals. An initial algorithm of 25 ICD-10 codes covering cirrhosis and its complications was applied to the EHR database to identify potential cirrhosis cases.

Subsequently, researchers manually reviewed each patient's records to confirm cirrhosis based on one or more of the following definitions: histological confirmation, radiological evidence of cirrhosis (coarse echotexture with irregular liver margins) on imaging (ultrasound, computed tomography, magnetic resonance imaging), evidence of portal hypertension on imaging (splenomegaly, varices, ascites) and/or endoscopy (esophageal and/or gastric varices), liver stiffness >13 kPa on elastography, or a clinical diagnosis of a decompensation event (ascites, bleeding varices, hepatic encephalopathy) [12]. Relevant demographic, clinical, and laboratory data of cirrhosis patients were retrieved using a standardized template and entered into a computerized database. Patients were excluded from the analysis if they did not fulfill the diagnosis of cirrhosis, were <18 years old, or were not admitted in the year 2018.

The etiologies of cirrhosis were determined using one or a combination of clinical history, laboratory results, and histology. Alcohol-associated liver disease was diagnosed if alcohol intake exceeded 21 units per week for men or 14 units for women. Metabolic dysfunction–associated steatotic liver disease (MASLD) was diagnosed with hepatic steatosis (radiologically or histologically) in the absence of significant alcohol use or other liver diseases, together with features of metabolic syndrome. Chronic hepatitis C and hepatitis B infections were determined from the presence of anti-hepatitis C antibody with positive hepatitis C viral RNA, and positive hepatitis B surface antigen for ≥ 6 months, respectively. Primary biliary cirrhosis was diagnosed by anti-mitochondrial antibodies or histology. Autoimmune hepatitis was diagnosed based on the detection of liver auto-antibodies and compatible histology. Primary sclerosing cholangitis was diagnosed based on cholestatic liver biochemistry, with characteristic cholangiography and histological findings. Other rarer causes (e.g., Wilson's disease, hemochromatosis, cardiac cirrhosis, etc.) [13] were identified by their respective diagnostic criteria and classified under "Others". The etiology was determined to be cryptogenic if no identifiable causes could be determined. For multiple etiologies of cirrhosis, the predominant etiology was determined by the treating physician. The severity of encephalopathy was graded according to the West Haven Criteria. Ascites were determined clinically and from imaging.

The performance of the consensus code set was evaluated by computing the sensitivity, specificity, PPV, and negative predictive value (NPV) to identify true cirrhosis among patients with potential diagnosis of cirrhosis. We also assessed individual ICD-10 codes and explored combinations that could enhance accuracy in identifying cirrhosis. This led to the development of a 10-code set, called the "SingHealth Chronic Liver Disease Registry-10 (SoLiDaRity-10) code set", and it was compared with Shearer's consensus code set. We then externally validated the SoLiDaRity-10 set in a separate cohort from Sengkang General Hospital (SKH), another general hospital in the SingHealth cluster.

2.2 | Statistical Analysis

We used software R version 4.3.3 for data analysis [14, 15]. Continuous variables were presented as medians with interquartile ranges (IQR), and categorical data as counts and percentages. Sensitivity is the proportion of true cirrhosis cases identified among all patients with true cirrhosis. Specificity is the proportion of true non-cirrhosis cases identified among all patients without cirrhosis. PPV is the proportion of true cirrhosis cases among those identified to have cirrhosis. NPV is the proportion of non-cirrhosis cases among those identified to have no cirrhosis. Results are presented with 95% confidence intervals.

We compared the diagnostic performance of the consensus code set and the SoLiDaRity-10 code set using McNemar's test, which is appropriate for paired binary data. The test evaluated differences in classification outcomes between the two algorithms, with statistical significance set at $p = 0.05$.

2.3 | Ethics

Approval was obtained from the hospital's centralized institutional review board (IRB number 2022/2239).

TABLE 1 | Initial broad 25-code ICD-10 code set which identified 1733 patients potentially with liver cirrhosis in Singapore General Hospital.

ICD-10 code	Description
C22.0	Liver cell carcinoma
E83.1	Disorders of iron metabolism
I85.0	Esophageal varices with bleeding
I85.9	Esophageal varices without bleeding
I86.4	Gastric varices
I98.2	Esophageal varices in diseases classified elsewhere without bleeding
I98.3	Esophageal varices in diseases classified elsewhere with bleeding
K65.0	Generalized peritonitis
K65.9	Peritonitis, unspecified
K70.3	Alcoholic cirrhosis of liver
K71.7	Toxic liver disease with fibrosis and cirrhosis of liver
K72.1	Chronic hepatic failure
K72.9	Hepatic failure, unspecified
K74.0	Hepatic fibrosis
K74.1	Hepatic sclerosis
K74.2	Hepatic fibrosis with hepatic sclerosis
K74.3	Primary biliary cirrhosis
K74.5	Biliary cirrhosis, unspecified
K74.6	Other unspecified cirrhosis of liver
K76.1	Chronic passive congestion of liver
K76.6	Portal hypertension
K76.7	Hepatorenal syndrome
P78.8	Other specified perinatal digestive system disorders
R17.0	Unspecified jaundice
R18.0	Ascites

TABLE 2 | Baseline characteristics of the study cohort ($n = 937$) with true liver cirrhosis at Singapore General Hospital.

Median age, years (IQR)		71.0 (64.0–78.0)
Males		615 (65.6)
Ethnicity	Chinese	705 (75.2)
	Indian	82 (8.8)
	Malay	87 (9.3)
	Others	63 (6.7)
Cirrhosis	Compensated	608/818 (74.3)
	Decompensated	210/818 (25.7)
Etiologies of cirrhosis	Chronic hepatitis B	276 (29.5)
	Metabolic dysfunction–associated steatotic liver disease	239 (25.5)
	Cryptogenic	160 (17.1)
	Alcohol	109 (11.6)
	Chronic hepatitis C	68 (7.3)
	Primary biliary cirrhosis	28 (3.0)
	Autoimmune hepatitis	15 (1.6)
	Primary sclerosing cholangitis	2 (0.2)
	Others	40 (4.3)
Ascites		222 (23.7)
Hepatocellular carcinoma		215 (24.9)
Child-pugh class	CP class A	523/922 (56.7)
	CP class B	316/922 (34.3)
	CP class C	83/922 (9.0)

Note: Data in table presented as number (percentage) unless otherwise specified.
Abbreviations: CP, child-pugh; IQR, interquartile range.

3 | Results

3.1 | Study Cohort

The initial broad algorithm of 25 ICD-10 codes (Table 1) identified 1733 patients with a potential diagnosis of liver cirrhosis who were admitted between January 1 and December 31, 2018 to SGH. Following the manual review of the EHRs, 937 (54.1%) cirrhosis cases were confirmed.

3.2 | Baseline Characteristics

The median age of the patients was 71.0 years (IQR: 64.0–78.0). There were 65.6% males. Ethnic distribution was 75.2% Chinese, 9.3% Malay, 8.8% Indian, and 6.7% others. At admission, 25.7% had decompensated cirrhosis. The main etiologies of cirrhosis were chronic hepatitis B infection (29.5%) and MASLD (25.5%). Significant complications present at diagnosis were ascites (23.7%) and hepatocellular carcinoma (24.9%). The distribution of the Child-Pugh score was 56.7% Class A, 34.3% Class B, and 9.0% Class C. (Table 2).

3.3 | External Validation of Consensus Code Set

The performance of Shearer's consensus code set for identifying patients with cirrhosis is presented in Table 3. Table 3 lists the nine individual ICD-10 codes used to identify cirrhosis cases. Patients with at least one of these nine ICD-10 codes were classified as having cirrhosis, while those without any of these codes were classified as not having cirrhosis.

These provided a sensitivity of 76.1% (95% CI: 73.2%–78.8%), specificity of 82.0% (95% CI: 79.2%–84.6%), PPV of 83.3% (95% CI: 80.6%–85.7%), and NPV of 74.4% (95% CI: 71.4%–77.3%) (Table 3).

3.4 | Design and Performance of the SoLiDaRity-10 Code Set

From the analysis of all the 25 ICD-10 codes, we observed that replacing "*K72.9 Hepatic failure, unspecified*" and "*K76.7 Hepatorenal syndrome*" from the consensus code set with "*K74.3 Primary biliary cirrhosis*" and "*K74.5 Biliary cirrhosis, unspecified*" improved the identification of patients with cirrhosis. In addition, we tested the performance of a variety of combinations of ICD-10 codes including "*R18.0 Ascites*" and "*C22.0 Liver cell carcinoma*" as well as "*I86.4 Gastric varices*" and "*C22.0 Liver cell carcinoma*". In summary, we identified a set of 10 ICD-10 codes (nine individual and one combination) that provided the most optimal performance in identifying cirrhosis patients in our study cohort (Table 4). We termed this the "SoLiDaRity-10 code set", which includes the replacement of two individual codes from the original consensus

TABLE 3 | Performance of Shearer et al.'s consensus code set to identify cirrhosis patients in Singapore General Hospital ($n = 1733$).

Shearer et al.'s consensus code set	
ICD-10 code	**Description**
I85.0	Esophageal varices with bleeding
I85.9	Esophageal varices without bleeding
I98.2	Esophageal varices in diseases classified elsewhere without bleeding
I98.3	Esophageal varices in diseases classified elsewhere with bleeding
K70.3	Alcoholic cirrhosis of the liver
K72.9	Hepatic failure, unspecified
K74.6	Other and unspecified cirrhosis of the liver
K76.6	Portal hypertension
K76.7	Hepatorenal syndrome

Performance of consensus code set in identifying patients with and without cirrhosis in analysis cohort			
	Patients' cirrhosis status		
	Cirrhosis	**No cirrhosis**	**Marginal number detected by algorithm**
Cirrhosis detected by algorithm	713	143	856
No cirrhosis detected by algorithm	224	653	877
Marginal number of patients in each status	937	796	Overall total: 1733

	Performance		**95% CI (%)**
Sensitivity	713/937	76.1%	73.2–78.8
Specificity	653/796	82.0%	79.2–84.6
Positive predictive value (PPV)	713/856	83.3%	80.6–85.7
Negative predictive value (NPV)	653/877	74.4%	71.4–77.3

Abbreviation: CI, confidence interval.

code set (*K74.3* and *K74.5* replacing *K72.9* and *K76.7*) and the introduction of a new combination code (*R18.0* and *C22.0*).

The SoLiDaRity-10 code set identified 838 patients with cirrhosis, of which 717 were patients with true cirrhosis among 937 patients with cirrhosis. Furthermore, the SoLiDaRity-10 code set identified 895 patients without cirrhosis, of which 675 patients were truly without cirrhosis among 796 patients without cirrhosis (Table 4). These provided a sensitivity of 76.5% (95% CI: 73.7%–79.2%), specificity of 84.8% (95% CI: 82.1%–87.2%), PPV of 85.6% (95% CI: 83.0%–87.9%), and NPV of 75.4% (95% CI: 72.5%–78.2%).

3.5 | Comparison of Performance Between the Consensus Code Set and SoLiDaRity-10 Code Set

McNemar's test was used to compare the performance of the consensus code set with the SoLiDaRity-10 code set in identifying patients with or without cirrhosis from the analysis cohort (Table 5). Notably, the specificity of the SoLiDaRity-10 code set was significantly higher than that of the consensus code set (84.8% vs. 82.0%, $p = 0.001$). The sensitivity, PPV, and NPV of the SoLiDaRity-10 code set were numerically higher than those of the consensus code set, but the differences were not statistically significant (Figure 1).

3.6 | External Validation of SoLiDaRity-10 Code Set

To address the inherent limitation of developing and validating the SoLiDaRity-10 code set from data collected from a single institution, we conducted a separate validation study using the same methodology on 578 patients potentially with cirrhosis from the Department of Gastroenterology and Hepatology of SKH, another tertiary hospital within the SingHealth cluster (Table S2). Based on the analysis of the SKH dataset, the SoLiDaRity-10 code set provided a sensitivity of 78.0%, specificity of 93.6%, PPV of 94.1%, and NPV of 76.4%. In comparison, the consensus code set provided a sensitivity of 76.2%, specificity of 93.6%, PPV of 94.0%, and NPV of 75.0% to identify the 328 patients with true cirrhosis in this SKH cohort. The SoLiDaRity-10 code set had improved sensitivity compared to the consensus code set ($p = 0.033$ by McNemar's test).

TABLE 4 | Performance of SoLiDaRity-10 code set to identify cirrhosis patients in Singapore General Hospital ($n = 1733$).

Solidary-10 code set	
ICD-10 code	**Description**
I85.0	Esophageal varices with bleeding
I85.9	Esophageal varices without bleeding
I98.2	Esophageal varices in diseases classified elsewhere without bleeding
I98.3	Esophageal varices in diseases classified elsewhere with bleeding
K70.3	Alcoholic cirrhosis of the liver
K74.3	Primary biliary cirrhosis[a]
K74.5	Biliary cirrhosis, unspecified[a]
K74.6	Other and unspecified cirrhosis of liver
K76.6	Portal hypertension
R18.0 & C22.0	Ascites & Liver cell carcinoma[b]

Performance of the SoLiDaRity10-code set in identifying patients with and without cirrhosis in the analysis cohort			
	Patients' cirrhosis status		
	Cirrhosis	**No cirrhosis**	**Marginal number detected by algorithm**
Cirrhosis detected by algorithm	717	121	838
No cirrhosis detected by algorithm	220	675	895
Marginal number of patients in each status	937	796	Overall total: 1733
	Performance		**95% CI (%)**
Sensitivity	717/937	76.5%	73.7–79.2
Specificity	675/796	84.8%	82.1–87.2
Positive predictive value (Ppv)	717/838	85.6%	83.0–87.9
Negative predictive value (NPV)	675/895	75.4%	72.5–78.2

Abbreviation: CI, confidence interval.
[a]ICD-10 codes that were replaced.
[b]ICD-10 code newly introduced in SoLiDaRity-10.

TABLE 5 | Paired observed frequencies of cirrhosis detection by the consensus code set and SoLiDaRity-10's code set in Singapore General Hospital.

Comparison of specificity: detection by two algorithms among patients without cirrhosis (796 patients)								
		SoLiDaRity-10						
		Cirrhosis detected	**Cirrhosis not detected**	**Total**		**Specificity**	***p***	**0.001**
Consensus code set	Cirrhosis detected	111	32	143	SoLiDaRity-10	675/796 = 84.8%		
	Cirrhosis not detected	10	643	653	Consensus code set	653/796 = 82.0%		
	Total	121	675	796				
Comparison of sensitivity: detection by two algorithms among patients with cirrhosis (937 patients)								
		SoLiDaRity-10						
		Cirrhosis detected	**Cirrhosis not detected**	**Total**		**Sensitivity**	***p***	**0.553**
Consensus code set	Cirrhosis detected	702	11	713	SoLiDaRity-10	717/937 = 76.5%		
	Cirrhosis not detected	15	209	224	Consensus code set	713/937 = 76.1%		
	Total	717	220	937				

Note: McNemar's test was used to compare the specificity and sensitivity of the code sets.

4 | Discussion

This is the first study to externally validate Shearer's consensus code set for identifying cirrhosis patients in a diverse multiethnic Asian population, which includes Chinese, Malays, Indians, Eurasians, and other ethnicities. Previous external validation studies reported PPVs ranging from 83% in a UK cohort to 89% in a US cohort [9]. Our study demonstrates that the consensus code set performs equally well in an Asian population, with a PPV of 83.3%. Despite variations in ethnicity and cirrhosis etiology, the consensus code set remains a reliable tool for identifying cirrhosis patients, supporting its use in international collaborative studies using EHRs.

Additionally, we found that modifying certain codes and adding a combination code for ascites and hepatocellular carcinoma (HCC) enhanced the performance of the consensus code set. This combination of 10 codes, the SoLiDaRity-10 code set, improved sensitivity from 76.1% to 76.5%, specificity from 82.0% to 84.8%, PPV from 83.3% to 85.6%, and NPV from 74.4% to 75.4%, compared to the consensus code set. Notably, the SoLiDaRity-10 code set showed significantly higher specificity, reducing the likelihood of incorrectly including non-cirrhosis patients in EHR-based databases. The improved PPV indicates better accuracy in identifying true cirrhosis cases, which could enhance case-finding for cirrhosis databases. This refined code set could help recruit more Asian cirrhosis patients for studies, improve patient audits, and ultimately lead to better patient care outcomes.

There are several reasons for the improved performance of the SoLiDaRity-10 code set to identify patients with cirrhosis in an Asian cohort. First, we found cases of cirrhosis linked to the codes "*K74.3 primary biliary cirrhosis*" (11 cirrhosis cases when used in combination, 3 when used alone) and "*K74.5 biliary cirrhosis, unspecified*" (4 cirrhosis cases when used in combination, 1 when used alone) that were missed by the consensus code set (Table S1). Interestingly, in the review by Shearer et al. [9], these codes were not observed in the review of US and UK cohorts, which led to their exclusion from the consensus code set. This exclusion represents a potential limitation, as it could fail to identify patients coded with "*K74.3 primary biliary cirrhosis*" and "*K74.5 biliary cirrhosis, unspecified*," especially in populations with a higher prevalence of primary biliary cirrhosis [16].

We also observed that in our study cohort the ICD-10 code "*K76.7 Hepatorenal syndrome*" identified very few cirrhosis cases (15 out of 1733 potential cirrhosis patients when part of a combination code) and no cases when used alone. (Table S1). Diagnosing hepatorenal syndrome in advanced liver disease is challenging, and it is often confounded by other causes of acute kidney injury, such as sepsis, dehydration, or medications which are common in patients with advanced liver disease and acute liver failure [17]. The poor performance of this code in identifying cirrhosis may reflect the underestimation of hepatorenal syndrome as a distinct diagnosis in our local hospital practice. However, all cirrhosis patients with hepatorenal syndrome were accurately captured under the "*K74.6 Other unspecified cirrhosis of liver*" code, demonstrating the SoLiDaRity-10 code set's overall accuracy in identifying cirrhosis patients in this population.

In addition, we observed that most of the patients who were diagnosed with "*K72.9 Hepatic failure, unspecified*" had acute liver

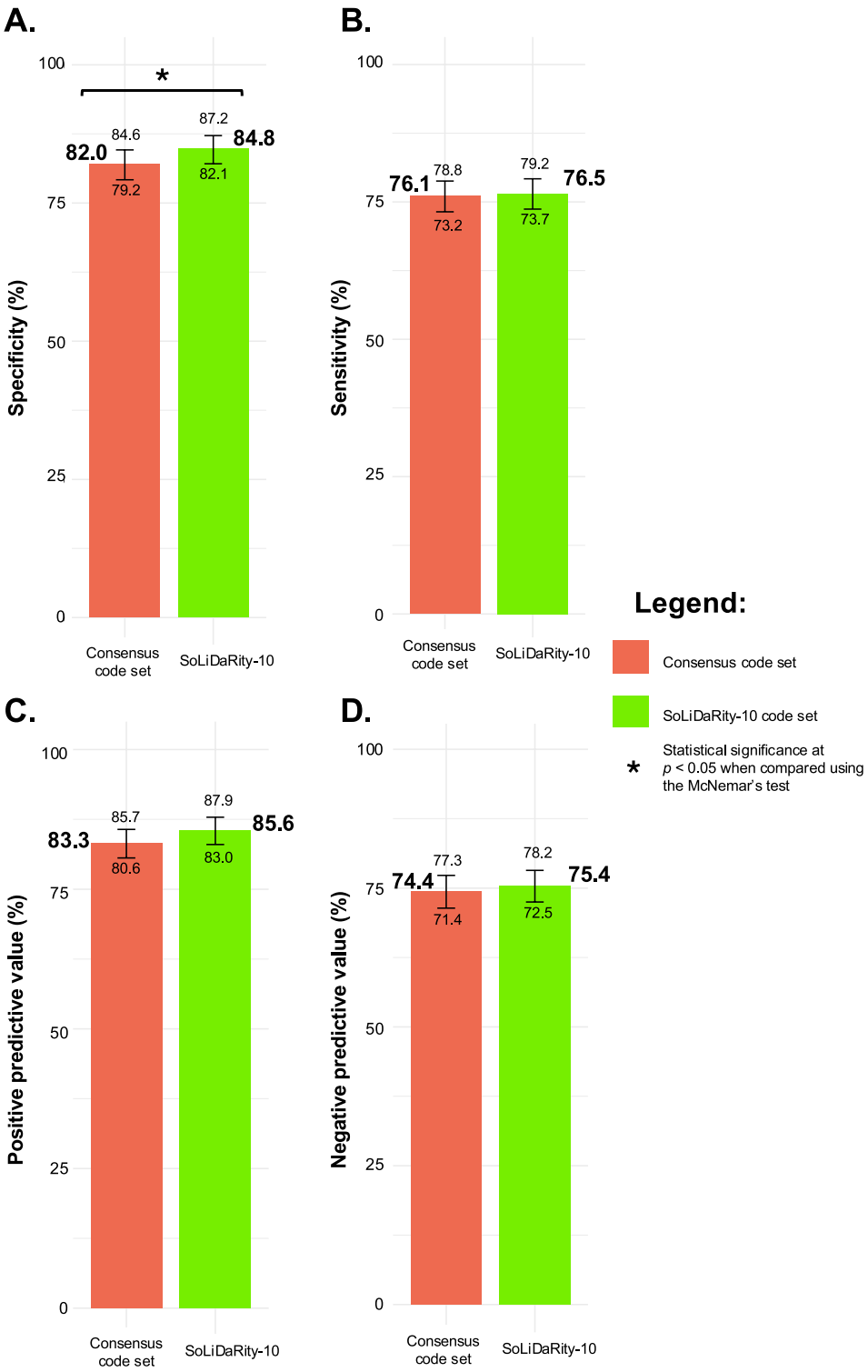

FIGURE 1 | Comparisons of the diagnostic performance of the consensus code set with the SoLiDaRity-10 code set in Singapore General Hospital.

failure but did not have liver cirrhosis (of 22 patients coded with K72.9, only 5 patients had cirrhosis) (Table S1). Most cases of liver failure in our cohort were due to acute-on-chronic hepatitis B flares in patients without evidence of underlying liver cirrhosis. Others were due to drug-induced liver injuries or autoimmune hepatitis, again without underlying liver cirrhosis. Hence the inclusion of the code for "*K72.9 Hepatic failure, unspecified*" in Shearer et al.'s consensus code set might lead to reduced specificity for the identification of cirrhosis as some of these patients may not have underlying liver cirrhosis.

Previous studies have established that the use of "*R18.0 Ascites*" as a standalone code is not useful for identifying cirrhosis, as ascites can result from conditions like cardiac or renal failure, or intra-abdominal malignancy [18, 19]. However, we observed that several patients who were diagnosed with ascites in combination

with HCC ("C22.0 Liver cell carcinoma") had underlying cirrhosis that would not have been otherwise identified as these patients were not diagnosed with cirrhosis during their initial presentation (79 patients with cirrhosis, out of 1733 patients potentially with cirrhosis) (Table S1). In the Asian population where chronic hepatitis B is endemic [20], many patients may unknowingly have chronic hepatitis B-related liver cirrhosis and only present to healthcare services when they experience decompensation symptoms, such as ascites [21]. Given the higher prevalence of HCC in the Asia-Pacific region as compared to the West [22], a significant proportion of patients present with HCC as their initial diagnosis. Hence, patients who present for the first time with ascites and are diagnosed on imaging to have HCC are often coded for ascites (*R18.0*) and HCC (*C22.0*) but not for liver cirrhosis, although most of these patients have underlying cirrhosis. In this setting, the use of the combination code of "*R18.0 Ascites"* and "*C22.0 Liver cell carcinoma"* improves the sensitivity and specificity of identifying patients with cirrhosis in our Asian population.

To address the limitation of single-center validation, we confirmed the reproducibility of our findings by performing a secondary validation using the same methodology on an independent cohort from a separate hospital, demonstrating the excellent performance of both the consensus code set and the SoLiDaRity-10 code set while confirming the improved performance of the latter. Nonetheless, the performance of the SoLiDaRity-10 code set requires validation in other external cohorts to confirm its reliability. Another limitation stems from focusing solely on ICD-10 codes from inpatient admissions, whereas other studies also included outpatient codes [23, 24]. Identifying cirrhosis patients from inpatient admissions proved more challenging due to the generation of multiple disease codes, which can create confusion regarding the primary diagnosis [24, 25]. This highlights the importance of a well-defined ICD-10 code set for accurately identifying cirrhosis patients, particularly in inpatient settings. For this reason, we believe that our findings are of merit to guide accurate case findings of patients with cirrhosis specifically in an inpatient setting.

4.1 | Conclusion

In conclusion, our study provides external validation of Shearer's consensus code set in a multi-ethnic Asian population and provides support for its use in international collaborative studies for accurate identification of cirrhosis patients from EHR systems based on ICD-10 codes. We propose that the SoLiDaRity-10 code set provides improved performance for case-finding of cirrhosis patients in the Asian population where rates of chronic hepatitis B and HCC are higher. Further multi-center studies comparing the performance of these two code sets would help to validate these findings in a broader Asian cohort.

Acknowledgments

The authors would like to acknowledge Andrea Teo, Kevin Ng, Chew Ming Yu, Nicky Wong, Choy Onn Leng, and Chong Wei Xuan from the Yong Loo Lin School of Medicine, National University of Singapore, who contributed to the data collection.

Disclosure

The authors have nothing to report.

Ethics Statement

Approval was obtained from the hospital's centralized institutional review board (IRB number: 2022/2239).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, Jason Pik-Eu Chang, upon reasonable request.

References

1. H. Devarbhavi, S. K. Asrani, J. P. Arab, Y. A. Nartey, E. Pose, and P. S. Kamath, "Global Burden of Liver Disease: 2023 Update," *Journal of Hepatology* 79, no. 2 (2023): 516–537, https://doi.org/10.1016/j.jhep.2023.03.017.

2. H.-Y. Lin, P. L. Loi, J. Ng, et al., "MELD3.0 Is Superior to MELDNa and MELD for Prediction of Mortality in Patients With Cirrhosis: An External Validation in a Multi-Ethnic Population," *JGH Open* 8, no. 6 (2024): e13098, https://doi.org/10.1002/jgh3.13098.

3. M. Senbekov, T. Saliev, Z. Bukeyeva, et al., "The Recent Progress and Applications of Digital Technologies in Healthcare: A Review," *International Journal of Telemedicine and Applications* 2020, no. 1 (2020): 8830200, https://doi.org/10.1155/2020/8830200.

4. J. A. Hirsch, G. Nicola, G. McGinty, et al., "ICD-10: History and Context," *AJNR. American Journal of Neuroradiology* 37, no. 4 (2016): 596–599, https://doi.org/10.3174/ajnr.A4696.

5. N. Gavrielov-Yusim and M. Friger, "Use of Administrative Medical Databases in Population-Based Research," *Journal of Epidemiology and Community Health* 68, no. 3 (2014): 283–287, https://doi.org/10.1136/jech-2013-202744.

6. S. Mapakshi, J. R. Kramer, P. Richardson, H. B. El-Serag, and F. Kanwal, "Positive Predictive Value of International Classification of Diseases, 10th Revision, Codes for Cirrhosis and Its Related Complications," *Clinical Gastroenterology and Hepatology* 16, no. 10 (2018): 1677–1678, https://doi.org/10.1016/j.cgh.2018.01.042.

7. A. Khalifa, J. S. Obeid, M. J. Gregoski, and D. C. Rockey, "Accurate Identification of Patients With Cirrhosis and Its Complications in the Electronic Health Record," *Digestive Diseases and Sciences* 68, no. 6 (2023): 2360–2369, https://doi.org/10.1007/s10620-023-07876-7.

8. N. S. Ramrakhiani, M. H. Le, Y. H. Yeo, A. K. Le, M. Maeda, and M. H. Nguyen, "Validity of International Classification of Diseases, Tenth Revision, Codes for Cirrhosis," *Digestive Diseases* 39, no. 3 (2021): 243–246, https://doi.org/10.1159/000510981.

9. J. E. Shearer, J. J. Gonzalez, T. Min, et al., "Systematic Review: Development of a Consensus Code Set to Identify Cirrhosis in Electronic Health Records," *Alimentary Pharmacology and Therapeutics* 55, no. 6 (2022): 645–657, https://doi.org/10.1111/apt.16806.

10. N. Jetté, H. Quan, B. Hemmelgarn, et al., "The Development, Evolution, and Modifications of ICD-10: Challenges to the International Comparability of Morbidity Data," *Medical Care* 48, no. 12 (2010): 1105–1110, https://doi.org/10.1097/MLR.0b013e3181ef9d3e.

11. D. Q. Huang, N. A. Terrault, F. Tacke, et al., "Global Epidemiology of Cirrhosis—Aetiology, Trends and Predictions," *Nature Reviews Gastroenterology and Hepatology* 20, no. 6 (2023): 388–398, https://doi.org/10.1038/s41575-023-00759-2.

12. H. Yoshiji, S. Nagoshi, T. Akahane, et al., "Evidence-Based Clinical Practice Guidelines for Liver Cirrhosis 2020," *Journal of Gastroenterology* 56, no. 7 (2021): 593–619, https://doi.org/10.1007/s00535-021-01788-x.

13. P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and P. S. Kamath, "Liver Cirrhosis," *Lancet* 398, no. 10308 (2021): 1359–1376, https://doi.org/10.1016/s0140-6736(21)01374-x.

14. R-Core-Team, "R: A Language and Environment for Statistical Computing," 2024.

15. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

16. J. Trivella, B. V. John, and C. Levy, "Primary Biliary Cholangitis: Epidemiology, Prognosis, and Treatment," *Hepatology Communications* 7, no. 6 (2023): e017, https://doi.org/10.1097/hc9.0000000000000179.

17. P. Ginès, E. Solà, P. Angeli, F. Wong, M. K. Nadim, and P. S. Kamath, "Hepatorenal Syndrome," *Nature Reviews Disease Primers* 4, no. 1 (2018): 23, https://doi.org/10.1038/s41572-018-0022-7.

18. M. S. Nehra, Y. Ma, C. Clark, R. Amarasingham, D. C. Rockey, and A. G. Singal, "Use of Administrative Claims Data for Identifying Patients With Cirrhosis," *Journal of Clinical Gastroenterology* 47, no. 5 (2013): 50–54, https://doi.org/10.1097/MCG.0b013e3182688d2f.

19. B. Bengtsson, J. Askling, J. F. Ludvigsson, and H. Hagström, "Validity of Administrative Codes Associated With Cirrhosis in Sweden," *Scandinavian Journal of Gastroenterology* 55, no. 10 (2020): 1205–1210, https://doi.org/10.1080/00365521.2020.1820566.

20. S. Wait, E. Kell, S. Hamid, et al., "Hepatitis B and Hepatitis C in Southeast and Southern Asia: Challenges for Governments," *Lancet Gastroenterology and Hepatology* 1, no. 3 (2016): 248–255, https://doi.org/10.1016/S2468-1253(16)30031-0.

21. M. Schwarz, C. Schwarz, L. Burghart, et al., "Late-Stage Presentation With Decompensated Cirrhosis Is Alarmingly Common but Successful Etiologic Therapy Allows for Favorable Clinical Outcomes," *PLoS One* 18, no. 8 (2023): e0290352, https://doi.org/10.1371/journal.pone.0290352.

22. K. A. McGlynn, J. L. Petrick, and H. B. El-Serag, "Epidemiology of Hepatocellular Carcinoma," *Hepatology* 73, no. Suppl 1 (2021): 4–13, https://doi.org/10.1002/hep.31288.

23. C. R. Cooke, M. J. Joo, S. M. Anderson, et al., "The Validity of Using ICD-9 Codes and Pharmacy Records to Identify Patients With Chronic Obstructive Pulmonary Disease," *BMC Health Services Research* 11, no. 1 (2011): 37, https://doi.org/10.1186/1472-6963-11-37.

24. J. Yoon and A. Chow, "Comparing Chronic Condition Rates Using ICD-9 and ICD-10 in VA Patients FY2014–2016," *BMC Health Services Research* 17, no. 1 (2017): 572, https://doi.org/10.1186/s12913-017-2504-9.

25. M. K. Szekendi, M. V. Williams, D. Carrier, L. Hensley, S. Thomas, and J. Cerese, "The Characteristics of Patients Frequently Admitted to Academic Medical Centers in the United States," *Journal of Hospital Medicine* 10, no. 9 (2015): 563–568, https://doi.org/10.1002/jhm.2375.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.