

# Validation of Algorithms Using ICD-10 Clinical Modification (CM) Codes in Claims Data to Identify People With Viral Hepatitis

**Ming-Jen Sheu**

Chi Mei Medical Center

**Tsung-Wei Chin**

National Cheng Kung University

**Fang-Ping Ku**

National Cheng Kung University

**Chung-Yi Li**

National Cheng Kung University

**Sen-Tung Li**

National Cheng Kung University

**Tsung-Hsueh Lu** (✉ [robertlu@mail.ncku.edu.tw](mailto:robertlu@mail.ncku.edu.tw))

National Cheng Kung University

---

## Research Article

**Keywords:** Hepatitis B virus, Hepatitis C virus, International Classification of Disease coding, Claims data, Algorithms, Validation study

**Posted Date:** October 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-917019/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The ICD-9 Clinical Modification (ICD-9-CM) coding system was transitioned into the ICD-10-CM on October 1, 2015, in the United States and on January 1, 2016, in Taiwan. Little is known on the performance of various algorithms using *ICD-10-CM* codes in claims data to identify people with hepatitis B virus (HBV) and hepatitis C virus (HCV).

**Methods:** A proportional systemic sampling of 10,000 patients aged  $\geq 20$  years in a health care system in Southern Taiwan were enrolled as study participants. According to the reference standards we confirmed 736 and 555 participants had HBV and HCV, respectively.

**Results:** The algorithms with higher number of outpatient (OP) visits with ICD codes had higher positive predictive value (PPV); for example, the PPV for HBV using algorithm 1 ( $\geq 1$  OP codes) was 72% and 86% according to ICD-9-CM and ICD-10-CM, respectively, and that of algorithm 3 ( $\geq 3$  OP codes) was 80% and 90%, respectively. Similarly, the PPV for HCV using algorithm 1 was 88% and 96%, respectively, and using algorithm 3 was 93% and 99%, respectively. However, the algorithms with higher PPV complemented with lower sensitivity.

**Conclusions:** In conclusion, algorithms using ICD-10-CM codes had better performance than those using ICD-9-CM codes in identifying people with HBV and HCV. Considering the tradeoff between PPV and sensitivity, the optimal algorithm is  $\geq 2$  OP visits or  $\geq 1$  inpatient visits with HBV or HCV ICD codes.

## Background

An increasing number of studies using real world data to assess the effectiveness and safety of anti-viral medication among people with viral hepatitis.<sup>1-5</sup> One critical step in using real world data is the use of the *International Classification of Diseases* (ICD) codes to identify people with a given disease, either as the main health outcome or as a covariate.<sup>6,7</sup> The checklist item 6.2 in REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) states that “any validation studies of the codes or algorithms used to select the population should be referenced.”<sup>8</sup> However, there were only five studies have examined the validity of algorithms using ICD codes to identify people with hepatitis B virus (HBV) or hepatitis C virus (HCV) infection.<sup>9-13</sup>

The ICD Ninth Revision Clinical Modification (ICD-9-CM) coding system was transitioned into the ICD Tenth Revision Clinical Modification (ICD-10-CM) on October 1, 2015, in the United States and on January 1, 2016, in Taiwan. Although some studies have examined the impact of this transition on changes in particular diseases,<sup>14-18</sup> no study has evaluated that in viral hepatitis. Therefore, this study aimed to compare the performance of algorithms using ICD-9-CM codes versus ICD-10-CM codes in identifying people with HBV and HCV.

## Methods

## Design and setting

This cross-sectional study was conducted at the Chi Mei health care system in Tainan, Taiwan. The Chi Mei health care system is the largest integrated health care system in southern Taiwan and includes one medical center (Yongkang), one regional hospital (Liouying), and one district hospital (Jiali). This study was approved by the Institutional Review Board of the Chi Mei Medical Center (number: 10901-015).

## Study participants

A proportional systemic sampling of 10,000 patients aged  $\geq 20$  years who had at least four visits to the Chi Mei health care system in 2015 were enrolled as study participants. We randomly sampled 6000 patients from medical center, 2500 patients from regional hospital, and 1500 patients from district hospital.

## Reference standard

The patients were defined as having HBV or HCV if they fulfilled the following criteria: 1) having a prescription of anti-HBV or HCV drugs reimbursed by the National Health Insurance (NHI); 2) positive results of HBsAg, HBeAg, or anti-HCV tests; or 3) clinical diagnosis of HBV or HCV in the discharge summary for inpatients or in problem list and past history for outpatients. A query system (Hyperion) was used to review all electronic medical records for each participant from 2005 through 2019.

## Coding algorithms

The ICD-9-CM and ICD-10-CM codes for HBV and HCV are listed in Table 1. We developed the following nine algorithms:

1.  $\geq 1$  outpatient codes
2.  $\geq 2$  outpatient codes
3.  $\geq 3$  outpatient codes
4.  $\geq 1$  inpatient codes
5.  $\geq 2$  inpatient codes
6.  $\geq 3$  inpatient codes
7.  $\geq 2$  outpatient or  $\geq 1$  inpatient codes
8.  $\geq 3$  outpatient or  $\geq 1$  inpatient codes
9.  $\geq 4$  outpatient or  $\geq 1$  inpatient codes

## Analysis

To assess the performance of ICD codes in identifying patients with HBV or HCV infection, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. We

compared the four validation indicators between ICD-9-CM and ICD-10-CM codes with respect to demographic characteristics.

## Results

Of the 10,000 patients enrolled in this study, 726 had confirmed HBV and 555 had confirmed HCV. Of the 726 and 555 confirmed diagnoses of HBV and HCV, 192 and 177, respectively, were based on drug prescriptions, 218 and 275, respectively, were based on laboratory test results, and 316 and 103, respectively, were based on textual clinical diagnosis (Fig. 1).

The demographic characteristics of all patients, with or without HBV/HCV, are summarized in Table 2. Three-fifths of patients with HBV and half of those with HCV were men. The HCV group had a higher proportion of older adults (aged  $\geq 70$  years) than the HBV group (51% versus 29%).

**Table 1** International Classification of Disease Clinical Modification Ninth and Tenth Revision (ICD-9-CM and ICD-10-CM) codes for hepatitis B and C virus

	Hepatitis B virus	Hepatitis C virus
ICD-9-CM codes	070.20, 070.21, 070.22, 070.23, 070.30, 070.31, 070.32, 070.33, V02.61	070.41, 070.44, 070.51, 070.54, 070.70, 070.71, V02.62
ICD-10-CM codes	B16.0, B16.1, B16.2, B16.9, B17.0, B18.0, B18.1, B19.1, Z22.51	B17.1, B18.2, B19.2, Z22.52

Table 2  
Characteristics of study participants

Characteristics	All participants (n=10,000)		HBV confirmed (n=726)		HCV confirmed (n=555)		Unconfirmed (n=8788)	
	No.	%	No.	%	No.	%	No.	%
Sex								
Men	4856	48.6	433	59.6	278	50.1	4186	47.6
Women	5144	51.4	293	40.4	277	49.9	4602	52.4
Age								
20-39 years	1184	11.8	45	6.2	9	1.6	1131	12.9
40-49 years	1209	12.1	113	15.6	34	6.1	1066	12.1
50-59 years	1731	17.3	185	25.5	84	15.1	1477	16.8
60-69 years	2311	23.1	176	24.2	146	26.3	2006	22.8
70-79 years	1691	16.9	133	18.3	132	23.8	1442	16.4
80+ years	1874	18.7	74	10.2	150	27.0	1666	19.0
Hospital								
Yongkang	6000	60.0	396	54.6	230	41.4	5400	61.4
Liouying	2500	25.0	220	30.3	237	42.9	2074	23.6
Jiali	1500	15.0	110	15.1	87	15.7	1314	15.0

Table 3 summarizes the performance of the nine algorithms in identifying people with HBV and HCV. The algorithms with higher number of visits with ICD codes had higher PPV; for example, the PPV for HBV using algorithm 1 ( $\geq 1$  OP codes) was 72% and 86% according to ICD-9-CM and ICD-10-CM, respectively, and that of algorithm 3 ( $\geq 3$  OP codes) was 80% and 90%, respectively. Similarly, the PPV for HCV using algorithm 1 was 88% and 96%, respectively, and using algorithm 3 was 93% and 99%, respectively. For each algorithm, the performance for identifying people with HCV was better than for identifying people with HBV.

Table 3  
Results of 9 coding algorithms to identify people with hepatitis B or C virus

Algorithm	ICD-9-CM				ICD-10-CM			
	Sen	Spe	PPV	NPV	Sen	Spe	PPV	NPV
Hepatitis B virus								
1. $\geq 1$ OP codes	62	98	72	97	54	99	86	96
2. $\geq 2$ OP codes	58	99	77	97	50	99	88	96
3. $\geq 3$ OP codes	53	99	80	96	45	100	90	96
4. $\geq 1$ IP codes	25	100	90	94	16	100	96	94
5. $\geq 2$ IP codes	14	100	93	94	7	100	100	93
6. $\geq 3$ IP codes	100	97	93	5	100	100	93	
7. $\geq 2$ OP or $\geq 1$ IP codes	63	98	76	97	56	99	88	97
8. $\geq 3$ OP or $\geq 1$ IP codes	59	99	79	97	53	100	91	96
9. $\geq 4$ OP or $\geq 1$ IP codes	57	99	80	97	51	100	91	96
Hepatitis C virus								
1. $\geq 1$ OP codes	63	100	88	98	56	100	96	97
2. $\geq 2$ OP codes	60	100	91	98	52	100	98	97
3. $\geq 3$ OP codes	56	100	93	97	48	100	99	97
4. $\geq 1$ IP codes	31	100	97	96	21	100	97	96
5. $\geq 2$ IP codes	17	100	99	95	12	100	100	95
6. $\geq 3$ IP codes	10	100	98	95	8	100	100	95
7. $\geq 2$ OP or $\geq 1$ IP codes	66	100	91	98	59	100	97	98
8. $\geq 3$ OP or $\geq 1$ IP codes	65	100	93	98	56	100	97	97
9. $\geq 4$ OP or $\geq 1$ IP codes	62	100	94	98	54	100	98	97
ICD-9-CM = International Classification of Disease Clinical Modification Ninth Revision, ICD-10-CM = International Classification of Disease Clinical Modification Tenth Revision, IP = inpatients, NPV = negative predictive value, OP = outpatients, PPV = positive predictive value, Sen = sensitivity, Spe = specificity								

However, as the PPV increased, the sensitivity declined. For example, the sensitivity of algorithm 1 for HBV according to ICD-9-CM and ICD-10-CM was 62% and 54%, respectively, and of algorithm 3 was 53% and 45%, respectively (Table 3). Considering the tradeoff between PPV and sensitivity, algorithm 7 ( $\geq 2$

OP or  $\geq 1$  IP codes) was deemed as the optimum for both HBV and HCV according to both ICD-9-CM and ICD-10-CM.

The validation indicators using algorithm 7 for different demographic characteristics are illustrated in Table 4. No significant differences between men and women in the indicators were noted. However, a different pattern of changes by age was found. For HBV, PPV decreased with age with both ICD-9-CM and ICD-10-CM codes. By contrast, the PPV increased with age for HCV with both ICD-9-CM and ICD-10-CM codes.

Table 4

Performance of algorithm 7 ( $\geq 2$  OP or  $\geq 1$  IP codes) for identifying people with hepatitis B or C virus stratified by demographic characteristics

Characteristics	ICD-9-CM				ICD-10-CM			
	Sen	Spe	PPV	NPV	Sen	Spe	PPV	NPV
Hepatitis B virus								
Sex								
Men	66	98	77	97	57	99	88	96
Women	59	99	75	98	55	100	88	97
Age								
20-39 years	47	100	88	98	42	100	100	98
40-49 years	50	99	81	95	57	100	93	96
50-59 years	68	98	82	96	58	99	89	95
60-69 years	73	98	74	98	65	99	86	97
70-79 years	67	98	74	97	56	99	88	96
80+ years	51	99	64	98	41	100	79	98
Hepatitis C virus								
Sex								
Men	66	99	89	98	60	100	96	98
Women	67	100	93	98	58	100	98	98
Age								
20-39 years	33	100	50	99	11	100	50	99
40-49 years	79	100	87	99	56	100	95	99
50-59 years	61	100	89	98	61	100	96	98
60-69 years	66	99	88	98	64	100	95	98
70-79 years	69	100	95	97	64	100	98	97
80+ years	67	100	93	97	51	100	100	96

ICD-9-CM = International Classification of Disease Clinical Modification Ninth Revision, ICD-10-CM = International Classification of Disease Clinical Modification Tenth Revision, IP = inpatients, NPV = negative predictive value, OP = outpatients, PPV = positive predictive value, Sen = sensitivity, Spe = specificity



## Discussion

The findings of this study suggest that the use of ICD-10-CM codes led to better performance than the use of ICD-9-CM codes for identifying people with HBV and HCV. The algorithms performed better in identifying people with HCV than in identifying people with HBV. Of the nine algorithms examined, a gain in PPV complemented with a loss in sensitivity; considering the tradeoff between PPV and sensitivity, the optimal algorithm was  $\geq 2$  OP or  $\geq 1$  IP codes.

The Centers for Disease Control and Prevention Chronic Hepatitis Cohort Study in the United States used an algorithm of two ICD-9 codes separated by  $\geq 6$  months, which had a PPV of 90% for HBV and 92% for HCV.<sup>11,12</sup> The PPV of algorithm 2 ( $\geq 2$  OP codes) in this study was 77% and 88% according to ICD-9-CM and ICD-10-CM, respectively, for HBV and was 91% and 98%, respectively, for HCV.

The better performance of algorithms according to ICD-10-CM than those according to ICD-9-CM is unlikely due to the differences in codes themselves because the classification scheme and number of codes do not differ much between the two revisions. A more likely explanation is the education and training offered by the NHI before the implementation of ICD-10-CM in January 2016, as well as some quality improvement programs later.

Our algorithms exhibited better performance in identifying people with HCV than people with HBV, especially when using ICD-10-CM codes; this was likely because the NHI has covered DAAs for people with HCV since January 24, 2017.<sup>19</sup> The physicians were required to provide ICD-10-CM codes for people with HCV for prescribing DAAs.

The performance of our algorithms was better than those of a previous Taiwanese study because of two possible reasons.<sup>13</sup> First, we used more data sources (drug prescription, laboratory results, clinical diagnosis) and included a longer study period (2005–2019) than in the previous study (laboratory results for one quarter in 2018). Some of the people with HBV or HCV ICD codes judged as false positive in previous studies might have been judged as true positive in this study because of more evidence. Second, this study was confined to one health care system with three hospitals with relatively high quality of coding, and the previous study covered thousands of hospitals and clinics in Taiwan.

One of the strengths of this study is large sample size. Unlike some previous studies using ICD codes to recruit patients, which allowed only PPV estimation,<sup>9,10</sup> in this study, by using systematic sampling, we could also calculate sensitivity. Second, we used a search engine to determine clinical diagnoses through large amounts of electronic medical records. Third, this study is the first to examine the performance of using ICD-10-CM code algorithms to identify people with HBV and HCV. Fourth, we used nine algorithms compared with only one in the previous Taiwan study.

Nevertheless, our study also had several limitations. First, this study was confined to a health care system in southern Taiwan, which might affect the generalization to other populations. However, the main findings (better performance of ICD-10-CM than ICD-9-CM and HCV than HBV) were affected by

contextual factors (education and training program and reimbursement of DAA). Therefore, we believe that these conclusions may be applicable to other clinical settings in Taiwan. Second, some of the patients might have positive results of laboratory tests in other hospitals but were not tested in this health care system, rendering them false negatives in this study. Third, similarly, some of the clinical diagnoses recorded by physicians might not be valid.

In conclusion, using the electronic medical records of proportional sampling of 10,000 patients in a health care system in south Taiwan, this study suggests that algorithms using ICD-10-CM codes had better performance than those using ICD-9-CM codes in identifying people with HBV and HCV. Considering the tradeoff between PPV and sensitivity, the optimal algorithm was determined to be as follows:  $\geq 2$  outpatient or  $\geq 1$  inpatient visits with HBV or HCV ICD codes. Furthermore, ICD codes can better identify people with HCV than people with HBV.

## Abbreviations

DAA: direct oral antiviral agents; HBV: hepatitis B virus; HCV: hepatitis C virus; ICD: International Classification of Diseases; ICD-9-CM: International Classification of Diseases Ninth Revision Clinical Modification; ICD-10-CM: International Classification of Diseases Tenth Revision Clinical Modification; IP: inpatient; NHI: National Health Insurance; OP: outpatient; PPV: positive predictive value; RECORD: Reporting of studies conducted using observational routinely-collected health data; Sen: sensitivity; Spe: specificity.

## Declarations

### Author details

<sup>1</sup> Division of Gastroenterology and Hepatology, Chi Mei Medical Center, Tainan, Taiwan. <sup>2</sup> Department of Medicinal Chemistry, Chia Nan University of Pharmacy and Science, Tainan, Taiwan. <sup>3</sup> Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan. <sup>4</sup> Department of Healthcare Administration, College of Medical and Health Science, Asia University, Taichung, Taiwan. <sup>5</sup> Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan. <sup>6</sup> Department of Industrial and Information Management, College of Management, National Cheng Kung University, Tainan, Taiwan.

### Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the Chi Mei Medical Center (number: 10901-015). The need for informed consent was exempted by the Institute Review Board (IRB) of Chi Mei Medical Center because of the IDs in the medical records were scrambled for analysis. All methods were carried out in accordance with relevant guidelines and regulations(eg. Helsinki declaration)

### Consent for publication

Not applicable.

## Availability of data and materials

The data are available from corresponding author upon reasonable request.

## Competing interests

None for all authors.

## Funding

This study was financially supported by Chi Mei Medical Center number CMFHR10911.

## Authors' contributions

M.J.S. and T.H.L. conceptualized the idea, supervised the study, wrote the first draft and critically reviewed and edited the manuscript. T.W.C. and F.P.K. analyzed the data and critically reviewed and edited the manuscript. C.Y.L. and S.T.L. critically reviewed and edited the manuscript. All authors read and approve the manuscript.

## Acknowledgements

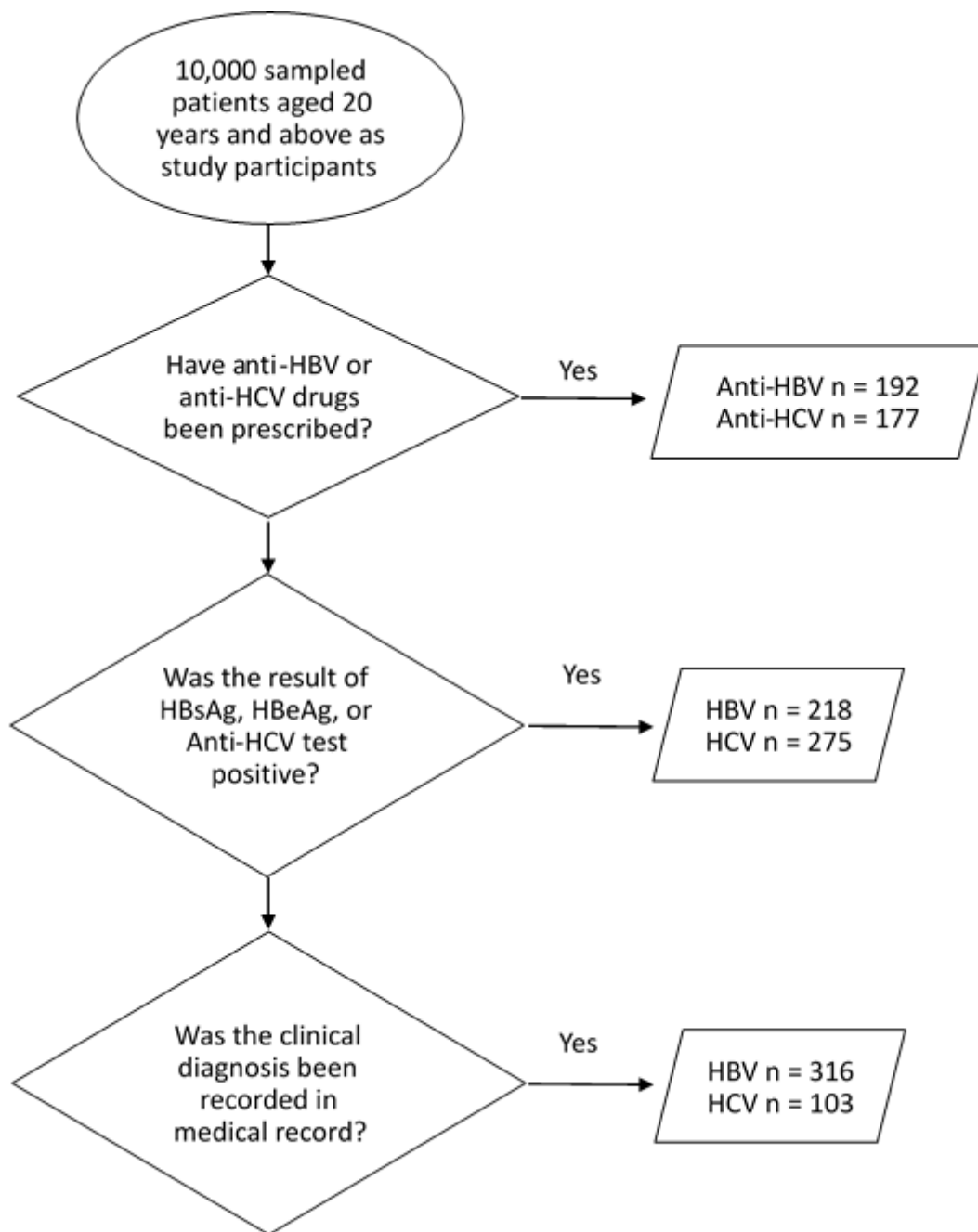
The authors would like to thank Huei-Lin Wu for help in administration coordination in this study.

## References

1. Lau CS, Chen YH, Lim K, de Longueville M, Arendt C, Winthrop K. Tuberculosis and viral hepatitis in patients treated with certolizumab pegol in Asia-Pacific countries and worldwide: real-world and clinical trial data. *Clin Rheumatol*. 2021;40:867–875.
2. Forns X, Feld JJ, Dylla DE, Pol S, et al. Safety of patients with hepatitis C virus treated with Glecaprevir/Pibrentasvir from clinical trials and real-world cohorts. *Adv Ther*. 2021;22:1–8.
3. Yotsuyanagi H, Kurosaki M, Yatsunami H, Lee IH, et al. Characteristics and healthcare costs in the aging hepatitis B population of Japan: A nationwide real-world analysis. *Dig Dis*. 2021; <https://doi.org/10.1159/000515854>
4. Wedemeyer H, Erren P, Naumann U, Rieke A, Stoehr A, Zimmermann T, Lohmann K, König B, Mauss S. Glecaprevir/pibrentasvir is safe and effective in hepatitis C patients with cirrhosis: Real-world data from the German Hepatitis C Registry. *Liver Int*. 2021;41:949–955.
5. Müllhaupt B, Semela D, Ruckstuhl L, Magenta L, Clerc O, Torgler R, Negro F, Semmo N. Real-world effectiveness and safety of glecaprevir/pibrentasvir therapy in patients with chronic hepatitis C virus infection in Switzerland. *Swiss Med Wkly*. 2021;151:w20399.
6. Ehrenstein V, Petersen I, Smeeth L, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. *Clin Epidemiol*. 2016;8:49.

7. Chun DS, Lund JL, Stürmer T. Pharmacoepidemiology and Drug Safety's special issue on validation studies. *Pharmacoepidemiol Drug Saf.* 2019;28:123–125.
8. Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 2015;12:e1001885.
9. Kramer JR, Davila JA, Miller ED, et al. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther.* 2008;27:274–282. doi.org/10.1111/j.1365-2036.2007.03572.x
10. Niu B, Forde KA, Goldberg DS. Coding algorithms for identifying patients with cirrhosis and hepatitis B or C virus using administrative data. *Pharmacoepidemiol Drug Saf.* 2015;24:107–111.
11. Mahajan R, Moorman AC, Liu SJ, et al. Use of the International Classification of Diseases, 9th revision, coding in identifying chronic hepatitis B virus infection in health system data: implications for national surveillance. *J Am Med Inform Assoc.* 2013;20:441–445. doi: 10.1136/amiajnl-2012-001558
12. Abara WE, Moorman AC, Zhong Y, et al. The predictive value of International Classification of Disease codes for chronic hepatitis C virus infection surveillance: The utility and limitations of electronic health records. *Popul Health Manag.* 2018;21(2):110–115. doi.org/10.1089/pop.2017.0004
13. Sheu MJ, Liang FW, Li ST, Li CY, Lu TH. Validity of ICD-10-CM codes used to identify patients with chronic hepatitis B and C virus infection in administrative claims data from the taiwan national health insurance outpatient claims dataset. *Clin Epidemiol.* 2020;12:185.
14. Panozzo CA, Woodworth TS, Welch EC, et al. Early impact of the ICD-10-CM transition on selected health outcomes in 13 electronic health care databases in the United States. *Pharmacoepidemiol Drug Saf.* 2018;27:839–847.
15. Caskey RN, Abutahoun A, Polick A, Barnes M, Srivastava P, Boyd AD. Transition to international classification of disease version 10, clinical modification: the impact on internal medicine and internal medicine subspecialties. *BMC Health Ser Res.* 2018;18:1–8.
16. Mainor AJ, Morden NE, Smith J, Tomlin S, Skinner J. ICD-10 coding will challenge researchers- caution and collaboration may reduce measurement error and improve comparability over time. *Med Care.* 2019;57:e42.
17. Ellis RP, Hsu HE, Song C, et al. Diagnostic category prevalence in 3 classification systems across the transition to the International Classification of Diseases, Tenth Revision, Clinical Modification. *JAMA Netw Open.* 2020;3:e202280-.
18. Kusnoor SV, Blasingame MN, Williams AM, DesAutels SJ, Su J, Giuse NB. A narrative review of the impact of the transition to ICD-10 and ICD-10-CM/PCS. *JAMIA Open.* 2020;3:126–131.
19. Wu GHM, Pwu RF, Chen SC. Achieving hepatitis C elimination in Taiwan—overcoming barriers by setting feasible strategies. *J Formos Med Assoc.* 2018;117:1044–1045.

## Figures



**Figure 1**

Flowchart of defining whether the study participants had hepatitis B or C