

Detecting non-alcoholic fatty liver disease and risk factors in health databases: accuracy and limitations of the ICD-10-AM

Kelly Lee Hayward ^{1,2}, Amy L Johnson,^{1,2} Leigh U Horsfall,^{1,2} Chris Moser,³ Patricia C Valery ,^{1,4} Elizabeth E Powell^{1,2}

To cite: Hayward KL, Johnson AL, Horsfall LU, *et al.* Detecting non-alcoholic fatty liver disease and risk factors in health databases: accuracy and limitations of the ICD-10-AM. *BMJ Open Gastro* 2021;**8**:e000572. doi:10.1136/bmjgast-2020-000572

PCV and EEP contributed equally.

Received 13 November 2020
Revised 22 December 2020
Accepted 6 January 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Centre for Liver Disease Research, The University of Queensland, Woolloongabba, Queensland, Australia

²Princess Alexandra Hospital, Metro South Hospital and Health Service, Woolloongabba, Queensland, Australia

³Statistical Services Branch, Queensland Government Department of Health and Ageing, Brisbane, Queensland, Australia

⁴QIMR Berghofer Medical Research Institute, Population Health, Herston, Queensland, Australia

Correspondence to Professor Elizabeth E Powell; e.powell@uq.edu.au

ABSTRACT

Objective The prevalence of non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH) cirrhosis is often underestimated in healthcare and administrative databases that define disease burden using International Classification of Diseases (ICD) codes. This retrospective audit was conducted to explore the accuracy and limitations of the ICD, Tenth Revision, Australian Modification (ICD-10-AM) to detect NAFLD, metabolic risk factors (obesity and diabetes) and other aetiologies of chronic liver disease.

Design/Method ICD-10-AM codes in 308 admitted patient encounters at two major Australian tertiary hospitals were compared with data abstracted from patients' electronic medical records. Accuracy of individual codes and grouped combinations was determined by calculating sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and Cohen's kappa coefficient (κ).

Results The presence of an ICD-10-AM code accurately predicted the presence of NAFLD/NASH (PPV 91.2%) and obesity (PPV 91.6%) in most instances. However, codes underestimated the prevalence of NAFLD/NASH and obesity by 42.9% and 45.3%, respectively. Overall concordance between clinical documentation and 'grouped alcohol' codes (κ 0.75) and hepatitis C codes (κ 0.88) was high. Hepatitis B codes detected false-positive cases in patients with previous exposure (PPV 55.6%). Accuracy of codes to detect diabetes was excellent (sensitivity 95.8%; specificity 97.6%; PPV 94.9%; NPV 98.1%) with almost perfect concordance between codes and documentation in medical records (κ 0.93).

Conclusion Recognition of the utility and limitations of ICD-10-AM codes to study the burden of NAFLD/NASH cirrhosis is imperative to inform public health strategies and appropriate investment of resources to manage this burgeoning chronic disease.

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) associated with obesity and type 2 diabetes mellitus (T2DM) is now the most common chronic liver disease (CLD) worldwide and is reported to be a rising cause of advanced liver disease, primary liver cancer and liver-related

Summary box

What is already known about this subject?

► Under-recording of non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH) and metabolic risk factors in population-based and administrative databases is widely recognised, although the reasons for this remain unclear. Understanding the limitations of International Classification of Diseases (ICD) codes is important to improve the reliability of health system databases for epidemiological studies and health services research.

What are the new findings?

► While specificity of the ICD, Tenth Revision, Australian Modification (ICD-10-AM) codes for NAFLD/NASH was high (97.7%), these codes underestimated the prevalence of NAFLD/NASH by 42.9%, despite explicit documentation in the medical record.

► The presence of an ICD-10-AM code for obesity accurately predicted obesity in most instances (specificity 95.9%); however, codes substantially underestimated obesity prevalence (sensitivity 54.7%). In most false-negative encounters (84.1%) this was due to lack of clear clinical documentation.

► In contrast, accuracy of codes to detect diabetes was excellent (sensitivity 95.8%; specificity 97.6%), with almost perfect concordance between codes and documentation in medical records.

How might it impact on clinical practice in the foreseeable future?

► These data suggest that changes may be required in the Australian Coding Standards and ICD-10-AM diagnosis codes, to better document the presence of NAFLD/NASH cirrhosis and obesity. There is also a need for clinician education to reinforce the importance of clear clinical documentation. Improving the accuracy of population-based data on the burden of NAFLD will be imperative to guide public health strategies and appropriate investment of resources to manage this burgeoning chronic disease.

mortality.¹ However, it is difficult to determine an accurate prevalence of NAFLD-related cirrhosis and complications because



it appears to be under-represented in databases that contain codes for disease aetiology and causes of death.

In a population-based study that examined International Classification of Diseases (ICD), Tenth Revision, Australian Modification (ICD-10-AM) codes among all people treated in hospital for cirrhosis in Queensland, Australia, during 2008–2016, only 4.8% of patients had a coded disease aetiology of NAFLD or non-alcoholic steatohepatitis (NASH).² While the proportion of patients admitted with NAFLD/NASH-related cirrhosis increased from 3.6% in 2008–2010 to 6.0% in 2014–2016 ($p < 0.00001$),² this is likely an under-representation of disease prevalence based on observational and cohort studies, as well as modelling of NAFLD in Australia.^{3,4}

Similar findings have been reported elsewhere. NAFLD prevalence is underestimated in US Medicare datasets and administrative databases that define NAFLD using ICD codes.^{5,6} In European primary healthcare databases, the pooled prevalence of NAFLD was 1.9%,⁷ far less than the expected community prevalence of 20%–30%. In order to reduce misclassification of NAFLD in population-based data, some authors have used an extended definition of NAFLD that included cryptogenic liver disease or cirrhosis in the presence of metabolic abnormality and the absence of other causes of liver disease⁵ or inferred data from the prevalence of obesity and T2DM.⁸ However, these methods also rely on the accuracy of clinical documentation and administrative coding, which may be influenced by country-specific codes and coding rules. In Australia, only 3.7% of admissions with cirrhosis had a coded diagnosis of obesity² despite a prevalence of obesity of 31% in Australian adults,⁹ illustrating the failure to capture or document important patient data that would help to determine liver disease aetiology or comorbidity.

Determining the accuracy of ICD coding in hospital admission data is necessary to understand the limitations and improve the reliability of health system databases for epidemiological studies and health services research. The primary aim of this study was to evaluate the performance and limitations of ICD-10-AM codes to identify NAFLD/NASH in admitted patient encounters with cirrhosis at a major tertiary hospital. The secondary aims were to investigate the accuracy of ICD-10-AM coding for NAFLD risk factors (obesity and T2DM) and other aetiologies of cirrhosis, namely alcohol-related liver disease (ALD), chronic hepatitis C virus (HCV) and hepatitis B virus (HBV) infection.

METHODS

Sample population and data collection

Patient cohorts

A retrospective audit was conducted in a sample of prospectively recruited patients with a diagnosis of cirrhosis to ascertain the level of concordance between select ICD-10-AM codes and documentation in patients' electronic medical records. Patients in the current study

were recruited between January 2016 and December 2018 to the CirCare Study, a multicentre observational study of patients with cirrhosis or a randomised controlled trial of an education intervention among outpatients with decompensated cirrhosis. The details of these studies have been previously described.^{10,11} All patients were adults aged ≥ 18 years with hepatic cirrhosis diagnosed by a hepatologist, based on liver histology, imaging or a combination of non-invasive markers and clinical assessment. To be included in the current study, patients had to have had at least one admission at the Princess Alexandra Hospital or the Royal Brisbane and Women's Hospital between 1 January 2016 and 30 June 2019.

ICD-10-AM codes

In the ICD-10-AM, two codes may be used to record NAFLD/NASH: K75.8 'Other specified inflammatory liver diseases (non-alcoholic steatohepatitis)' and K76.0 'Fatty (change of) liver, not elsewhere classified (non-alcoholic fatty liver disease)'. NAFLD/NASH cirrhosis may also attract code K74.6 'Other and unspecified cirrhosis of liver'. Obesity may be represented by ICD-10-AM codes under the E66 code block or by the chronic condition supplementary code U78.1. T2DM codes are contained in the E11 code block. Codes for other causes of CLD are outlined in table 1.

All ICD-10-AM codes were obtained from the Queensland Hospital Admitted Patient Data Collection registry (QHAPDC) for every hospital encounter within the study cohort during a minimum 12-month follow-up period. For the current study, one encounter was randomly selected for each patient to be included in the audit sample. In a subset of patients who had ≥ 1 encounter with and ≥ 1 encounter without code K74.6, a second encounter was selected (one encounter containing code K74.6 and one encounter that did not contain code K74.6) to ensure the sample contained a representative number of patients with possible NAFLD.

In Australia, assignment of ICD-10-AM diagnosis codes is impacted by the Australian Coding Standards, which stipulate the condition must be either the chief reason for admission or required commencement, alteration or adjustment of therapeutic treatment, diagnostic procedures or increased clinical care and/or monitoring during the encounter. T2DM and HCV are exceptions to this assignment process, as these conditions are always coded when documented. In addition, a supplementary 'U' coding capability was introduced on 1 July 2015, to capture conditions (such as obesity) that contribute to a patient's health status during the admission, but do not otherwise meet the criteria for coding. The index pathway used by coders for coding of cirrhosis, CLD aetiology, obesity and T2DM remained consistent during the audited time period.

Clinical data

Three clinicians (EEP, LUH and ALJ) blinded to QHAPDC coding conducted a comprehensive review of

Table 1 ICD-10-AM codes for aetiologies of chronic liver disease

ICD-10-AM code	Encounters with code	PPV (95% CI)	NPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Kappa
Alcohol-related liver disease						
F10.1 Harmful use of alcohol	111	0.95 (0.89 to 0.98)	0.63 (0.57 to 0.70)	0.59 (0.52 to 0.66)	0.95 (0.91 to 0.98)	0.51
K70.1 Alcoholic hepatitis	9	1.00	0.44 (0.43 to 0.45)	0.05 (0.03 to 0.09)	1.00	0.04
K70.2 Alcoholic fibrosis and sclerosis of liver	0	–	–	–	–	–
K70.3 Alcoholic cirrhosis of liver	146	0.93 (0.88 to 0.96)	0.74 (0.67 to 0.80)	0.76 (0.69 to 0.82)	0.92 (0.87 to 0.96)	0.67
K70.4 Alcoholic hepatic failure	65	0.95 (0.89 to 0.99)	0.53 (0.46 to 0.59)	0.35 (0.28 to 0.42)	0.98 (0.94 to 0.99)	0.30
K70.9 Alcoholic liver disease, unspecified	0	–	–	–	–	–
Grouped alcohol codes*†	158	0.94 (0.89 to 0.97)	0.81 (0.74 to 0.86)	0.84 (0.78 to 0.89)	0.92 (0.87 to 0.96)	0.75
Hepatitis B‡						
B18.2 Chronic viral hepatitis C	85	0.95 (0.89 to 0.98)	0.95 (0.92 to 0.97)	0.88 (0.80 to 0.94)	0.98 (0.96 to 0.99)	0.88
Z22.52 Carrier of viral hepatitis C	0	–	–	–	–	–
Hepatitis B‡						
B18.0 Chronic viral hepatitis B with delta-agent	0	–	–	–	–	–
B18.1 Chronic viral hepatitis B without delta-agent	18	0.56 (0.33 to 0.77)	1.00	1.00	0.97 (0.95 to 0.99)	0.70
Z22.51 Carrier of viral hepatitis B	0	–	–	–	–	–
NAFLD/NASH						
K75.8 Other specified inflammatory diseases of the liver	48	0.92 (0.82 to 0.97)	0.82 (0.77 to 0.86)	0.48 (0.38 to 0.59)	0.98 (0.96 to 0.99)	0.54
K76.0 Fatty (change of) liver, not elsewhere classified	10	0.90 (0.63 to 0.99)	0.72 (0.67 to 0.77)	0.10 (0.05 to 0.17)	0.96 (0.98 to 1.00)	0.13
Grouped NAFLD/NASH codes†	57	0.91 (0.82 to 0.97)	0.84 (0.80 to 0.89)	0.57 (0.47 to 0.67)	0.98 (0.95 to 0.99)	0.62

*‘Grouped alcohol’ codes included: E24.4 ‘Alcohol induced pseudo-Cushing’s syndrome’, F10 code block ‘Mental and behavioural disorders due to use of alcohol’, G31.2 ‘Degeneration of nervous system due to alcohol’, G62.1 ‘Alcoholic polyneuropathy’, G72.1 ‘Alcoholic myopathy’, I42.6 ‘Alcoholic cardiomyopathy’, K29.2 ‘Alcoholic gastritis’, K29.20 ‘Alcoholic gastritis without mention of haemorrhage’, K29.21 ‘Alcoholic gastritis with haemorrhage’, K70 code block ‘Alcoholic liver disease’, K86.0 ‘Alcohol-induced chronic pancreatitis’, T51.0 ‘Toxic effect of alcohol, ethanol’, Z50.2 ‘Alcohol rehabilitation’, Z71.4 ‘Counselling and surveillance for alcohol use disorder’, Z72.1 ‘Alcohol use’.

†More than one code may have been assigned to the audited encounter.

‡Codes for acute viral hepatitis (including code blocks B16 ‘Acute hepatitis B’ and B17 ‘Other acute viral hepatitis’) were not included. ICD-10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NPV, negative predictive value; PPV, positive predictive value.



patients' medical records and extracted data for each audited encounter using an agreed template. A subset of medical records was reviewed by all investigators to ensure consistency in the approach to data extraction, and any discordant results were adjudicated by a hepatologist (EEP). The presence of cirrhosis, obesity or body mass index (BMI) $\geq 30.0 \text{ kg/m}^2$ ($\geq 27.5 \text{ kg/m}^2$ if patient of Asian descent), T2DM and liver disease aetiology was collected for each audited encounter. These conditions were considered present if they were documented by a treating clinician or if BMI was recorded in the patients' medical record. The accuracy of documented diagnoses and recorded BMI was not audited. Clinical information was retrieved from prior admissions and outpatient encounters to corroborate information where available, although only data from the audited encounter were included in the abstraction.

Data analysis

Data are presented as counts and proportions. The accuracy of ICD-10-AM codes to predict the aetiology of CLD and presence of metabolic risk factors (obesity and T2DM) was determined by calculating sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) (see online supplemental file). Concordance between data abstracted from the patient's medical records (gold standard) and ICD-10-AM data from QHAPDC was assessed using Cohen's kappa coefficient (κ) of agreement. κ values < 0.20 indicated poor agreement, 0.21 – 0.40 fair, 0.41 – 0.60 moderate, 0.61 – 0.80 substantial and 0.81 – 1.0 almost perfect agreement between the codes and medical records.¹²

RESULTS

A total of 312 encounters among 271 patients were selected; however, four encounters were excluded ($n=2$ data not available and $n=2$ post liver transplant admissions). Therefore, the final sample contained 308 encounters among 267 patients (41 patients contributed two encounters each).

Medical record review

Among the total 308 audited admissions, at least one liver disease aetiology was documented in 289 (93.8%) encounters. ALD was documented in 177 encounters (57.5%), NAFLD/NASH in 91 (29.5%), current or treated HCV infection in 92 (29.9%), HBV in 10 (3.2%) and 'other' in 28 (9.1%) (including drug-induced liver injury, alpha-1 antitrypsin deficiency, autoimmune hepatitis, familial intrahepatic cholestasis, primary biliary cholangitis, primary sclerosing cholangitis and haemochromatosis). In 104 (33.8%) admissions, more than one liver disease aetiology was documented, most commonly ALD and HCV in 61 (19.8%) and ALD and NAFLD/NASH in 32 (10.4%) encounters. Most encounters ($n=289$; 93.8%) had cirrhosis documented during the admission including one patient with cryptogenic cirrhosis.

Obesity was documented in 46 (14.9%) admissions. In a further 93 (30.2%) admissions, obesity was present (as determined by BMI) but not clearly documented in the audited encounter. T2DM was documented in 96 (31.2%) admissions. An additional six encounters included documentation of 'other specified diabetes' ($n=3$ steroid-induced diabetes and $n=3$ diabetes secondary to pancreatic insufficiency following a Whipple procedure) but these were not included in the subgroup with T2DM.

ICD-10-AM code accuracy

Aetiology

An ICD-10-AM code for NAFLD/NASH was present in 57 (18.5%) encounters (table 1). Overall concordance between codes and patients' medical records was good (κ 0.62). However, while specificity (97.7%), PPV (91.2%) and NPV (84.4%) were high, codes for NAFLD/NASH underestimated the prevalence of NAFLD/NASH by 42.9% (39 of 91 encounters did not contain a code despite documentation of NAFLD/NASH in the medical record; sensitivity 57.1%). The five false-positive codes were in patients with ALD ($n=4$ including two with HCV as a cofactor) and HBV ($n=1$).

At least one alcohol-related ICD-10-AM code was present in 158 (51.3%) encounters. Concordance between alcohol-related codes and patients' medical records was substantial (κ 0.75). Codes had good overall accuracy to detect ALD with high sensitivity (83.6%), specificity (92.4%), PPV (93.7%) and NPV (80.7%).

An HCV-related ICD-10-AM code was present in 85 (27.6%) encounters. Accuracy of codes to detect current or treated HCV infection was almost perfect (κ 0.88), with high sensitivity (88.0%), specificity (98.1%), PPV (95.3%) and NPV (95.1%).

Current HBV (hepatitis B surface antigen (HBsAg) positive) was present in a minority of patients ($n=10$) and was identified with an ICD code in all cases (sensitivity 100%; specificity 97.3%; NPV 100%; κ 0.70). However, an additional eight patients with prior exposure to HBV (HBsAg negative, hepatitis B surface antibody (HBsAb) and/or hepatitis B core antibody (HBcAb) positive) had a false-positive ICD-10-AM code for HBV (PPV 55.6%).

Obesity

Eighty-three encounters (26.9%) included an ICD-10-AM code for obesity (table 2). While the presence of a code accurately predicted obesity in most instances (PPV 91.6%; specificity 95.9%), codes substantially underestimated obesity prevalence (sensitivity 54.7%; NPV 72.0%). In most false-negative encounters ($n=53$; 84.1%) this was likely due to lack of clear clinical documentation (ie, obesity only identified by BMI during medical record review). As a consequence, overall accuracy of ICD-10-AM codes to detect obesity was modest (κ 0.52). When obesity was clearly documented in the medical record ($n=46$ encounters), a code was assigned in 78.3% of cases.

Table 2 Relationship between ICD-10-AM codes and patients' medical records

		On medical record review		
		Present	Absent	Total
ALD Grouped codes*	+	148	10	158
	–	29	121	150
	Total	177	131	
HCV B18.2	+	81	4	85
	–	11	212	223
	Total	92	216	
HBV B18.1	+	10	8	18
	–	0	290	290
	Total	10	298	
NAFLD/NASH K75.8, K76.0	+	52	5	57
	–	39	212	251
	Total	91	217	
Obesity E66.0, E66.1, E66.2, E66.8, E66.9, U78.1	+	76	7	83
	–	63	162	225
	Total	139	169	
T2DM E11	+	92	5	97
	–	4	207	211
	Total	96	212	

+ Encounter includes ≥ 1 ICD-10-AM code.

– Encounter does not include an ICD-10-AM code.

*Refer to table 1 for 'grouped alcohol' codes.

ALD, alcohol-related liver disease; HBV, hepatitis B virus; HCV, hepatitis C virus; ICD-10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; T2DM, type 2 diabetes mellitus.

Diabetes

One-third of encounters (31.5%) included an ICD-10-AM code for T2DM. Accuracy of codes to detect diabetes was excellent (sensitivity 95.8%; specificity 97.6%; PPV 94.8%; NPV 98.1%) with almost perfect concordance between codes and documentation in medical records (κ 0.93).

K74.6 'Other and unspecified cirrhosis'

The ICD-10-AM code K74.6 'Other and unspecified cirrhosis of liver' was present in 126 (40.9%) of admissions.



Figure 1 Heatmap depicting prevalence and clustering of CLD aetiologies and cirrhosis according to clinical review in 126 encounters that contained code K74.6 'Other and unspecified cirrhosis of liver'. Columns represent individual encounters. ALD, alcohol-related liver disease; CLD, chronic liver disease; HBV, hepatitis B virus; HCV, hepatitis C virus; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis.

Most encounters (95.2%) were in patients who had cirrhosis documented in the audited admission. The most common aetiology among encounters containing code K74.6 was NAFLD/NASH (42.9%); however, over one-third of these encounters were in patients who also had ALD, HCV or an 'other' concurrent aetiology (figure 1). Three encounters included code K74.6 in the absence of a documented CLD aetiology. These were all day procedures with minimal documentation in the medical record (n=1 endoscopy and n=2 large volume paracentesis). A further three encounters were assigned this code in the absence of documented cirrhosis (n=2 day admissions for liver biopsy and n=1 admission for ascites of cardiac origin).

ICD codes in patients with NAFLD/NASH

Of 91 encounters in patients with NAFLD/NASH identified by medical chart review, 59.3% included code K74.6, 57.1% had an NAFLD/NASH code, 29.7% had an ALD code and 11.0% had an HCV code (figure 2A). Thirty-one encounters (34.1%) included an obesity code in combination with a T2DM code; however, most of these (n=22; 71.0%) were in encounters that already contained an NAFLD/NASH code.

There were 47 encounters identified by medical chart review in patients with NAFLD/NASH without comorbid liver disease. Of these 47 encounters, 68.1% included code K74.6, 55.3% included an NAFLD/NASH code, 68.1% had a diabetes code and 51.1% had an obesity code (figure 2B). None of the 47 encounters contained a code for ALD, HCV, HBV or 'other' aetiology.

DISCUSSION

Under-recording of NAFLD in population-based or administrative databases is widely recognised, although the reasons for this remain unclear. In this study of patient encounters with cirrhosis at two major tertiary hospitals, ICD-10-AM codes had a high specificity but low sensitivity for NAFLD/NASH and similarly had high specificity but low sensitivity for identifying obesity, although for different reasons. In contrast, accuracy of codes to detect T2DM was excellent. Determining the accuracy of ICD coding in hospital admission data for NAFLD and the factors that influence this are necessary to understand the limitations and improve the reliability of health



Figure 2 Prevalence and clustering of ICD-10-AM codes in 91 encounters among patients with NAFLD/NASH on clinical review (A) and 47 encounters among patients with NAFLD/NASH without a concurrent aetiology of CLD on clinical review (B). Columns represent individual encounters. *Refer to table 1 for ‘grouped alcohol’ codes. CLD, chronic liver disease; HCV, hepatitis C virus; ICD-10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; T2DM, type 2 diabetes mellitus.

system databases for epidemiological studies and health services research for this patient group.

Our data show that, although a clinical diagnosis of NAFLD was made and recorded in the medical record, misclassification of patients in the ICD-10-AM coding process was prevalent. This may be related to under-recognition of the condition by clinical coders or lack of optimal codes to differentiate this increasingly common liver disease. The absence of a specific code for NASH cirrhosis may contribute to inadequate capture of this information. At present, clinical coding of this diagnosis requires the combination of two imprecise codes: K74.6 ‘Other and unspecified cirrhosis of liver’ and K75.8 ‘Other specified inflammatory liver diseases (non-alcoholic steatohepatitis)’ or K76.0 ‘Fatty (change of) liver, not elsewhere classified (non-alcoholic fatty liver disease)’. In addition to misclassification, it is also possible that other patients in the cohort may have undiagnosed NAFLD, though assessment of the accuracy of diagnoses recorded in the medical record was outside the scope of the current study. Our data support findings from a prior study that found a combination of ICD diagnosis codes had high PPV but low sensitivity to identify documented cases of NASH cirrhosis.¹³

Interestingly in this ‘real-world’ study, a substantial proportion of patients with NAFLD (48.4%) had a concurrent liver disease, particularly alcohol excess or chronic HCV. Although the currently accepted nomenclature by the American and European Associations for the Study of Liver Diseases excludes excessive alcohol consumption or other liver diseases in the definition of NAFLD,^{14 15} there is increasing acceptance that NAFLD often coexists with other hepatic disorders.^{16 17} In fact, the Asian Pacific Association for the Study of the Liver has recently endorsed a proposal to redefine NAFLD based on the detection of steatosis together with the presence of overweight/obesity, or T2DM, or clinical evidence of metabolic dysfunction (‘metabolic-associated fatty liver

disease’ or ‘MAFLD’), regardless of alcohol consumption or other concomitant liver disease.¹⁸ Recognition of dual aetiology is important because coexistent NAFLD has a synergistic role in liver disease progression^{16 17} and may increase the risk of cardiometabolic problems. However, lack of clarity about the terminology (NAFLD vs MAFLD) or its diagnosis in the presence of coexisting liver diseases may result in under-reporting of the true burden of NAFLD and highlights the need for consensus within the hepatology community and appropriate dissemination of this information.

Our data show that notation of ‘obesity’ in the medical records was poor (based on recorded BMI in clinical documentation) and was largely responsible for the low sensitivity of obesity codes. These findings support a recent study from Australia that examined the agreement between medical records and ICD-10-AM comorbidity codes in trauma patients.¹⁹ The authors found that, based on clinician documentation, the prevalence of obesity was only 9.3% (compared with the Australian population prevalence of 31%⁹), with an even lower prevalence based on administrative data.¹⁹ Multiple previous studies have also shown that the completeness of ICD diagnosis coding for overweight/obesity is low compared with its prevalence in medical records based on clinical weight/height or BMI measurements.^{20–23} Although it is a potential risk factor for many diseases, obesity is not usually the primary cause of admission and may therefore be overlooked by clinicians when registering diagnoses. This systemic under-reporting currently limits the value of using administrative coding for obesity as part of an ‘extended definition’ of NAFLD in patients admitted with cirrhosis. Further review of coding practices and improved clinician documentation of BMI and obesity is urgently required in order to improve reporting of obesity in healthcare administrative databases.

In our patient cohort, there was almost perfect concordance between administrative codes and documentation

of T2DM in medical records. Recent studies examining hospital data from a trauma centre¹⁹ and cancer outcomes registry²⁴ in Australia have also reported ‘excellent’ and ‘substantial’ agreement, respectively, between administrative data and medical records for T2DM. This high concordance is likely a consequence of the mandatory requirement to code T2DM when it is documented, as well as the need to monitor and treat this condition during the hospital admission.

Compared with NAFLD/NASH codes, overall accuracy of ICD-10-AM codes for other aetiologies of CLD was high, similar to findings in prior studies.^{25–27} While most patients with ALD cirrhosis were detected using codes F10.1 ‘Harmful use of alcohol’, K70.3 ‘Alcoholic cirrhosis of liver’ and K70.4 ‘Alcoholic hepatic failure’, ‘grouped alcohol’ codes had the best overall concordance (κ 0.75). Viral hepatitis codes B18.2 ‘Chronic viral hepatitis C’ and B18.1 ‘Chronic viral hepatitis B without delta-agent’ also had high concordance ($\kappa \geq 0.70$), though the code for HBV identified several false-positive patients with previous exposure to HBV. We were not able to assess the accuracy of other viral hepatitis codes because they were not present in our sample.

Strengths of our study include the selection of a well-characterised cohort of patients with a diagnosis of cirrhosis confirmed by a healthcare provider. While it was outside the scope of the current study to assess the accuracy of clinical documentation, data abstraction from medical records was conducted by clinicians experienced in the management of CLD, including a hepatologist. The study was conducted using encounters at two tertiary hospitals that use electronic medical records. Therefore, a limitation of the study is that quality of clinical documentation and coding may differ from other smaller hospitals in regional areas that may use paper charts. However, regular internal auditing processes at a hospital, health service and jurisdictional level are regularly conducted for quality assurance purposes, which are supported by national data validation activities. Therefore, we are confident that our data represent an accurate sample of patients with cirrhosis in Southeast Queensland, Australia. The relatively small number of patients with HBV also limits conclusions.

Population-based data on the epidemiology and natural history of NAFLD are crucial to assess the true burden of this liver disease. Our data suggest that misclassification of NAFLD and obesity in the ICD-10-AM coding process may be related to a lack of appropriate codes or inadequate clinical documentation due to under-recognition or under-recording of these conditions. There may also be a lack of clarity about the diagnosis of NAFLD in the presence of coexisting liver diseases such as alcohol excess or viral hepatitis. Recognition of the utility and limitations of ICD-10-AM codes to study the burden of NAFLD/NASH cirrhosis in Australia is imperative to inform public health strategies and appropriate investment of resources to manage this burgeoning chronic disease.

Contributors PCV and EP conceived and planned the study. PCV and KLH identified the study sample and obtained International Classification of Diseases, Tenth Revision, Australian Modification data from Queensland Hospital Admitted Patient Data Collection. EP, ALJ and LUH collected the clinical data. KLH analysed the data. PCV, EP and CM interpreted study findings. PCV, EP and KLH drafted the manuscript. All authors reviewed and approved the final version for publication and have agreed to be accountable for all aspects of the work.

Funding KLH was supported by a Health Innovation, Investment and Research Office Clinical Research Fellowship. PCV was supported by an Australian National Health and Medical Research Council Career Development Fellowship (no. 1083090).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethics approval was obtained from the Human Research Ethics Committees of the Metro South Health (HREC/16/QPAH/628 and HREC/15/QPAH/688) and QIMR Berghofer Medical Research Institute (P2207).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data have not been made publicly available due to requirements of research ethics and governance approvals. Requests to collaborate and share data may be directed to the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Kelly Lee Hayward <http://orcid.org/0000-0001-6115-3420>

Patricia C Valery <http://orcid.org/0000-0002-8823-3006>

REFERENCES

- 1 Younossi ZM, Koenig AB, Abdelatif D, *et al*. Global epidemiology of nonalcoholic fatty liver disease – meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
- 2 Valery PC, McPhail S, Stuart KA, *et al*. Changing prevalence of aetiological factors and comorbidities among Australians hospitalised for cirrhosis. *Intern Med J* 2020. doi:10.1111/imj.14809. [Epub ahead of print: 25 Feb 2020].
- 3 Adams LA, Roberts SK, Strasser SI, *et al*. Nonalcoholic fatty liver disease burden: Australia, 2019–2030. *J Gastroenterol Hepatol* 2020;35:1628–35.
- 4 Mahady SE, Adams LA. Burden of non-alcoholic fatty liver disease in Australia. *J Gastroenterol Hepatol* 2018;33:1–11.
- 5 Paik JM, Henry L, Golabi P, *et al*. Presumed nonalcoholic fatty liver disease among Medicare beneficiaries with HIV, 2006–2016. *Open Forum Infect Dis* 2020;7:ofz509.
- 6 Sayiner M, Arshad T, Golabi P, *et al*. Extrahepatic manifestations and healthcare expenditures of non-alcoholic fatty liver disease in the Medicare population. *Hepatol Int* 2020;14:556–66.
- 7 Alexander M, Loomis AK, Fairburn-Beech J, *et al*. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Med* 2018;16:130.
- 8 Pimpin L, Cortez-Pinto H, Negro F, *et al*. Burden of liver disease in Europe: epidemiology and analysis of risk factors to identify prevention policies. *J Hepatol* 2018;69:718–35.
- 9 Australian Institute of Health and Welfare. Overweight & obesity [Online], 2020. Available: <https://www.aihw.gov.au/reports-data/behaviours-risk-factors/overweight-obesity/overview> [Accessed 22 Sep 2020].

- 10 Hayward KL, Patel PJ, Valery PC, *et al.* Medication-related problems in outpatients with decompensated cirrhosis: opportunities for harm prevention. *Hepatol Commun* 2019;3:620–31.
- 11 Valery PC, Bernardes CM, Stuart KA, *et al.* Development and evaluation of the supportive needs assessment tool for cirrhosis (SNAC). *Patient Prefer Adherence* 2020;14:599–611.
- 12 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
- 13 Danford CJ, Lee JY, Strohbehn IA, *et al.* Development of an algorithm to identify cases of nonalcoholic steatohepatitis cirrhosis in the electronic health record. *Dig Dis Sci* 2020. doi:10.1007/s10620-020-06388-y. [Epub ahead of print: 13 Jun 2020].
- 14 European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016;64:1388–402. doi:10.1016/j.jhep.2015.11.004
- 15 Chalasani N, Younossi Z, Lavine JE, *et al.* The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American association for the study of liver diseases. *Hepatology* 2018;67:328–57.
- 16 Powell EE, Jonsson JR, Clouston AD. Steatosis: co-factor in other liver diseases. *Hepatology* 2005;42:5–13.
- 17 Powell EE, Ali A, Clouston AD, *et al.* Steatosis is a cofactor in liver injury in hemochromatosis. *Gastroenterology* 2005;129:1937–43.
- 18 Eslam M, Sarin SK, Wong VW-S, *et al.* The Asian Pacific association for the study of the liver clinical practice guidelines for the diagnosis and management of metabolic associated fatty liver disease. *Hepatol Int* 2020;14:889–919.
- 19 Daly S, Nguyen TQ, Gabbe BJ, *et al.* Agreement between medical record and administrative coding of common comorbidities in orthopaedic trauma patients. *Injury* 2019;50:1277–83.
- 20 Gribsholt SB, Pedersen L, Richelsen B, *et al.* Validity of ICD-10 diagnoses of overweight and obesity in Danish hospitals. *Clin Epidemiol* 2019;11:845–54.
- 21 Pantalone KM, Hobbs TM, Chagin KM, *et al.* Prevalence and recognition of obesity and its associated comorbidities: cross-sectional analysis of electronic health record data from a large US integrated health system. *BMJ Open* 2017;7:e017583.
- 22 Mocarski M, Tian Y, Smolarz BG, *et al.* Use of international classification of diseases, ninth revision codes for obesity: trends in the United States from an electronic health record-derived database. *Popul Health Manag* 2018;21:222–30.
- 23 Samadoulougou S, Idzerda L, Dault R, *et al.* Validated methods for identifying individuals with obesity in health care administrative databases: a systematic review. *Obes Sci Pract* 2020;6:677–93.
- 24 Sheriffdeen A, Millar JL, Martin C, *et al.* (Dis)concordance of comorbidity data and cancer status across administrative datasets, medical charts, and self-reports. *BMC Health Serv Res* 2020;20:858.
- 25 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;27:274–82.
- 26 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;24:107–11.
- 27 Bengtsson B, Askling J, Ludvigsson JF, *et al.* Validity of administrative codes associated with cirrhosis in Sweden. *Scand J Gastroenterol* 2020;55:1205–10.

Detecting non-alcoholic fatty liver disease and risk factors in health databases: accuracy and limitations of the ICD-10-AM

Authors:

Kelly L. Hayward, Amy L. Johnson, Leigh U. Horsfall, Chris Moser, Patricia C. Valery, Elizabeth E. Powell.

Correspondence to:

Prof. Elizabeth E. Powell

Level 5 West Wing, Translational Research Institute, 37 Kent Street, Woolloongabba QLD 4102, Australia

P: +6 1 7 3443 8015

E: e.powell@uq.edu.au

Additional Methods

QHAPDC

The Queensland Hospital Admitted Patient Data Collection (QHAPDC) registry contains clinical and socio-demographic information on all hospital episodes of care for patients admitted to hospital in Queensland, Australia. The state of Queensland adopted ICD-10-AM from ICD-9-AM in the year 1999. Data relating to admitted episodes of care are coded using the current version of ICD-10-AM; Australian Classification of Health Interventions; and Australian Coding Standards, “as a means by which admitted patient activity can be reported, monitored, evaluated, planned for and researched, thereby allowing improved and objective decision-making” (Statistical Services Branch, Queensland Health. QHAPDC standard: Clinical coding. Brisbane, 2020).

Calculating Sensitivity, Specificity, Positive and Negative Predictive Values

Supplementary Table 1A. Measures of accuracy between ICD-10-AM codes and the medical record

Measure	Calculation
Sensitivity <i>Proportion of patients who have the disease/condition that are correctly identified by the ICD-10-AM code(s)</i>	True positive / (true positive + false negative)
Specificity <i>Proportion of patients who do not have the disease/condition that do not have the ICD-10-AM code(s)</i>	True negative / (false positive + true negative)
PPV <i>Proportion of encounters with ICD-10-AM code(s) that correctly identify patients with the disease/condition</i>	True positive / (true positive + false positive)
NPV <i>Proportion of encounters without ICD-10-AM code(s) that correctly identify patients without the disease/condition</i>	True negative / (true negative + false negative)

Supplementary Table 1B. Classification table cells and definitions

		Disease / condition (on medical record review)	
		Present (+)	Absent (-)
Test (ICD-10-AM code)	Code present (+)	True Positive	False Positive
	Code absent (-)	False Negative	True Negative