Title

Administrative Coding in Electronic Health Care Record-Based Research of NAFLD: An Expert Panel Consensus Statement

Journal

Hepatology, 74(1)

ISSN

0270-9139

Authors

Hagström, Hannes
Adams, Leon A
Allen, Alina M
et al.

Publication Date

2021-07-01

DOI

10.1002/hep.31726

Peer reviewed

Published in final edited form as:

Hepatology. 2021 July ; 74(1): 474–482. doi:10.1002/hep.31726.

Administrative coding in electronic health care record-based research of NAFLD: an expert panel consensus statement

Correspondence and reprint requests: Hannes Hagström, C1:77, Division of Hepatology, Karolinska University Hospital, SE-141 86, Stockholm, Sweden, Phone: +46 (0) 8 5858 2328, hannes.hagstrom@ki.se.

Author's contributions: Study conception and design: HH, FÅ
Acquisition of data: All
Statistical analysis: N/A
Analysis and interpretation of data: All
Drafting of manuscript: HH
Critical revision: All

Guarantor of article: HH

All authors approved the final version of the article, including the authorship list.

Writing Assistance: None.

Disclosure / Conflict of interest declaration:
HH: Research grants from Astra Zeneca, Gilead, Intercept. Board advisory for Gilead, Bristol Myers-Squibb.
LA: No conflicts of interest.
AMA: Research grants from Pfizer, Galmed, Target Pharma; advisory/consulting for Ionis, Theratechnologies.
CDB: No conflicts of interest.
YC: No conflicts of interest.
HG: Research grants from ARLA, Abbvie, Intercept, NOVO Nordick Foundation. Advisory board at Ipsen, and speaker at Norgine.
MI: No conflicts of interest.
PJ: No conflicts of interest.
FK: No conflicts of interest.
JK: No conflicts of interest.
JVL: No conflicts of interest.
MTL: Research grants from Echosens Corporation and Gilead Sciences. Advisory board for Ionis Pharmaceuticals.
RL: RL serves as a consultant or advisory board member for Anylam/Regeneron, Amgen, Arrowhead Pharmaceuticals, AstraZeneca, Bristol-Myer Squibb, CohBar, Eli Lilly, Galmed, Gilead, Glympse bio, Inipharm, Intercept, Ionis, Janssen Inc., Merck, Metacrine, Inc., NGM Biopharmaceuticals, Novartis, Novo Nordisk, Pfizer, Sagimet, 89 bio, and Viking Therapeutics. In addition, his institution has received grant support from Allergan, Boehringer-Ingelheim, Bristol-Myers Squibb, Eli Lilly, Galectin Therapeutics, Galmed Pharmaceuticals, Genfit, Gilead, Intercept, Inventiva, Janssen, Madrigal Pharmaceuticals, Merck, NGM Biopharmaceuticals, Pfizer, and Siemens. He is also co-founder of Liponexus, Inc.
PNN: PNN reports, on behalf of the University of Birmingham, grant/research support from Pharmaxis, Boehringer Ingelheim, Echosens and Novo Nordisk, and on behalf of the University of Birmingham consulting fees from BMS, Boehringer Ingelheim, Gilead, Novo Nordisk, Pfizer, and Poxel.
IAR: Personal fees from Abbvie and Roche.
SR: No conflicts of interest
JMS: JMS serves as a consultant or advisory board member for Boehringer Ingelheim, Bristol-Myer Squibb, Echosens, Gilead, Genfit, Intercept, IQVIA, Madrigal, Nordic Bioscience, Novartis, Novo Nordisk, Pfizer, Siemens Healthineers. In addition, his institution has received grant support from Gilead.
MS: Consulting fees from Gilead Sciences
NS: No conflicts of interest.
TGS: Consultant to Aetion
EBT: No conflicts of interest.
SW: Advisory board Gilead Sciences, research grants Novo Nordisk.
VW: Consultant or advisory board member for 3V-BIO, AbbVie, Allergan, Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens, Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk, Perspectum Diagnostics, Pfizer, ProSciento, Sagimet Biosciences, TARGET PharmaSolutions, and Terns; and speaker for AbbVie, Bristol-Myers Squibb, Echosens, and Gilead Sciences. Research grant from Gilead Sciences
YY: Research grants from Gilead, Biocodex. Speaker at Gilead, Echosens.
SZS: No conflicts of interest
FÅ: No conflicts of interest.

Hannes Hagström[1,2,3], **Leon A Adams**[4], **Alina M. Allen**[5], **Christopher D. Byrne**[6,7], **Yoosoo Chang**[8], **Henning Grønbæk**[9], **Mona Ismail**[10,11], **Peter Jepsen**[9], **Fasiha Kanwal**[12], **Jennifer Kramer**[12], **Jeffrey V. Lazarus**[13], **Michelle T. Long**[14], **Rohit Loomba**[15], **Philip N. Newsome**[16,17], **Ian A. Rowe**[18], **Seungho Ryu**[8,19], **Jörn M. Schattenberg**[20], **Marina Serper**[21], **Nick Sheron**[22], **Tracey G. Simon**[23,24], **Elliot B. Tapper**[25], **Sarah Wild**[26], **Vincent Wai-Sun Wong**[27], **Yusuf Yilmaz**[28,29], **Shira Zelber-Sagi**[30], **Fredrik Åberg**[31,32]

[1]Division of Hepatology, Department of Upper GI, Karolinska University Hospital, Stockholm, Sweden

[2]Clinical Epidemiology Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden

[3]Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

[4]Medical School, University of Western Australia, Perth Australia

[5]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

[6]Nutrition and Metabolism, Faculty of Medicine, University of Southampton, UK

[7]Southampton National Institute for Health Research Biomedical Research Centre, University Hospital Southampton, Southampton General Hospital, Tremona Road, Southampton, UK

[8]Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[9]Department of Hepatology and Gastroenterology, Aarhus University Hospital, Aarhus, Denmark

[10]Division of Gastroenterology, Department of Internal Medicine, King Fahad Hospital of the University, Al-Khobar, Saudi Arabia.

[11]College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

[12]Baylor College of Medicine and Michael E. DeBakey VA Medical Center, Houston TX, USA

[13]Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain

[14]Department of Medicine, Section of Gastroenterology, Boston University School of Medicine, Boston, MA, USA

[15]NAFLD Research Center, Division of Gastroenterology and Epidemiology, University of California at San Diego, La Jolla, California, USA

[16]National Institute for Health Research Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust, University of Birmingham, UK,

[17]Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham, UK

[18]Leeds Institute for Medical Research, University of Leeds, UK

[19]Department of Clinical Research Design and Evaluation, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

[20]Metabolic Liver Research Program, I. Department of Medicine, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

[21]Division of Gastroenterology and Hepatology, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[22]Foundation for Liver Research, London, UK

[23]Division of Gastroenterology and Hepatology, Massachusetts General Hospital, Boston, Massachusetts, USA

[24]Harvard Medical School, Boston, MA, USA

[25]Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor, USA

[26]Usher Institute, University of Edinburgh, UK

[27]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong

[28]Liver Research Unit, Institute of Gastroenterology, Marmara University, Turkey

[29]Department of Gastroenterology, School of Medicine, Marmara University, Istanbul, Turkey

[30]School of Public Health, University of Haifa, Haifa, Israel

[31]Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki, Finland

[32]University of Helsinki, Helsinki, Finland

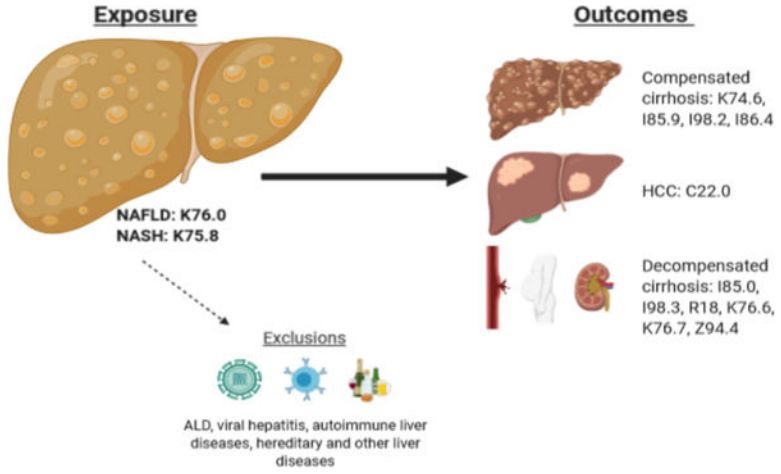
Abstract

Background and aims: Electronic health record (EHR)-based research allows the capture of large amounts of data, which is necessary in nonalcoholic fatty liver disease (NAFLD), where the risk of clinical liver outcomes is generally low. The lack of consensus on which International Classification of Disease (ICD) codes should be used as exposures and outcomes limits comparability and generalizability of results across studies. We aimed to establish consensus among a panel of experts on ICD codes that could become the reference standard and provide guidance around common methodological issues.

Approach and results: Researchers with an interest in EHR-based NAFLD research were invited to collectively define which administrative codes are most appropriate for documenting exposures and outcomes. We used a modified Delphi approach to reach consensus on several commonly encountered methodological challenges in the field. After two rounds of revision, a high level of agreement (>67%) was reached on all items considered. Full consensus was achieved on a comprehensive list of administrative codes to be considered for inclusion and exclusion criteria in defining exposures and outcomes in EHR-based NAFLD research. We also provide suggestions on how to approach commonly encountered methodological issues and identify areas for future research.

Conclusions: This expert panel consensus statement can help harmonize and improve generalizability of EHR-based NAFLD research.

Graphical abstract



Keywords

cirrhosis; epidemiology; health information; hepatocellular carcinoma; MAFLD; NAFLD; registry-based research; steatosis

Introduction

Research using routinely collected data from registries on electronic healthcare records (EHR) is becoming increasingly common with the ongoing digitalization of healthcare and can make valuable contributions to many fields (1). For instance, EHR-based research generally enables the inclusion of a vast number of study participants from multiple sites and, depending on the setting, allows for prospective long-term follow-up or historical cohort data. In nonalcoholic fatty liver disease (NAFLD), there are many examples of EHR-based studies that have provided important insight into the natural history of the disease (2–7) and have helped inform international guidelines (8, 9). This type of research can also be used to identify patients with liver diseases not traditionally seen at hepatology clinics, such as patients with NAFLD seen in primary care or by other hospital-based specialists including endocrinologists and cardiologists (10–14). In addition, population-based health examination studies, which fall under EHR-based research, allow for the identification of persons in the general population who may be unaware of their disease status (15–17). Furthermore, EHR-based studies could be used to assess the safety and real-world effectiveness of new NAFLD drug therapies on patient outcomes outside of clinical trial settings. However, little effort has been made to standardize the definitions of exposures and outcomes in EHR-based NAFLD research, resulting in many different definitions being used by distinct research groups. For instance, no clear guidance exists on how to define the progression of NAFLD to cirrhosis, (i.e. which administrative codes to use), or how to define nonalcoholic steatohepatitis (NASH), and current guidelines do not list suggestions for coding (8, 9). The lack of consensus on definitions threatens the comparability and generalizability of results from different studies. Several methodological considerations, such as how to deal with the risk of misclassification between NAFLD and alcohol-related liver disease (ALD), are also dealt with differently by distinct research

groups. Importantly, few validation studies have been performed to assess the validity of ICD-codes representative of NAFLD/NASH, while codes associated with cirrhosis and hepatocellular carcinoma (HCC) have been validated in different settings (18–23). The aim of this study was to survey researchers, including those engaged in EHR-based research, to understand their views regarding the definitions used to document exposures and outcomes in EHR-based NAFLD research with a focus on liver-related outcomes and to establish a consensus on which codes are most appropriate to use. We chose to focus on diagnostic coding, as many databases do not routinely record data on more detailed parameters such as laboratory data. We have also made some general methodological recommendations for future EHR-based studies in this field.

Materials and methods

A group of researchers with expertise in large-scale database studies in NAFLD field were invited to collaborate. Initial collaborators were then asked to suggest other key researchers to establish a large panel from diverse countries and settings. All invited collaborators that accepted the invitation were asked to supply feedback on an Excel spreadsheet with diagnostic codes relevant to exposures and outcomes in NAFLD research (Table 1). Administrative codes were defined as codes representing the different stages of NAFLD including NASH, fibrosis without cirrhosis, compensated cirrhosis, decompensated cirrhosis, hepatocellular carcinoma (HCC), liver transplantation, and a list of important specific liver diseases other than NAFLD relevant to the studies in this field. We defined all diagnoses using International Classification of Disease (ICD) coding, versions 8–11. Outdated versions of ICD (version 8 and for some countries version 9) were also included as retrospective studies using historical data might benefit from using such versions, and the upcoming ICD-11 version was used as it is likely to replace older versions in the coming years (available at: https://www.who.int/classifications/icd/en/). Additional information sources such as data on procedure codes and coding for specific pharmacotherapies were also considered.

An online survey, utilizing the platform www.sogosurvey.com, was used to ask methodological questions. After all collaborators had given feedback, answers were collated and anonymized and results were shared using a modified Delphi approach (24). Next, an updated spreadsheet was distributed for additional feedback and the survey was re-distributed giving participants an option to change their replies after feedback from the group. Adding additional questions to the survey could be requested by any collaborator. In the end, the questions were defined as statements that generally could be answered with a “yes” or “no” answer. Consensus was defined as greater than 67% agreement with a statement (25). We highlight statements where consensus was between 67% and 90%, as these might be topics for future studies.

Ethical considerations

No patient-level data were included in this study; thus no ethical approval was required.

Results

A total of 27 collaborators were invited, of which 25 (93%) agreed to collaborate, while one did not reply, and one declined to participate. 90% of participants reported using primarily ICD-10 codes, while the remainder reported using the ICD-9 version. No other coding system was used. The final list of suggested ICD-codes is presented in Table 1. Codes are truncated to the left for diagnoses that include several diagnostic alternatives. For instance, when excluding alcohol-related liver disease in the ICD-10 system, all codes starting with "K70" should be excluded. Therefore, we did not present a more detailed level of codes.

The table lists possible exposure and outcome diagnoses as well as exclusion criteria but does not list an order of how to use these, as the usage of codes should be tailored to specific research questions and data availability across cohorts. All collaborators agreed on the final set of codes.

After two rounds of revision, a >67% consensus was established for all 14 recommendations considered in the final version of the survey, with 6/14 questions having a >90% agreement. The recommendations and the percentage of collaborators that agreed with each statement are listed below. The replies to the first round of the survey are presented in the Appendix (eTable 1). Some questions from the first round of the survey were redundant and some were rephrased or added after suggestions from collaborators.

General recommendations

1. Ideally, specific validation studies should be performed for diagnostic codes across different settings. Preferably, validation studies should obtain random patient charts with the diagnostic code in question, and the gold standard should be defined as a pre-specified set of criteria to calculate positive predictive values (e.g. histological or radiological signs of cirrhosis for code K74.6 [cirrhosis of unspecified origin]). It is particularly important to validate codes representative of NAFLD/NASH. (100% agreement)

2. The order of coding in administrative databases should be actively considered. Many systems use primary and contributing diagnoses. A primary diagnosis usually defines the main event of a hospitalization, with contributing diagnoses that might or might not be associated with that event. Coding for cirrhosis can thus be a primary diagnosis while a contributing diagnosis can be made for the etiologic disease, such as NAFLD. For many research questions, primary codes could be sufficient, but when a more granular level of data is required, it could be more appropriate to use combinations of primary and contributing causes, depending on the research question. For some research questions, only using primary coding could lead to a study unable to consider different causes of cirrhosis. (68% agreement)

3. Where available, the setting of the diagnosis can be used to enhance its validity. For example, a code for cirrhosis might have a higher validity when diagnosed in a hepatology clinic versus a primary care center (18). (67% agreement)

Defining NAFLD/NASH

4. Misclassification of other liver diseases can be common in NAFLD, perhaps most commonly occurring with alcohol-related liver disease (ALD). For studies that examine the risk of "pure" NAFLD, exclusion of concurrent liver diseases should be made using diagnosis codes given before or concurrently with the diagnosis of NAFLD. Suggested codes for these diseases are listed in Table 1. Concurrent liver diseases can also be diagnosed after a diagnosis of NAFLD, but the methodology for how to treat these depends on the research question. (74% agreement)

5. To define NASH, the ICD-10 code K75.8 should be primarily used. However, local / national practices should be considered. (82% agreement)

6. Defining and separating NASH from simple steatosis can be difficult, and in general EHR-based research cannot accurately distinguish NASH from simple steatosis in the absence of biopsy. Studies that try to accomplish a high specificity of correctly identifying NASH could use procedure coding for liver biopsy to increase specificity of the NASH diagnosis (i.e. a code of K75.8 + a code for liver biopsy). In general, this approach should also report the time interval between biopsy and NASH diagnosis to increase transparency. (84% agreement)

Ascertaining progression to cirrhosis

7. When ascertaining progression to cirrhosis, specific coding for cirrhosis needs to be present to define progression to cirrhosis. Studies should not count events, such as hospitalizations, where only coding for NAFLD/NASH is available, as defining cirrhosis. (100% agreement)

8. Liver disease outcomes are generally rare in most populations. A composite outcome, including diagnoses related to complications of cirrhosis of high validity, should generally be applied when ascertaining outcomes associated with cirrhosis. Our recommendations for these are specified in Table 1. For instance, for a person where bleeding esophageal varices are found but there is no diagnosis of cirrhosis or competing cause such as portal vein thrombosis, they should be counted as having cirrhosis. However, the etiology of cirrhosis should be individually determined. Special considerations are also needed for potential liver-related outcomes that could be caused by other underlying conditions (see statement #11). (94% agreement)

Aspects on coding for cirrhosis

9. Specific coding for NAFLD/NASH in patients with cirrhosis might be missing in some registries/databases. Coding for metabolic syndrome components (diabetes type 2, obesity, hypertension, or hyperlipidemia) at or before the cirrhosis event may be considered to improve sensitivity of capturing NAFLD-related cirrhosis outcomes. Such an approach should also be validated in accordance with statement #1. (79% agreement)

10. For hepatocellular carcinoma outcomes, the specific code for HCC should be used (ICD-10: C22.0), as coding for “Liver cancer, unspecified” (ICD-10: C22.9) is likely to have a low specificity for HCC. However, the C22.9 code can be used for sensitivity analyses. (94% agreement)

11. Ascites may be caused by cirrhosis, but there are other causes of non-hepatic ascites such as heart failure or malignancy. In studies of patients with known chronic liver disease status (such as NAFLD, ALD, or cirrhosis), ascites can be used as a liver-related outcome, but in general in population cohorts, ascites should be combined with coding for chronic liver disease to achieve an acceptable specificity. (94% agreement)

12. Hepatic encephalopathy (HE) does not have a specific code in ICD-10. To define HE, coding for compensated or decompensated cirrhosis can be combined with coding for prescriptions (such as Anatomical Therapeutic Chemical Classification System [ATC] codes) of lactulose or rifaximin to increase specificity, in settings where such data are available. For instance, a person with a diagnosis of cirrhosis and a prescription of rifaximin can be considered as having HE. (78% agreement)

13. Coding for liver failure might represent cirrhosis, but there are alternate causes and specificity might be low. Unless validation studies are performed in the system of the study setting, chronic liver failure codes (ICD-10: K72.1, K72.9) should only be considered as defining cirrhosis in sensitivity analyses, and acute liver failure codes should not be used when defining cirrhosis. (95% agreement)

14. Procedure codes can in general be used to ascertain outcomes related to cirrhosis and can be used alone even if specific coding for cirrhosis is lacking. For example, a case with coding for banding of bleeding esophageal varices, but no formal coding for varices per se, can help define decompensated cirrhosis. However, in contrast to the diagnostic ICD-systems, there are a multitude of different procedure code systems and a list of procedure codes should be defined by the individual research group depending on the setting. (83% agreement)

Procedure codes that could be interesting would depend on the research question. Liver transplantation is a commonly used outcome that can be studied in itself, or as part of a composite outcome. There are certain ICD-codes for the presence of a liver transplant (ICD-10: Z94.4), but procedure coding could also be added to define liver transplantation and would be better to define the date of the transplantation.

Procedures that are strongly associated with cirrhosis include banding or ligation of esophageal or gastric varices, paracentesis, and transjugular intrahepatic porto-systemic shunt placement.

Discussion

We present a comprehensive list of ICD codes that can serve as guidance for future studies of NAFLD when using EHR-data, for instance how to define NAFLD in study cohorts

and how to ascertain progression to cirrhosis. This guidance could be used both when using administrative and clinical databases, as well as upcoming phase 4-studies in the NAFLD/NASH field. Harmonization of definitions will make it easier to compare and contrast study results from different cohorts, leading to an improved understanding of the consequences of NAFLD. These recommendations could also provide guidance for clinicians or administrators responsible for coding disease status in patients with NAFLD.

We also give guidance on some key methodological questions in the field.

However, no guidance can cover all possible research questions or study settings, which is why these recommendations should not be a mandate but rather a suggestion of relevant codes that can be used to define exposure or outcome variables. Some registries/databases might contain more detailed data, such as laboratory data or results from radiology, which allow for more granular definitions than made here. Moreover, some regions may have different coding systems or coding practices so that our recommendations could be irrelevant.

This should not be seen as a complete guide on how to perform research in this field. Different research questions and study settings will require distinct strategies. For instance, examining liver-related outcomes in an EHR-derived general population with unknown NAFLD status is different than examining liver-related endpoints in clinical cohorts with known NAFLD status. EHR-derived population-based cohorts often have a low risk for cirrhosis and have the possibility of having liver diseases other than NAFLD, while the definition of outcomes could be made narrower by combining codes through pre-specified algorithms. An example of this is requiring a code for chronic liver disease when ascites is diagnosed, as there are other causes of ascites apart from cirrhosis, thus risking false-positive findings. This can be compared to EHR-based follow-up studies of persons with known NAFLD, where it is less likely that ascites would be related to competing causes. Again, definitions should be tailored to the specific research question. Also, liver-related outcomes are relatively rare in a general population (26, 27), and consideration of examining more common causes of mortality should be actively considered, depending on the research question.

When assessing the overall impact of NAFLD on clinical liver outcomes in the general population, there is a need to recognize that NAFLD often co-exists with alcohol-related liver disease, and that metabolic risk factors and alcohol use have bi-directional interaction effects on liver disease with in general higher risk for liver-related outcomes in persons with mixed etiologies (28, 29). In such research settings, the exclusion of competing and co-existent causes for liver disease may lead to lower risk estimates associated with NAFLD. For instance, excluding diagnoses of alcohol-related outcomes (ICD-10: K70.3) will reduce the sensitivity of the impact of NAFLD in persons with co-existing NAFLD and ALD, and thought need to be given to the possible co-existence of NAFLD and ALD when designing EHR-based studies depending on the research question. This is true for most use of algorithms using ICD-coding. The requirement of multiple codes to enhance specificity will almost always lead to a lower sensitivity and under-appreciation of the prevalence of the target diagnosis, which should be considered depending on the research question(s).

It should also be acknowledged that NAFLD is severely under-coded in the general population (30, 31), although there are methods to enhance detection of NAFLD in databases with high level of details (32, 33). Consequently, there is often misclassification bias (due to inaccurate coding of subjects in 'control' groups), resulting in attenuation of any association towards the null, but also selection bias (cases diagnosed with NAFLD might be the more severe cases, resulting in inflated risk estimates).

Preferably, validation studies of codes or algorithms used to increase the positive predictive value of the target diagnosis should be undertaken before starting an EHR-based study, if possible. Examples of these are (18, 19, 34). Results of prior validation studies in hepatology in general suggest that while some diagnoses have a high positive predictive value (PPV), there are several non-specific diagnoses where considerations need to be made. An example is ascites which can be found in persons without cirrhosis, but combining coding for ascites with coding for cirrhosis leads to acceptable PPVs (18). Another example is hepatic encephalopathy, where combining drug prescription coding with cirrhosis increases PPVs to acceptable levels (19). One must also take into account the setting where the diagnosis is made. For instance, a diagnosis of "chronic liver failure" made by a hepatologist might have higher validity when compared to the same diagnosis made in a general hospital (18).

Areas of controversy and considerations for future research

Statements where less than 90% of collaborators agreed can be considered to be somewhat controversial and may represent areas for future studies. For instance, how to best classify disease etiology in persons in a general population setting that develop cirrhosis can be difficult when there is no etiological coding, meaning that differentiating between e.g. ALD and NAFLD as the cause of cirrhosis can be difficult.

Similarly, how to deal with study subjects with multiple etiologic codes (e.g. NAFLD and ALD) at different timepoints during study follow-up was considered to be too nuanced and study dependent to make a general recommendation and could be a topic for a future study. As NAFLD is severely under-coded, future studies should also consider how to best use available ICD-coding such as combinations of metabolic syndrome components to achieve an acceptable accuracy for estimation of NAFLD in the general population.

Additionally, full consensus could not be achieved on how to best classify HE in the absence of a specific ICD-code. Some collaborators thought that defining HE by codes for cirrhosis and a prescription code for lactulose risked misclassifying persons where lactulose might be used for primary prevention.

Finally, we chose to limit our discussion to liver-related outcomes, and how to best examine the role of NAFLD in cardiovascular disease using EHR-data remains an open question.

Conclusion

We defined a list of ICD-codes that can be considered by investigators examining NAFLD in electronic healthcare records-based research and reached consensus statements addressing several methodological questions. This guidance is intended to streamline future studies in

this field, leading to increased generalizability of study results, with the aim to improve our understanding of NAFLD prognosis pertaining to liver related outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Grant Support (Funding):

None specific for this study.

RL receives funding support from NIEHS (5P42ES010337), NCATS (5UL1TR001442), NIDDK (U01DK061734, R01DK106419, P30DK120515, R01DK121378, R01DK124318), NHLBI (P01HL147835), and DOD PRCRP (W81XWH-18–2–0026).

AA receives funding support from NIH (DK115594).

Abbreviations:

AIH	autoimmune hepatitis
ATC	Anatomical Therapeutic Chemical Classification System
ALD	alcohol-related liver disease
EHR	electronic healthcare records
HCC	hepatocellular carcinoma
HE	hepatic encephalopathy
ICD	international classification of diseases
NAFLD	nonalcoholic fatty liver disease
NASH	nonalcoholic steatohepatitis
PBC	primary biliary cholangitis
PPV	positive predictive value
PSC	primary sclerosing cholangitis

References

1. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. Hepatology 2006;44:1075–1082. [PubMed: 17058242]
2. Ekstedt M, Hagstrom H, Nasr P, Fredrikson M, Stal P, Kechagias S, Hultcrantz R. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. Hepatology 2015;61:1547–1554. [PubMed: 25125077]
3. Feldstein AE, Charatcharoenwitthaya P, Treeprasertsuk S, Benson JT, Enders FB, Angulo P. The natural history of non-alcoholic fatty liver disease in children: a follow-up study for up to 20 years. Gut 2009;58:1538–1544. [PubMed: 19625277]

4. Hagstrom H, Nasr P, Ekstedt M, Hammar U, Stal P, Hultcrantz R, Kechagias S. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265–1273. [PubMed: 28803953]
5. Jepsen P, Vilstrup H, Mellemkjaer L, Thulstrup AM, Olsen JH, Baron JA, Sørensen HT. Prognosis of patients with a diagnosis of fatty liver--a registry-based cohort study. Hepatogastroenterology 2003;50:2101–2104. [PubMed: 14696473]
6. Kaps L, Labenz C, Galle PR, Weinmann-Menke J, Kostev K, Schattenberg JM. Non-alcoholic fatty liver disease increases the risk of incident chronic kidney disease. United European Gastroenterol J 2020;8:942–948.
7. Wild SH, Walker JJ, Morling JR, McAllister DA, Colhoun HM, Farran B, McGurnaghan S, et al. Cardiovascular Disease, Cancer, and Mortality Among People With Type 2 Diabetes and Alcoholic or Nonalcoholic Fatty Liver Disease Hospital Admission. Diabetes Care 2018;41:341–347. [PubMed: 29167212]
8. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016;64:1388–1402. [PubMed: 27062661]
9. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, 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;67:328–357. [PubMed: 28714183]
10. Alexander M, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, Pasqua A, et al. Non-alcoholic fatty liver disease and risk of incident acute myocardial infarction and stroke: findings from matched cohort study of 18 million European adults. Bmj 2019;367:l5367. [PubMed: 31594780]
11. Alexander M, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, Pasqua A, et al. Risks and clinical predictors of cirrhosis and hepatocellular carcinoma diagnoses in adults with diagnosed NAFLD: real-world study of 18 million patients in four European cohorts. BMC Med 2019;17:95. [PubMed: 31104631]
12. Bengtsson B, Stal P, Wahlin S, Bjorkstrom NK, Hagstrom H. Characteristics and outcome of hepatocellular carcinoma in patients with NAFLD without cirrhosis. Liver Int 2019;39:1098–1108. [PubMed: 30829446]
13. Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year-community study. Hepatology 2018;67:1726–1736. [PubMed: 28941364]
14. Allen AM, Van Houten HK, Sangaralingham LR, Talwalkar JA, McCoy RG. Healthcare Cost and Utilization in Nonalcoholic Fatty Liver Disease: Real-World Data From a Large U.S. Claims Database. Hepatology 2018;68:2230–2238. [PubMed: 29774589]
15. Åberg F, Puukka P, Salomaa V, Männistö S, Lundqvist A, Valsta L, Perola M, et al. Risks of Light and Moderate Alcohol Use in Fatty Liver Disease: Follow-Up of Population Cohorts. Hepatology 2020;71:835–848. [PubMed: 31323122]
16. Unalp-Arida A, Ruhl CE. Patatin-Like Phospholipase Domain-Containing Protein 3 I148M and Liver Fat and Fibrosis Scores Predict Liver Disease Mortality in the U.S. Population. Hepatology 2020;71:820–834. [PubMed: 31705824]
17. Kanwal F, Kramer JR, Mapakshi S, Natarajan Y, Chayanupatkul M, Richardson PA, Li L, et al. Risk of Hepatocellular Cancer in Patients With Non-Alcoholic Fatty Liver Disease. Gastroenterology 2018;155:1828–1837.e1822. [PubMed: 30144434]
18. Bengtsson B, Askling J, Ludvigsson JF, Hagström H. Validity of administrative codes associated with cirrhosis in Sweden. Scand J Gastroenterol 2020:1–6.
19. Tapper EB, Korovaichuk S, Baki J, Williams S, Nikirk S, Waljee AK, Parikh ND. Identifying Patients With Hepatic Encephalopathy Using Administrative Data in the ICD-10 Era. Clin Gastroenterol Hepatol 2019.
20. Lapointe-Shaw L, Georgie F, Carlone D, Cerocchi O, Chung H, Dewit Y, Feld JJ, et al. Identifying cirrhosis, decompensated cirrhosis and hepatocellular carcinoma in health administrative data: A validation study. PLoS One 2018;13:e0201120. [PubMed: 30133446]

21. Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V 3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. Pharmacoepidemiol Drug Saf 2012;21:765–769. [PubMed: 22674685]
22. Lo Re V 3rd, Lim JK, Goetz MB, Tate J, Bathulapalli H, Klein MB, Rimland D, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. Pharmacoepidemiol Drug Saf 2011;20:689–699. [PubMed: 21626605]
23. Kaplan DE, Dai F, Aytaman A, Baytarian M, Fox R, Hunt K, Knott A, et al. Development and Performance of an Algorithm to Estimate the Child-Turcotte-Pugh Score From a National Electronic Healthcare Database. Clin Gastroenterol Hepatol 2015;13:2333–2341.e2331–2336. [PubMed: 26188137]
24. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2:i–iv, 1–88.
25. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, Wales PW. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014;67:401–409. [PubMed: 24581294]
26. Hagstrom H, Talback M, Andreasson A, Walldius G, Hammar N. Ability of Noninvasive Scoring Systems to Identify Individuals in the Population at Risk for Severe Liver Disease. Gastroenterology 2019.
27. Hagstrom H, Tynelius P, Rasmussen F. High BMI in late adolescence predicts future severe liver disease and hepatocellular carcinoma: a national, population-based cohort study in 1.2 million men. Gut 2018;67:1536–1542. [PubMed: 28320770]
28. Åberg F, Färkkilä M. Drinking and Obesity: Alcoholic Liver Disease/Nonalcoholic Fatty Liver Disease Interactions. Semin Liver Dis 2020;40:154–162. [PubMed: 32069503]
29. Liu B, Balkwill A, Reeves G, Beral V, Million Women Study C. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. British Medical Journal 2010;340:7.
30. Blais P, Husain N, Kramer JR, Kowalkowski M, El-Serag H, Kanwal F. Nonalcoholic fatty liver disease is underrecognized in the primary care setting. Am J Gastroenterol 2015;110:10–14. [PubMed: 24890441]
31. Alexander M, Loomis AK, Fairburn-Beech J, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. BMC Med 2018;16:130. [PubMed: 30099968]
32. Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB, Kanwal F. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Aliment Pharmacol Ther 2014;40:949–954. [PubMed: 25155259]
33. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and Validation of an Algorithm to Identify Nonalcoholic Fatty Liver Disease in the Electronic Medical Record. Dig Dis Sci 2016;61:913–919. [PubMed: 26537487]
34. Driver RJ, Balachandrakumar V, Burton A, Shearer J, Downing A, Cross T, Morris E, et al. Validation of an algorithm using inpatient electronic health records to determine the presence and severity of cirrhosis in patients with hepatocellular carcinoma in England: an observational study. BMJ Open 2019;9:e028571.

Table 1.

List of suggested ICD-codes for versions 8–11 with applicable codes for different diagnoses.

Diagnosis	ICD-8	ICD-9	ICD-10	ICD-11 (version 04/19)
Defining NAFLD at baseline				
NAFLD, all	No available code	571.8	K76.0	DB92.0, DB92.Z DB92.Y
NASH	No available code	No available code	K75.8	DB92.1
Excluding other liver diseases at/before baseline				
ALD	571,00, 571,01	571.0–571.3	K70	DB94
Viral hepatitis	999,2, 070	070	B16, B17, B18, B19	1E50, 1E51, 1E5Z
Autoimmune liver disease (AIH, PBC, PSC)	No available code	571.6, 576.1	K83.0A, K83.0F, K74.3, K75.4	DB96
Hemochromatosis	273,2	275.0	E83.1	5C64.1
Wilson	273,3	275.1	E83.0B	5C64.0
Alpha-1-antitrypsin deficiency	No available code	277.6	E88.0A, E88.0B	5C5A
Budd-Chiari	No available code	453.0	I82.0, K76.5	DB98.5
Chronic hepatitis, unspecified	570	571.4	K73.9, K73.2	DB97.2
Secondary or unspecified biliary cirrhosis	No available code	571.6	K74.4, K74.5	DB93.2
Excluding alcohol/drug use disorder at/before baseline				
Codes associated with alcohol use disorder	303	303, 305.0	F10	6C40
Codes associated with somatic consequences of alcohol (except ALD)	291, 980,00, 980,01, 980,99	291, 357.5, 425.5, 535.3, 980.1, 980.9	E24.4, G62.1, I42.6, K29.2, G31.2, G72.1, K85.2, K86.0, T51.0, T51.9, Y57.3, X65, Z50.2, Z71.4, Z72.1	5A70.2, 8D44.0, 8D44.1, BC43.01, DA42.80, 8A03.30, 8D44.Y, DC31.1, DC32.3, QB95.2, QE10, 6D72.10, DA51.50, 6D84.0
Codes associated with drug use disorders except nicotine/caffeine	No available code	305.1–9	F11-F14, F16, F18, F19	6C41–6C47, 6C4B-H
Compensated cirrhosis				
Cirrhosis, compensated	571,9	571.5	K74.6	DB93.1
Esophageal varices, not bleeding	No available code	456.1, 456.21	I85.9, I98.2	DA26.01, DA26.0Z
Gastric varices, not bleeding	No available code	No available code	I86.4	DA43.0
Decompensated cirrhosis				
Esophageal varices, bleeding	456.0	456.0, 456.20	I85.0, I98.3	DA26.00

Diagnosis	ICD-8	ICD-9	ICD-10	ICD-11 (version 04/19)
Ascites	785.3	789.5	R18	ME04
Hepatic encephalopathy	No available code	572.2	Code for cirrhosis + prescription for lactulose/rifaximin	DB99.5
Hepatorenal syndrome	No available code	572.4	K76.7	DB99.2
Portal hypertension	571,9	572.3	K76.6	DB98.7
Liver transplantation				
Liver transplantation status	No available code	V427	Z94.4	QB63.3
Liver cancer outcomes				
HCC	155,01	155.0	C22.0	2C12.0
Liver cancer, unspecified	No available code	155.2	C22.9	2C12.Z
“Unspecific” codes that might be relevant for some studies				
Chronic or unspecified liver failure	573	572.8	K72.1, K72.9	DB99.7, DB99.8
Acute or subacute liverfailure	570	570	K72.0	DB91
Portal vein thrombosis	No available code	No available code	I81.9, K75.1	DB98.0
Hepatic fibrosis or sclerosis or fibrosis with sclerosis	No available code	571,9	K74.0, K74.1, K74.2	DB93.0

Abbreviations: AIH, autoimmune hepatitis; ALD, alcohol-related liver disease; HCC, hepatocellular carcinoma; ICD, International Classification of Disease; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cholangitis: PSC, primary sclerosing cholangitis.