

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

3

4 Short title: Identifying NAFLD using EHR

5

6 Hannes Hagström^{1,2,3}, Leon A Adams⁴, Alina M. Allen⁵, Christopher D. Byrne^{6,7},

7 Yoosoo Chang⁸, Henning Grønbaek⁹, Mona Ismail^{10,11}, Peter Jepsen⁹, Fasiha

8 Kanwal¹², Jennifer Kramer¹², Jeffrey V. Lazarus¹³, Michelle T. Long¹⁴, Rohit

9 Loomba¹⁵, Philip N. Newsome^{16,17}, Ian A. Rowe¹⁸, Seungho Ryu^{8,19}, Jörn M.

10 Schattenberg²⁰, Marina Serper²¹, Nick Sheron²², Tracey G. Simon^{23,24}, Elliot B.

11 Tapper²⁵, Sarah Wild²⁶, Vincent Wai-Sun Wong²⁷, Yusuf Yilmaz^{28,29} Shira Zelber-

12 Sagi³⁰, Fredrik Åberg^{31,32}

13

14 ¹ Division of Hepatology, Department of Upper GI, Karolinska University Hospital,

15 Stockholm, Sweden

16 ² Clinical Epidemiology Unit, Department of Medicine, Solna, Karolinska Institutet,

17 Stockholm, Sweden

18 ³ Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

19 ⁴ Medical School, University of Western Australia, Perth Australia

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

21 ⁶ Nutrition and Metabolism, Faculty of Medicine, University of Southampton, UK

22 ⁷ Southampton National Institute for Health Research Biomedical Research Centre,

23 University Hospital Southampton, Southampton General Hospital, Tremona Road,

24 Southampton, UK

1 ⁸ Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital,
2 Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
3 ⁹ Department of Hepatology and Gastroenterology, Aarhus University Hospital,
4 Aarhus, Denmark
5 ¹⁰ Division of Gastroenterology, Department of Internal Medicine, King Fahad
6 Hospital of the University, Al-Khobar, Saudi Arabia.
7 ¹¹ College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi
8 Arabia
9 ¹² Baylor College of Medicine and Michael E. DeBakey VA Medical Center, Houston
10 TX, USA
11 ¹³ Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of
12 Barcelona, Barcelona, Spain
13 ¹⁴ Department of Medicine, Section of Gastroenterology, Boston University School of
14 Medicine, Boston, MA, USA
15 ¹⁵ NAFLD Research Center, Division of Gastroenterology and Epidemiology,
16 University of California at San Diego, La Jolla, California, USA
17 ¹⁶ National Institute for Health Research Biomedical Research Centre at University
18 Hospitals Birmingham NHS Foundation Trust, University of Birmingham, UK,
19 ¹⁷ Centre for Liver and Gastrointestinal Research, Institute of Immunology and
20 Immunotherapy, University of Birmingham, UK
21 ¹⁸ Leeds Institute for Medical Research, University of Leeds, UK
22 ¹⁹ Department of Clinical Research Design and Evaluation, SAIHST, Sungkyunkwan
23 University, Seoul, Republic of Korea
24 ²⁰ Metabolic Liver Research Program, I. Department of Medicine, University Medical
25 Center of the Johannes Gutenberg-University, Mainz, Germany

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

3 ²² Foundation for Liver Research, London, UK

4 ²³ Division of Gastroenterology and Hepatology, Massachusetts General Hospital,
5 Boston, Massachusetts, USA

6 ²⁴ Harvard Medical School, Boston, MA, USA

7
8 ²⁵ Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor,
9 USA

10 ²⁶ Usher Institute, University of Edinburgh, UK

11 ²⁷ Department of Medicine and Therapeutics, The Chinese University of Hong Kong,
12 Hong Kong

13 ²⁸ Liver Research Unit, Institute of Gastroenterology, Marmara University, Turkey

14 ²⁹ Department of Gastroenterology, School of Medicine, Marmara University,
15 Istanbul, Turkey

16 ³⁰ School of Public Health, University of Haifa, Haifa, Israel

17 ³¹ Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki,
18 Finland

19 ³² University of Helsinki, Helsinki, Finland

20

21 Correspondence and reprint requests:

22 Hannes Hagström, C1:77, Division of Hepatology, Karolinska University Hospital,

23 SE-141 86 Stockholm, Sweden

24 Phone: +46 (0) 8 5858 2328

25 E-mail: hannes.hagstrom@ki.se

1 **Keywords:** cirrhosis; epidemiology; health information; hepatocellular carcinoma;
2 MAFLD; NAFLD; registry-based research; steatosis

3

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

10

11 **Total word count: 4 241 (introduction – references). Abstract: 212. Tables: 1.**

12 **Figures: 1. Character count for title: 123**

13

14 **Disclosure / Conflict of interest declaration:**

15 HH: Research grants from Astra Zeneca, Gilead, Intercept. Board advisory for Gilead,
16 Bristol Myers-Squibb.

17 LA: No conflicts of interest.

18 AMA: Research grants from Pfizer, Galmed, Target Pharma; advisory/consulting for
19 Ionis, Theratechnologies.

20 CDB: No conflicts of interest.

21 YC: No conflicts of interest.

22 HG: Research grants from ARLA, Abbvie, Intercept, NOVO Nordick Foundation.
23 Advisory board at Ipsen, and speaker at Norgine.

24 MI: No conflicts of interest.

1 PJ: No conflicts of interest.

2 FK: No conflicts of interest.

3 JK: No conflicts of interest.

4 JVL: No conflicts of interest.

5 MTL: Research grants from Echosens Corporation and Gilead Sciences. Advisory
6 board for Ionis Pharmaceuticals.

7 RL: RL serves as a consultant or advisory board member for Anylam/Regeneron,
8 Amgen, Arrowhead Pharmaceuticals, AstraZeneca, Bristol-Myer Squibb, CohBar, Eli
9 Lilly, Galmed, Gilead, Glympse bio, Inipharm, Intercept, Ionis, Janssen Inc., Merck,
10 Metacrine, Inc., NGM Biopharmaceuticals, Novartis, Novo Nordisk, Pfizer, Sagimet,
11 89 bio, and Viking Therapeutics. In addition, his institution has received grant support
12 from Allergan, Boehringer-Ingelheim, Bristol-Myers Squibb, Eli Lilly, Galectin
13 Therapeutics, Galmed Pharmaceuticals, Genfit, Gilead, Intercept, Inventiva, Janssen,
14 Madrigal Pharmaceuticals, Merck, NGM Biopharmaceuticals, Pfizer, and Siemens. He
15 is also co-founder of Liponexus, Inc.

16 PNN: PNN reports, on behalf of the University of Birmingham, grant/research
17 support from Pharmaxis, Boehringer Ingelheim, Echosens and Novo Nordisk, and on
18 behalf of the University of Birmingham consulting fees from BMS, Boehringer
19 Ingelheim, Gilead, Novo Nordisk, Pfizer, and Poxel.

20 IAR: Personal fees from Abbvie and Roche.

21 SR: No conflicts of interest

22 JMS: JMS serves as a consultant or advisory board member for Boehringer
23 Ingelheim, Bristol-Myer Squibb, Echosens, Gilead, Genfit, Intercept, IQVIA,

1 Madrigal, Nordic Bioscience, Novartis, Novo Nordisk, Pfizer, Siemens Healthineers.

2 In addition, his institution has received grant support from Gilead.

3 MS: Consulting fees from Gilead Sciences

4 NS: No conflicts of interest.

5 TGS: Consultant to Aetion

6 EBT: No conflicts of interest.

7 SW: Advisory board Gilead Sciences, research grants Novo Nordisk.

8 VW: Consultant or advisory board member for 3V-BIO, AbbVie, Allergan,

9 Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens,

10 Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk,

11 Perspectum Diagnostics, Pfizer, ProSciento, Sagimet Biosciences, TARGET

12 PharmaSolutions, and Terns; and speaker for AbbVie, Bristol-Myers Squibb,

13 Echosens, and Gilead Sciences. Research grant from Gilead Sciences

14 YY: Research grants from Gilead, Biocodex. Speaker at Gilead, Echosens.

15 SZS: No conflicts of interest

16 FÅ: No conflicts of interest.

17

18 **Grant Support (Funding):** None specific for this study.

19 RL receives funding support from NIEHS (5P42ES010337), NCATS

20 (5UL1TR001442), NIDDK (U01DK061734, R01DK106419, P30DK120515,

21 R01DK121378, R01DK124318), NHLBI (P01HL147835), and DOD PRCRP

22 (W81XWH-18-2-0026).

23 AA receives funding support from NIH (DK115594).

24

- 1 **Author's contributions:** Study conception and design: HH, FÅ
- 2 Acquisition of data: All
- 3 Statistical analysis: N/A
- 4 Analysis and interpretation of data: All
- 5 Drafting of manuscript: HH
- 6 Critical revision: All
- 7 **Guarantor of article:** HH
- 8 All authors approved the final version of the article, including the authorship list.
- 9 **Writing Assistance:** None.

1 **Abstract**

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

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

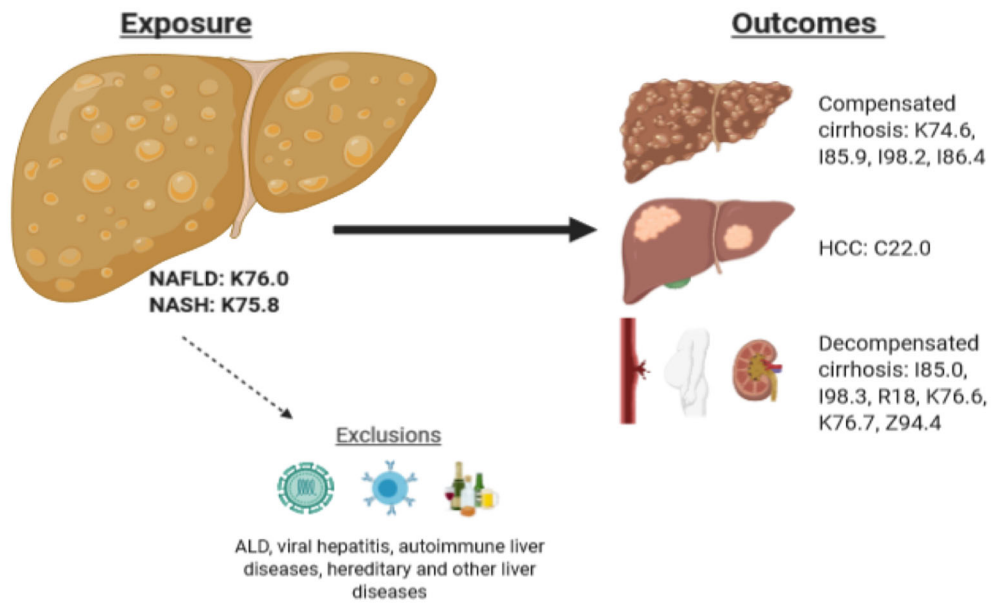
20 Conclusions: This expert panel consensus statement can help harmonize and improve
21 generalizability of EHR-based NAFLD research.

22

23

1 **Graphical abstract**

2



3

4

1 Introduction

2 Research using routinely collected data from registries on electronic healthcare
3 records (EHR) is becoming increasingly common with the ongoing digitalization of
4 healthcare and can make valuable contributions to many fields (1). For instance,
5 EHR-based research generally enables the inclusion of a vast number of study
6 participants from multiple sites and, depending on the setting, allows for prospective
7 long-term follow-up or historical cohort data. In nonalcoholic fatty liver disease
8 (NAFLD), there are many examples of EHR-based studies that have provided
9 important insight into the natural history of the disease (2-7) and have helped inform
10 international guidelines (8, 9). This type of research can also be used to identify
11 patients with liver diseases not traditionally seen at hepatology clinics, such as
12 patients with NAFLD seen in primary care or by other hospital-based specialists
13 including endocrinologists and cardiologists (10-14). In addition, population-based
14 health examination studies, which fall under EHR-based research, allow for the
15 identification of persons in the general population who may be unaware of their
16 disease status (15-17). Furthermore, EHR-based studies could be used to assess the
17 safety and real-world effectiveness of new NAFLD drug therapies on patient
18 outcomes outside of clinical trial settings.

19 However, little effort has been made to standardize the definitions of exposures and
20 outcomes in EHR-based NAFLD research, resulting in many different definitions
21 being used by distinct research groups. For instance, no clear guidance exists on how
22 to define the progression of NAFLD to cirrhosis, (i.e. which administrative codes to
23 use), or how to define nonalcoholic steatohepatitis (NASH), and current guidelines do
24 not list suggestions for coding (8, 9). The lack of consensus on definitions threatens
25 the comparability and generalizability of results from different studies. Several

1 methodological considerations, such as how to deal with the risk of misclassification
2 between NAFLD and alcohol-related liver disease (ALD), are also dealt with
3 differently by distinct research groups. Importantly, few validation studies have been
4 performed to assess the validity of ICD-codes representative of NAFLD/NASH, while
5 codes associated with cirrhosis and hepatocellular carcinoma (HCC) have been
6 validated in different settings (18-23).

7 The aim of this study was to survey researchers, including those engaged in EHR-
8 based research, to understand their views regarding the definitions used to document
9 exposures and outcomes in EHR-based NAFLD research with a focus on liver-related
10 outcomes and to establish a consensus on which codes are most appropriate to use.
11 We chose to focus on diagnostic coding, as many databases do not routinely record
12 data on more detailed parameters such as laboratory data. We have also made some
13 general methodological recommendations for future EHR-based studies in this field.

14 **Materials and methods**

15 A group of researchers with expertise in large-scale database studies in NAFLD field
16 were invited to collaborate. Initial collaborators were then asked to suggest other key
17 researchers to establish a large panel from diverse countries and settings. All invited
18 collaborators that accepted the invitation were asked to supply feedback on an Excel
19 spreadsheet with diagnostic codes relevant to exposures and outcomes in NAFLD
20 research (**Table 1**). Administrative codes were defined as codes representing the
21 different stages of NAFLD including NASH, fibrosis without cirrhosis, compensated
22 cirrhosis, decompensated cirrhosis, hepatocellular carcinoma (HCC), liver
23 transplantation, and a list of important specific liver diseases other than NAFLD

1 relevant to the studies in this field. We defined all diagnoses using International
2 Classification of Disease (ICD) coding, versions 8-11. Outdated versions of ICD
3 (version 8 and for some countries version 9) were also included as retrospective
4 studies using historical data might benefit from using such versions, and the
5 upcoming ICD-11 version was used as it is likely to replace older versions in the
6 coming years (available at: <https://www.who.int/classifications/icd/en/>). Additional
7 information sources such as data on procedure codes and coding for specific
8 pharmacotherapies were also considered.

9

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

20 Ethical considerations

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

23

24

1 **Results**

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

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

14

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

22

23

24

25

1 **General recommendations**

2

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

10

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

22

23 3. Where available, the setting of the diagnosis can be used to enhance its
24 validity. For example, a code for cirrhosis might have a higher validity when

1 diagnosed in a hepatology clinic versus a primary care center (18). (67%
2 agreement)

4 **Defining NAFLD/NASH**

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

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

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

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)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

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

1 acute liver failure codes should not be used when defining cirrhosis. (95%
2 agreement)

3

4 14. Procedure codes can in general be used to ascertain outcomes related to
5 cirrhosis and can be used alone even if specific coding for cirrhosis is lacking.
6 For example, a case with coding for banding of bleeding esophageal varices,
7 but no formal coding for varices per se, can help define decompensated
8 cirrhosis. However, in contrast to the diagnostic ICD-systems, there are a
9 multitude of different procedure code systems and a list of procedure codes
10 should be defined by the individual research group depending on the setting.
11 (83% agreement)

12

13 Procedure codes that could be interesting would depend on the research question.

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

18

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

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,	5A70.2, 8D44.0, 8D44.1, BC43.01, DA42.80, 8A03.30, 8D44.Y, DC31.1,

			T51.9, Y57.3, X65, Z50.2, Z71.4, Z72.1	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
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 liver failure	570	570	K72.0	DB91
Portal vein thrombosis	No available code	No available code	I81.9, K75.1	DB98.3
Hepatic fibrosis or sclerosis or fibrosis with sclerosis	No available code	571,9	K74.0, K74.1, K74.2	DB93.0

1

2 **Table 1.** List of suggested ICD-codes for versions 8-11 with applicable codes for different diagnoses.

3 Abbreviations: AIH, autoimmune hepatitis; ALD, alcohol-related liver disease; HCC, hepatocellular carcinoma; ICD, International Classification of Disease;

4 NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

5

1 Discussion

2 We present a comprehensive list of ICD codes that can serve as guidance for future
3 studies of NAFLD when using EHR-data, for instance how to define NAFLD in study
4 cohorts and how to ascertain progression to cirrhosis. This guidance could be used
5 both when using administrative and clinical databases, as well as upcoming phase 4-
6 studies in the NAFLD/NASH field. Harmonization of definitions will make it easier
7 to compare and contrast study results from different cohorts, leading to an improved
8 understanding of the consequences of NAFLD. These recommendations could also
9 provide guidance for clinicians or administrators responsible for coding disease status
10 in patients with NAFLD.

11

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

20 This should not be seen as a complete guide on how to perform research in this field.
21 Different research questions and study settings will require distinct strategies. For
22 instance, examining liver-related outcomes in an EHR-derived general population
23 with unknown NAFLD status is different than examining liver-related endpoints in

1 clinical cohorts with known NAFLD status. EHR-derived population-based cohorts
2 often have a low risk for cirrhosis and have the possibility of having liver diseases
3 other than NAFLD, while the definition of outcomes could be made narrower by
4 combining codes through pre-specified algorithms. An example of this is requiring a
5 code for chronic liver disease when ascites is diagnosed, as there are other causes of
6 ascites apart from cirrhosis, thus risking false-positive findings. This can be compared
7 to EHR-based follow-up studies of persons with known NAFLD, where it is less
8 likely that ascites would be related to competing causes. Again, definitions should be
9 tailored to the specific research question. Also, liver-related outcomes are relatively
10 rare in a general population (26, 27), and consideration of examining more common
11 causes of mortality should be actively considered, depending on the research question.

12

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

1 always lead to a lower sensitivity and under-appreciation of the prevalence of the
2 target diagnosis, which should be considered depending on the research question(s).

3
4 It should also be acknowledged that NAFLD is severely under-coded in the general
5 population (30, 31), although there are methods to enhance detection of NAFLD in
6 databases with high level of details (32, 33). Consequently, there is often
7 misclassification bias (due to inaccurate coding of subjects in ‘control’ groups),
8 resulting in attenuation of any association towards the null, but also selection bias
9 (cases diagnosed with NAFLD might be the more severe cases, resulting in inflated
10 risk estimates).

11
12 Preferably, validation studies of codes or algorithms used to increase the positive
13 predictive value of the target diagnosis should be undertaken before starting an EHR-
14 based study, if possible. Examples of these are (18, 19, 34). Results of prior validation
15 studies in hepatology in general suggest that while some diagnoses have a high
16 positive predictive value (PPV), there are several non-specific diagnoses where
17 considerations need to be made. An example is ascites which can be found in persons
18 without cirrhosis, but combining coding for ascites with coding for cirrhosis leads to
19 acceptable PPVs (18). Another example is hepatic encephalopathy, where combining
20 drug prescription coding with cirrhosis increases PPVs to acceptable levels (19). One
21 must also take into account the setting where the diagnosis is made. For instance, a
22 diagnosis of “chronic liver failure” made by a hepatologist might have higher validity
23 when compared to the same diagnosis made in a general hospital (18).

24

1 Areas of controversy and considerations for future research

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

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

15

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

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

23

1 Conclusion

2 We defined a list of ICD-codes that can be considered by investigators examining
3 NAFLD in electronic healthcare records-based research and reached consensus
4 statements addressing several methodological questions. This guidance is intended to
5 streamline future studies in this field, leading to increased generalizability of study
6 results, with the aim to improve our understanding of NAFLD prognosis pertaining to
7 liver related outcomes.

1 References

- 2 1. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a
3 critical comparison of trials with nonrandomized studies. *Hepatology*
4 2006;44:1075-1082.
- 5 2. Ekstedt M, Hagstrom H, Nasr P, Fredrikson M, Stal P, Kechagias S, Hultcrantz
6 R. Fibrosis stage is the strongest predictor for disease-specific mortality in
7 NAFLD after up to 33 years of follow-up. *Hepatology* 2015;61:1547-1554.
- 8 3. Feldstein AE, Charatcharoenwithaya P, Treeprasertsuk S, Benson JT, Enders
9 FB, Angulo P. The natural history of non-alcoholic fatty liver disease in children:
10 a follow-up study for up to 20 years. *Gut* 2009;58:1538-1544.
- 11 4. Hagstrom H, Nasr P, Ekstedt M, Hammar U, Stal P, Hultcrantz R, Kechagias S.
12 Fibrosis stage but not NASH predicts mortality and time to development of
13 severe liver disease in biopsy-proven NAFLD. *J Hepatol* 2017;67:1265-1273.
- 14 5. Jepsen P, Vilstrup H, Mellemkjaer L, Thulstrup AM, Olsen JH, Baron JA,
15 Sørensen HT. Prognosis of patients with a diagnosis of fatty liver--a registry-
16 based cohort study. *Hepatogastroenterology* 2003;50:2101-2104.
- 17 6. Kaps L, Labenz C, Galle PR, Weinmann-Menke J, Kostev K, Schattenberg JM.
18 Non-alcoholic fatty liver disease increases the risk of incident chronic kidney
19 disease. *United European Gastroenterol J* 2020;8:942-948.
- 20 7. Wild SH, Walker JJ, Morling JR, McAllister DA, Colhoun HM, Farran B,
21 McGurnaghan S, et al. Cardiovascular Disease, Cancer, and Mortality Among
22 People With Type 2 Diabetes and Alcoholic or Nonalcoholic Fatty Liver Disease
23 Hospital Admission. *Diabetes Care* 2018;41:341-347.
- 24 8. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-
25 alcoholic fatty liver disease. *J Hepatol* 2016;64:1388-1402.
- 26 9. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison
27 SA, et al. The diagnosis and management of nonalcoholic fatty liver disease:
28 Practice guidance from the American Association for the Study of Liver Diseases.
29 *Hepatology* 2018;67:328-357.
- 30 10. Alexander M, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D,
31 Ansell D, Pasqua A, et al. Non-alcoholic fatty liver disease and risk of incident
32 acute myocardial infarction and stroke: findings from matched cohort study of
33 18 million European adults. *Bmj* 2019;367:l5367.
- 34 11. Alexander M, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D,
35 Ansell D, Pasqua A, et al. Risks and clinical predictors of cirrhosis and
36 hepatocellular carcinoma diagnoses in adults with diagnosed NAFLD: real-world
37 study of 18 million patients in four European cohorts. *BMC Med* 2019;17:95.
- 38 12. Bengtsson B, Stal P, Wahlin S, Bjorkstrom NK, Hagstrom H. Characteristics
39 and outcome of hepatocellular carcinoma in patients with NAFLD without
40 cirrhosis. *Liver Int* 2019;39:1098-1108.
- 41 13. Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS.
42 Nonalcoholic fatty liver disease incidence and impact on metabolic burden and
43 death: A 20 year-community study. *Hepatology* 2018;67:1726-1736.
- 44 14. Allen AM, Van Houten HK, Sangaralingham LR, Talwalkar JA, McCoy RG.
45 Healthcare Cost and Utilization in Nonalcoholic Fatty Liver Disease: Real-World
46 Data From a Large U.S. Claims Database. *Hepatology* 2018;68:2230-2238.

- 1 15. Åberg F, Puukka P, Salomaa V, Männistö S, Lundqvist A, Valsta L, Perola M, et
2 al. Risks of Light and Moderate Alcohol Use in Fatty Liver Disease: Follow-Up of
3 Population Cohorts. *Hepatology* 2020;71:835-848.
- 4 16. Unalp-Arida A, Ruhl CE. Patatin-Like Phospholipase Domain-Containing
5 Protein 3 I148M and Liver Fat and Fibrosis Scores Predict Liver Disease
6 Mortality in the U.S. Population. *Hepatology* 2020;71:820-834.
- 7 17. Kanwal F, Kramer JR, Mapakshi S, Natarajan Y, Chayanupatkul M, Richardson
8 PA, Li L, et al. Risk of Hepatocellular Cancer in Patients With Non-Alcoholic Fatty
9 Liver Disease. *Gastroenterology* 2018;155:1828-1837.e1822.
- 10 18. Bengtsson B, Askling J, Ludvigsson JF, Hagström H. Validity of administrative
11 codes associated with cirrhosis in Sweden. *Scand J Gastroenterol* 2020:1-6.
- 12 19. Tapper EB, Korovaichuk S, Baki J, Williams S, Nikirk S, Waljee AK, Parikh ND.
13 Identifying Patients With Hepatic Encephalopathy Using Administrative Data in
14 the ICD-10 Era. *Clin Gastroenterol Hepatol* 2019.
- 15 20. Lapointe-Shaw L, Georgie F, Carlone D, Cerocchi O, Chung H, Dewit Y, Feld JJ,
16 et al. Identifying cirrhosis, decompensated cirrhosis and hepatocellular
17 carcinoma in health administrative data: A validation study. *PLoS One*
18 2018;13:e0201120.
- 19 21. Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V, 3rd. Validation of three
20 coding algorithms to identify patients with end-stage liver disease in an
21 administrative database. *Pharmacoepidemiol Drug Saf* 2012;21:765-769.
- 22 22. Lo Re V, 3rd, Lim JK, Goetz MB, Tate J, Bathulapalli H, Klein MB, Rimland D, et
23 al. Validity of diagnostic codes and liver-related laboratory abnormalities to
24 identify hepatic decompensation events in the Veterans Aging Cohort Study.
25 *Pharmacoepidemiol Drug Saf* 2011;20:689-699.
- 26 23. Kaplan DE, Dai F, Aytaman A, Baytarian M, Fox R, Hunt K, Knott A, et al.
27 Development and Performance of an Algorithm to Estimate the Child-Turcotte-
28 Pugh Score From a National Electronic Healthcare Database. *Clin Gastroenterol*
29 *Hepatol* 2015;13:2333-2341.e2331-2336.
- 30 24. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J,
31 Marteau T. Consensus development methods, and their use in clinical guideline
32 development. *Health Technol Assess* 1998;2:i-iv, 1-88.
- 33 25. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, Wales
34 PW. Defining consensus: a systematic review recommends methodologic criteria
35 for reporting of Delphi studies. *J Clin Epidemiol* 2014;67:401-409.
- 36 26. Hagstrom H, Talback M, Andreasson A, Walldius G, Hammar N. Ability of
37 Noninvasive Scoring Systems to Identify Individuals in the Population at Risk for
38 Severe Liver Disease. *Gastroenterology* 2019.
- 39 27. Hagstrom H, Tynelius P, Rasmussen F. High BMI in late adolescence predicts
40 future severe liver disease and hepatocellular carcinoma: a national, population-
41 based cohort study in 1.2 million men. *Gut* 2018;67:1536-1542.
- 42 28. Åberg F, Färkkilä M. Drinking and Obesity: Alcoholic Liver
43 Disease/Nonalcoholic Fatty Liver Disease Interactions. *Semin Liver Dis*
44 2020;40:154-162.
- 45 29. Liu B, Balkwill A, Reeves G, Beral V, Million Women Study C. Body mass index
46 and risk of liver cirrhosis in middle aged UK women: prospective study. *British*
47 *Medical Journal* 2010;340:7.

- 1 30. Blais P, Husain N, Kramer JR, Kowalkowski M, El-Serag H, Kanwal F.
2 Nonalcoholic fatty liver disease is underrecognized in the primary care setting.
3 *Am J Gastroenterol* 2015;110:10-14.
- 4 31. Alexander M, Loomis AK, Fairburn-Beech J, van der Lei J, Duarte-Salles T,
5 Prieto-Alhambra D, Ansell D, et al. Real-world data reveal a diagnostic gap in
6 non-alcoholic fatty liver disease. *BMC Med* 2018;16:130.
- 7 32. Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB,
8 Kanwal F. Nonalcoholic fatty liver disease (NAFLD) in the Veterans
9 Administration population: development and validation of an algorithm for
10 NAFLD using automated data. *Aliment Pharmacol Ther* 2014;40:949-954.
- 11 33. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and Validation of an
12 Algorithm to Identify Nonalcoholic Fatty Liver Disease in the Electronic Medical
13 Record. *Dig Dis Sci* 2016;61:913-919.
- 14 34. Driver RJ, Balachandrakumar V, Burton A, Shearer J, Downing A, Cross T,
15 Morris E, et al. Validation of an algorithm using inpatient electronic health
16 records to determine the presence and severity of cirrhosis in patients with
17 hepatocellular carcinoma in England: an observational study. *BMJ Open*
18 2019;9:e028571.
- 19