ORIGINAL ARTICLE

Liver INTERNATIONAL   WILEY

Alcohol-related hepatocellular carcinoma is a heterogenous condition: Lessons from a latent class analysis

Charlotte E. Costentin[1,2] | Mélanie Minoves[3,4] | Sylvain Kotzki[3] | Olivier Farges[5] |
Nathalie Goutté[6] | Thomas Decaens[1,2] | Sébastien Bailly[3]

[1]Institute for Advanced Biosciences, INSERM U1209/CNRS UMR, Grenoble Alpes University, Grenoble, France

[2]Hepatology, gastro-enterology and digestive oncology, Grenoble Alpes University Hospital, La Tronche, France

[3]HP2 Laboratory, INSERM U1300 and Grenoble Alpes University Hospital, Grenoble Alpes University, Grenoble, France

[4]Department of Pharmacy, Grenoble Alpes University Hospital, La Tronche, France

[5]Hepato-Biliary Surgery Department, Beaujon Hospital, Clichy, France

[6]INSERM UMRS-1193, DHU Hépatinov and Centre Hépatobiliaire, Paul-Brousse Hospital, Assistance Publique – Hôpitaux de Paris Villejuif, Paris-Saclay University et DHU Hépatinov par FHU Hépatinov, Villejuif, France

Correspondence
Charlotte E. Costentin, Centre Hospitalier Universitaire Grenoble Alpes, Avenue Maquis du Grésivaudan, 38700 La Tronche, France.
Email: ccostentin@chu-grenoble.fr

Abstract

Background: Alcohol-associated hepatocellular carcinoma (AL-HCC) poor prognosis has been attributed to diagnosis at a later stage. However, host factors and specific health trajectories have been associated with severe outcomes in alcohol-related liver disease. We hypothesize AL-HCC is not a homogeneous condition but encompasses subgroups yielding different outcomes.

Aims: Our aim was to provide a first attempt at a clinical phenotyping of AL-HCC.

Methods: We analysed data for the calendar years 2007–2013 from the French nationwide administrative hospital database. We selected patients with AL-HCC only. Clustering of AL-HCC phenotypes was performed by latent class analysis (LCA).

Results: The study included 11 363 patients with AL-HCC, mainly male (89.6%), median age 67 years [IQR: 61; 74] of which 71.2% had at least one metabolic comorbidity. Five phenotypes were identified. Phenotype 1 (41.4%) displayed high rates of unrecognized cirrhosis prior to HCC diagnosis (81%), low rates of metabolic comorbidities (diabetes 13%), and mostly compensated liver disease at HCC diagnosis while the four other phenotypes displayed high rates of metabolic comorbidities (diabetes up to 100%), various patterns of liver disease trajectories and overall 42% unrecognized cirrhosis. In adjusted survival analysis, compared to phenotype 1, risk of death after HCC diagnosis was significantly different for all phenotypes.

Conclusion: LCA uncovers AL-HCC is a heterogeneous condition with distinct phenotypes yielding specific survival outcomes. Frequent unrecognized cirrhosis prior to HCC underlines the urgent need for implementing strategies to identify the underlying liver disease prior to HCC onset in patients with documented alcohol use disorders and metabolic comorbidities.

KEYWORDS
alcohol, cluster, liver cancer, survival

1 | INTRODUCTION

Hepatocellular carcinoma (HCC) is a leading cause of cancer-related death.[1] Worldwide, hepatitis B and C infections are the two main causes of HCC.[1] Alcohol and metabolic-associated fatty liver disease (MAFLD) are the other two major risk factors for developing HCC.[2] Importantly, the burden of alcohol and MAFLD compared to viral aetiologies is likely to increase because of the recent improvement of antiviral treatment, the ongoing diabetes and obesity epidemic, and the increase of alcohol consumption in some parts of the world.[3] In France, HCC accounts for approximately 10 000 deaths per year. Despite a steady decrease in alcohol consumption per capita over 5 decades, alcohol still accounts for 70% of HCC[4] and the prognosis is one of the poorest compared to all cancers, with a median overall survival of 9.4 months.[5,6]

Consistent data suggest that alcohol-associated hepatocellular carcinoma (AL-HCC) yields poorer prognosis compared to HCC related to other risk factors.[5–8] This difference has been mostly attributed to diagnosis at a later stage, less likely to be eligible for curative treatment. Interestingly, patients with alcohol-associated HCC are more likely to have severe underlying liver disease at the time of HCC diagnosis, precluding curative therapeutic options such as resection or ablation.[9] If the patient, in addition, does not reach abstinence, listing for liver transplantation is not possible and only supportive care is available.[9] Along with tumour stage and alcohol consumption, other factors might contribute to the overall survival impairment of alcohol-associated HCC compared to other aetiologies. Diabetes and body mass index (BMI) as well as genetic factors seem to confer increased susceptibility to HCC in patients with alcohol-associated liver disease (ALD).[10,11] Obesity or diabetes were also associated with a higher risk of developing severe liver disease in patients with excessive alcohol consumption.[12] Also, specific health trajectories can impact survival. In the setting of cirrhosis, Ratib et al. showed that hospital admission marks a turning point in the clinical course of cirrhosis, associated with poorer survival compared to ambulatory-only health trajectories.[13]

Based on this information, alcohol-associated hepatocellular carcinoma might not be a homogeneous condition, and variations of host factors or history of liver disease could lead to different outcomes. If this hypothesis is true, a new delineation of specific subgroups of alcohol-associated HCC is critical for the development of precision and personalized medicine in this field, to better prevent disease progression and severe outcomes.

In France, the administrative hospital database (acronym PMSI for « Programme de Médicalisation des Systèmes d'Information ») provides longitudinal data on hospital activities, nationwide. This database can be leveraged to conduct large-scale epidemiological investigations focusing on specific diseases such as primary liver cancer.[6,14]

The aim of this study was to provide a first attempt at clinical phenotyping of alcohol-associated HCC including host factors, comorbidities and the dynamic course of the underlying liver disease along with associated survival outcomes. In order to identify homogeneous classes of alcohol-associated HCC, we conducted a cluster analysis in a nationwide series of alcohol-associated HCC cases from the French administrative hospital database (PMSI).

Lay summary

• In the French nationwide administrative hospital database, latent class analysis identified five phenotypes of alcohol-related hepatocellular carcinoma associated with specific survival outcomes.

• This study shows a high prevalence of metabolic comorbidities in patients developing alcohol-related HCC and high rates of unrecognized cirrhosis prior to HCC occurrence.

• High rates of unrecognized cirrhosis prior to HCC underline the urgent need for implementing strategies to identify the underlying liver disease prior to HCC onset in patients with documented alcohol use disorders and metabolic comorbidities.

2 | PATIENTS AND METHODS

2.1 | Study design

We analysed data for the calendar years 2007 to 2013 from the French administrative hospital database (PMSI). PMSI provides data for over 65 million persons with at least one hospital stay and includes ICD-10 codes (International Classification of Diseases, Tenth Revision). The study was approved by the Commission Nationale de l'Informatique et des Libertés (CNIL), the French national commission for data protection.

2.1.1 | Definition of HCC

International Classification of Diseases tenth revision (ICD-10) codes were used to identify patients with HCC. We first identified all patients with at least one hospital stay with a primary, related or associated diagnosis of liver cancer according to ICD-10 codes C22.0 « Liver cell carcinoma ». Then, we excluded patients with at least one ICD code for other liver carcinomas such as intrahepatic bile duct carcinoma (C22.1), hepatoblastoma (C22.2), angiosarcoma (C22.3), other sarcoma (C22.4), other specified carcinomas (C22.7), C22.9 « Malignant neoplasm: liver, unspecified », or with history of liver transplantation prior to the first occurrence of an ICD code related to liver cancer (C22.0).

2.1.2 | Case selection

An incident case was defined by a first occurrence of an HCC-related ICD-10 code (date of diagnosis, DD) within the period of interest (from 2007, July 1 to 2012, December 31). Patients were included in the study if they met the following criteria: (1) At least one hospital stay more than 3 months before the diagnosis and (2) one rolling year

of follow-up from the first hospital stay preceding the diagnosis to ensure a minimum depth into the patient's history. Aetiology was ascertained according to ICD-10 codes. Viral aetiology was identified from B1 root-codes and K77 codes for viral disease unclassified elsewhere; alcohol-related liver disease was determined from K70 root-codes and codes for alcohol use disorders (any combination of K700, K701, K702 K703, K704, K709, F100-109, Z502; excluding patients with F100 alone: Mental and behavioural disorders because of use of alcohol, acute intoxication) while other aetiologies were retrieved from cosdes for liver diseases (Tables S1 and S2). For the purpose of this study, only patients with alcohol only-associated HCC were included. Patients with mixed aetiologies (alcohol associated with any other lisver disease) were excluded.s

2.1.3 | Relevant variables and definitions

Patient characteristics and follow-up data were retrieved: date of diagnosis, age, sex, chronic comorbidities (diabetes, obesity, arterial hypertension, dyslipidaemia, obstructive sleep apnoea) and cirrhosis complications. The ICD-10 code algorithm to define comorbidities are provided in Table S3. For each patient and each hospital stay, all primary, related and associated diagnosis codes were extracted. For each diagnosis ICD-10 code, we retrieved the list of associated medical procedures (CCAM codes) from the national health insurance website (AMELI, ameli.fr).

2.1.4 | Data aggregation

With the intent of assessing liver-disease trajectory, we defined for each patient three different periods of follow-up according to the date of HCC diagnosis: (1) 'prior HCC diagnosis' is the period ending 3 months before the diagnosis date (history-variables); (2) 'contemporary' is the period extending from 3 months prior to 3 months after HCC diagnosis (contemporary-variables) and (3) 'post-HCC diagnosis' is the period that begins 3 months after HCC diagnosis. For each period, we merged all the data extracted from each stay summary, eliminating any duplicates which resulted, into three lists of diagnosis and medical interventions for each patient (prior, contemporary and post-HCC diagnosis) (Figure 1).

For chronic comorbidities as listed above, we assumed that if related ICD-10 codes were observed during the 'prior HCC diagnosis' period, they were still present during the contemporary period.

Finally, treatment was categorized into 'curative therapy' and 'non-curative therapy'. 'Curative therapy' was either surgical resection, radiofrequency or liver transplantation and merged into a unique variable while other strategies (chemotherapy or intrahepatic injection of chemotherapeutics) were merged into a 'non-curative therapy' variable. Importantly, it was not possible to identify patients receiving biotherapies such as sorafenib as oral medications are not referenced in the PMSI database. Therefore, the 'no treatment recorded' variable accounts for biotherapies and/or best supportive care.
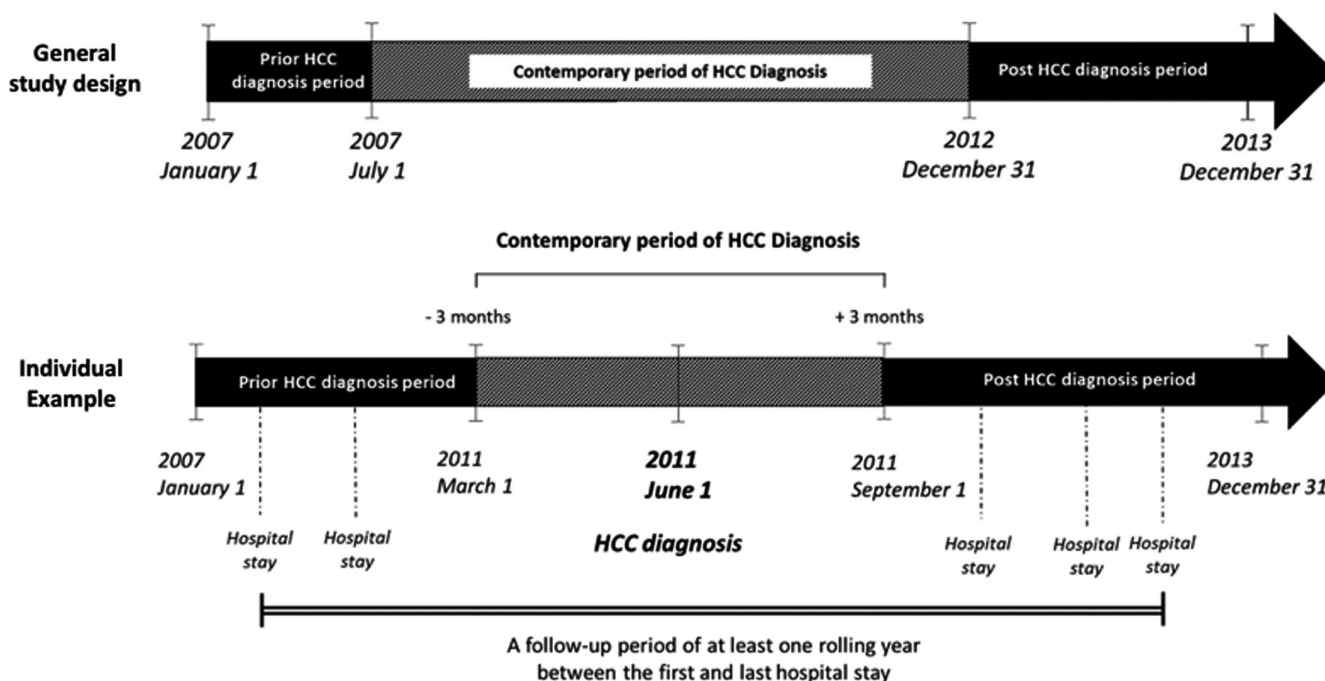

FIGURE 1 Data aggregation temporality with respect to date of HCC diagnosis. PMSI data from 1 January 2007 to 31 December 2013 were considered for the study. Only the period from 1 July 2007 to 31 December 2012 was considered to identify a HCC diagnosis (contemporary period of HCC diagnosis). The individual sample shows the case for a patient with a HCC diagnosis performed on 1 June 2011. The data from 3 months before and 3 months after are considered as contemporary to the HCC diagnosis. Before this period is the history of the patient before the HCC diagnosis and after this period is the event occurred after the HCC diagnosis

2.2 | Statistical analysis

Clustering of alcohol-associated HCC was performed by latent class analysis (LCA).[15] LCA is a probabilistic clustering method which allows homogeneous subgroups of phenotypes, called latent classes, to be identified from a larger heterogeneous population. Patients are classified into clusters based on their higher probability of belonging to one cluster than to another, which is directly estimated by the model. Finally, clusters are constituted of patients who share similar characteristics. The assumption of local independence for variables introduced in the LCA was verified using Cramer's V correlation coefficient. No missing values were observed for the variables considered for the clustering. To avoid inflation and collinearity of features involved in the LCA, only variables with occurrence rate ≥ 2% were included in the analysis (Table S4). The cluster analysis was performed by a two-step process: (1) an initial clustering was performed by including all selected variables (Table S4) and (2) a final clustering was performed by considering only the 10 most discriminant variables identified from the first step: two specific comorbidities (diabetes and arterial hypertension), four variables pertaining to liver-related medical history prior to HCC diagnosis (history-cirrhosis, portal hypertension, oesophageal varices, ascites, liver failure) and three variables related to the liver disease at the time of HCC diagnosis (contemporary-liver failure, ascites and oesophageal varices). The optimal number of clusters was determined using the integrated completed likelihood criterion (ICL).[15] An internal validation was performed to ensure the choice of cluster number by performing a n-fold cross-validation: 100 random samples of 10 000 patients were constituted and for each sample, a latent class analysis was performed with a range of 3 to 7 clusters. For each sample, the optimal number of clusters was identified by using the ICL statistics. The median of the optimal number of clusters was 5 and was retained. The final set of clusters was further described using number and percentage for qualitative variables and median with first and third quartiles for quantitative variables. The probability of a patient belonging to the cluster to which he/she was assigned is presented. Comparisons between clusters were performed using non-parametric Kruskal–Wallis tests for quantitative variables and Chi-square or Fisher exact tests for qualitative variables. Bonferroni correction was applied for multiple comparisons when comparing variables across clusters.

To assess the relation between clusters and 12-month survival, a Kaplan–Meier estimate was used for survival curves and a cluster comparison was performed using the log-rank test, considering patient survival time greater than 0 month.

To assess the hazard ratio of 12-month death, a semi-parametric Cox model was performed. As clusters are built on variables differentiating all clusters, we adjusted only on age and gender. Moreover, to account for treatment after HCC diagnosis as confounding factors, the curative and non-curative treatments were considered as a time-dependent covariate. Statistical analyses were performed with both SAS v9.4 and R v 3.6.1 software, and the R package VarSelLCM was used for LCA.[16]

3 | RESULTS

3.1 | Study population

A total of 67 547 patients with a diagnosis of HCC within the calendar period 2007–2013 were identified. After exclusion of patients without any medical stay prior to HCC diagnosis (n = 20 776), patients with another type of primary liver cancer (n = 9287), patients with HCC diagnosis outside the study period time-frame (n = 1742), patients with less than one rolling year of follow-up after first hospital stay (n = 8698) and patients with other aetiology than alcohol (n = 15 681), 11 363 patients with alcohol-associated HCC were included in this study (Figure 2). Alcohol-associated HCC patients were mainly male (89.6%), and the median age of 67 years (interquartile range: [61; 74]). History of portal hypertension was observed in 15.2% of the cases; history of cirrhosis decompensation in 21% and 43% had decompensated cirrhosis at the time of HCC diagnosis. Overall 71.2% of patients had at least one metabolic comorbidity (diabetes 41.3%, arterial hypertension 52.6%, dyslipidaemia 16.1%, obesity 19.7%, OSA 4.5%), 22.2% received curative treatment, 40.5% received non-curative treatment and 48.4% received neither in the overall follow-up period. At 1 year, and after a median of 10.8 months of follow-up after HCC diagnosis [2.4; 24], 3782 deaths were observed. The raw 1-year survival was 33.3%.

3.2 | Phenotypes of alcohol-associated HCC

By considering the statistical criteria ICL, the optimal number of classes was 5. The heatmap illustrating the final clustering considering the 10 most discriminant variables is presented in Figure 3. The median probability a patient belongs to the cluster to which he/she was assigned ranged from 0.70 to 1 (Table S5).

3.3 | Alcohol-associated HCC phenotypes

Detailed descriptions of the phenotypes are given in Table S6, and a summarized description is provided in Figure 4.

- *Phenotype 1 = compensated alcohol-associated liver disease (ALD) unrecognized prior to HCC diagnosis (4703 patients – 41.4%):* larger group, median age: 68 years [61; 75]), 90,3% males; mostly unrecognized cirrhosis prior to HCC diagnosis, compensated at HCC diagnosis (low rates of ascites or liver failure); significant rates of arterial hypertension (37.9%) but low rates of other metabolic comorbidities: diabetes 13%; obesity 12.8%, dyslipidaemia 11.2%; OSA 3.8%.
- *Phenotype 2 = compensated both alcohol and metabolic-associated liver disease (AMALD) unrecognized prior to HCC diagnosis (1960 patients – 17.2%):* median age: 70 years [64; 76]), 93% males; mostly unrecognized (80.5%) cirrhosis prior to

FIGURE 2   Study flow chart

FIGURE 2   Study flow chart
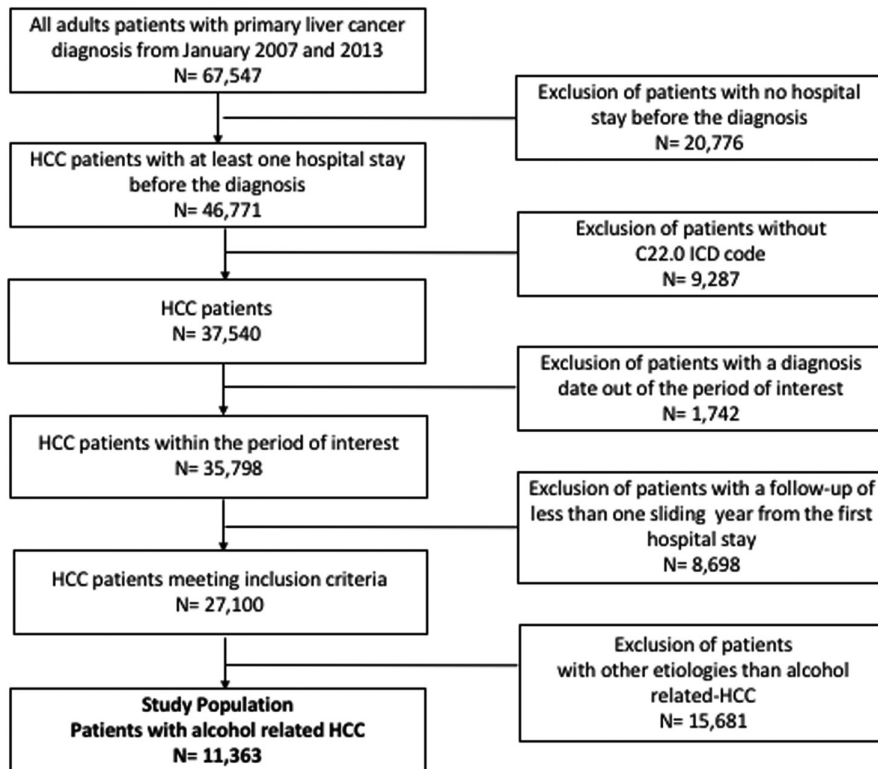
	Class 1 N=4703 (41.4 %)	Class 2 N=1960 (17.2 %)	Class 3 N=2319 (20.4 %)	Class 4 N=1152 (10.1 %)	Class 5 N=1229 (10.8%)
Diabetes	13	100	44.3	47.2	44.9
Arterial hypertension	37.9	100	44.5	50.6	50.2
H. Ascites	0.3	0.9	69.9	0	2
H. Liver failure	1.6	1.6	45.5	8.2	1.3
H. Portal HT	0.8	1.2	47.7	48.4	0.4
H. Cirrhosis	19	19.5	97.2	88	17.2
H. Esophageal varices	0.7	0	59.2	86.6	2.9
Co. Ascites	11.9	13.2	69.1	4.7	90.3
Co. Liver failure	6.2	7.6	50.4	9.3	63.9
Co. Esophageal varices	7.8	10.1	42.5	57.4	59.5

H: History
Co: Contemporary
HT: Hypertension

FIGURE 3   Heatmap: illustrating the final clustering considering the 10 most discriminant variables (%)

HCC diagnosis; compensated at HCC diagnosis (low rates of ascites or liver failure); 100% metabolic comorbidities (diabetes 100%; arterial hypertension 100%, obesity 36.1%, dyslipidaemia 36.3%; OSA 9%).

- Phenotype 3 = AMALD with history of liver events (2319 patients – 20.4%): median age: 64 years [58; 71]); 85.5% males; cirrhosis identified prior to HCC diagnosis (97.2%) with frequent liver events (portal hypertension and/or ascites and/or liver failure) prior to and at the time of HCC diagnosis; frequent metabolic comorbidities (diabetes 44.3%; arterial hypertension 44.5%, obesity 19.4%, dyslipidaemia 11.8%; OSA 2.4%).

- Phenotype 4 = compensated AMALD (1152 patients – 10.1%) recognized prior to HCC diagnosis: median age: 65 years [60; 72]), 89.1% males; cirrhosis identified prior to HCC diagnosis (88%); high rates of portal hypertension complications, but few events related to liver failure; mostly compensated at HCC diagnosis (low rates of ascites or liver failure); frequent metabolic comorbidities (diabetes 47.2%, arterial hypertension 50.6%, obesity 20.4%, dyslipidaemia 14.1%; OSA 4.2%).

- Phenotype 5 = recently complicated AMALD unrecognized prior to HCC diagnosis (1229 patients – 10.8%): median age: 67 years [61; 73]); 90.3% males; mostly unrecognized (82.8%) cirrhosis prior to
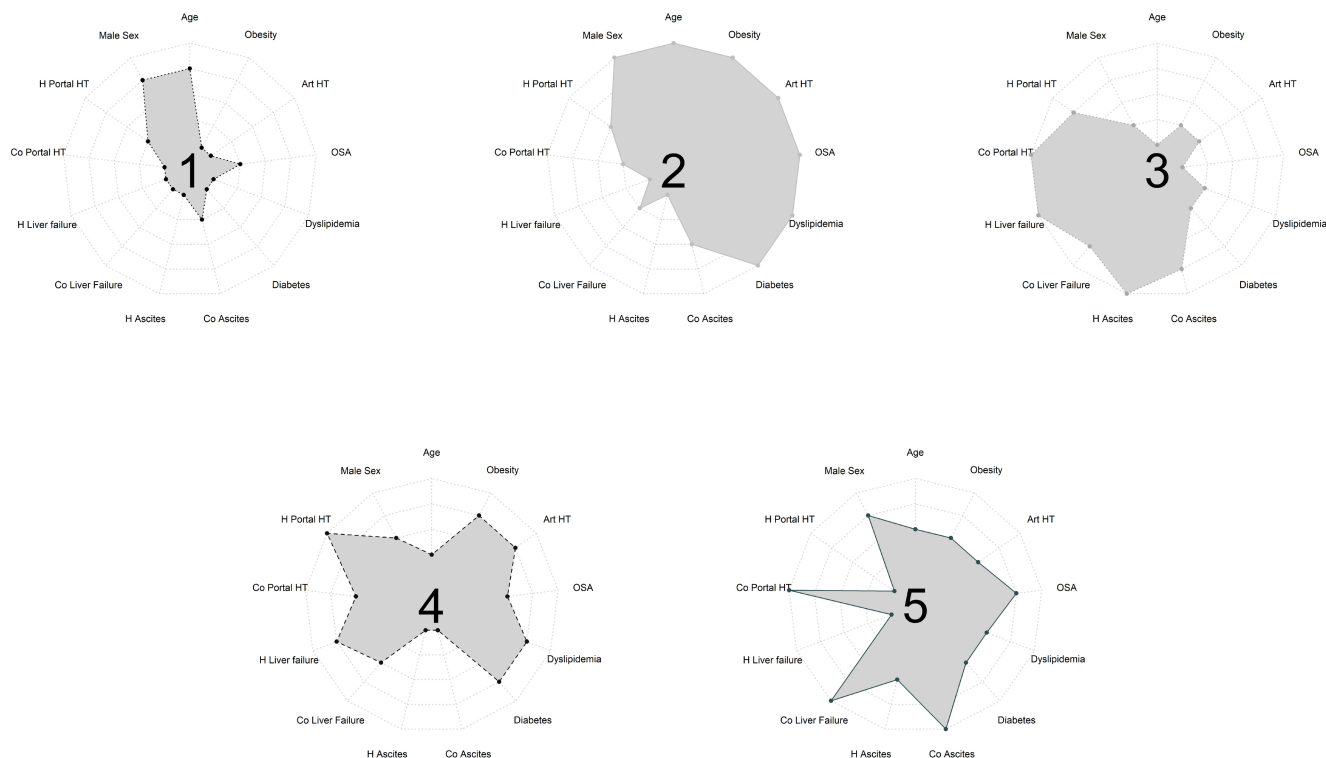
FIGURE 4 Phenotypes ID cards: clustering variables ranked according to the observed percentage for each variable in each phenotype compared to the others, from 1 (smallest percentage) to 5 (higher percentage); H, history; Co, contemporary; Portal HT, portal hypertension; OSA, obstructive sleep apnoea; Art HT, arterial hypertension)

HCC diagnosis with low rates of past liver events; complicated cirrhosis at the time of HCC diagnosis with both liver failure and portal hypertension (90% ascites at diagnosis); frequent metabolic comorbidities (diabetes 44.9%, arterial hypertension 50.2%, obesity 19.5%, dyslipidaemia 12.5%; OSA 4.4%).

3.4 | Treatment allocation according to alcohol-associated HCC phenotypes

Phenotypes 3 and 4 had higher rates of curative treatment (25.4% and 32.8% respectively) compared to phenotypes 1, 2 and 5 (21.6%, 20.8% and 10.4% respectively). Liver transplantation ranged from 2.5% in phenotype 2 to 10.7% in phenotype 3. Among curative treatment performed, ablation was the preferred option in all phenotypes (Table S7).

3.5 | Survival according to alcohol-associated HCC phenotypes

Figure 5 shows Kaplan–Meier survival curves stratified by phenotype. Decreasing rates of survival at 12 months after diagnosis were observed when moving from phenotypes 1 and 4 and then phenotypes 2, 3 and 5. After Bonferroni correction for multiple test, all pairs of clusters were significantly different for 12-month survival ($p < 0.01$) except phenotypes 2 and 4 ($p = 1$). (Table S8).

The multivariable survival Cox analysis including treatment as a time-dependent covariate, showed that compared to phenotype 1, the probability of death at 1 year was not statistically different for phenotype 4 (HR 1.094, 95%CI 0.942–1.27; $p = 0.24$) and significantly increased for phenotypes 2 (HR 1.184, 95%CI 1.054–2.331; $p < 0.01$), 3 (HR 1.933, 95%CI 1.737–2.152; $p < 0.01$) and 5 (HR 3.157, 95%CI 2.805–3.553; $p < 0.01$) (Table S9).

4 | DISCUSSION

By conducting a cluster analysis in a large administrative database, we were able to identify five phenotypes of alcohol-related HCC associated with survival outcomes. Our results provide new perspectives on potential determinants of HCC prognosis in the setting of alcohol-associated HCC, and new insights to improve early identification of alcohol-related liver disease.

The first important finding was that a majority of patients with alcohol-associated HCC combine risk factors for metabolic-associated fatty liver disease (in our study 71% had at least one among diabetes, arterial hypertension, obesity, dyslipidaemia, OSA). Therefore, alcohol-associated HCC cases are, in fact, most frequently both alcohol- and metabolic-associated HCC. This is concordant with previous reports from the French CHANGH cohort (diabetes was present in 30% of alcohol-associated HCC) as well as the Italian cohort.[5,8] Nevertheless, in these studies, patients with alcohol-associated HCC were considered as a unique entity when
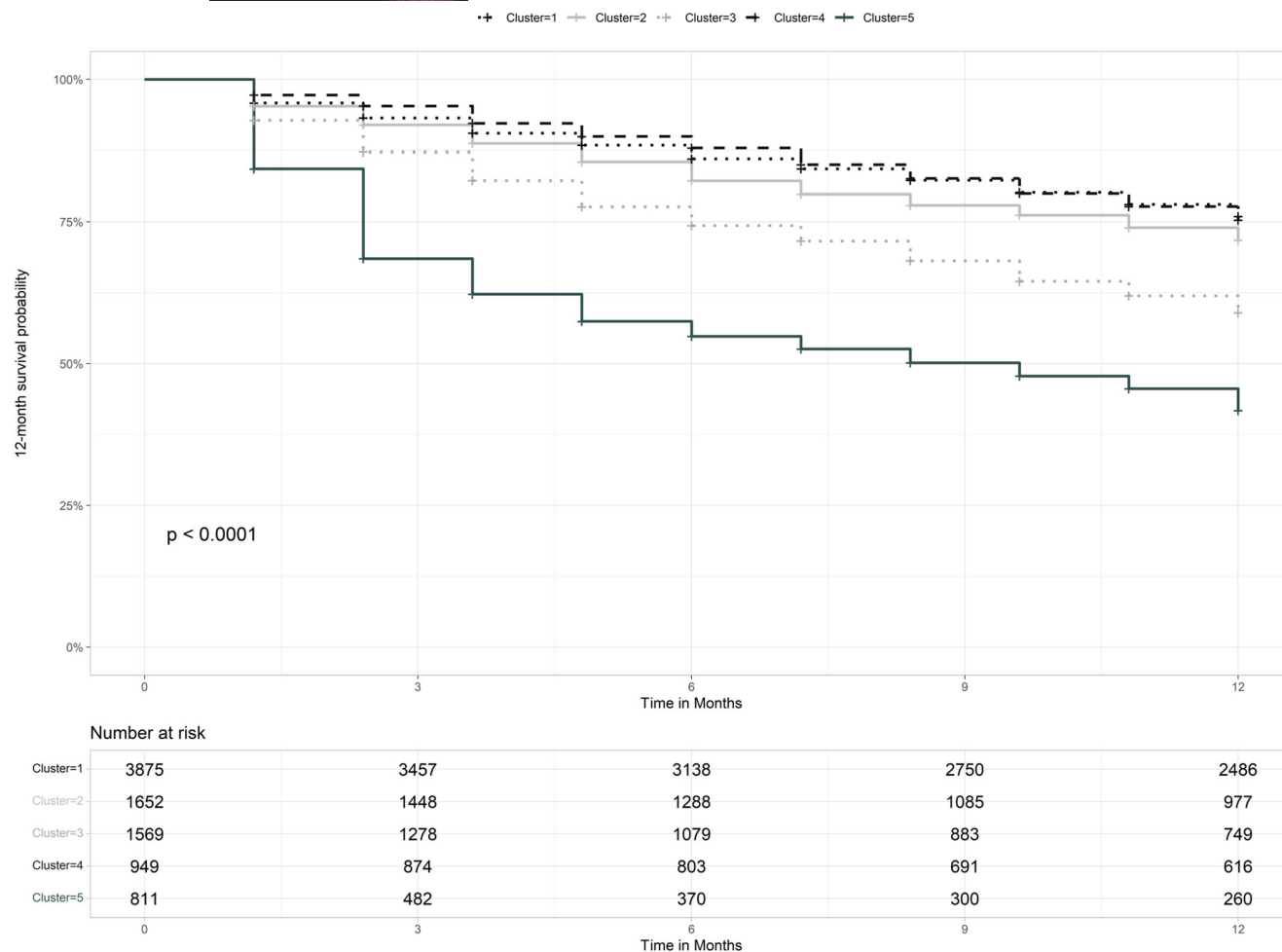
FIGURE 5 Kaplan–Meier curve of overall survival at 1 year by phenotype. *p* value: log-rank test. Foot note: 12-month survival rates for each phenotype are as follows: phenotype 1: 74.4%; phenotype 2: 70.5%; phenotype 3: 58.3%; phenotype 4: 74.7%; phenotype 5: 43.5%

compared to other aetiologies, whereas subgroup analysis could have been more informative, especially to further scrutinize the drivers of survival discrepancies. Interestingly, cardiovascular and metabolic comorbidities were associated with improved survival in the primary analysis of alcohol-associated HCC within the French PMSI database.[5,8] In the present study, one phenotype (phenotype 1) displayed low rates of metabolic comorbidities except for arterial hypertension, which could, in this case, be alcohol-related secondary arterial hypertension.[17] We hypothesize phenotype 1 corresponds to patients with HCC solely related to alcohol. This phenotype accounted for less than half (41.4%) of the total cohort of 'alcohol-related HCC'.

Second, more than half (58,2%) of the patients overall had unrecognized cirrhosis at HCC diagnosis overall, but this proportion was down to 42% in AMALD groups (phenotypes 2 to 5), compared to 81% in phenotype 1. This is concordant with previous reports of alcohol and metabolic-associated HCCs aetiologies being more likely to be associated with unrecognized cirrhosis prior to HCC diagnosis, suggesting identification of advanced liver fibrosis in patients with identified alcohol abuse or metabolic comorbidities is not optimal.[18] Our findings suggest alcohol-only underlying liver disease might be even more

likely to progress unrecognized in France. Unfortunately, when cirrhosis is not recognized, HCC screening programs are not implemented, leading to diagnosis of HCC at a later stage, associated with poorer survival.[6-8] Importantly, among patients with compensated cirrhosis at HCC diagnosis (phenotypes 1, 2 and 4), rate of curative treatment was greatest for phenotype 4, with cirrhosis identified before HCC diagnosis, whereas in phenotypes 1 and 2, cirrhosis was mostly unrecognized. Our results strongly support the urgent need for implementing innovative strategies to improve screening for liver disease in patients with documented alcohol abuse, but also to raise awareness among health practitioners dealing with patients with metabolic comorbidities in order to screen for alcohol use disorders as well as liver fibrosis prior to the occurrence of life-threatening complications.

Third, the liver disease trajectory prior to HCC diagnosis formed a critical component of the clustering associated with different outcomes. Unsurprisingly, both phenotypes featuring patients with ongoing decompensated liver disease at the time of HCC diagnosis (phenotypes 3 and 5) had the lowest 1-year survival rates, as liver failure precludes two curative treatment options (resection and ablation). However, it is interesting that phenotype 5, with low rates of history of liver decompensation, had a worse

survival compared to phenotype 3, with history of complicated liver disease. We hypothesize that phenotype 3 patients have survived past decompensation and may have recovered from liver failure, possibly through abstinence, facilitating screening strategies and access to curative treatment such as liver transplantation. Unfortunately, liver function cannot be thoroughly evaluated in this database without biological data available. In addition, such databases lack information on alcohol intake. Nevertheless, the impact of disease trajectory, with poorer prognosis associated with decompensated liver disease underlines the critical importance of identifying alcohol-related liver disease prior to liver decompensation leading to premature death.[13] Indeed, early management of alcohol use disorders targeting risk reduction or abstinence, and management of metabolic comorbidities can decrease HCC incidence or at least improve liver function which increases chances to access curative treatment options for HCC.[5,19]

Lastly, although phenotypes 1 and 2 yield very similar liver disease trajectories (barely symptomatic prior to HCC diagnosis) and similar rates of overall treatment allocation (10–12% curative and 60–66% no treatment recorded), after adjustment for age, sex and treatment at 6 months, phenotype 2 (AMALD 100% diabetes and hypertension) had reduced 1-year survival probabilities compared to phenotype 1 (alcohol-only) suggesting competitive risk of death in a context of high prevalence of metabolic risk factors. Overall, the impact of disease trajectory, with poorer prognosis associated with decompensated liver disease underlines the importance of identifying alcohol-related liver disease prior to liver decompensation to prevent liver events, associated with poorer prognosis.

The strength of our study relies on a large database capturing all cases of alcohol-associated HCC at a national level, with a wide range of diagnostic information based on ICD-10 codes. We used a probabilistic method for clustering to identify five homogeneous phenotypes with high probabilities a patient belongs to the cluster he/she was assigned.

Our study also has limitations inherent in administrative databases. Recording for some comorbidities may have been incomplete. It is well reported that some conditions like obesity suffer from low record rates in administrative databases,[20] and other conditions such as sleep apnoea suffer both low diagnostic rates and low recording rates when present.[21] However, this is unlikely to have biased the results as these two variables were not included in the final set of clustering variables. For other variables, the recording procedure was the same throughout France and misclassification should be balanced among all clusters. Second, as the PMSI database provides data only on hospital stays, health trajectories as reported might not be fully representative of medical histories. Of note, severe liver events such as liver failure or significant portal hypertension, which are critical components of the clustering, usually require in-hospital management and therefore were captured in our analysis. Also, PMSI database does not allow for recovering biological, pathological and imaging data, precluding quantitative analysis of liver function and assessment of tumour stage. In addition, PMSI does not provide information on alcohol consumption history

which is critical to better understand discrepancies in trajectories between groups. The linkage between PMSI and hospital data will need to be conducted to address this issue and strengthen the relevance of the five phenotypes. Last, we intended to illustrate heterogeneity in the group of patients labelled 'alcohol-related HCC' with the a priori hypothesis that disease trajectories were heterogeneous and potentially linked with heterogeneity in outcomes. To achieve this goal, we needed to select a study population with minimal depth into patient' history prior to HCC diagnosis. We acknowledge that excluding patients who did not have one-rolling year of complete follow-up after the first hospital stay may lead to overestimate survival in patients with alcohol-related HCC. However, our work did not intend to focus on the survival of alcohol-related HCC in France, as we previously reported overall poorer survival of alcohol-related HCC compared to HCV-related HCC (5). Instead, in this study, we illustrate how disease trajectory prior to HCC onset impacts outcomes after HCC diagnosis.

The most striking finding of our study is the low proportion of patients with alcohol as the only identified risk factor for the underlying liver disease (less than 50%), and the reclassification from ALD to AMALD of most patients. Our study stresses out that patients with alcohol-related HCC are not a homogeneous group. Recognizing this novel fact is critical to implement precision medicine taking into account phenotypic differences. High rates of unrecognized cirrhosis in phenotype 1 suggest missed opportunities to screen for advanced fibrosis in patients with recognized harmful alcohol consumption. Moreover, early management of alcohol use disorders is critical to decrease the incidence of HCC patients with alcohol-related liver disease.[19] In the meantime, targeting metabolic features to screen for liver fibrosis and alcohol consumption is a promising way to uncover AMALD patients. To achieve this goal, liver specialists must contribute to increase awareness among primary care practitioners and specialists in charge of patients with metabolic comorbidities about liver fibrosis and alcohol consumption assessment.

Also, our study raises the question of the risk factors assessment accuracy to determine the cause of the underlying liver disease. Interestingly, substantial discrepancies in alcohol-associated HCC burden compared to other aetiologies are observed across the globe. For instance, France and South Korea have a similar alcohol consumption per capita (16.7 vs 16 Litre of pure alcohol/capita/year respectively)[3] and yet, the reported proportion of alcohol-related HCC among all HCCs is very different: 70% vs 10% respectively.[5,8,22] This could be explained by low reporting or assessment of alcohol consumption when other risk factors, such as HBV, HCV or even metabolic comorbidities are identified. Conversely, the absence of identified risk factors is a common way to define metabolic-associated liver disease in research conducted in administrative health databases.[23] One can hypothesize that alcohol habits could be under-reported and under-evaluated based on cultural stigma associated with alcohol consumption. Our study provides a framework to improve the classification of liver diseases when conducting research in such

databases. Indeed, the clustering approach provides the opportunity to further study patients with HCC categorized as 'unknown aetiology'. Under this denomination fall patients with metabolic-associated fatty liver disease (MAFLD) who are to date mostly recognized after exclusion of all other causes of liver diseases, as well as possibly AMALD patients with un-coded alcohol use disorders. Our results should encourage other investigators to assess if, within different country settings, the respective burden of alcohol vs alcohol-and-metabolic associated HCC or the respective burden of unrecognized/recognized cirrhosis in both categories are similar, in order to adjust local health policies to improve the management of risk factors and the underlying cirrhosis earlier in the course of the liver disease. Moreover, external validation of this clustering analysis would be useful to assess the reproducibility of these clusters and allow comparison between countries or periods. Importantly, assessing these phenotypes in databases allowing assessment of liver function, tumour stage and history of alcohol use disorders would also increase understanding of alcohol-related liver disease natural history.

In conclusion, five phenotypes of alcohol-related HCC associated with specific survival outcomes were identified by means of LCA. Our results provide new perspectives on determinants of HCC prognosis in the setting of alcohol-associated HCC. By highlighting the high prevalence of metabolic comorbidities in patients developing alcohol-related HCC and high rates of unrecognized cirrhosis prior to HCC occurrence, our work strongly supports such findings should be leveraged to raise awareness among health professionals dealing with metabolic comorbidities in order to improve the early identification alcohol-related liver disease.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare related to this study.

AUTHORS' CONTRIBUTIONS

Costentin, Minoves, Kotzki, Decaens and Bailly had full access to all data in the study and take responsibility for the integrity of data and the accuracy of data analysis. And also they carried out study concept and design, and analysis and interpretation of data. Farges and Goutté carried out acquisition of data. Kotzki carried out data preprocessing. Costentin, Minoves, Kotzki and Bailly contributed to study supervision and drafting of the manuscript. Bailly was also involved in statistical analysis. Costentin, Minoves, Kotzki, Farges, Goutté, Decaens and Bailly contributed to the critical revision and approval of the manuscript.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analysed during the current study are available from the Système National des Données de Santé (https://www.snds.gouv.fr/SNDS/Actualites/Actu-10).

ORCID

Charlotte E. Costentin https://orcid.org/0000-0001-6115-8512
Thomas Decaens https://orcid.org/0000-0003-0928-0048

REFERENCES

1. Bray F, Ferlay J, Laversanne M, et al. Cancer incidence in five continents: inclusion criteria, highlights from volume X and the global status of cancer registration. *Int J Cancer*. 2015;137(9):2060-2071.
2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301-1314.
3. World Health Organization. Management of Substance Abuse Team. *Global status report on alcohol and health*. World Health Organization; 2018.
4. 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(3):718-735.
5. Costentin CE, Mourad A, Lahmek P, et al. Hepatocellular carcinoma is diagnosed at a later stage in alcoholic patients: results of a prospective, nationwide study. *Cancer*. 2018;124(9):1964-1972.
6. Costentin CE, Sogni P, Falissard B, et al. Geographical disparities of outcomes of hepatocellular carcinoma in France: the heavier burden of alcohol compared to hepatitis C. *Dig Dis Sci*. 2020;65(1):301-311.
7. Schutte K, Bornschein J, Kahl S, et al. Delayed diagnosis of HCC with chronic alcoholic liver disease. *Liver Cancer*. 2012;1(3–4):257-266.
8. Bucci L, Garuti F, Camelli V, et al. Comparison between alcohol- and hepatitis C virus-related hepatocellular carcinoma: clinical presentation, treatment and outcome. *Aliment Pharmacol Ther*. 2016;43(3):385-399.
9. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182-236.
10. Matsushita H, Takaki A. Alcohol and hepatocellular carcinoma. *BMJ Open Gastroenterol*. 2019;6(1):e000260.
11. Ganne-Carrie N, Nahon P. Hepatocellular carcinoma in the setting of alcohol-related liver disease. *J Hepatol*. 2019;70(2):284-293.
12. Aberg F, Helenius-Hietala J, Puukka P, Farkkila M, Jula A. Interaction between alcohol consumption and metabolic syndrome in predicting severe liver disease in the general population. *Hepatology*. 2018;67(6):2141-2149.
13. Ratib S, Fleming KM, Crooks CJ, Aithal GP, West J. 1 and 5 year survival estimates for people with cirrhosis of the liver in England, 1998-2009: a large population study. *J Hepatol*. 2014;60(2):282-289.
14. Goutte N, Sogni P, Bendersky N, Barbare JC, Falissard B, Farges O. Geographical variations in incidence, management and survival of hepatocellular carcinoma in a Western country. *J Hepatol*. 2017;66(3):537-544.
15. Hagenaars JA, McCutcheon AL. *Applied latent class analysis*. Cambridge University Press; 2002.
16. Marbac M, Sedki M. VarSelLCM: an R/C++ package for variable selection in model-based clustering of mixed-data with missing values. *Bioinformatics*. 2019;35(7):1255-1257.
17. Puddey IB, Mori TA, Barden AE, Beilin LJ. Alcohol and hypertension-new insights and lingering controversies. *Curr Hypertens Rep*. 2019;21(10):79.
18. Walker M, El-Serag HB, Sada Y, et al. Cirrhosis is under-recognised in patients subsequently diagnosed with hepatocellular cancer. *Aliment Pharmacol Ther*. 2016;43(5):621-630.
19. Rodriguez M, Gonzalez-Dieguez ML, Varela M, et al. Impact of alcohol abstinence on the risk of hepatocellular carcinoma in patients with alcohol-related liver cirrhosis. *Am J Gastroenterol*. 2021;116:2390-2398.
20. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008;43(4):1424-1441.
21. McIsaac DI, Gershon A, Wijeysundera D, Bryson GL, Badner N, van Walraven C. Identifying obstructive sleep apnea in

administrative data: a study of diagnostic accuracy. *Anesthesiology.* 2015;123(2):253-263. Epub July 23, 2015.

22. Kim BH, Park JW. Epidemiology of liver cancer in South Korea. *Clin Mol Hepatol.* 2018;24(1):1-9.

23. 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(6):2230-2238.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.