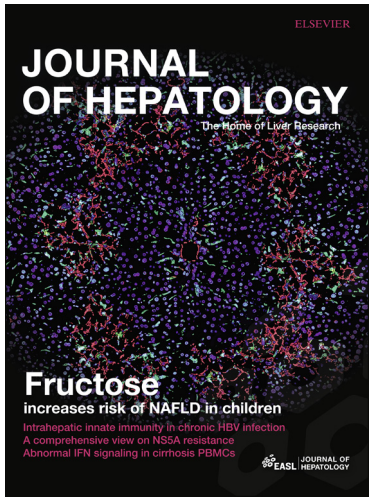
Accepted Manuscript



Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD

Hannes Hagström, Patrik Nasr, Mattias Ekstedt, Ulf Hammar, Per Stål, Rolf Hultcrantz, Stergios Kechagias

PII: S0168-8278(17)32202-X
DOI: http://dx.doi.org/10.1016/j.jhep.2017.07.027
Reference: JHEPAT 6622

To appear in: *Journal of Hepatology*

Received Date: 30 April 2017
Revised Date: 16 June 2017
Accepted Date: 21 July 2017

Please cite this article as: Hagström, H., Nasr, P., Ekstedt, M., Hammar, U., Stål, P., Hultcrantz, R., Kechagias, S., Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD, *Journal of Hepatology* (2017), doi: http://dx.doi.org/10.1016/j.jhep.2017.07.027

Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD

Hannes Hagström[1,2*], Patrik Nasr[3*], Mattias Ekstedt[3], Ulf Hammar[4], Per Stål[1,5], Rolf Hultcrantz[1,5], Stergios Kechagias[3]

[1] Centre for Digestive Diseases, Division of Hepatology, Karolinska University Hospital, Stockholm, Sweden and Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

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

[3] Department of Gastroenterology and Hepatology, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

[4] Unit of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet

[5] Department of Medicine, Huddinge, Karolinska Institutet

[*] Both authors contributed equally.

Correspondence and reprint requests:

Hannes Hagström, Centre for Digestive Diseases, Division of Hepatology, Karolinska University Hospital, 141 86 Stockholm, Sweden

Phone: +46 (0) 8 5858 2305, Fax: +46 (0) 8 5858 2335

E-mail: hannes.hagstrom@ki.se

Keywords: Steatosis, cirrhosis, epidemiology

Total word count: 4368. **Abstract: 314**. **Tables: 4**. **Figures: 3**. **Character count for title: 118**

Disclosure / Conflict of interest declaration: None

Grant Support (Funding): HH was funded by grants from the Swedish Royal Academy of Sciences, the Bengt Ihre scholarship and by the Swedish Gastroenterology Fund.

Author's contributions: Study conception and design: HH, PN, ME, PS, RH, SK

Acquisition of data: HH, PN, ME

Statistical analysis: HH, UH

Analysis and interpretation of data: HH, PN, ME, UH, PS, RH, SK

Drafting of manuscript: HH, PN

Critical revision: HH, PN, ME, UH, PS, RH, SK

Guarantor of article: HH

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

Writing Assistance: None.

Abstract

Background and Aims: Non-alcoholic fatty liver disease (NAFLD) is highly common in the general population, but identifying patients with increased risk for mortality and liver-specific morbidity remains a challenge. Non-alcoholic steatohepatitis (NASH) is thought to enhance this risk; therefore, resolution of NASH is a major endpoint in current pharmacologic studies.

Methods: We conducted a retrospective cohort study of 646 biopsy-proven NAFLD patients. Each case was matched for age, sex and municipality to ten controls. Outcomes on mortality and severe liver disease, defined as cirrhosis, liver decompensation/failure or hepatocellular carcinoma, were evaluated using population-based registers. Cox regression models adjusted for age, sex and type 2 diabetes were used to examine the long-term risk per fibrosis stage. Likelihood ratio tests were used to study if adding NASH to these models increased the predictive capacity. Laplace regression was used to estimate the time to severe liver disease per stage of fibrosis.

Results: During a follow-up of in mean 20 years (range 0-40) equivalent to 139 163 person-years, 12% of NAFLD patients and 2.2% of controls developed severe liver disease ($p<0.001$). Compared to controls, the risk of severe liver disease increased per stage of fibrosis (HR ranging from 1.9 in F0 to 104.9 in F4). Adding presence of NASH did not change these estimates significantly (likelihood ratio test >0.05 for all stages of fibrosis). Similar results were seen for overall mortality. The lower end of the 95% CI for the 10^{th} percentile of time to development of severe liver disease was

22-26 years in F0-1, 9.3 years in F2, 2.3 years in F3, and 0.9 years to liver decompensation in F4.

Conclusions: In this largest ever study of biopsy-proven NAFLD, presence of NASH did not increase the risk of liver-specific morbidity or overall mortality. Knowledge of time to development of severe liver disease per fibrosis stage can be used in individual patient counselling and for public health decisions.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease worldwide [1-3], affecting up to 25% of the global population [4].
NAFLD ranges from isolated hepatic steatosis (non-alcoholic fatty liver; NAFL) to non-alcoholic steatohepatitis (NASH), the latter characterized by presence of lobular inflammation and hepatocyte ballooning, with or without fibrosis [5]. A subset of patients with NAFLD develop progressive fibrosis, with risk of progression to cirrhosis and hepatocellular carcinoma (HCC) [6]. It has been shown that patients with NAFL can develop NASH and progressive fibrosis [7], thus NAFL cannot be entitled a completely benign entity. The definition of NASH has changed during the years. Previously, fibrosis stage was included as part of the diagnosis. Therefore, it has been unclear if NASH defined as per today is associated with adverse outcomes. NAFLD patients have an increased all-cause and liver related mortality [8, 9]. Novel studies have indicated that fibrosis stage, but no other histological characteristics, predicts all-cause and disease specific mortality in NAFLD [10, 11].

The natural history of NAFLD is still somewhat unclear. As NAFLD in most cases is a slowly progressive disease [12], a prolonged follow-up period in a cohort of sufficient size is needed to ascertain outcomes in patients with sub-clinical disease, including cases with NAFL that might progress to NASH and/or clinically significant fibrosis. Moreover, to draw relevant conclusions, a control population should be used to investigate the risk of NAFLD compared to the general population. Previous studies have either been small [10] or have not included a control population [13], why the estimates do not allow for risk calculations in patients with lower stages of

fibrosis, which might make it hard to put results into context. In addition, the methodology in our previous study [10] did not allow us to estimate other outcomes than mortality, whilst the study by Angulo and co-workers [13] has been criticised for a high loss to follow-up (16%), and a high proportion of patients with fibrosis stage 0 (52%) at baseline, making it improbable for this group to experience a liver-related event during the relative short follow-up time of this study.

The aim of our study was to investigate the long-term prognosis in a large cohort of NAFLD patients with a prolonged follow-up time, and to study the specific effect of NASH on the outcomes of mortality and liver-specific morbidity.

Material and methods

Subjects

We conducted a retrospective cohort study including all patients diagnosed with biopsy-proven NAFLD, at the Karolinska University Hospital, Huddinge and Linköping University Hospital, during 1971 to 2009. All biopsies were categorized by a pathologist at the time of biopsy using the systemized nomenclature of medicine (SNOMED).[14] The code for hepatic steatosis (M50080) has not changed since the start of the study period and was used to identify all liver biopsies with steatosis (N=2644). All patients' medical charts were scrutinized in detail. Patients with other causes for steatosis than NAFLD or diagnosed with any concurrent liver disease during follow-up were excluded. Patients that reported a daily alcohol consumption of more than 30 g for men or 20 g for women at baseline or during follow-up were categorized as having alcoholic fatty liver disease and were excluded. Patients that

reported binge drinking defined as five or more units of alcohol for men and four or more units for women on the same occasion were also excluded, as were patients with any concurrent liver disease diagnosed at the time of biopsy or during follow-up. Patients on treatment with drugs associated with hepatic steatosis or hepatotoxicity at the time of biopsy were also excluded. Furthermore, ten patients with a diagnosis of either HCC or decompensated liver disease at or within six months from baseline were excluded. A flow-chart for patient inclusion and exclusion is presented in Fig. 1.

Variables

Histopathological evaluation

All liver biopsies were stored in archives after the initial assessment. The old slides were usually well preserved but in those with fading staining new sections and staining were performed. Biopsies of insufficient quality (e.g. macerated during the staining procedure) or size (< 7 portal tracts), or where no steatosis was present were excluded from the cohort (N=57). Sixty-nine liver biopsies were not available for reassessment but had previously been reassessed by another experienced liver pathologist as part of a prior follow-up study [8], these were used for fibrosis staging but not for analysis of activity scores as the reproducibility for necroinflammatory changes between the two pathologists were low (κ=0.062). One expert liver pathologist (R.H) reviewed all available biopsies, which were scored according to the NAFLD activity score (NAS).[15] Lobular inflammation and steatosis were scored on a 0-3 scale and ballooning and portal inflammation on a 0-2 scale.[15] The FLIP algorithm was used to define presence of NASH.[16, 17] Fibrosis stage was scored according to the Kleiner classification on a 5-point scale (F0-F4). [15]

Baseline characteristics

All diagnoses at the time of liver biopsy were registered from patient charts. Type 2 diabetes mellitus (T2DM) was defined as a registered diagnosis in patient charts, a non-fasting glucose value of ≥ 180 mg/dL or a fasting glucose value of ≥126 mg/dL, or having any anti-diabetic medication prescribed. Hypertension was defined as a registered diagnosis in patient charts, a resting blood pressure of ≥ 140/90 mmHg or having any anti-hypertensive medication prescribed. Cardiovascular disease was defined as having previous ischemic heart disease or known angina pectoris, a previous stroke or intermittent claudication. Hyperlipidaemia was defined as having any anti-lipidaemic drug prescribed or a fasting total cholesterol value of ≥ 240 mg/dL. Smoking was defined as being a current smoker or having smoked previously. Weight and height were objectively measured by hospital staff and used to calculate body mass index (BMI) as kg/m^2.

Biochemical variables

Routine biochemical variables within one month of liver biopsy were registered and included alanine and aspartate aminotransferase, albumin, bilirubin, alkaline phosphatase, and gamma-glutamyl transferase levels, complete blood count, fasting cholesterol and triglycerides, fasting glucose, autoantibodies and α_1-antitrypsin levels. In cases with missing data, multiple imputation was used. Analysis for detection of hepatitis B surface antigen was performed in all cases and anti-hepatitis C virus (HCV) antibodies were analysed in cases evaluated after 1991 when testing became available.

Follow-up

The personal identification number (PIN) is a unique ten-digit code provided to all Swedish citizens.[18] The PIN was first used to create a control population derived from Statistics Sweden using ten controls per case (N=6345 after exclusions). Matching was performed for sex, age and municipality at the time of biopsy. Patients and controls (N=6991) were then linked to national, population-based registers. We used outcome data from the National Patient Register of Hospital Discharges (NPR), from the Cause of Death Register (CDR) and from the Swedish Cancer Register (SCR). The validity of hospital discharge diagnoses obtained from the NPR is between 85-95% depending on diagnosis.[19] The CDR contains data regarding the causes of death of all Swedish inhabitants, including if the person died abroad. It is mandatory for the responsible physician to report the underlying cause of death (e.g. HCC) and any disease that could have contributed to the death of the individual (e.g. liver cirrhosis). The SCR contains data on verified solid and non-solid tumours, irrespective of the diagnostic modality. The completeness of the register is around 96%.[20]

Outcomes

The registries were used to ascertain all causes of death and all cases of severe liver disease during follow-up. Severe liver disease was defined as an ICD-code for liver failure, cirrhosis, HCC or decompensated liver disease. Decompensated liver disease was in turn defined as an ICD-code for esophageal varices (bleeding or not bleeding), ascites or hepatic encephalopathy, while liver failure was defined as having a specific ICD-code for liver failure only. A table of all diagnoses used to define the outcomes is presented in supplementary table 1.

Statistical analysis

Data is expressed as means with standard deviations (SD), or as total numbers with percentages where applicable. The Mann-Whitney test was used to compare continuous data between groups, and the Chi squared test was used for categorical data.

Cox regression models were used to estimate hazard ratios for mortality and severe liver disease, respectively. For the comparison of cases with NAFLD against controls, the distribution of sex, age and residence was balanced due to the matching. Thus, no further adjustments were made. Also, data on other variables was lacking in the control population.

Models that investigated within-group differences in NAFLD cases were stratified on fibrosis stage or presence of NASH, and adjusted using an a priori defined model, with age at biopsy, sex and presence of type 2 diabetes mellitus at baseline as covariates. Additionally, models that were stratified on fibrosis stage were adjusted for presence of NASH, and models that were stratified on NASH were adjusted for fibrosis stage. Here, a likelihood ratio test was used to test if presence of NASH or fibrosis stage significantly adjusted the predictive capacity of the model. Laplace regression [21] was used to estimate the time to the first event of severe liver disease for controls and per fibrosis stage for F0-F3 for the 10^{th}-percentile of patients. For patients with F4 at baseline, time to development of decompensated liver disease (ascites, esophageal varices, encephalopathy) or HCC was estimated. Time was modelled on the log scale, and fibrosis, age, BMI, sex and diabetes were used to model the covariance matrix. Death from other causes was treated as censored in the Laplace analysis. Laplace regression was used instead of Cox regression as it allows for a better estimation of time to event.

Also, we specifically investigated the risk of NASH in patients with lower stages of fibrosis (stage 0-2), thus excluding cases with stage 3-4 from this model.

In addition, specific models to investigate the risk of development of severe liver disease and overall mortality per individual histological feature (fat, lobular inflammation, ballooning and portal inflammation) were performed, adjusted for age, sex, diabetes type 2 and fibrosis stage. The end of follow-up was December 31 2014, or time of event (whichever occurred earliest). All analyses were made using STATA 13 (StataCorp. 2013. College Station, TX).

A number of sensitivity analyses were performed. Firstly, a competing risk regression [22] for the outcome severe liver disease was performed. Secondly, we excluded cases where hepatitis C testing had not been performed at baseline or during follow-up. Thirdly, cases with follow-up of less than 6 months were excluded from the analyses.

Ethical considerations

The regional ethics committees at Karolinska Institutet and Linköping University approved the study. (*Dnr 2011/905-31/2* and *2015/1591-32*).

Results

The cohort with data on fibrosis stage included 646 cases and consisted of 402 men (62.2%). Mean age at baseline was 48.2 (SD 13.7) years, and mean BMI was 28.3 (SD 4.0) kg/m^2. There was a strong collinearity between higher stages of fibrosis and higher NAS and presence of NASH, with 82% of patients with stage 4 having a NAS

of 5-8 and NASH in 94% compared to a NAS of 5-8 in 24% and NASH in 35% in patients with stage 0 (p<0.001, respectively). Baseline characteristics of the cohort are listed in table 1 and distribution of NAS and NASH per fibrosis stage is presented in the supplementary table 4.

The cohort was followed for a mean period of 19.9 years (SD 8.7, range 0.4-40.0 years), corresponding to 139.163 person-years in the entire cohort and 12.631 person-years in NAFLD cases. Data on presence of NASH could be ascertained in 577 of these cases.

Patients in which data on NAS and presence of NASH could not be ascertained (N=69) were comparable to the rest of the cohort (for instance, 13.0 vs. 12.0 % of patients had fibrosis stage 3-4, p=0.79 and mean age was 50.3 vs 47.9 years, p=0.18).

Mortality

During follow-up, a total of 214 cases and 1903 controls died (p=0.10). Liver-related (7.9% in cases vs 1.4% in controls, p<0.001) and endocrine-related mortality including diabetes (5.1% vs 2.7%, p=0.02) were significantly more common in NAFLD cases than controls. Causes of death in cases and controls are listed in table 2.

Regression analyses

Compared to controls

The entire group of NAFLD patients showed a trend towards higher risk for mortality than controls (HR 1.14, 95%CI 0.99-1.32, p=0.07). Patients with fibrosis stage 0 (HR 0.86, 95%CI 0.65-1.16, p=0.33) and stage 1 (HR 0.88, 95%CI 0.70-1.12, p=0.30) did not have an increased mortality, while patients with stage 2 (HR 1.36, 95%CI 1.02-

1.80, p=0.03), stage 3 (HR 2.54, 95%CI 1.79-3.60, p<0.001) and stage 4 (HR 5.19, 95%CI 3.06-8.79, p<0.001) did with increasing HR:s per stage of fibrosis. Patients with NASH had a slight increase in overall mortality (1.22, 95%CI 1.02-1.46, p=0.03).

NAFLD cases

Using NAFLD cases with fibrosis stage 0 as controls, and adjusting for age, sex and T2DM, a significant increase in mortality was seen for patients with fibrosis stage 3 (HR 1.76, 95%CI 1.02-3.06, p=0.04) and stage 4 (HR 3.75, 95%CI 1.81-7.73, p<0.001). Adding presence of NASH to the model with fibrosis stage as the independent variable did not add to the predictive capacity of the model for any stage of fibrosis (LR test p > 0.05 for each stage of fibrosis), while adding information on fibrosis stage significantly improved the model with NASH as an independent variable (LR test p<0.001). Hazard ratios for overall mortality for the respective models are presented in table 3, and a Kaplan-Meier plot stratified on fibrosis stage and compared to controls is presented in Fig. 2.

Severe liver disease

Time to severe liver disease and decompensated liver disease

During follow-up, 76 patients (11.8%) and 139 controls (2.2%) developed severe liver disease (p<0.001). There were 12 cases of HCC in the NAFLD group (1.9%) compared to 18 in the control group (0.3%) (p<0.001). When stratified on fibrosis stage, severe liver disease was diagnosed in 2.2% of the control group, 7.4% in patients with F0, 6.3% in F1, 12.1% in F2 and 25.9% in F3. Decompensated liver disease occurred in 1.8% of controls, 3.7% in F0, 4.3% in F1, 8.7% in F2, 12.1% in

F3 and in 45% of patients with F4. Laplace regression revealed that the time until the first 10% of the patients had developed severe liver disease was 30.5 years in F0 (95%CI 21.5-39.6), 35.6 years in F1 (95%CI 25.6-45.4), 19.4 years in F2 (95%CI 9.3-29.5) and 6.0 in F3 (95%CI 2.3-9.6). The time until 10% of the patients had developed liver decompensation was 33.4 years for F0 (95%CI 24.2-42.6), 34.1 years for F1 (95%CI 25.1-43.2), 22.7 years for F2 (95%CI 13.7-31.7), 11.8 years for F3 (95%CI 4.3-19.4) and 5.6 years for F4 (95%CI 0.9-10.3).

Influence of NASH

There was no significant difference in the number of severe liver disease cases in patients without and with NASH (9.8% vs. 12.8%, p=0.29). Patients with NASH developed severe liver disease slightly earlier than patients without (mean time to first event 17.7 years in NASH vs 19.4 years, p=0.02). Adding presence of NASH to the model with fibrosis stage as the independent variable did not add to the predictive capacity of the model for any stage of fibrosis (LR test $p > 0.05$ for each stage of fibrosis), while adding information on fibrosis stage significantly improved the model with NASH as an independent variable (LR test p=0.002).

Regression analyses

Compared to controls

Cases with NAFLD were significantly more likely than controls to develop severe liver disease (HR 4.25, 95%CI 3.09-5.84, $p<0.001$), and fibrosis stage was highly predictive of this. Using the control population as a reference, fibrosis stage was associated with an increased risk for severe liver disease with increasing HRs per increasing stage, with significant results found for stage 2 (HR 5.48, 95%CI 3.10-

9.70, p<0.001), stage 3 (HR 14.28, 95%CI 7.90-25.80, p<0.001) and stage 4 (HR 104.52, 95%CI 57.20-191.10, p<0.001). Patients with fibrosis stage 0-2 and NASH had a slightly higher risk for severe liver disease compared to their controls (HR 2.79, 95%CI 1.68-4.62, p<0.001) than patients with F0-2 without NASH (HR 2.19, 95%CI 1.12-4.31) (p for difference 0.01). This was largely attributed to a higher proportion of subjects with F2 in the NASH group (36.8 vs 11.6 %, p<0.001).

NAFLD cases

Within the NAFLD group and using fibrosis stage 0 as the reference category, after adjustment for age, sex and T2DM, a significantly increased risk was found for fibrosis stage 3 (HR 4.31, 95%CI 1.63-11.36, p=0.003) and for stage 4 (HR 42.41, 95%CI 15.60-115.34, p<0.001) but not for stage 2 (HR 2.00, 95%CI 0.78-5.12, p=0.15). Importantly, adding NASH as a covariate to the model did not affect these estimates significantly (LR test p >0.05 for each stage of fibrosis). Additionally, presence of NASH was associated with an increased risk for severe liver disease in univariate analysis (HR 2.04, 95%CI 1.01-4.11, p=0.05), but not after adjustments for age, sex, T2DM and fibrosis stage (HR 0.62, 95%CI 0.28-1.40, p=0.25). Patients with F0-2 and NASH did not have an increased risk of severe liver disease compared to patients with F0-2 without NASH either in univariate (HR 1.28, 95%CI 0.57-2.89) or multivariate analysis (HR 1.24, 95%CI 0.54-2.84, p=0.62).

Data on the risk of development of severe liver disease is presented in table 4, and a Kaplan-Meier plot stratified on fibrosis stage and compared to controls are presented in Fig. 3.

Specific histological features

No individual histological feature was associated with an increased risk for overall mortality or severe liver disease after adjustment for age, sex, T2DM and fibrosis stage. These data are presented in supplementary appendix, table 3.

Sensitivity analyses

In competing risk regression, presence of NASH was not associated with an increased risk for severe liver disease after adjustment for age, sex, T2DM and fibrosis stage (HR 0.72, 95%CI 0.32-1.66, p=0.45). Fibrosis stage 3-4 was highly predictive of severe liver disease after adjustment for age, sex, T2DM and presence of NASH (HR 5.37, 95%CI 2.73-10.56, p<0.001).

Hepatitis C testing had been performed at baseline or during follow-up in 492 cases (76%). Of the 154 cases in whom HCV had not been actively excluded, only 12 cases had been investigated after 1991. There was no difference in the number of clinical events regarding severe liver disease in the HCV-tested group versus the untested group (11.4% vs 13.0% experienced an event, p=0.59). No significant changes in the regression models was found (data not shown).

Removing cases with a follow-up of less than six months (N=48, 0.7%) did not change any of the main results (data not shown).

Discussion

This is the largest study investigating long-term outcomes in patients with biopsy-proven NAFLD. We show that after a follow-up time of up to 40 years, presence of NASH did not affect the long-term outcome of NAFLD significantly.

Importantly, we found that patients with lower stages of fibrosis (stage 0-1) had the potential to progress to severe liver disease and this occurred in roughly 7% of these patients, independent of presence of NASH. The lower end of the 95% CI for the 10th percentile of liver decompensation was 0.9 years in manifest cirrhosis (5.6, 95%CI 0.9-10.3). The lower end of the 95% CI for the 10th percentile of time to severe liver disease was 2.3 years in F3 (6.0, 95%CI 2.3-9.6), 9.3 years in F2 (19.4, 95%CI 9.3-29.5) and around 22-26 years in F0-1 (30.5 years in F0, 95%CI 21.5-39.6, and 35.6 years in F1, 95%CI 25.6-45.4).

Thus, we can infer that this is the minimum time needed to develop clinically relevant hepatic endpoints in NAFLD per stage of fibrosis, which can be used to guide individual patient prognosis and when constructing screening and follow-up programs.

Although patients with NASH seemed to have a small increase in risk for overall mortality and liver-specific morbidity in univariate analyses compared to a reference population, this was generally not significant when adjusting for confounders including fibrosis stage, and no added risk was seen for presence of NASH after adjusting for confounders. Furthermore, in patients with similar stages of fibrosis, presence of NASH did not affect the estimates in any specific manner.

There was a clear collinearity between NASH and higher stages of fibrosis, with 94% of patients with F4 having NASH compared to 35% of patients with F0, and 17% of patients with NASH having stage 3-4 fibrosis compared to 2% of patients without NASH. It is highly likely that the excess risk of NASH on mortality seen in this and other studies is due to collinearity between NASH and higher stages of fibrosis. Adjusting the model with fibrosis stage as an independent variable for NASH did not change the estimates in a significant manner, while adjusting models with NASH as an independent variable for fibrosis stage significantly reduced the estimates.

Previous studies

Our results support data from previous studies, including our recent publication that used a smaller subset of this cohort [10], and other studies [13, 23], that primarily higher stages of fibrosis are predictive of mortality in NAFLD. In a recent meta-analysis, partly built on these cohorts, also lower stages of fibrosis were associated with overall and liver-specific mortality [24]. However, as data on confounders such as age were missing in the analysis, these data should be interpreted cautiously. Indeed, in our study, fibrosis stage 2 was associated with an increased risk of mortality compared to controls in univariate analysis, but this risk was obviated when adjusting for covariates including age, indicating that it is primarily higher stages of fibrosis that best predict future mortality.

Strengths and limitations

The main strengths of this study are that all patients were diagnosed with NAFLD using gold standard, i.e. liver biopsy. All histological slides included in the analysis were of good quality, and all cases were examined by the same expert pathologist,

eliminating inter-observer variability. The cohort is the hitherto largest of its kind (N=646), and the follow-up time is the longest ever documented in biopsy-proven NAFLD. This is critical in allowing for enough time to pass in patients with minimal disease at baseline to develop cirrhosis, as the time needed for this was at least 22 years in this study.

Furthermore, due to high-quality Swedish registers all outcomes could be ascertained. The selection of robust outcome criteria such as mortality or variceal bleeding allowed us to capture all cases with severe liver disease. However, due to the fact that outpatient visits in specialist care were not entered into the NPR before 2001, we probably have missed cases with compensated cirrhosis only seen in outpatient care before this, why our estimates might be falsely low. Estimates for mortality and decompensated liver disease are not affected by this. An important strength is also the extensive chart review process not only at baseline but also during the entire follow-up period, which allowed us to identify cases of alcoholic liver disease that originally were classified as NAFLD.

Limitations include that selection of patients were made from two university hospitals, indicating that selection bias could be present. Indeed, 66.4 % of patients had NASH and 12% had advanced fibrosis (F3-4), which is higher than in a general population [25]. The use of liver biopsy was more common in the past decades, and is reflected by that 81.6% of cases in this study underwent biopsy before the year 2000.

Knowledge of HCV infection was not available in cases included before 1991, when testing became available. However, during the chart review and histological review processes, we excluded cases with atypical histological appearance, or with typical risk factors for HCV infection, such as previous intravenous drug abuse. Information

on HCV status was available in 76% of patients, and excluding cases with unknown HCV status did not change the main results.

Additionally, apart from register data and patient charts, we did not obtain detailed individual patient data, including new histologic and biochemical data as well as BMI during follow-up. This was not deemed as feasible for this study, also a clear selection bias would have been introduced as we would not have been able to gather data on deceased patients. Additionally, given the high prevalence of NAFLD in the general population, there are likely many cases with NAFLD in the control group. This means that our estimates might in fact be falsely low since the effect of NAFLD could be diluted.

Implications

The timelines for development of severe liver disease obtained from this study could have implications for patient follow-up, but depends on local resources and potential future effective treatments. In general, NAFLD patients with F3 and F4 should be closely monitored by a hepatologist in order to assess liver-related complications. Patients with F2 should be re-examined within a period not more than ten years to identify cases with progressive fibrosis. In patients with F0 and F1, the large majority do not develop severe liver disease within 20 years. However, re-evaluation is likely beneficial to identify cases with progressive fibrosis in selected patients. Of importance, absence of NASH should not influence time to follow-up.

Currently, the market for pharmacological treatment of NAFLD is highly valued and large phase III trials are under way. The main outcomes used in these trials is

reducing the rate of fibrosis progression, with resolution of NASH as a surrogate endpoint. Our results suggest that the presence of NASH is not significantly associated with an increased risk of either overall mortality or liver-specific morbidity. Although our study cannot determine if resolution of NASH per se is associated with a reduced risk of mortality or severe liver disease, the size of the estimates for NASH are generally small compared to those for fibrosis stage, and it is unlikely that resolution of NASH without impact on fibrosis stage will have an impact on these outcomes. Future studies should examine if resolution of NASH per se is associated with a long-term reduction in these risks, independent of simultaneous reduction in fibrosis stage.

Conclusions

This is the largest study including NAFLD patients diagnosed with gold standard i.e. liver biopsy, with the longest documented follow-up time. Time to cirrhosis and liver decompensation per stage of fibrosis is reported. Fibrosis stage was the most robust marker for future mortality and liver-specific morbidity and this was robust when adjusting for important confounders, while adding data on NASH added very little to the prognosis. We suggest a deeper emphasis on reducing fibrosis stage as an endpoint for future clinical trials in NASH.

Tables

Parameter	Complete data, N	Mean/Frequency (SD/percent)
NASH (yes)	577	383 (66.4)
Fibrosis stage 0	646	164 (25.4)
Fibrosis stage 1		256 (23.6)
Fibrosis stage 2		149 (23.1)
Fibrosis stage 3		58 (9.0)
Fibrosis stage 4		20 (3.1)
Steatosis grade 1	580	228 (39.3)
Steatosis grade 2		149 (25.7)
Steatosis grade 3		203 (35.0)
Lobular inflammation 0	579	52 (9.0)
Lobular inflammation 1		245 (42.3)
Lobular inflammation 2		220 (38.0)
Lobular inflammation 3		62 (10.7)
Ballooning 0	577	189 (32.8)
Ballooning 1		207 (35.9)
Ballooning 2		181 (31.4)
Portal inflammation 0	579	254 (43.9)
Portal inflammation 1		226 (39.0)
Portal inflammation 2		99 (17.1)
Hypertension (yes)	646	196 (30.3)
T2DM (yes)	646	93 (14.4)
Smoking (never)	569	355 (55.0)
Smoking (previous)		136 (21.0)
Smoking (current)		155 (24.0)
BMI (kg/m2)	546	28.3 (4.1)
Sex (male)	646	402 (62.2)
Age (years)	646	48.2 (13.7)
AST (U/L)	631	50 (34)
ALT (U/L)	632	84 (52)
GGT (U/L)	540	109 (127)
ALP (U/L)	625	91 (47)
Bilirubin (mg/dL)	606	0.69 (0.45)
INR	505	1.0 (0.1)
Albumin (g/L)	573	4.2 (0.4)
Hemoglobin (g/dL)	562	14.8 (1.2)
Platelet count ($x10^9$)	551	247 (73)

Total cholesterol (mg/dL)	462	233 (54)
Triglycerides (mg/dL)	430	208 (146)
Fasting glucose (mg/dL)	432	108 (40)
Ferritin (µg/L)	355	237 (249)

Table 1. Demographic, clinical and histopathological characteristics of study subjects at baseline. Data are presented as mean values with standard deviations for continuous parameters, and as frequencies and percentages for categorical parameters. Complete data indicates the number of patients where data was available per presented parameter. Hypertension was defined as blood pressure ≥ 140/90 mmHg or requiring treatment, type 2 diabetes as a fasting plasma glucose ≥ 126 mg/dL or a non-fasting plasma glucose ≥ 180 mg/dL, or requiring treatment. Abbreviations: T2D, type 2 diabetes; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; ALP, alkaline phosphatase; INR, international normalized ratio; SD, standard deviation.

Cause of death	**Cases, N**	**%**	**Controls, N**	**%**	
	N=646		N=6345		P-value
Psychiatric disorder (including suicide)	2	0.9	85	4.5	**0.02**
Infections	4	1.9	19	1.0	0.18
Kidney disease	4	1.9	20	1.1	0.21
External trauma	4	1.9	38	1.9	0.87
Gastrointestinal	5	2.4	53	2.8	0.87
Nervous system	6	2.8	67	3.5	0.76
Other	9	4.2	130	6.9	0.23
Endocrine (including T2DM)	11	5.1	52	2.7	**0.02**
Liver-related	17	7.9	26	1.4	**<0.001**
Respiratory disease	18	8.4	121	6.4	0.13
Extrahepatic malignancy	55	25.7	544	28.6	0.96
Cardiovascular	79	36.9	748	39.3	0.74
Total	214	33.1	1903	30.0	0.1

Table 2. Causes of death in NAFLD cases and controls. Abbreviation: T2DM, type 2 diabetes mellitus.

Table 3				
Hazard ratios for overall mortality				
3a NAFLD vs controls	N Cases	HR	95%CI	P
NAFLD (yes)	646	1.14	0.99-1.32	0.07
Fibrosis stage				
Controls	6345	(ref)		
Cases				
F0	163	0.87	0.65-1.16	0.33
F1	255	0.88	0.70-1.12	0.3
F2	149	1.36	1.02-1.80	0.03
F3	58	2.54	1.79-3.60	<0.001
F4	20	5.19	3.06-8.79	<0.001
Controls	6345	(ref)		
Case, NAFL	194	0.97	0.75-1.25	0.83
Case, NASH	383	1.22	1.02-1.46	0.03

3b within NAFLD group		**Model 1**			**Model 2 (+age. sex. T2DM)**			**Model 3 (+NASH)**			**LR test***
Fibrosis stage	N Cases	HR	95%CI	P	HR	95%CI	P	HR	95%CI	P	P
0	163	(ref)									0.57
1	255	1.01	0.70-1.46	0.95	0.97	0.67-1.40	0.87	0.91	0.61-1.36	0.66	0.34
2	149	1.6	1.09-2.39	0.02	1.12	0.75-1.69	0.57	1.11	0.70-1.76	0.66	0.14
3	58	3.04	1.94-4.78	<0.001	1.79	1.12-2.86	0.01	1.76	1.02-3.06	0.04	0.41
4	20	6.53	3.55-12.03	<0.001	2.86	1.53-5.32	0.001	3.75	1.81-7.73	<0.001	0.51
		Model 1			**Model 2 (+age, sex, T2DM)**			**Model 3 (+fibrosis stage)**			**LR test***
NASH (yes)	383	1.27	0.94-1.73	0.12	1.04	0.76-1.42	0.82	0.83	0.58-1.17	0.29	<0.001

Table 3. Hazard ratios (univariate and multivariate) for overall mortality. In table 3a, comparisons are made against matched controls (age, sex and municipality). In table 3b, comparisons are made against fibrosis stage 0 or absence of NASH. Model 1: crude. Model 2: Adjusted for age, sex and type 2 diabetes. Model 3: further adjusted for fibrosis stage or NASH. * Likelihood ratio test for adding NASH or fibrosis stage to model 2.

Table 4					
Hazard ratios for development of severe liver disease	**N Cases**	**N Failures**			
4a NAFLD vs controls			HR	95%CI	P
	646	186	4.25	3.09-5.84	<0.001
Fibrosis					
Controls	6345	139	(ref)		
Cases					
F0	163	12	1.92	0.90-4.10	0.09
F1	256	16	1.65	0.84-3.24	0.15
F2	149	18	5.48	3.10-9.70	<0.001
F3	58	15	14.28	7.90-25.8	<0.001
F4	20	15	104.52	57.2-191.1	<0.001
Controls	6345	139	(ref)		
Case, NAFL	194	19	2.44	1.28-4.64	0.007
Case, NASH	383	49	5.18	3.58-7.48	<0.001

4b within NAFLD group			**Model 1**			**Model 2 (adjusted for age, sex, T2DM)**			**Model 3 (+NASH)**			**LR test***
Fibrosis stage	N cases	N failures	HR	95%CI	P	HR	95%CI	P	HR	95%CI	P	P
F0	163	12	(ref)			(ref)						0.16
F1	256	16	0.86	0.32-2.32	0.77	0.82	0.30-2.20	0.69	0.76	0.26-2.25	0.63	0.46
F2	149	18	2.83	1.12-7.12	0.03	2.00	0.78-5.12	0.15	2.08	0.70-6.14	0.19	0.47
F3	58	15	7.34	2.88-18.75	<0.001	4.31	1.63-11.36	0.003	3.82	1.15-12.66	0.03	0.06
F4	20	15	52.81	20.05-139.-05	<0.001	42.41	15.60-115.34	<0.001	58.12	17.29-195.39	<0.001	-
				Model 1		**Model 2 (adjusted for age, sex, T2DM)**			**Model 3 (+Fibrosis stage)**			**LR test***
No NASH	194	19	(ref)			(ref)						
NASH	383	49	2.04	1.01-4.11	0.05	1.76	0.86-3.60	0.12	0.62	0.28-1.40	0.25	0.002

Table 4: Hazard ratios (univariate and multivariate) for severe liver disease. In table 4a, comparisons are made against matched controls (age, sex and municipality). In table 4b, comparisons are against F0 or absence of NASH. Model 1: crude. Model 2: Adjusted for age, sex and type 2 diabetes. Model 3: further adjusted for fibrosis stage or NASH. * Likelihood ratio test for adding NASH or fibrosis stage to model 2.

Figure legends

Figure 1. Flowchart for patient inclusion.

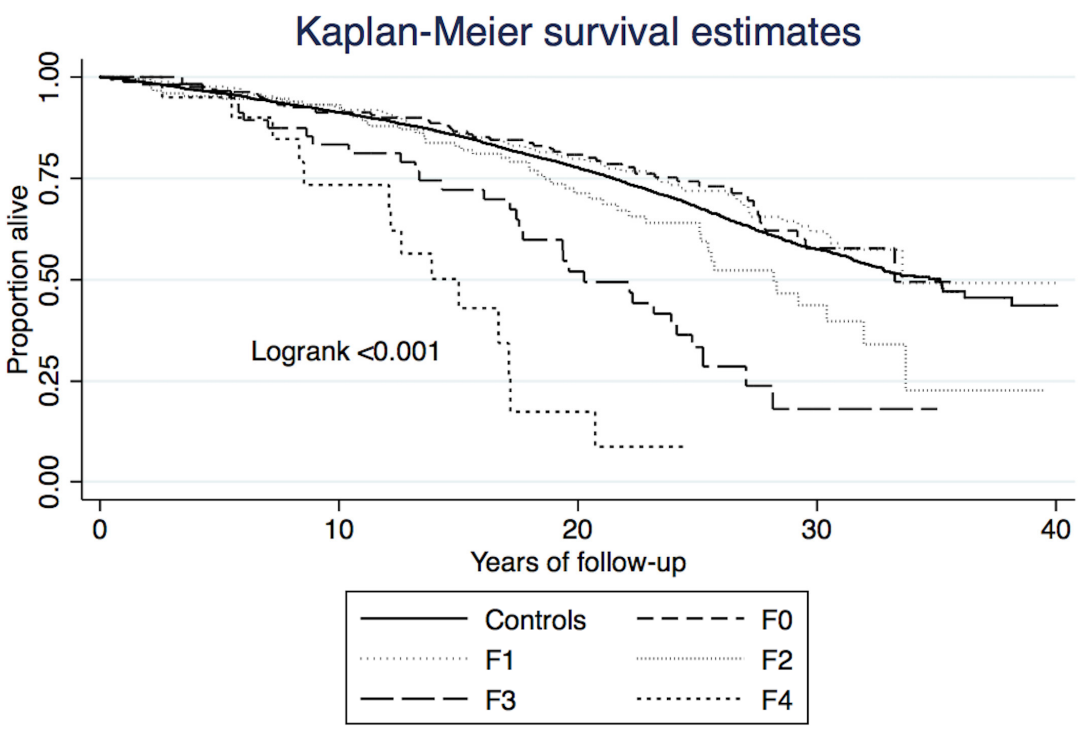
Figure 2. Overall mortality stratified on fibrosis stage compared to matched controls. Logrank test p <0.001.

Figure 3. Development of severe liver disease stratified on fibrosis stage compared to matched controls. Logrank test p <0.001.

References

[1] Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. Clin Gastroenterol Hepatol 2011;9:524-530.e521; quiz e560.
[2] Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013;10:686-690.
[3] Setiawan VW, Stram DO, Porcel J, Lu SC, Le Marchand L, Noureddin M. Prevalence of chronic liver disease and cirrhosis by underlying cause in understudied ethnic groups: The multiethnic cohort. Hepatology 2016;64:1969-1977.
[4] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73-84.
[5] Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012;55:2005-2023.
[6] Farrell GC, Larter CZ. Nonalcoholic fatty liver disease: from steatosis to cirrhosis. Hepatology 2006;43:S99-S112.
[7] McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. Journal of hepatology 2015;62:1148-1155.
[8] Ekstedt M, Franzen LE, Mathiesen UL, Thorelius L, Holmqvist M, Bodemar G, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. Hepatology 2006;44:865-873.
[9] Söderberg C, Stål P, Askling J, Glaumann H, Lindberg G, Marmur J, et al. Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. Hepatology 2010;51:595-602.
[10] Ekstedt M, Hagstrom H, Nasr P, Fredrikson M, Stal P, Kechagias S, et al. 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.
[11] Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2015;149:389-397.e310.
[12] Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis Progression in Nonalcoholic Fatty Liver vs Nonalcoholic Steatohepatitis: A Systematic Review and Meta-analysis of Paired-Biopsy Studies. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2014.
[13] Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver Fibrosis, but No Other Histologic Features, Associates With Long-Term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2015.

[14] Cote RA, Robboy S. Progress in medical information management. Systematized nomenclature of medicine (SNOMED). Jama 1980;243:756-762.
[15] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313-1321.
[16] Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology 2012;56:1751-1759.
[17] Bedossa P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology 2014;60:565-575.
[18] Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. European journal of epidemiology 2009;24:659-667.
[19] Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC public health 2011;11:450.
[20] Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta oncologica (Stockholm, Sweden) 2009;48:27-33.
[21] Bottai M, Zhang J. Laplace regression with censored data. Biometrical journal Biometrische Zeitschrift 2010;52:487-503.
[22] Jepsen P, Vilstrup H, Andersen PK. The clinical course of cirrhosis: The importance of multistate models and competing risks analysis. Hepatology 2015;62:292-302.
[23] Younossi ZM, Stepanova M, Rafiq N, Makhlouf H, Younoszai Z, Agrawal R, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011;53:1874-1882.
[24] Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased risk of mortality by fibrosis stage in non-alcoholic fatty liver disease: Systematic Review and Meta-analysis. Hepatology 2017.
[25] Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 2011;140:124-131.



JHEPAT-D-17-00894R1

Highlights

- Identifying NAFLD patients with high risk for adverse outcomes is a challenge
- We here present data from the hitherto largest cohort study of NAFLD patients
- Presence of NASH did not increase the risk of adverse outcomes
- Time to severe liver disease per stage of fibrosis is reported