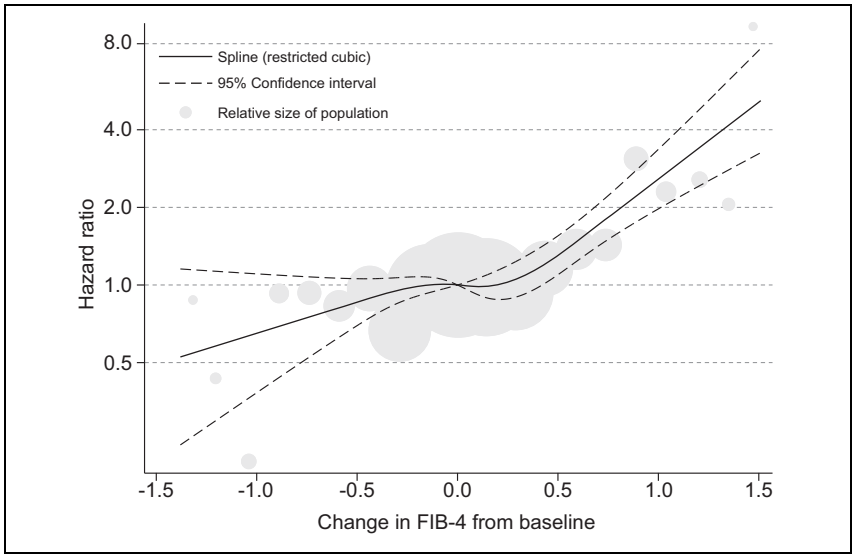
Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL
OF HEPATOLOGY

Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease

Graphical abstract



Highlights

- An increase in FIB-4 over time is associated with risk of severe liver disease.
- Repeating FIB-4 tests can help to identify those at risk of severe liver disease.
- 50% of severe liver disease outcomes had consistently low or intermediate FIB-4.
- About one-third of the cohort had intermediate or high FIB-4 at one of the tests.
- FIB-4 is likely insufficient for screening of fibrosis in the general population.

Authors

Hannes Hagström, Mats Talbäck, Anna Andreasson, Göran Walldius, Niklas Hammar

Correspondence

hannes.hagstrom@ki.se
(H. Hagström).

Lay summary

The fibrosis-4 scoring system is often used to estimate the risk of advanced fibrosis in liver diseases. Herein, we found that changes in this score over time are associated with the risk of future severe liver disease in a population-based cohort. However, even if the prediction is improved by repeated testing, the overall ability of the score to predict future events is relatively low.

https://doi.org/10.1016/j.jhep.2020.06.007
 J. Hepatol. 2020, 73, 1023–1029

Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL OF HEPATOLOGY

Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease

Hannes Hagström[1,2,3,*], Mats Talbäck[4], Anna Andreasson[5], Göran Walldius[4], Niklas Hammar[4]

[1]Division of Hepatology, Department of Upper GI Diseases, Karolinska University Hospital, Stockholm, Sweden; [2]Clinical Epidemiology Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden; [3]Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden; [4]Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [5]Stress Research Institute, Stockholm University, Stockholm, Sweden

Background & Aims: It is unclear whether the identification of individuals at risk of cirrhosis using non-invasive tests can be improved by repeated measurements. Herein, we tested whether repeated measurements of fibrosis-4 index (FIB-4) could improve the identification of individuals at risk of severe liver disease.

Methods: Data were derived from the population-based Swedish AMORIS cohort with baseline examinations from 1985-1996. FIB-4 was calculated at 2 time points within 5 years. Thereafter, we associated changes in FIB-4 with outcomes. Incident severe liver disease data was ascertained through linkage to Swedish national registers until 2011. Hazard ratios (HRs) and CIs for outcomes were calculated using Cox regression.

Results: Of 126,942 individuals with available FIB-4 data, 40,729 (32.1%) underwent a second test within 5 years (mean interval 2.4 years). During 613,376 person-years of follow-up, 581 severe liver disease events were documented (0.95/1,000 person-years). An increase of 1 unit in FIB-4 was associated with an elevated risk of severe liver disease (adjusted hazard ratio [aHR] 1.81; 95% CI 1.67–1.96). Transitioning from a low- or intermediate- to a high-risk group was associated with an increased risk of severe liver disease compared with those consistently in the low-risk group (aHR 7.99 and 8.64, respectively). A particularly increased risk of severe liver disease was found in individuals defined as high risk at both tests (aHR 17.04; 95% CI 11.67–24.88). However, almost half of all events occurred in those consistently in the low-risk group.

Conclusions: Repeated testing of FIB-4 within 5 years improves the identification of individuals at an increased risk of severe liver disease in the general population. However, the sensitivity is comparatively low and improved tests are needed for screening in a general population or primary care setting.

Lay summary: The fibrosis-4 scoring system is often used to estimate the risk of advanced fibrosis in liver diseases. Herein, we found that changes in this score over time are associated with the risk of future severe liver disease in a population-based cohort. However, even if the prediction is improved by repeated testing, the overall ability of the score to predict future events is relatively low.

Introduction

Advanced fibrosis (stage 3-4 by liver biopsy) is the major predictor of clinically significant liver-related outcomes.[1–3] Thus, defining the presence or absence of advanced fibrosis is key in determining prognosis in individuals with known or suspected chronic liver disease. Individuals without advanced fibrosis have a low risk of progression to cirrhosis within a 10-15-year time frame.[2,4] Conversely, those with advanced fibrosis more frequently experience severe liver-related endpoints and have higher overall mortality[1–3] The gold standard for diagnosing fibrosis is liver biopsy, which is not reasonable to use as a screening tool in larger populations, expressly in a general population or primary care setting. Several non-invasive scores have been developed to identify individuals with prevalent advanced fibrosis.[5–7] These scores have all been made from selected populations exposed to liver biopsy with a high prevalence of advanced fibrosis; their use in general population settings with a much lower prevalence of advanced fibrosis is limited. Recently, we showed that the capacity of 5 non-invasive scores to predict incident severe liver disease in a general population setting was modest.[8]

Whether repeated measures of the available non-invasive screening tools would improve the usefulness of these tools and whether improvement or worsening in these measures is associated with an improved or worsened prognosis has not been determined.

Herein, we tested the general hypothesis that repeated measurements of the commonly used fibrosis-4 index (FIB-4) would improve the identification of individuals at risk of severe liver disease compared with a single measurement. Our specific aims were to i) investigate the association of changes in FIB-4 measured at 2 time points with incident severe liver disease in the general population and ii) examine the natural course of FIB-4 in the same population.

Material and methods

Study population

We used data from the Swedish Apolipoprotein MOrtality RISk (AMORIS) cohort. AMORIS is a general population cohort that

Keywords: NAFLD; Cirrhosis; Epidemiology; Prognosis; FIB-4.
Received 17 March 2020; received in revised form 4 May 2020; accepted 3 June 2020; available online 1 July 2020

* Corresponding author. Address: C1:77, Division of Hepatology, Karolinska University Hospital Huddinge, 141 86 Stockholm, Sweden. Tel.: +46 (0) 8 5858 2305, fax: +46 (0) 8 5858 2335.
E-mail address: hannes.hagstrom@ki.se (H. Hagström).

https://doi.org/10.1016/j.jhep.2020.06.007

underwent blood sampling between 1985 and 1996.[9] The cohort includes 812,073 individuals who were either taking part in yearly routine health check-ups through occupational health screening or outpatients in primary care referred for laboratory testing. No individuals were hospitalized at the time of blood sampling. All individuals of the AMORIS cohort were residents of Sweden and predominantly living in Stockholm County (67%) at the time of blood sampling. During the testing period, the total population of Stockholm County was about 1.6 million inhabitants. Thus, the AMORIS cohort constituted a substantial part of the total population of Stockholm County during this period. A detailed cohort description is available elsewhere.[9]

Individuals were included in the study if the data required to calculate FIB-4 was available at 2 time points. We chose to focus on FIB-4 in that it was one of the best-performing scores in our previous analyses.[8] In addition, data were available for a large proportion of the initial cohort and FIB-4 is one of the most commonly used scores in clinical practice.[10]

Because FIB-4 has been found not to perform well in younger and older populations,[11] we excluded individuals below 35 and above 79 years of age. We also excluded individuals with an ICD-based diagnosis of any specific liver disease (*e.g.*, alcohol-related liver disease) at or before baseline, except for NAFLD. We also excluded individuals with a history of severe liver disease (see definition below and in the supplementary information) or any diagnosis of drug or alcohol abuse at or before baseline. Finally, we excluded people with secondary tests only within 3 months after the first test. This was done to reduce the risk of selecting people with baseline significant liver disease that led to the second test or those with falsely high laboratory tests. People diagnosed with a specific liver disease other than NAFLD or a drug- or alcohol-related disorder during follow-up were censored at the time of diagnosis. A list of all diagnoses and ICD codes used in the current study is presented in Table S1A,B.

Variables

Blood sampling and laboratory analyses

Information on all biomarkers was available from the health examinations in 1985–1996. All laboratory analyses were conducted on fresh blood serum samples (53% after overnight fasting) at CALAB Medical Laboratories, Stockholm, Sweden using a uniform and well-documented methodology. Technical specifications for the applied methods are listed in the supplementary materials. The FIB-4 was calculated as in.[5]

We categorized people into low-, intermediate- and high-risk groups for advanced fibrosis based on the following suggested cut-offs: <1.30 (low risk), 1.30–2.67 (intermediate risk and >2.67 (high risk). However, we did not change the lower cut-off for people ≥65 years of age to 2.0, as has been suggested.[11] This approach was introduced to reduce false-positive findings in individuals ≥65 years, but how this should be applied using repeated measurements has not been evaluated and is not entirely straightforward. For instance, a 64-year old person with a score of 1.9 at the first test (intermediate risk) would be re-categorized as low risk when he or she reached 65 years of age, provided that aspartate aminotransferase (AST), alanine aminotransferase (ALT) and platelets remained stable.

As a person's first test, we selected the record for which FIB-4 could be calculated for the first time. As the second test for the same person, we used the last record within a 5-year time frame. This tactic was used as we had previously shown that the prediction of incident severe liver disease is best in a shorter time frame.[8] In a sensitivity analysis we included every person with a second test within the full study period, giving a theoretical time between tests of 12 years. We chose the second test with the longest possible duration from the first test. For instance, if a person had a second test in year 3 and an additional test in year 4, the year 4 test was chosen as the time of the second test in the main analysis.

Information on covariates

The Swedish personal identification number is a unique 12-digit code provided to all Swedish residents.[12] The personal identification number was used to link the laboratory data from the study cohort to Swedish national registers and other databases to obtain information on BMI, presence of type 2 diabetes mellitus (T2DM) and other covariates[9] in those for whom such data were available.

Information on BMI was retrieved from the baseline health examinations where available but also from the Swedish Medical Birth Register, national quality of care registers and research cohorts at Karolinska Institutet previously linked to the AMORIS cohort.[10] We allowed BMI to be used if data were present within 4 years before the first test. T2DM was defined as present if the person had a serum glucose of >126 mg/dl (fasting) or >200 mg/dl (non-fasting) at baseline testing, or was listed in the Swedish National Diabetes Register or had a self-reported T2DM diagnosis from a linked research cohort, or if an ICD code corresponding to diabetes was present in the National Patient Register at or before baseline.[10] In all cases the age at first diagnosis of T2DM had to be ≥35 years to reduce the risk of misclassifying people with type 1 diabetes.

Information about socioeconomic status was obtained from the national population and housing censuses for 1970-1990.[13] Socioeconomic status was classified as blue- or white-collar workers.

Follow-up

Follow-up started at the date of the second test and ended at an outcome event, emigration, death, a diagnosis of a specific liver disease other than NAFLD (*e.g.*, hepatitis C) or end of follow-up (December 31, 2011), whichever came first. To ascertain outcomes, we used the personal identification number which is linked to nationwide Swedish registers. A description of the registers used for outcome ascertainment is available in the supplementary information. The completeness and overall quality of the registers are considered high.[13–16] Severe liver disease was defined as an ICD code corresponding to a diagnosis of cirrhosis, liver failure, hepatocellular carcinoma, liver transplantation, decompensated liver disease or death in liver disease as the main cause of death. Decompensated liver disease was defined as coding for esophageal varices, ascites, hepatorenal syndrome or hepatic encephalopathy. ICD codes used to define outcomes are listed in Table S1A.

Analyses

First, we investigated transitions from one risk group to another from the first to the second test. In the proportional hazards regression analyses those classified as low risk at both tests were used as the reference group. We also analyzed the hazard ratio (HR) associated with a 1-unit change over time in FIB-4 as a continuous variable.

Second, we estimated sensitivity, specificity, negative and positive predictive values (NPVs and PPVs) and overall test accuracy for the development of severe liver disease based on transitioning between tests. This analysis used individuals classified as low risk at both tests as the comparator group; a second group was established from those classified as intermediate in the second test; a third group was constructed from those classified as high in the second test; and a fourth group was created from those classified as high at both tests. These analyses excluded individuals that transitioned from the high- or intermediate-risk groups to the low-risk group. We also compared key characteristics of individuals included in the study to those that only had a single testing occasion where FIB-4 could be calculated.

Statistical analysis

Participant characteristics were described using means, percentages, medians and IQRs. The incidence proportion of severe liver disease was calculated as the number of events during follow-up divided by the number of individuals at risk at baseline during the defined study period. Cox proportional hazards models, with attained age as the time scale, were used to estimate HRs together with 95% CIs. Three models were estimated: model 1 adjusted for age, model 2 additionally adjusted for sex and socioeconomic status and model 3 additionally adjusted for the time between tests.

In the analysis in which the FIB-4 had been grouped into 3 risk categories at the respective time points (low, intermediate and high risk) the low-low group was used as the reference category. In the analysis in which the FIB-4 was treated as a continuous variable we used the baseline score together with the change in score between the 2 time points. The change in the FIB-4 over a 5-year period was calculated using the difference between an individual's baseline value and the last measurements between 3 months and 5 years after baseline. The average yearly change in the FIB-4 was then calculated by fitting a least-squares regression line with 95% CI to the mean of the differences for each 30-day period after baseline. In addition, we calculated the specificity, sensitivity, PPV, NPV and general test accuracy for the development of severe liver disease during the follow-up. Statistical analyses were conducted using STATA version 15.1 (StataCorp LLC, College Station, Texas, USA).

Ethical considerations

The study was approved by the Regional Ethics Committee in Stockholm (Dnr. 2010/1047-31/1). Informed consent was waived by the board because the study was strictly register-based.

For further details regarding the materials and methods used, please refer to the CTAT table and supplementary information.

Results

There were 126,942 individuals in whom the FIB-4 could be calculated at least once during the study period. We excluded individuals where FIB-4 could only be calculated once (n = 79,705). To reduce the risk of including individuals with a high probability of a falsely high FIB-4 on first testing, we also excluded 2,862 individuals who had a second test performed within 3 months of the first test, but never again after that period.

From the remaining 44,375 individuals (35.0% of the full FIB-4 cohort), 40,729 (91.8%) had the second test within 5 years from the first test. These 40,729 individuals constituted the study population for the main analysis, whereas the 44,375 individuals with a second test at any time during the 12-year baseline study period were included in a sensitivity analysis.

Table 1. Characteristics of the cohort with FIB-4 measured at 2 time points within 5 years at the time of the first and last available measurement.

Variable	First test	Last measurement
Person-years at risk	18.9 (14.8–22.0)	16.2 (12.1–19.2)
Male	16,792 (41.2%)	16,792 (41.2%)
Attained age at inclusion	54.5 (45.5–65.1)	57.1 (48.0–67.9)
Attained age at exit	72.9 (64.8–82.0)	72.94 (64.8–82.0)
Number of events after the last measurement	–	581 (1.43%)
Time between tests, years	–	2.4 (1.2–3.9)
FIB-4 value	0.91 (0.67–1.24)	0.96 (0.70–1.32)
FIB-4 Low	31,680 (77.8%)	30,210/(74.2%)
FIB-4 Intermediate	8,444 (20.7%)	9,704 (23.8%)
FIB-4 High	605 (1.5%)	815 (2.0%)
Change in FIB-4 from the first test	–	0.05 (−0.13 to 0.24)
ALT, IU/L	21 (15–30)	22 (16–31)
AST, IU/L	20 (16–25)	20 (16–25)
Platelets, 10^9	261 (222–306)	251 (213–292)
GGT, IU/L	20 (14–32)	22 (15–36)
Total cholesterol, mg/dl*	224 (197–255)	228 (201–255)
Triglycerides, mg/dl*	97 (71–150)	106 (71–159)
Glucose, mg/dl*	88 (81–97)	90 (83–99)
Blue-collar worker*	21,380 (54.9%)	21,265 (54.3%)

Categorical data presented as n (%), continuous data presented as median (IQR).
*Missing data in about 5% of the cohort. ALT, alanine aminotransferase; AST, aspartate aminotransferase; FIB-4, fibrosis-4 index; GGT, gamma-glutamyltransferase.

After the second test, the cohort was followed for a median time of 16.2 years (IQR 12.1–19.2), corresponding to 613,376 person-years. We identified 11,929 (29.29%) deaths and 581 severe liver disease events (1.43%) during the follow-up. In all, 1,212 individuals (2.98%) emigrated from Sweden and 2,871 (7.05%) were diagnosed with a specific liver disease other than NAFLD and were censored.

The median age at the first test was 54.5 years (IQR 45.5–65.1) and 41.2% were male. The median value of the FIB-4 at the first test was 0.91 (IQR 0.67–1.24) and the proportions of people in the low-, intermediate- and high-risk groups were 77.8%, 20.7% and 1.5%, respectively.

Characteristics of the cohort at the time of the first and second tests are presented in Table 1, while corresponding information stratified by risk groups based on the first and second tests is shown in Table S3. Differences in key parameters between individuals included in this study compared to those that only had a single test (n = 79,705) are presented in Table S4. In brief, those included were slightly older (55.0 *vs.* 52.4 years) but the overall risk of severe liver disease was similar (mean difference 0.09%; 95% CI −0.04 to 0.24).

The median time between tests was 2.4 years (IQR 1.2–3.9). The mean annual change in the FIB-4 over 5 years was 0.020 units (95% CI 0.016–0.023). Men had a faster progression rate (mean annual change 0.030; 95% CI 0.025–0.035) compared with women (0.013; 95% CI 0.009–0.018) (Fig. 1). This increase was similar using data from all tests during the 5-year period (estimated annual change 0.027; 95% CI 0.024–0.031) and slightly higher in the sensitivity analysis using data from the full 12-year follow-up (mean annual change 0.024; 95% CI 0.022–0.025).

The rate of change was also associated with age, with a somewhat faster progression in individuals ≥65 years old for both men (mean annual change 0.032 *vs.* 0.029) and women (0.018 *vs.* 0.011) (Fig. S1). Of the 40,729 individuals included, 30,435 (74.7%) were below 65 at the time of the first test and of these, 2,295 (7.5%) were 65 or older at the time of the second test.

Transition between risk groups

The number and proportion of individuals that were stable or changed risk groups based on FIB-4 are presented in Table 2, along with the total number of severe liver disease events, incidence rates and corresponding HRs. About 25% of all individuals changed risk group from the first to the second test. Transitioning was less common in the group defined as low risk at the first test (13.3%) *vs.* the intermediate- (36.9%) and high-risk group (58.7%) (Table 2).

In individuals classified as low risk at both tests, also used as the reference group (n = 27,466 [67.4%]), there were 281 severe liver disease events (1.0% of exposed individuals in that group, corresponding to 48.4% of all events). Compared with this group, an increased risk of severe liver disease was found for all other categories, except for those initially classified as intermediate risk who transitioned to low risk. In that group (n = 2,661 [6.5%], 1.1% experienced an event) the risk was comparable with the reference group (adjusted HR [aHR] 0.97; 95% CI 0.66–1.43). The highest risk was found in those classified as high risk at both time points (n = 250 [0.6%]; 13.2% experienced an event; aHR 17.04; 95% CI 11.67–24.88).

A 1-unit increase in the FIB-4 between the 2 tests was also associated with an elevated risk of severe liver disease (aHR 1.81; 95% CI 1.67–1.96). A restricted cubic spline model of the risk of severe liver disease associated with an increase in the FIB-4, modelled as a continuous predictor, is depicted in Fig. 2. Using a Kaplan-Meier analysis, the risk of severe liver disease stratified by the 9 subgroups is presented in Fig. 3, with median time to event also presented in Table 2.

General test characteristics (sensitivity, specificity, NPV, PPV and general test accuracy) for the pre-specified transitioning groups are listed in Table 3. For those in the high-risk group at the second test, the sensitivity for predicting future severe liver disease was 0.21, specificity 0.97, NPV 0.99 and PPV 0.09, yielding a general test accuracy of 0.96. For those at high risk at both tests, sensitivity was 0.10, specificity 0.99, NPV 0.99 and PPV 0.13, resulting in a general test accuracy of 0.98.

Sensitivity analysis

Using a second test at any time during the 12-year baseline follow-up period produced similar results as the main analysis. For instance, the risk of a 1-unit change in the FIB-4 between the 2 tests was 1.82 in the sensitivity analysis *vs.* 1.81 in the main analysis. Detailed data are given in Table S3.

Discussion

In this study, conducted in a general population setting, we found that repeating the FIB-4 within a 5-year period can, in comparison with a single measurement, help to identify individuals who are at higher risk of developing severe liver disease, a clinically relevant endpoint. An increase in the FIB-4 over time was associated with higher risk while a decrease in the FIB-4 was associated with reduced risk. However, even if there was a

Table 2. Transitioning between risk groups based on FIB-4 and a numeric change in FIB-4 measured up until 5 years after the first test (as a continuous parameter) and incident severe liver disease after the second test.

1st test	2nd test	N (%)	Median FIB-4		Events, total	% of all events	% events in group	Incidence per 1,000 pyr	Median time to event (years, IQR)	HR (95% CI)		
			1st test	2nd test						HR[1]	HR[2]	HR[3]
Low risk	Low risk	27,466 (67.4%)	0.76	0.80	281	48.36	1.02	0.63 (0.56–0.71)	16.8 (15.1–19.7)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Intermediate risk	4,100 (10.1%)	1.07	1.50	81	13.94	1.98	1.41 (1.14–1.76)	15.6 (9.5–19.1)	1.63 (1.26–2.11)	1.61 (1.25–2.09)	1.63 (1.26–2.11)
	High risk	114 (0.3%)	1.04	3.10	7	1.20	6.14	6.60 (3.15–13.9)	7.9 (1.5–16.0)	8.22 (3.87–17.43)	7.91 (3.72–16.81)	7.99 (3.76–16.97)
Intermediate risk	Low risk	2,661 (6.5%)	1.49	1.09	30	5.16	1.13	0.83 (0.58–1.19)	15.5 (9.0–18.1)	0.98 (0.67–1.44)	0.98 (0.67–1.43)	0.97 (0.66–1.43)
	Intermediate risk	5,332 (13.1%)	1.63	1.71	101	17.38	1.89	1.53 (1.26–1.86)	13.7 (7.4–17.0)	1.63 (1.27–2.09)	1.60 (1.25–2.06)	1.60 (1.24–2.05)
	High risk	451 (1.1%)	1.93	3.03	35	6.02	7.76	8.41 (6.04–11.7)	8.9 (3.5–14.6)	8.79 (6.07–12.72)	8.57 (5.91–12.41)	8.64 (5.96–12.52)
High risk	Low risk	83 (0.2%)	3.35	0.95	3	0.52	3.61	2.87 (0.93–8.91)	14.8 (8.8–17.6)	4.00 (1.28–12.47)	3.95 (1.27–12.34)	3.88 (1.24–12.13)
	Intermediate risk	272 (0.7%)	3.00	2.00	10	1.72	3.68	3.66 (1.97–6.81)	10.1 (4.4–15.2)	3.93 (2.07–7.45)	3.84 (2.03–7.29)	3.80 (2.00–7.20)
	High risk	250 (0.6%)	3.41	3.52	33	5.68	13.20	16.47 (11.7–23.2)	7.3 (2.7–12.5)	17.81 (12.22–25.95)	17.34 (11.88–25.30)	17.04 (11.67–24.88)
Change in FIB-4 between tests		40,729 (100%)	0.91	0.96	581	100.00	1.43	0.95 (0.87–1.03)	–	1.82 (1.68–1.96)	1.81 (1.67–1.96)	1.81 (1.67–1.96)

All models used attained age as the timescale: model 1 adjusted for age, model 2 additionally adjusted for sex and socioeconomic status and model 3 additionally adjusted for the time between tests. FIB-4, fibrosis-4 index; HR, hazard ratio; pyr, person-years.

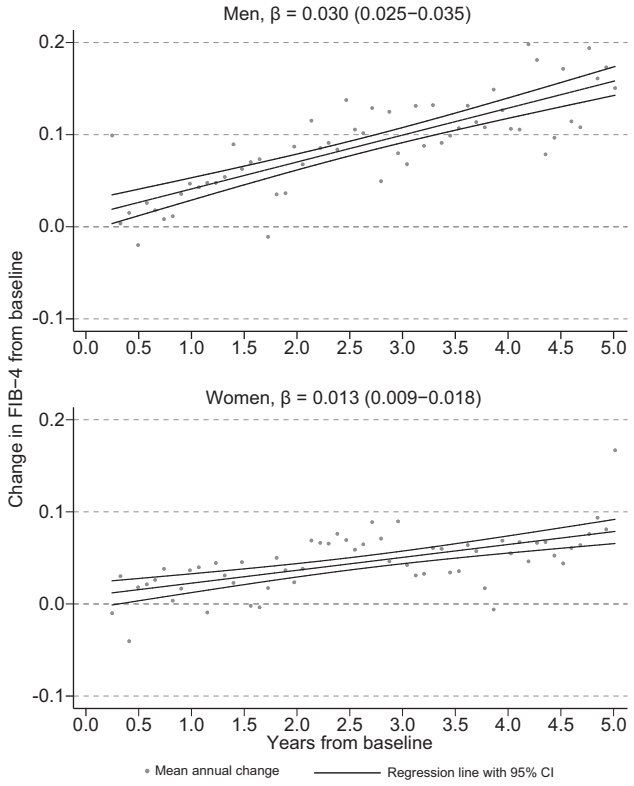


Fig. 1. Mean changes in the FIB-4 with 95% CIs during the 5-year study period in the full cohort stratified by sex using least-squares regression. FIB-4, fibrosis-4 index.

clear association between higher risk based on FIB-4 and the 581 severe liver disease events, 281 (48.4%) of these events were observed in individuals classified as low risk at both tests. This finding is an improvement on a single test, where 74.6% of those that eventually developed severe liver disease were in the low-risk group,[8] but is also a clear indication of the need for improved non-invasive scores for the general population.

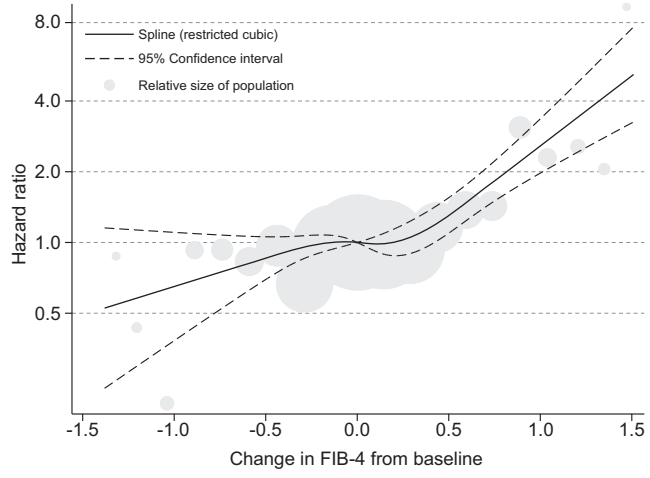
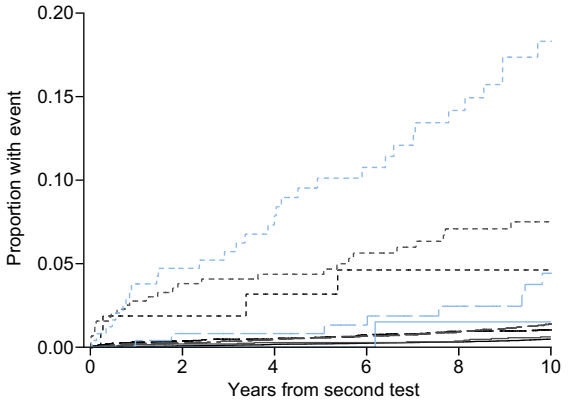


Fig. 2. Restricted cubic spline reflecting the risk of severe liver disease and change in the FIB-4 between 2 time points. FIB-4, fibrosis-4 index.



N° at risk						
Group = 11	27,466	26,588	25,870	25,158	24,359	23,575
Group = 12	4,100	3,841	3,661	3,437	3,239	3,010
Group = 13	114	81	72	64	57	55
Group = 21	2,661	2,485	2,355	2,215	2,064	1,914
Group = 22	5,332	4,946	4,599	4,269	3,882	3,466
Group = 23	451	359	326	292	242	201
Group = 31	83	70	68	66	64	59
Group = 32	272	233	213	182	163	138
Group = 33	250	198	168	138	115	81

Fig. 3. Kaplan-Meier curve of the risk of severe liver disease stratified on the 9 subgroups from the time of the second test during the first 10 years of follow-up. Clarification: Group 1 signifies low risk, group 2 intermediate risk and group 3 high risk, with the first figure being the risk group at the first testing occasion and the second figure being the risk group at the time of the second test. *E.g.* group 11 denotes people defined as low risk at both testing occasions.

About one-third of the population was classified as intermediate or high risk at 1 of the 2 tests, but only 1.43% developed severe liver disease in up to 27 years of follow-up. This finding suggests that if used as a general population screening tool and requiring everyone with an intermediate or high test to undergo additional testing such as transient elastography,[17] a large proportion of the tested individuals would have been referred because of false-positive findings, potentially straining healthcare systems and causing undue exposure to physical and psychological stress for many healthy individuals.

The absolute risk of incident severe liver disease was low (below 2%) in individuals that were classified as low or intermediate risk at any of the tests; in contrast, the absolute risk was considerably higher (from 6–13%) in those defined as high risk at either of the 2 tests. This observation suggests that those classified as high risk should be referred for additional evaluation to verify the 'high risk' classification.

There was no clinically significant increase in prediction when comparing individuals at high risk on 1 test to those at high risk on both tests. While a strategy to test individuals at high risk on both tests would lead to an improved specificity and a lower number of false positives, this was not a major problem and would likely be counteracted by capturing a lower number of individuals that would develop severe liver disease, *i.e.* producing more false-negative tests. These data support the strategy that individuals at high risk should undergo additional diagnostics (*e.g.*, elastography) directly and that a 'wait-and-see' strategy is not advisable.

Table 3. Test characteristics of individuals defined as intermediate or high risk on the second (final) measurement and those defined as high risk on both tests.

Risk group	N exposed	n with outcome	n without outcome	
Low	27,466	281	27,185	NPV = 99.0
Intermediate at *second* test	9,704	192	9,512	PPV = 2.0
		Sensitivity = 40.6	Specificity = 74.1	Accuracy = 73.7
Low	27,466	281	27,185	NPV = 99.0
High at *second* test	815	75	740	PPV = 9.2
		Sensitivity = 21.1	Specificity = 97.4	Accuracy = 96.4
Low	27,466	281	27,185	NPV = 99.0
High at *both* tests	250	33	217	PPV = 13.2
		Sensitivity = 10.5	Specificity = 99.2	Accuracy = 98.2

Each group was compared with individuals defined as low risk on both tests based on transitioning between risk groups between tests. Low: individuals defined as low risk at both tests. Intermediate: individuals defined as intermediate at the second test. High at last test: individuals defined as high at the second test. High at both tests: individuals defined as high at both tests. NPV, negative predictive value. PPV, positive predictive value.

The change across risk groups with time was considerable but transitioning from a low- to high-risk classification was rare within a 5-year period (only 0.4%) and still uncommon when transitioning from an intermediate to high risk (5.3%). However, we cannot exclude the possibility that the improvement in FIB-4 was largely due to a falsely high score at the first test and subsequently a result of regression towards the mean. Indeed, individuals at high risk on the first test had the highest probability of a change in score.

We present data from a large population-based cohort study on the natural history of the development of FIB-4 over time, with a mean of 0.020 units per year but markedly affected by age and sex. These findings could be an important reference point for future studies.

These results can be compared with some previous studies. For instance, Vergniol *et al.* showed that delta values of FIB-4 predicted mortality significantly better than just a baseline value in patients with hepatitis C.[18] Improvement in FIB-4 has been found to associate with improved fibrosis using gold standard liver biopsy in a clinical trial of patients with non-alcoholic steatohepatitis,[19] and worsening of FIB-4 has been associated with histological progression of fibrosis in a landmark dual-biopsy study, with a median of 6.6 years between biopsies.[20] A 2018 American Diabetes Association meeting abstract reported that in a large T2DM population about 0.7% progressed from low to high risk after approximately 4 years, which can be compared with 0.4% in our study. However, the main results of that study are yet to be published.[21] That finding gives some indication that, compared with the general population, the rate of fibrosis progression is faster in patients with diabetes, which is an important risk factor for incident severe liver disease.[22]

The data in the present study are derived from a large population-based cohort and thus generalizability to western countries (such as Sweden) should be high. All laboratory tests were performed using the same methods over time and with a low coefficient of variation (good precision), yielding well-defined and comparably high-quality exposure data with a low misclassification of exposure. The high-quality Swedish national registers allowed us to identify outcomes with little loss to follow-up. We selected 'hard' outcomes (*i.e.* outcomes that are important to patients and that can be objectively and independently measured) and unlikely to be misclassified. Any misclassification of events is unlikely to be associated with the exposure (FIB-4) and thus non-differential and should not bias the main findings of this study.

Some limitations should be mentioned. First, we do not know the reason for the inclusion of transaminases or platelets at either of the 2 testing occasions. Nonetheless, a large part of the cohort was sampled as part of routine health care in occupational care and not due to symptomatic disease. In addition, we excluded those with known (diagnosed) liver disease before the first baseline examination or with secondary tests only within a 3-month period after the first test to reduce the risk of selecting people with baseline significant liver disease that led to the second test, or people with falsely high laboratory tests. Also, the general risk for severe liver disease was not significantly higher than in those with only a single measurement of FIB-4 which suggests a low risk of selection bias. Second, we cannot be sure that all events are due to NAFLD, although we did censor anyone with a specific liver disease other than NAFLD or with coding for alcohol-related cirrhosis or alcohol use disorders at baseline or follow-up. Still, we did not have access to data on alcohol consumption. There may be undiagnosed or wrongly coded cases with cirrhosis or decompensated cirrhosis (*e.g.*, bleeding varices coded as a peptic ulcer), which would drive our estimates towards the null and the risk of severe liver disease might be higher. Moreover, the selected 'hard' outcomes are likely to lead to contact with specialized care, which would explain why the ascertained cases should have a low likelihood of misclassification. Finally, the cohort was sampled approximately 30 years from today. Such a cohort should have a lower prevalence of obesity and likely a lower prevalence of NAFLD than an equivalent present-day cohort.

Based on these data, it seems likely that, in the general population, adding a second measurement of FIB-4 can enhance the identification of individuals at risk of severe liver disease later in life. The absolute risk of severe liver disease in individuals classified as low or intermediate risk at both tests, however, was below 2% within 27 years of follow-up. And we previously showed that the risk of severe liver disease within 5 years is very low in those at low (0.18%) or intermediate risk (0.38%) per the FIB-4.[8] Therefore, our data support the contention that individuals defined as intermediate risk could be considered for repeated testing and lifestyle modification (*e.g.*, weight loss, physical activity), with repeated testing within 5 years. In contrast, those defined as high risk should undergo additional diagnostic testing (*e.g.*, elastography) directly, without repeated testing of FIB-4.[23] Future research is needed to evaluate the significance of a change in FIB-4 (or other scores) in other populations, in particular, those at a higher risk of liver disease. When used in the general population, a definition of new cut-off levels for FIB-4 could be considered. Even more attractive would be the construction of new scores designed for use in the general population. Such scores should ideally be inexpensive and convenient and based on readily available data to allow for use in primary care.

A second measurement of FIB-4 within 5 years of the first was found to improve the identification of individuals at risk of future severe liver disease in this population-based study. However, there were considerable changes in the risk classification over time, with one-third of the population being defined as at intermediate or high risk of having advanced fibrosis on at least 1 of the 2 tests. In particular, for those in the intermediate risk group, the absolute risk of severe liver disease was low and although repeated testing improves identification of at-risk individuals, this may lead to an increase in false positives. New and improved non-invasive scores are needed for population-level screening.

Abbreviations

aHR, adjusted HR; ALT, alanine aminotransferase; AMORIS, Apolipoprotein-related Mortality Risk; AST, aspartate aminotransferase; FIB-4, fibrosis-4 index; GGT, gamma-glutamyltransferase; HR, hazard ratio; NAFLD, non-alcoholic fatty liver disease; NPV, negative predictive value; PPV, positive predictive value; T2DM, type 2 diabetes mellitus.

Financial support

This study was supported by an independent grant from AstraZeneca to HH's institution. HH was further supported by grants from the Stockholm County Council (clinical postdoctoral appointment). MT, GW and NH were supported by the Jungner Foundation of Laboratory Medicine.

Conflict of interest

This study was supported by a research grant from Astra Zeneca to HH's institution. The funder had no role in the design and conduct of the study, nor in obtaining or analyzing the data, nor any significant contribution to the analysis and interpretation of the results, nor drafting of the manuscript. HH has served as a consultant for Novo Nordisk, Gilead, IQVIA and Intercept Pharmaceuticals. HH's institution has received research grants from Gilead Sciences Inc and Intercept Pharmaceuticals. HH has served as an advisory board member at Bristol Myers-Squibb and Gilead. None of these has relevance for the current study.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Study conception and design: HH, AA, MT, GW, NH. Acquisition of data: GW, NH. Statistical analysis: MT. Analysis and interpretation of data: All. Drafting of manuscript: HH. Critical revision: All. Guarantors of the article: HH, NH. All authors approved the final version of the article, including the authorship list.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2020.06.007.

References

Author names in bold designate shared co-first authorship

[1] **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.

[2] **Hagstrom H**, **Nasr P**, Ekstedt M, Hammar U, Stal P, Hultcrantz R, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265–1273.

[3] 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.e10.

[4] 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. Clin Gastroenterol Hepatol 2015;13:643–654.e9. quiz e639-640.

[5] Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317–1325.

[6] Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007;45:846–854.

[7] Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38:518–526.

[8] Hagstrom H, Talback M, Andreasson A, Walldius G, Hammar N. Ability of noninvasive scoring systems to identify individuals in the population at risk for severe liver disease. Gastroenterology 2020;158:200–214.

[9] Walldius G, Malmstrom H, Jungner I, de Faire U, Lambe M, Van Hemelrijck M, et al. Cohort profile: the AMORIS cohort. Int J Epidemiol 2017;46:1103. 1103i.

[10] Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: clinical prediction rules and blood-based biomarkers. J Hepatol 2018;68:305–315.

[11] McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol 2017;112:740–751.

[12] Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol 2009;24:659–667.

[13] Ludvigsson JF, Svedberg P, Olen O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. Eur J Epidemiol 2019;34:423–437.

[14] 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.

[15] Brooke HL, Talback M, Hornblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish cause of death register. Eur J Epidemiol 2017;32:765–773.

[16] Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta Oncol 2009;48:27–33.

[17] Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705–1713.

[18] Vergniol J, Boursier J, Coutzac C, Bertrais S, Foucher J, Angel C, et al. Evolution of noninvasive tests of liver fibrosis is associated with prognosis in patients with chronic hepatitis C. Hepatology 2014;60:65–76.

[19] Chalasani N, Abdelmalek MF, Loomba R, Kowdley KV, McCullough AJ, Dasarathy S, et al. Relationship between three commonly used non-invasive fibrosis biomarkers and improvement in fibrosis stage in patients with non-alcoholic steatohepatitis. Liver Int 2019;39:924–932.

[20] 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. J Hepatol 2015;62:1148–1155.

[21] Filozof CJS, Goldstein BJ. Liver Fibrosis as Assessed by the FIB-4 Index in Patients with Type 2 Diabetes (T2DM). Chicago: ADA; 2018.

[22] Bjorkstrom K, Franzen S, Eliasson B, Miftaraj M, Gudbjornsdottir S, Trolle-Lagerros Y, et al. Risk factors for severe liver disease in patients with type 2 diabetes. Clin Gastroenterol Hepatol 2019;17:2769–2775.e4.

[23] Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1717–1730.