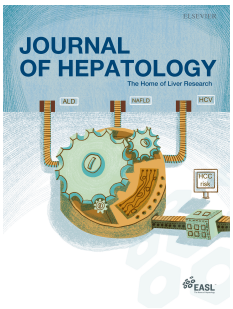


Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease

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

PII: S0168-8278(20)30378-0

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

Reference: JHEPAT 7802

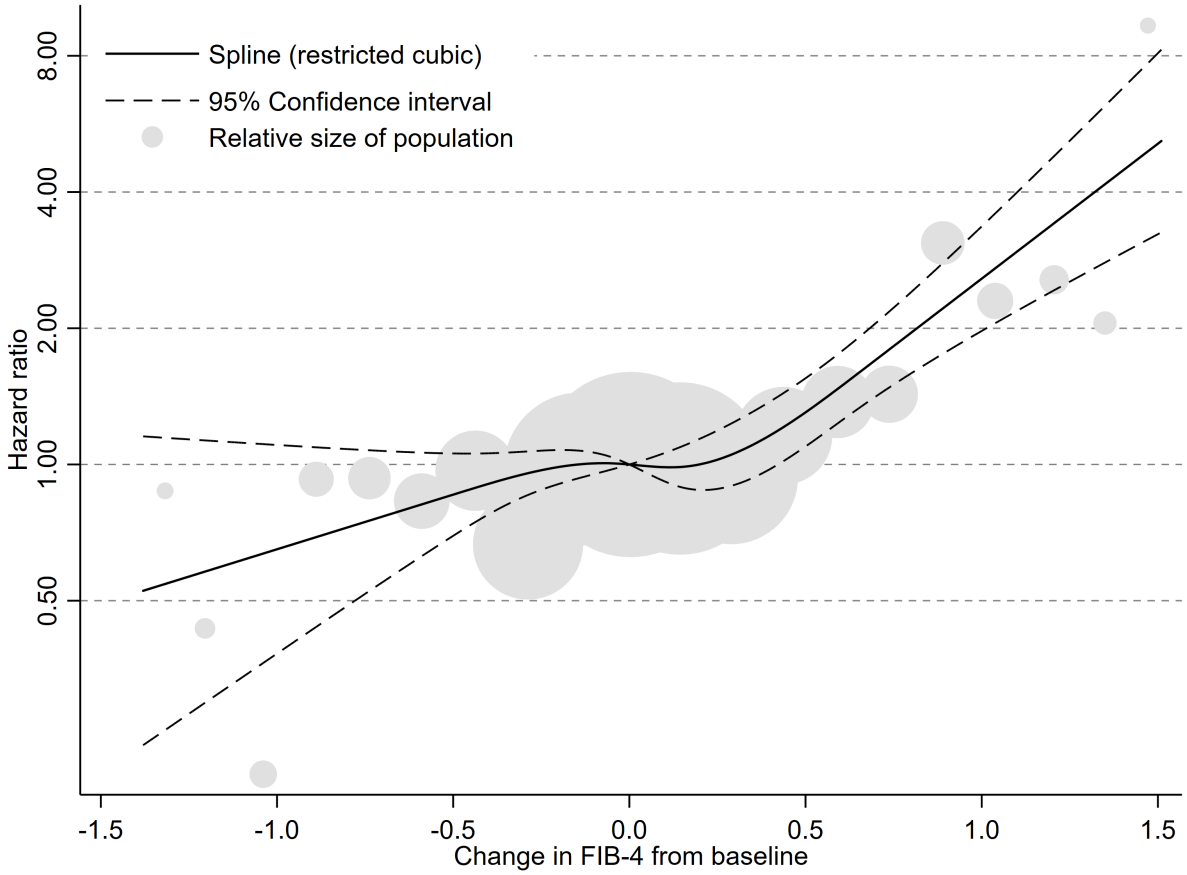
To appear in: *Journal of Hepatology*

Received Date: 17 March 2020

Revised Date: 4 May 2020

Accepted Date: 3 June 2020

Please cite this article as: Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N, Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease, *Journal of Hepatology* (2020), doi: https://doi.org/10.1016/j.jhep.2020.06.007.


Spline (restricted cubic)
95% Confidence interval
Relative size of population
Hazard ratio
0.50
1.00
2.00
4.00
8.00
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Change in FIB-4 from baseline

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

Short title: Repeated measurements of FIB-4 and severe liver disease

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

Correspondence and reprint requests:

Associate professor Hannes Hagström, C1:77, Division of Hepatology, Karolinska University Hospital Huddinge, 141 86 Stockholm, Sweden

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

E-mail: hannes.hagstrom@ki.se

Keywords: NAFLD; cirrhosis; epidemiology; prognosis; FIB-4

Abbreviations: ALT, alanine aminotransferase. AMORIS, Apolipoprotein-related Mortality Risk. AST, aspartate aminotransferase. BMI, body mass index. CI, confidence interval. FIB-4, Fibrosis-4 index. g-GT, gamma-glutamyltransferase. HR, hazard ratio. ICD, international classification of disease. IQR, interquartile range. NAFLD, nonalcoholic fatty liver disease. NPV, negative predictive value. PPV, positive predictive value. T2DM, type 2 diabetes mellitus.

Total word count: 4258. **Abstract: 257**. **Tables: 3**. **Figures: 3**. **Character count for title: 170**

Disclosure / Conflict of interest declaration: 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.

Grant Support (Funding): This study was supported by an independent grant from Astra Zeneca 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.

Author's 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.

Writing Assistance: None.

Abstract

Background & Aims: It is unclear as to whether the identification of individuals at risk of cirrhosis using noninvasive tests can be improved by repeated measurements.

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

Results: Of 126,942 persons 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 events of severe liver disease were documented (0.95/1,000 person-years). An increase of one unit in FIB-4 was associated with an elevated risk of severe liver disease (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 persons 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 in the general population at an increased risk of severe liver disease. 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. Here, we found that changes in this score over time is 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 outcomes [1-3]. Thus, defining the presence or absence of advanced fibrosis is key in making a prognosis in persons with known or suspected chronic liver disease. Persons without advanced fibrosis have a low risk of progression to cirrhosis within a 10-15-year time frame [2, 4]. Conversely, persons 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 five non-invasive scores to predict incident severe liver disease in a general population setting was modest [8].

It is not well described whether repeated measures of the available noninvasive 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.

Here, we tested the general hypothesis that repeated measurements of the commonly used FIB-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

1) investigate the association of changes in FIB-4 measured at two time points with incident severe liver disease in the general population and 2) 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 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 with information to calculate FIB-4 at two time points were included in the study. 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 persons below 35 and above 79 years.

We also excluded persons 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 persons with a history of severe liver disease (see definition below and in the Supplementary Appendix) or any diagnosis of drug or alcohol abuse at or before

baseline. Finally, we excluded persons with secondary tests only within 3 months after the first test. This exclusion was done to reduce the risk of selecting persons with baseline significant liver disease that led to the second test or persons with falsely high lab tests. Persons 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 Supplementary Table 1a-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 Appendix. The FIB-4 was calculated as:

- (Age x aspartate aminotransferase, AST) / (Platelets x √(alanine aminotransferase, ALT) [5].

We categorized persons into low-, intermediate- and high-risk groups for advanced fibrosis based on the following suggested cut-offs: <1.30 (low risk), 1.3-2.67 (intermediate risk and >2.67 (high risk). However, we did not change the lower cut-off for persons ≥65 years of age to 2.0, as has been suggested [11]. This approach was introduced to reduce false-positive findings in persons ≥65 years, but how this should be applied using repeated measurements has not been evaluated and is not entirely straightforward. For instance, a person at the age of 64 years at a first test with a score

of 1.9 (intermediate risk) would be re-categorized as low risk when he or she reached 65 years, provided that AST, 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 previously showed 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 body mass index (BMI), presence of type 2 diabetes mellitus (T2DM) and other covariates [9] in persons for which 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

from a baseline testing of >126 mg/dl (fasting) or >200 mg/dl (non-fasting) 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 persons 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 linkage to nationwide Swedish registers using the personal identification number was conducted. A description of the registers used for outcome ascertainment is available in the Supplementary Appendix. 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 Supplementary Table 1a.

Analyses

First, we investigated transitions from one risk group to another from the first to the second test. In the proportional hazards regression analyses persons classified as low risk at both tests were used as the reference group. We also analyzed the hazard ratio (HR) associated with a one-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 persons classified as low risk at both tests as the comparator group; a second group was established from persons classified as intermediate in the second test; a third group was constructed from persons classified as high in the second test; and a fourth group was created from persons classified as high at both tests. These analyses excluded persons that transitioned from the high- or intermediate-risk groups to the low-risk group. We also compared key characteristics of the persons 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 interquartile ranges (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% confidence intervals (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 three 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 two 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.

Results

There were 126,942 individuals in which 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 persons with a high probability of a falsely high FIB-4 at the first testing time, we also excluded 2,862 individuals who had a second test done 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 persons with a second test at any time during the 12-year baseline study period were included in a sensitivity analysis.

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 ascertained 11,929 (29.29%) deaths and 581 events of severe liver disease (1.43%) during the follow-up. In all, 1,212 persons (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 persons 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 Supplementary Table 3. Differences in key parameters between the persons included in this study compared to persons that only had a single

testing occasion (n=79,705) are presented in the Supplementary Table 4. In brief, the included persons 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 – 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). 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 persons ≥65 years in both men (mean annual change 0.032 vs. 0.029) and women (0.018 vs. 0.011) (Supplementary Figure 1). Of the 40,729 included persons, 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 persons that were stable or changed risk groups based on the FIB-4, total events of severe liver disease, incidence rates and corresponding HRs are presented in Table 2. About 25% of all persons changed the risk group from the first to the second test. Transitioning was less common in persons 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 persons classified as low risk at both tests, also used as the reference group (n=27,466 [67.4%]), there were 281 events of severe liver disease (1.0% of exposed persons 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 persons 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 persons 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 one-unit increase in the FIB-4 between the two 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 Figure 2. Using a Kaplan-Meier analysis, the risk of severe liver disease stratified on the nine subgroups is presented in Figure 3, with median time to event presented also 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 persons 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 persons 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 one-unit change in the FIB-4 between the two tests was 1.82 in the sensitivity analysis vs. 1.81 in the main analysis. Detailed data are given in Supplementary Table 3.

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 persons that 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 were a clear association between higher risk based on FIB-4 from the ascertained 581 events of severe liver disease, 281 (48.4%) of these events were found in persons classified as low risk at both tests. This finding, however, is better compared with only using a single test, where 74.6% of persons that eventually developed severe liver disease were found in the low-risk group [8], but also a clear indication of the need for improved noninvasive scores of liver disease risk and progression in the general population.

About one third of the population was classified as intermediate or high risk at one of the two 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 all persons with an intermediate or high test to undergo additional testing such as transient elastography [17], a large proportion of the tested persons would have been referred because of false-positive findings, potentially straining healthcare systems and undoing exposure of physical and psychological stress for many healthy individuals.

The absolute risk of incident severe liver disease was low (below 2%) in persons 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 persons defined as high risk at any of

the two tests. This observation suggests that persons classified as high risk should be referred to additional evaluation to verify the 'high risk' classification.

There was no clinically significant increase in prediction when comparing persons at high risk on one test occasion compared with persons at high risk on both tests. While a strategy to test persons 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 likely counteracted by capturing a lower number of persons that developed severe liver disease, i.e. producing more false-negative tests. These data support the strategy that persons at high risk should undergo additional diagnostics (e.g., elastography) directly and that a 'wait-and-see' strategy is not recommendable.

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 in 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, persons 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. The findings of this study can be an important reference point in identifying individuals in the general population at risk of severe liver disease in future studies.

Comparison with previous 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 in median 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 persons with diabetes, which is an important risk factor for incident severe liver disease [22].

Strengths and limitations

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 two 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 persons with baseline significant liver disease that led to the second test, or persons with falsely high lab 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 any person with a specific liver disease other than NAFLD or with coding for alcohol-related cirrhosis or alcohol use disorders at baseline or follow-up, which is why most events are likely due to NAFLD. 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 compared to today.

Implications

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 persons 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 persons defined as low (0.18%) or intermediate risk (0.38%) per the FIB-4 [8]. Therefore, our data support the contention that persons 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, persons 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.

Conclusions

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 27-year follow-up study of more than 40,000 persons. 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 one of the two 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 scores are needed if the use of noninvasive scores in the general population were to be considered for screening purposes.

Figure legends

Figure 1. Mean changes in the FIB-4 with 95% confidence intervals during the 5-year study period in the full cohort stratified by sex using least squares regression.

Figure 2. Restricted cubic spline reflecting the risk of severe liver disease and change in the FIB-4 between two time points.

Figure 3. Kaplan-Meier curve of the risk of severe liver disease stratified on the nine 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 persons defined as low risk at both testing occasions.

Tables

Table 1. Characteristics of the cohort with FIB-4 measured at two time points within 5 years at the time of the first and last available measurement. * Missing data in about 5% of the cohort. Abbreviations: ALT, alanine aminotransferase. AST, aspartate aminotransferase. FIB-4, fibrosis-4 index. Gamma-GT, gamma-glutamyltransferase. IQR, interquartile range.

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

Table 2. Associations of 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. 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.

Abbreviations: CI, confidence interval. FIB-4, fibrosis-4 index. HR, hazard ratio. pyr, person-years.

			Median FIB-4							HR (95% CI)		
First test	**Second test**	**N (%)**	**First test**	**Second test**	**Events, total**	**% of all events**	**% events in group**	**Incidence per 1,000 pyr**	**Median time to event (years, IQR)**	**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**	**1.00**	**1.00**
	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)

Table 3. Test characteristics of persons defined as at intermediate and high risk at the second (final) measurement and persons defined as high risk at ***both*** tests. Each group was compared with persons defined at low risk at both tests based on transitioning between risk groups between tests. Low: Persons defined as low risk at both tests. Intermediate: Persons defined as intermediate at the second test. High at last test: persons defined as high at the second test. High at both tests: persons defined as high at both tests. Abbreviations: NPV, negative predictive value. PPV, positive predictive value.

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

Risk group	N exposed	N with outcome	N without outcome	
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

Risk group	N exposed	N with outcome	N without outcome	
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

References

[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. Journal of hepatology 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.e310.

[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. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2015;13:643-654.e641-649; 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. Journal of hepatology 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. The American journal of gastroenterology 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. European journal of epidemiology 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. European journal of epidemiology 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. European journal of epidemiology 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 oncologica (Stockholm, Sweden) 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 in Medicine and Biology 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 international : official journal of the International Association for the Study of the Liver 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. Journal of hepatology 2015;62:1148-1155.
[21] Filozof C JS, Goldstein BJ. Liver Fibrosis as Assessed by the FIB-4 Index in Patients with Type 2 Diabetes (T2DM). ADA. Chicago; 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. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2019;17:2769-2775.e2764.
[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.

Taking repeated measurements of FIB-4 can improve identification of individuals at risk of severe liver disease: A population-based follow-up study of 40,729 individuals

Supplementary Appendix

Description of laboratory analyses conducted at baseline

Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) were determined with an enzymatic UV test and Gamma-glutamyltransferase (GGT) by an enzymatic colorimetric test using a Technicon DAX 96 Multichannel Analyzer with a total imprecision of <6.0% coefficient of variation (CV). Platelets were determined by a fully automated hematology analyzer using the Coulter principle with a total imprecision of 2.1-5.6% CV. Total cholesterol and triglycerides were measured by enzyme techniques. Glucose levels were analyzed with an enzyme colorimetric technique (glucose oxidase/peroxidase, GOD-PAP) using automated multichannel analyzers [AutoChemist-PRISMA® (New Clinicon, Stockholm, Sweden) and Technicon DAX® TM 96 (Technicon Instruments Corp., Tarrytown, NY, USA)]. Creatinine levels were analyzed with the non-kinetic alkaline picrate method (Jaffé) using an AutoChemist-PRISMA from 1985 through 1992 and a DAX 96 analyzer from 1993 through 1996. The CV was <3% for all laboratory tests.

Description of Swedish National Registers

The National Patient Register (NPR) contains data on all hospitalizations regionally since 1964 and nationally since 1987 and on outpatient visits in specialized care since

2001. The validity of diagnoses of relevance for this study obtained ranges from 85-95%, depending on diagnosis [1]. Primary care is not included in the NPR.

The Cause of Death Register contains data on the causes of death of all Swedish inhabitants, including whether the person died abroad. The responsible physician must report the underlying cause of death (e.g., hepatocellular carcinoma) and any disease that could have contributed to the death of the individual (e.g., liver cirrhosis) [2].

The Swedish Cancer Register contains data on verified solid and non-solid tumors since 1958, irrespective of the diagnostic modality. Reporting is mandatory by law for all confirmed (diagnosed) cases to this register. The completeness of the register is estimated to be about 96% [3].

1

Diagnosis	ICD-10 (1997-)	ICD-9 (1987-1996)	ICD-8 (1969-1986)	ICD-7 *
Severe liver disease				
Liver failure, acute or subacute	K72.0	570	570	
Ascites	R18.9	789.5	785.3	
Esophageal varices, bleeding	I85.0, I98.3	456.0, 456.20	456.0	
Esophageal varices, non-bleeding	I85.9, I98.2	456.1, 456.21	456.0	
Hepatorenal syndrome	K76.7	572.4		
Liver failure, chronic	K72.1	572.8	573	
Liver cirrhosis	K74.6	571.5	571.9	
Liver encephalopathy		572.2	573.02	
Liver failure not otherwise defined	K72.9			
Portal hypertension	K76.6	572.3	571.9	
Hepatocellular carcinoma	C22.0	155.0	155.01	155.0
Procedure codes				
Liver transplantation	JJC00, JJC10, JJC20, DJ005, DJ006, JJC30, JJC40	5200	5200	
Laparocentesis	TJA10	4041	4041	

2

3 **Supplementary Table 1a**. ICD codes used to define endpoints. * ICD-7 was only used in the Swedish Cancer Register.

4

Liver disease	ICD-10 (1997-)	ICD-9 (1987-1996)	ICD-8 (1969-1986)
AAT deficiency	E88.0A, E88.0B	277.6	
Alcohol-related liver disease	K70	571.0-3	571.00, 571.01
Autoimmune hepatitis	K75.4		
Budd-Chiari syndrome	I82.0, K76.5	453.0	
Hemochromatosis	E83.1	275.0	273.20
PBC	K74.3, K74.5	571.6	
PSC	(K50 or K51) + K83.0	(555 or 556) + 576.1	563 + 575.05
Wilson	E83.0B	275.1	273.30
Viral hepatitis	B15, B16, B17, B18, B19	070, 571.4	070, 999.20
Alcohol/drug use disorders			
Alcohol-related diagnoses	E24.4, F04.9, F10, G31.2, G62.1, G72.1, I42.6, K29.2, K85.2, K86.0, O35.4, X65, Y15, Y91	255, 294.0, 291, 303, 305.0, 357.5, 359.4, 425.5, 535.3, 577, 655.4	258, 291.1, 299,
Other drug use disorders	F11-F19	292, 305	

Supplementary Table 1b. ICD codes used to define liver diseases other than NAFLD and diagnoses associated with alcohol or drug use disorders. AAT, alpha-1-antitrypin. PBC, primary biliary cholangitis. PSC, primary sclerosing cholangitis.

Parameter	Full Cohort	Low-Low	Low-Intermediate	Low-High	Intermediate-Low	Intermediate-Intermediate	Intermediate-High	High-Low	High-Intermediate	High-High
Number of persons	40,729	27,466	4,100	114	2,661	5,332	451	83	272	250
Person-years at risk (years)	613,376	442,704	57,308	1,060	36,198	66,168	4,160	1,044	2,730	2,003
Person-years at risk, median (IQR)	16.21 (12.11-19.16)	16.81 (15.13-19.71)	15.63 (9.46-19.12)	7.94 (1.55-15.99)	15.51 (8.95-18.09)	13.69 (7.39-17.04)	8.92 (3.53-14.57)	14.84 (8.79-17.57)	10.14 (4.37-15.25)	7.30 (2.74-12.50)
Male, n (%)	16,792 (41.23)	10,734 (39.08)	1,917 (46.76)	77 (67.54)	1,123 (42.20)	2,386 (44.75)	237 (52.55)	41 (49.40)	139 (51.10)	138 (55.20)
Attained age at inclusion, median (IQR)	57.12 (47.98-67.89)	51.92 (45.31-60.57)	65.14 (58.14-72.42)	65.20 (53.69-73.67)	65.39 (56.37-72.91)	72.13 (65.22-76.91)	73.77 (67.60-78.41)	57.06 (48.74-69.25)	73.24 (65.78-78.07)	73.06 (65.88-78.24)
Attained age at exit, median (IQR)	72.94 (64.81-82.01)	68.78 (62.56-76.82)	80.17 (72.40-86.13)	73.94 (66.08-83.19)	79.17 (71.46-85.72)	84.07 (78.28-88.91)	82.72 (76.20-88.08)	70.58 (63.66-80.23)	82.69 (75.42-87.98)	80.68 (72.14-87.23)
Number of events, n (%)	581 (1.43)	281 (1.02)	81 (1.98)	7 (6.14)	30 (1.13)	101 (1.89)	35 (7.76)	3 (3.61)	10 (3.68)	33 (13.20)
Years between 1st and 2nd tests, median (IQR)	2.41 (1.23-3.89)	2.29 (1.17-3.77)	2.86 (1.55-4.18)	2.53 (1.46-4.20)	2.44 (1.24-3.93)	2.64 (1.35-4.03)	2.92 (1.61-4.02)	1.52 (0.78-3.43)	2.25 (1.18-3.76)	2.11 (0.98-3.46)
FIB-4 at baseline, median (IQR)	0.96 (0.70-1.32)	0.80 (0.62-0.99)	1.50 (1.38-1.70)	3.10 (2.84-3.75)	1.09 (0.93-1.19)	1.71 (1.49-2.00)	3.03 (2.82-3.51)	0.95 (0.76-1.10)	2.00 (1.62-2.30)	3.52 (3.02-4.72)
Change in FIB-4 from 1st test, median (IQR)	0.05 (-0.13-0.24)	0.03 (-0.10-0.17)	0.49 (0.32-0.72)	2.18 (1.77-3.16)	-0.45 (-0.70--0.29)	0.06 (-0.17-0.31)	1.20 (0.79-1.63)	-2.42 (-3.34--1.94)	-1.12 (-1.68--0.68)	0.09 (-0.46-0.87)
ALT (IU/L), median (IQR)	21.76 (15.88-31.17)	22.35 (16.47-31.76)	21.18 (15.29-31.76)	47.35 (17.06-122.35)	21.18 (15.88-28.82)	20.59 (15.29-28.23)	22.94 (14.12-42.35)	22.94 (17.06-33.53)	21.18 (14.70-30.88)	28.53 (17.65-55.29)
AST (IU/L), median (IQR)	20.00 (16.47-24.70)	18.82 (15.29-22.94)	24.12 (20.00-30.59)	64.41 (31.76-128.82)	18.82 (15.88-22.35)	22.94 (19.41-28.23)	32.94 (25.29-52.35)	19.41 (16.47-25.29)	24.12 (20.00-30.59)	37.64 (26.47-70.00)
Platelets (10^9), median (IQR)	251 (213-292)	266 (231-306)	220 (191-253)	179 (148-212)	251 (218-287)	209 (181-240)	167 (141-199)	261 (221-305)	198 (165-233)	142 (114-175)
gamma-GT (IU/L), median (IQR)	21.60 (15.00-35.99)	21.60 (14.40-34.79)	22.80 (15.60-40.79)	61.19 (26.39-185.96)	22.20 (15.60-37.19)	21.60 (15.00-37.19)	29.99 (16.20-91.18)	25.19 (17.40-40.79)	23.40 (16.80-53.39)	52.79 (21.60-127.77)
Total cholesterol (mg/dL)*, median (IQR)	228 (201-255)	224 (197-255)	232 (205-259)	224 (189-266)	232 (205-259)	232 (205-259)	224 (193-255)	237 (205-263)	224 (197-251)	216 (189-247)
Triglycerides (mg/dL)*, median (IQR)	106.20 (70.80-159.30)	106.20 (70.80-159.30)	106.20 (79.65-159.30)	123.90 (79.65-185.85)	115.05 (79.65-168.15)	106.20 (70.80-150.45)	97.35 (70.80-150.45)	123.90 (79.65-185.85)	115.05 (79.65-168.15)	106.20 (79.65-168.15)
Glucose (mg/dL)*, median (IQR)	90.09 (82.88-99.10)	90.09 (82.88-99.10)	90.09 (82.88-100.90)	91.89 (82.88-100.90)	91.89 (84.68-100.90)	90.09 (82.88-100.90)	91.89 (82.88-106.31)	93.69 (82.88-106.31)	95.50 (84.68-107.21)	95.50 (86.49-113.51)
Blue-collar worker*, n (%)	21,265 (54.31)	14,145 (53.26)	2,114 (53.74)	58 (55.77)	1,400 (55.29)	2,994 (59.31)	236 (56.32)	41 (53.25)	139 (55.60)	138 (58.72)

1

Supplementary Table 2. Characteristics of the cohort with FIB-4 measured at two time points within 5 years at the time of the first measurement and stratified by risk groups at the first and second tests. For example, 'Low-Low' means the subgroup of the cohort was defined as low risk at both tests. *Missing data in about 5% of the cohort. Abbreviations: ALT, alanine aminotransferase. AST, aspartate aminotransferase. FIB-4, fibrosis-4 index. Gamma-GT, gamma-glutamyltransferase. IM, intermediate. IQR, interquartile range.

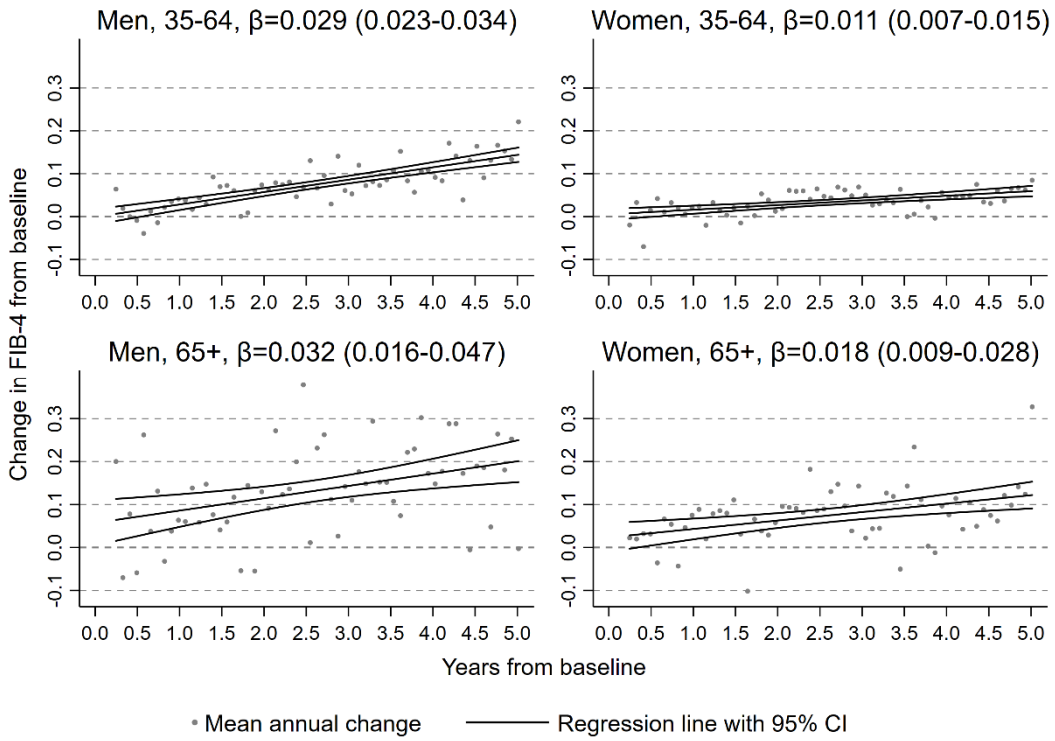
Supplementary Table 3. Last measurement of FIB-4 within 12 years from baseline. Associations of transitioning between risk groups based on FIB-4 and numeric change in FIB-4, measured as a continuous parameter, and incident severe liver disease after the second test. All models used attained age as the timescale: Model 1 adjusted for age, model 2 additionally adjusted for sex and socioeconomic status in addition to age and model 3 additionally adjusted for the time between tests. Abbreviations: CI, confidence interval. FIB-4, fibrosis-4 index. HR, hazard ratio. pyr, person-years.

				Median FIB-4						HR (95% CI)		
First test	**Second test**	**N**	**%**	**First test**	**Second test**	**Events**	**%**	**Events/N**	**Incidence per 1,000 pyr**	**HR[1]**	**HR[2]**	**HR[3]**
Low risk	Low risk	29,798	67.15	0.75	0.80	298	47.23	1.00%	0.65 (0.58-0.73)	**1.00**	**1.00**	**1.00**
	Intermediate risk	4,845	10.92	1.06	1.51	103	16.32	2.13%	1.64 (1.36-2.00)	1.88 (1.48-2.37)	1.86 (1.47-2.35)	1.86 (1.47-2.36)
	High risk	175	0.39	1.02	3.24	9	1.43	5.14%	6.27 (3.26-12.05)	7.58 (3.90-14.76)	7.42 (3.81-14.46)	7.47 (3.83-14.56)
Intermediate risk	Low risk	2,685	6.05	1.49	1.08	30	4.75	1.12%	0.86 (0.60-1.23)	1.02 (0.70-1.49)	1.01 (0.69-1.48)	1.01 (0.69-1.48)
	Intermediate risk	5,682	12.80	1.62	1.72	100	15.85	1.76%	1.52 (1.25-1.85)	1.62 (1.26-2.08)	1.60 (1.25-2.05)	1.60 (1.24-2.05)
	High risk	566	1.28	1.89	3.04	44	6.97	7.77%	9.35 (6.96-12.57)	9.94 (7.12-13.89)	9.75 (6.97-13.64)	9.77 (6.98-13.67)
High risk	Low risk	79	0.18	3.35	0.93	4	0.63	5.06%	4.19 (1.57-11.15)	6.01 (2.24-16.11)	5.97 (2.23-16.03)	5.94 (2.21-15.95)
	Intermediate risk	280	0.63	3.03	1.98	8	1.27	2.86%	3.05 (1.52-6.09)	3.28 (1.61-6.66)	3.22 (1.58-6.54)	3.20 (1.57-6.52)
	High risk	265	0.60	3.39	3.59	35	5.55	13.21%	17.61 (12.6-24.5)	19.21 (13.32-27.69)	18.81 (13.02-27.16)	18.71 (12.94-27.05)
Change in FIB-4 between tests		44,375	100.00	0.90	0.97	631	100.00	1.42%	1.00 (0.92-1.08)	1.82 (1.70-1.95)	1.82 (1.70-1.95)	1.82 (1.70-1.95)

Supplementary Table 4. Differences in key parameters between the persons included in this study (n=40,729) compared to persons that only had a single testing occasion (n=79,705)

Parameter	Persons with repeated measurements of FIB-4 (n=40,729) (Mean)	Persons without repeated measurements of FIB-4 (n=79.705) (Mean)	Mean difference (95%CI)
Age (years)	55.0	52.4	2.6 (2.4 – 2.7)
Sex (male, %)	41.2	45.9	4.6 (4.0 – 5.2)
ALT (IU/L)	26	28	1.4 (0.96 – 1.83)
AST (IU/L)	23	22	0.22 (-0.004 – 0.44
Platelets ($x10^9$)	268	261	6.7 (5.9 – 7.5)
FIB-4 score (continuous)	1.03	0.97	0.063 (0.057 – 0.070)
FIB-4 category (%)			
Low	77.8	82.1	4.3 (3.8 – 4.8)
Intermediate	20.7	16.6	4.1 (3.7 – 4.6)
High	1.5	1.3	0.2 (0.04 – 0.31)
Persons with outcome during follow-up (%)	1.42	1.33	0.09 (-0.04 – 0.24)
Time to event, years	17.6	16.4	1.2 (1.1 – 1.2)

Supplementary Figure 1. Mean changes in FIB-4 with 95%CIs for 5 years in men and women stratified by age using a least-squares regression.

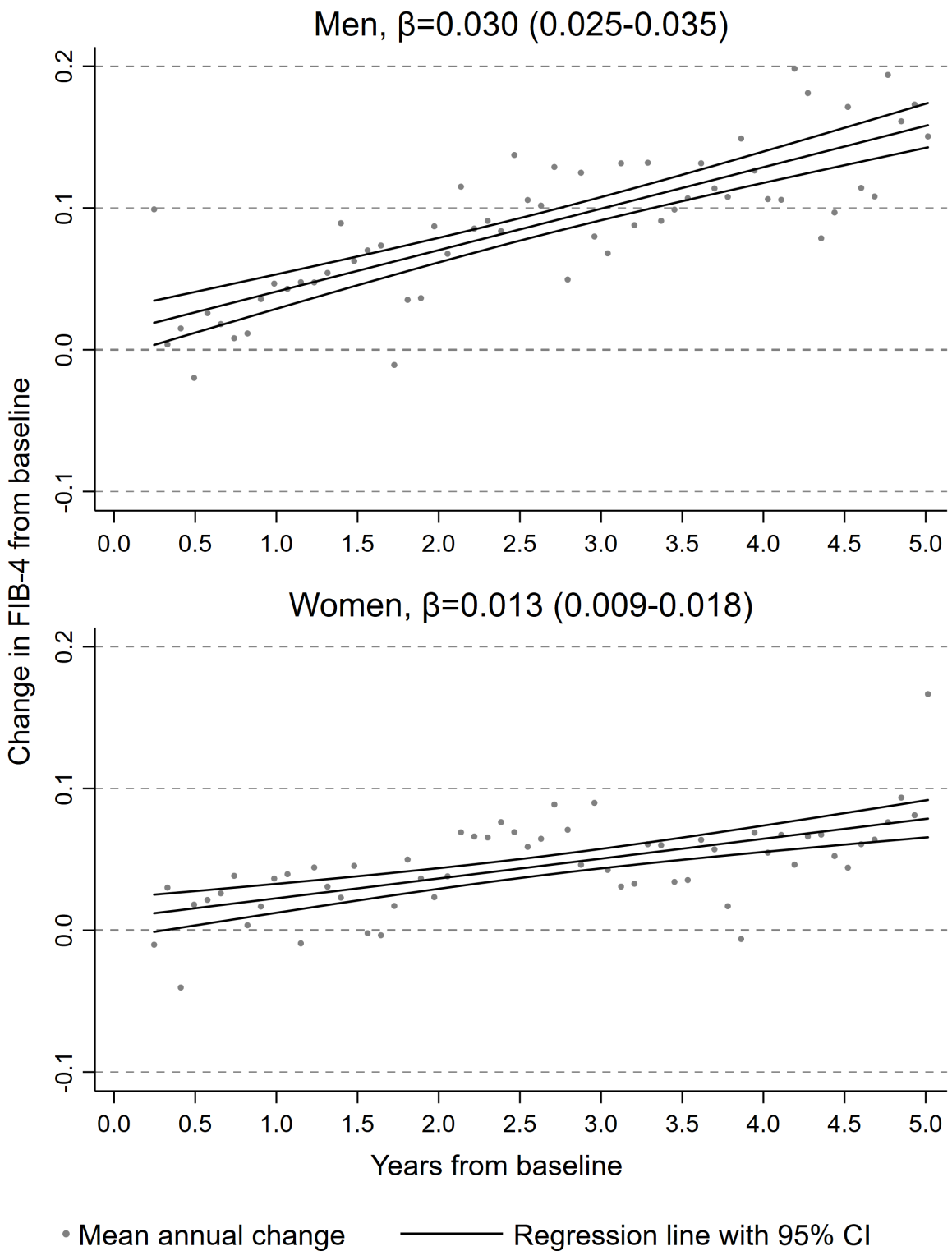


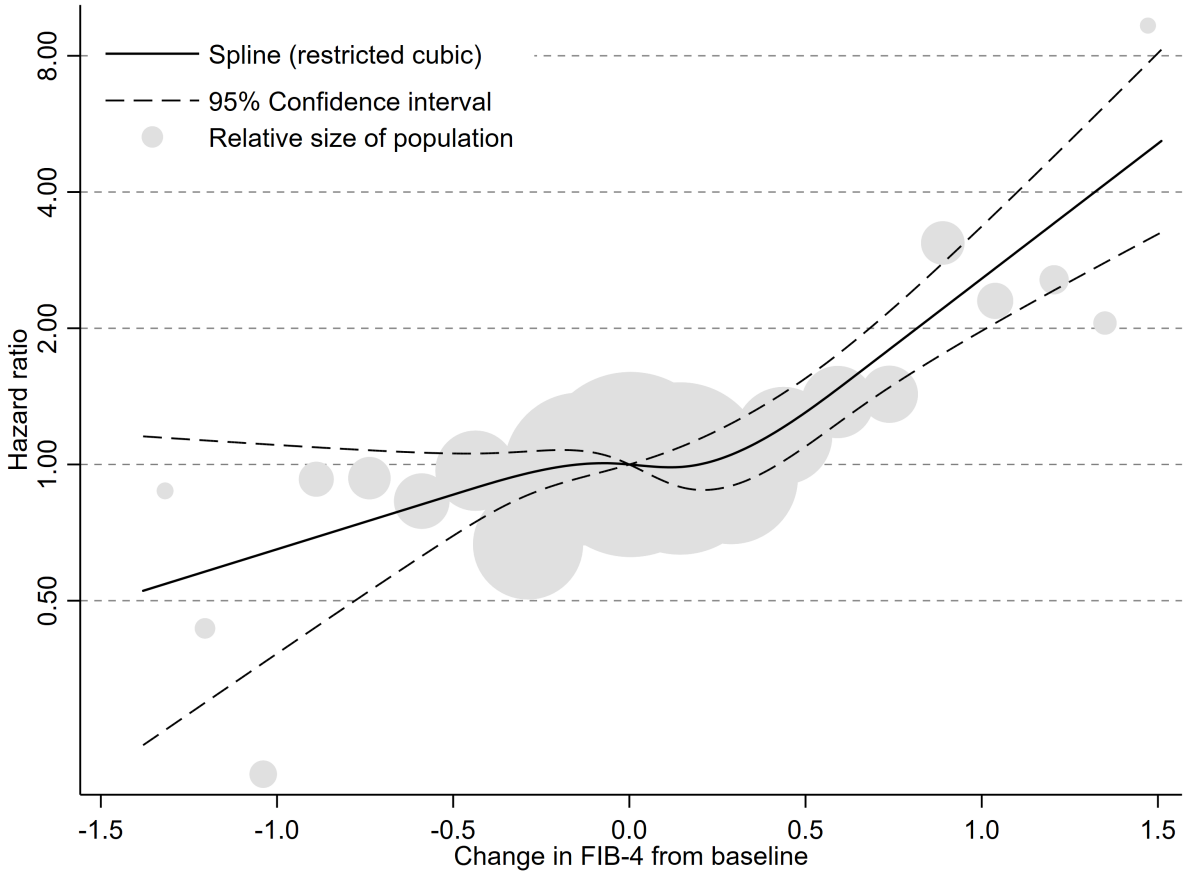
References

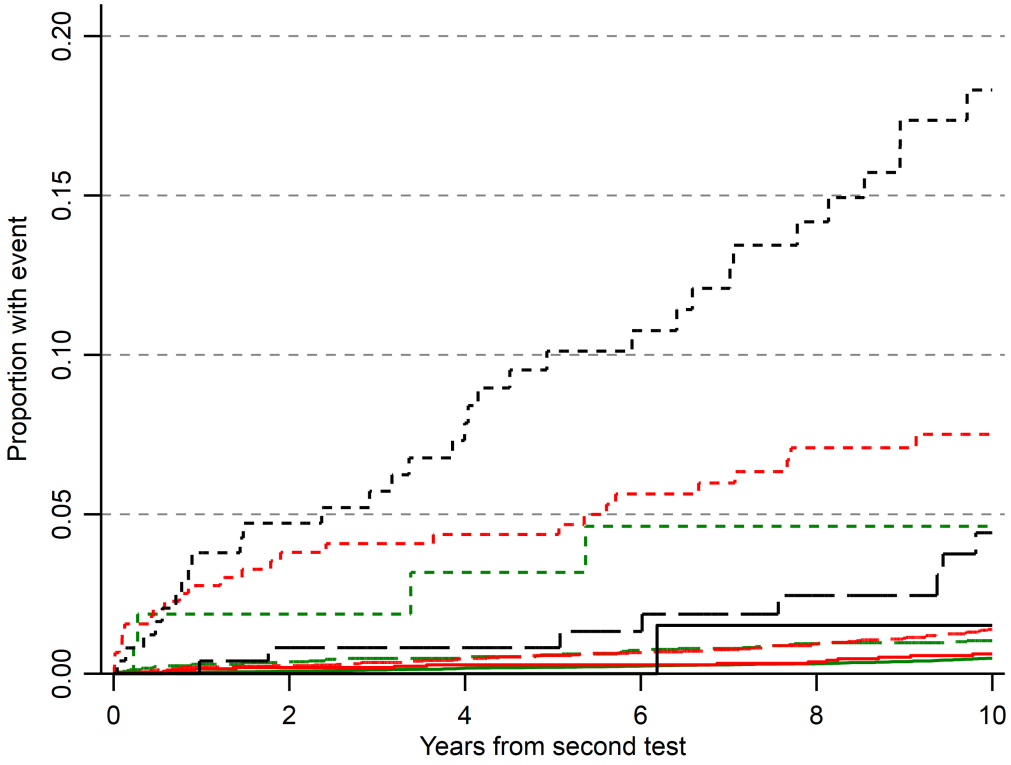
[1] Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish National Inpatient Register. BMC public health 2011;11:450.
[2] Brooke HL, Talback M, Hornblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish cause of death register. European Journal of Epidemiology 2017;32:765-773.
[3] 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.

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 for severe liver disease
- 50% of severe liver disease outcomes had consistently low or intermediate FIB-4
- About 1/3 of the cohort had intermediate or high FIB-4 at one of the tests
- FIB-4 is likely insufficient for screening for fibrosis in the general population


Men, β=0.030 (0.025-0.035)
Women, β=0.013 (0.009-0.018)
Change in FIB-4 from baseline
-0.1
0.0
0.1
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Years from baseline
Mean annual change
Regression line with 95% CI


Spline (restricted cubic)
95% Confidence interval
Relative size of population
Hazard ratio
0.50
1.00
2.00
4.00
8.00
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Change in FIB-4 from baseline



Number 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

1.Low - 1.Low 1.Low - 2.Int. 1.Low - 3.High
2.Int. - 1.Low 2.Int. - 2.Int. 2.Int. - 3.High
3.High - 1.Low 3.High - 2.Int. 3.High - 3.High