A machine learning-based classification of adult-onset diabetes identifies patients at risk of liver-related complications

Authors

Lukas Otero Sanchez, **Clara-Yongxiang Zhan**, Carolina Gomes da Silveira Cauduro, Laurent Crenier, Hassane Njimi, Gael Englebert, Antonella Putignano, Antonia Lepida, Delphine Degré, Nathalie Boon, Thierry Gustot, Pierre Deltenre, Astrid Marot, Jacques Devière, Christophe Moreno, Miriam Cnop, Eric Trépo

Correspondence

Lukas.otero.sanchez@ulb.be (L. Otero Sanchez), Eric.trepo@ulb.be (E. Trépo).

Graphical abstract
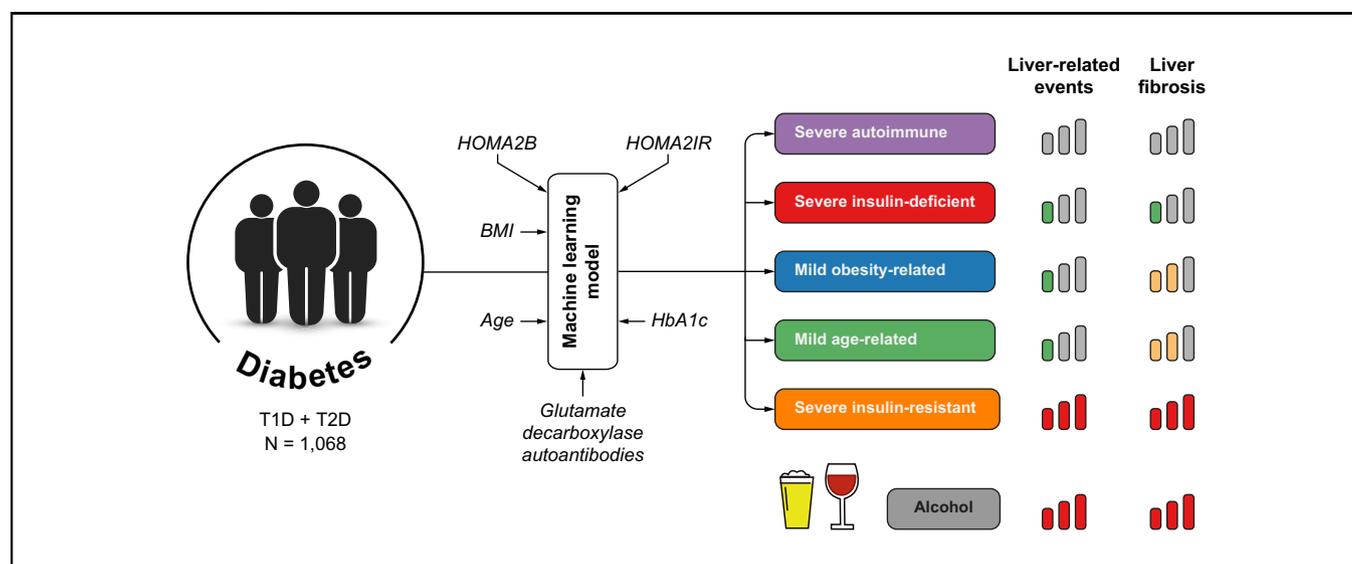
Highlights

- A novel machine learning-based classification identifies patients with diabetes at-risk of liver-related complications.

- Patients with severe insulin resistance had the highest risk of liver-related outcomes and fibrosis progression.

- Excessive alcohol consumption at the diagnosis of diabetes was the strongest risk factor for developing liver-related events.

Impact and implications

Diabetes represents a major risk factor for NAFLD development and progression. This study examined the ability of a novel machine-learning approach to identify at-risk diabetes subtypes for liver-related complications. Our results suggest that patients that had severe insulin resistance had the highest risk of liver-related outcomes and fibrosis progression. Moreover, excessive alcohol consumption at the diagnosis of diabetes was the strongest risk factor for developing liver-related events.

A machine learning-based classification of adult-onset diabetes identifies patients at risk of liver-related complications

Lukas Otero Sanchez,[1,2,*,†] Clara-Yongxiang Zhan,[1,†] Carolina Gomes da Silveira Cauduro,[3,4] Laurent Crenier,[3] Hassane Njimi,[5] Gael Englebert,[1] Antonella Putignano,[1,2] Antonia Lepida,[1,2] Delphine Degré,[1,2] Nathalie Boon,[1,2] Thierry Gustot,[1,2,6,7] Pierre Deltenre,[1,8] Astrid Marot,[9] Jacques Devière,[1,2] Christophe Moreno,[1,2] Miriam Cnop,[3,4] Eric Trépo[1,2,*]

[1]Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, CUB Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium; [2]Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium; [3]Department of Endocrinology, CUB Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium; [4]ULB Center for Diabetes Research, Medical Faculty, Université Libre de Bruxelles, Brussels, Belgium; [5]Biomedical Statistics, Universite Libre de Bruxelles, Brussels, Belgium; [6]Inserm Unité 1149, Centre de Recherche sur l'inflammation (CRI), Paris, France; [7]UMR S_1149, Université Paris Diderot, Paris, France; [8]Department of Gastroenterology and Hepatology, Clinique St Luc, Bouge, Belgium; [9]Department of Gastroenterology and Hepatology, CHU UCL Namur, Université Catholique de Louvain, Yvoir, Belgium

Background & aims: Diabetes mellitus is a major risk factor for fatty liver disease development and progression. A novel machine learning method identified five clusters of patients with diabetes, with different characteristics and risk of diabetic complications using six clinical and biological variables. We evaluated whether this new classification could identify individuals with an increased risk of liver-related complications.

Methods: We used a prospective cohort of patients with a diagnosis of type 1 or type 2 diabetes without evidence of advanced fibrosis at baseline recruited between 2000 and 2020. We assessed the risk of each diabetic cluster of developing liver-related complications (*i.e.* ascites, encephalopathy, variceal haemorrhage, hepatocellular carcinoma), using competing risk analyses.

Results: We included 1,068 patients, of whom 162 (15.2%) were determined to be in the severe autoimmune diabetes subgroup, 266 (24.9%) had severe insulin-deficient diabetes, 95 (8.9%) had severe insulin-resistant diabetes (SIRD), 359 (33.6%) had mild obesity-related diabetes, and 186 (17.4%) were in the mild age-related diabetes subgroup. In multivariable analysis, patients in the SIRD cluster and those with excessive alcohol consumption at baseline had the highest risk for liver-related events. The SIRD cluster, excessive alcohol consumption, and hypertension were independently associated with clinically significant fibrosis, evaluated by liver biopsy or transient elastography. Using a simplified classification, patients assigned to the severe and mild insulin-resistant groups had a three- and twofold greater risk, respectively, of developing significant fibrosis compared with those in the insulin-deficient group.

Conclusions: A novel clustering classification adequately stratifies the risk of liver-related events in a population with diabetes. Our results also underline the impact of the severity of insulin resistance and alcohol consumption as key prognostic risk factors for liver-related complications.

Impact and implications: Diabetes represents a major risk factor for NAFLD development and progression. This study examined the ability of a novel machine-learning approach to identify at-risk diabetes subtypes for liver-related complications. Our results suggest that patients that had severe insulin resistance had the highest risk of liver-related outcomes and fibrosis progression. Moreover, excessive alcohol consumption at the diagnosis of diabetes was the strongest risk factor for developing liver-related events.

Introduction

NAFLD is the predominant cause of chronic liver disease worldwide.[1] The rising prevalence of NAFLD is closely related to the global increases in overweight/obesity and diabetes mellitus that have occurred in the past few decades.[2,3]

Type 2 diabetes (T2D) is a major risk factor for NAFLD development and progression.[4] The estimated prevalence of NAFLD among patients with T2D is nearly 70%.[5] Current joint guidelines from the American Association of Clinical Endocrinology, the American Association for the Study of Liver Diseases,

the European Association for the Study of the Liver, the European Association for the Study of Diabetes, and the European Association for the Study of Obesity recommend screening for liver fibrosis in this specific population.[6,7] In T2D, liver fibrosis pathogenesis and NAFLD progression are strongly associated with insulin resistance.[8] Although type 1 diabetes (T1D) is pathophysiologically distinct from T2D, the population with T1D also shares similar risk factors related to liver fibrosis, including obesity, features of metabolic syndrome, and insulin resistance. These similarities suggest that some patients with T1D might also be at risk for NAFLD.[9] The diabetes population is highly heterogenous; it is, therefore, not surprising that the current classification fails to adequately identify at-risk patients with unfavourable outcomes, including liver-related complications.

Alcohol use disorder remains the major risk factor associated with complications of liver disease in patients with T2D.[10] However, very few studies to date have examined the association between the level of alcohol consumption and the risk of liver-related complications in the diabetes population.

Using a machine-learning approach, Ahlqvist et al.[11] identified five clusters of adult-onset diabetes reflecting different pathological profiles, and associated with different treatment responses and complication risks. Six key clinical parameters at diagnosis (i.e. age, BMI, HbA1c, glycaemia, C-peptide, and islet autoantibodies) were used to classify patients. Since, several studies have now confirmed the association between this clustering method and the risk for specific diabetes-related complications, including chronic kidney disease, polyneuropathy, diabetic retinopathy, and cardiovascular disease. Only a few studies have explored the association between this novel classification and NAFLD.[12–14] None have evaluated it in patients with diabetes with significant alcohol consumption. The association between cluster assignment and the development of liver-related events or clinically significant fibrosis has not been investigated.

In this study, we evaluated the risk of liver-related events and the risk of clinically significant fibrosis in a cohort of patients with diabetes without evidence of advanced fibrosis at baseline using this novel clustering classification.

Patients and methods
Data collection
The cohort of patients with diabetes from the endocrinology department of CUB Hôpital Erasme has been systematically followed since 2000. The following data were collected at diabetes diagnosis: age, weight, height, fasting glycaemia, and concomitant C-peptide, HbA1C, glutamic acid decarboxylase antibodies, most recent liver imaging, if available, Fibrosis-4 (FIB4), alcohol consumption (units/day), tobacco consumption (units/day), presence of arterial hypertension (systolic blood pressure ≥140 mmHg or treated hypertension), presence of hypercholesterolaemia (LDL >100 mg/dl or treated). Excessive alcohol consumption was defined as more than three drinks/day (or ≥40 g pure alcohol/day), and defined as chronic heavy alcohol consumption.[15] T1D or T2D was defined according to the diabetologist's diagnosis.

The local institutional review board approved the development of this prospective database and the study protocol (Erasme-ULB Ethics Committee, reference: B2020/580).

Study population
All adult patients diagnosed with diabetes included in the endocrinology department database between January 2000 and December 2020 were evaluated for inclusion. Exclusion criteria related to diabetes were: gestational, secondary, monogenic, or mitochondrial diabetes; missing data for clustering analysis, including age at diagnosis, weight, height, fasting glycaemia, and concomitant C-peptide, HbA1c, and glutamic acid decarboxylase antibodies. For each of the previous variables, the closest value from the diagnosis was selected to perform the clustering. Glycaemia and C-peptide were used to calculate homeostasis model assessment (HOMA), estimates of beta-cell function (HOMA2B), and insulin resistance (HOMA2IR) with the HOMA2 calculator from the University of Oxford (Oxford, UK).[16] Although useful for large cohorts, it should be acknowledged that HOMA estimates derived from fasting measures are less sensitive than measures from dynamic, stimulated tests, that is meal-stimulated or oral or intravenous glucose tolerance tests.[17–19] Patients diagnosed before 2000 for whom clinical/biological baseline data were available were also included. Other exclusion criteria were: the presence of hepatitis B surface antigen, positive hepatitis C virus PCR or human immunodeficiency virus antigen/antibodies, haematological or solid-organ transplantation, concomitant immunosuppressive treatment, and clinical cirrhosis or any sign of chronic liver disease at baseline. To exclude patients with a high probability of advanced fibrosis at diabetes diagnosis, we included only patients with FIB4 <2.67; the FIB4 value closest to diabetes diagnosis was used. The timeframe for FIB4 assessment was within a year of diabetes diagnosis. A total number of 1,612 patients with diabetes mellitus recorded from January 2000 to December 2020 in the endocrinology department database were evaluated for inclusion. After removing those with exclusion criteria, 1,068 patients with diabetes mellitus were included (Fig. S1).

Outcome measures
All outcomes were recorded until March 2022.

Liver-related outcomes
We defined liver-related events as the occurrence of at least one of the following complications during the follow-up period: ascites, encephalopathy, variceal haemorrhage, or hepatocellular carcinoma.

Liver fibrosis was assessed with either liver biopsy (percutaneous or trans-jugular) or fasting transient elastography using the FibroScan® system. The transient elastography value was validated if there were more than 10 valid measures with a success rate above 80%.[20] Clinically significant fibrosis (≥F2) was defined by a liver stiffness ≥9.1 kPa or histologically by the presence of portal and perisinusoidal fibrosis.[21] Advanced fibrosis was histologically characterised by the presence of bridging fibrosis (≥F3) or cirrhosis (≥F4) or by an liver stiffness ≥12.1 kPa (≥F3) or >18.5 kPa (≥F4).[22] Only patients in whom FIB4 and/or imaging and/or clinical and blood test results were suggestive of fatty liver disease were invited to have liver fibrosis assessment by FibroScan/liver biopsy during the follow-up period.

Liver steatosis was assessed by: liver biopsy with the presence of steatosis in ≥5% of hepatocytes; or fasting transient elastography with a controlled attenuation parameter (CAP) >255 dB/

min; or the presence of hepatic steatosis on imaging including abdominal ultrasound, magnetic resonance imaging, and abdominal computed tomography.[6] The presence of steatosis at baseline was defined as the presence of steatosis at imaging closest to and at the latest 5 years after diabetes diagnosis. Steatosis at the last-follow up was defined as steatosis present at the last available liver imaging during the follow-up. NASH was defined, on liver biopsy, by the presence of steatosis, hepatocyte ballooning, and lobular inflammation and a NAFLD activity score (NAS) ≥5.[6,21]

Statistical analysis

We applied the sex-specific classification algorithm published by Ahlqvist *et al.*[11] using the nearest centroid approach. The following variables at baseline were used for the cluster analysis: age, sex, BMI, HbA1c, HOMA2B, HOMA2IR, and glutamic acid decarboxylase antibody status. Each patient was assigned to one of the following clusters, mild age-related diabetes (MARD), mild obesity-related diabetes (MOD), severe insulin-deficient diabetes (SIDD), and severe insulin-resistant diabetes (SIRD). Patients who were glutamic acid decarboxylase antibody-positive were assigned to the severe autoimmune diabetes (SAID) cluster. Based on key clinical and biological features shared by some diabetes subgroups (*e.g.* metabolic control [HbA1c] level, beta cell/insulin-resistance scores, age, BMI), we secondarily pooled SAID with SIDD (creating an insulin-deficient cluster) and MARD with MOD (mild-insulin resistant cluster). This simplified classification is hence composed of three subgroups: severe insulin-resistant (SIR), mild insulin-resistant (MIR), and insulin-deficient (ID) diabetes.

Continuous variables are described as means (95% CI) or medians (IQR). The Kruskal–Wallis rank sum test, one-way ANOVA test, or Fischer's exact test were performed to compare clinical and demographical variables between groups. Follow-up time was defined as the period from diabetes diagnosis to the last follow-up visit. Data for patients were censored at the time of the last follow-up visit. We used a multistate approach using a competing risk model and coded (0-1-2) as follows: 0, alive; 1, non-fatal outcome of interest; and 2, death. Death was considered a competing risk event for the occurrence of non-fatal outcomes. Gray's test was used to compare cause-specific cumulative incidence function between groups. To assess factors predicting outcomes of interest, univariate and multivariable regression analyses were performed using the Fine and Gray competing risks regression model. Multivariable models were built using a forward elimination technique, and we introduced as candidates in multivariable models only factors with a value of p <0.1 in univariate analysis. Variables already included in the clustering analysis were not introduced as covariates in the multivariable models containing diabetes subtypes to avoid bias related to the effect of collinearity.[23] We also compared the predictive accuracy for liver fibrosis over time between insulin resistance score (HOMA2IR), and clusters. The different models were compared using the AUC and the Brier score over follow-up time. The prediction accuracy of each model was defined using a cross-validation.[24] All tests were two-tailed, and p ≤0.05 indicated a significant difference. All statistical analyses were done using R version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Characteristics of the study population and diabetes clusters

The study included 1,068 adult patients, of whom 878 (82.2%) had T2D, and 190 (17.3%) had T1D. Baseline clinical and biological characteristics are stratified by cluster and summarised in Table 1. A total of 162 (15.2%) patients with glutamic acid decarboxylase antibodies were allocated to the SAID cluster. They had the lowest BMI, the lowest insulin secretion/resistance scores (HOMA2B/HOMA2IR), and the lowest age at diabetes onset. Patients with SIDD (n = 266, 24.9%) had the poorest glycaemic control – with a low insulin secretion level (low HOMA2B index) – and the highest level of excessive alcohol consumption (>3 units/day). Patients with SIRD (n = 95, 8.8%) had the highest insulin resistance/secretion indexes (HOM2IR/HOMA2B) and 72

Table 1. Study population characteristics.

Characteristic	1/SAID, n = 162[†]	2/SIDD, n = 266[†]	3/SIRD, n = 95[†]	4/MOD, n = 359[†]	5/MARD, n = 186[†]	p value[*]
Age (years)	37 (14)	46 (11)	55 (11)	44 (10)	60 (9)	<0.001
Sex						<0.001
F	74 (46%)	88 (33%)	30 (32%)	197 (55%)	66 (35%)	
M	88 (54%)	178 (67%)	65 (68%)	162 (45%)	120 (65%)	
BMI (kg/m²)	24 (21–27)	29 (26–32)	32 (30–38)	34 (31–38)	27 (25–29)	<0.001
HbA1c (%)	9.20 (7.12–11.97)	10.50 (9.40–12.30)	7.10 (6.10–8.40)	6.90 (6.20–7.90)	6.70 (6.20–7.27)	<0.001
HOMA2B	24 (9–56)	31 (21–48)	142 (88–180)	84 (56–116)	76 (52–100)	<0.001
HOMA2IR	0.82 (0.57–1.67)	1.93 (1.18–3.00)	5.99 (3.88–7.72)	2.69 (1.88–3.79)	2.19 (1.48–2.74)	<0.001
C-peptide (nmol/L)	0.23 (0.00–0.55)	0.66 (0.40–0.96)	2.03 (1.53–2.56)	1.07 (0.76–1.49)	0.89 (0.61–1.09)	<0.001
Glycaemia (mg/dl)	178 (121–256)	202 (151–250)	131 (106–213)	137 (112–164)	131 (113–156)	<0.001
Alcohol consumption						0.022
No	129 (80%)	203 (76%)	66 (69%)	295 (82%)	138 (74%)	
Occasionally	19 (12%)	23 (8.6%)	18 (19%)	37 (10%)	24 (13%)	
1–3 units/day	8 (4.9%)	17 (6.4%)	5 (5.3%)	13 (3.6%)	16 (8.6%)	
>3 units/day	6 (3.7%)	23 (8.6%)	6 (6.3%)	14 (3.9%)	8 (4.3%)	
Tobacco	51 (32%)	57 (22%)	18 (19%)	53 (15%)	24 (13%)	<0.001
Hypertension	21 (13%)	85 (32%)	57 (60%)	131 (36%)	91 (49%)	<0.001
Hypercholesterolaemia	45 (28%)	87 (33%)	50 (53%)	145 (40%)	78 (42%)	<0.001

Excessive alcohol consumption was defined as >3 units/day. Group comparisons were made using the Kruskal–Wallis rank sum test, or Pearson's X² test. Values of p <0.05 were considered as significant.
MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.
[*] Kruskal-Wallis rank sum test; Pearson's X² test.
[†] Mean (SD); n (%); median (IQR).

(75.8%) were obese. The MOD cluster (n = 359, 33.6%) had the highest BMI without notable insulin resistance. Finally, patients with MARD (n = 186, 17.4%) were older and had better metabolic control than those assigned to other subgroups.

Cumulative incidence and factors associated with liver-related events in patients with diabetes without evidence of advanced fibrosis at diagnosis

The median follow-up for the whole cohort was 14.76 years (IQR: 8.2–22.06). For the whole population, the cumulative incidence of liver-related events at 20 and 30 years was 1.5% (95% CI: 0.7–2.8) and 3.4% (95% CI: 1.4–5.4), respectively (Fig. S2). The highest cumulative incidence of liver-related events at 20 years was observed in the SIRD cluster (7.8%, 95% CI: 2.2–18.2), followed by MOD (2.3%, 95% CI: 0.6–6.1), MARD (1.5%, 95% CI: 0.1–7.1), SIDD (0.5%, 95% CI: 0.0–2.4), and SAID 0% (Fig. S3). Results of univariate and multivariable analysis of factors associated with liver-related events are summarised in Table 2. Independent predictors of liver-related events in multivariable models were: excessive alcohol consumption at baseline (subdistribution hazard ratio [sHR], 11.30; 95% CI: 3.07–41.05, p <0.001) and allocation to the SIRD cluster (sHR, 7.62; 95% CI: 1.89–30.80, p = 0.001). Moreover, in a competing risk regression model, we included all clinical variables used to subclassify patients with diabetes. Independent predictors of liver-related events in univariate models were: excessive alcohol consumption at baseline, HOM2IR, age, and hypertension. Independent predictors of liver-related events in multivariable models remained: excessive alcohol consumption at baseline (sHR, 7.44; 95% CI: 2.12–26.1, p = 0.002), and HOMA2IR score (sHR, 1.43; 95% CI: 1.20–1.69, p <0.001) (Table S1). The median age of patients with diabetes who developed liver-related events was 68 (IQR: 63–75). Only one patient encountered liver-related events under 50 years (a patient with SIRD with excessive alcohol consumption).

Association between diabetes cluster and liver fibrosis

In this study, 359 (33.6%) patients were screened for liver fibrosis by FibroScan, and 126 (11.8%) had a liver biopsy. To assess the association between the cluster assignment and the development of clinically significant fibrosis during the follow-up, we restricted our analysis to the subgroup of patients that have been evaluated for liver fibrosis (n = 418). At 20 years, the highest cumulative incidence of diagnosis of clinically significant fibrosis (≥F2) was observed in patients with SIRD with 63.3% (95% CI:

41.3–85.3), followed by MOD with 44.6% (95% CI: 32.9–55.6) and MARD with 40.3% (95% CI: 23.6–56.5) (p <0.001) (Fig. 1). The lowest risk of significant fibrosis (i.e. F2–F4) was observed in patients allocated to the SAID and SIDD clusters, with 5.0% (95% CI: 0.003–21.1) and 26.3% (95% CI: 15.6–38.2), respectively. sHRs for the risk of clinically significant fibrosis for the whole follow-up period stratified by cluster were: 2.89 (95% CI: 1.67–5.01) for SIRD, 1.60 (95% CI: 1.03–2.48) for MOD, and 2.05 (95% CI: 1.19–3.53) for patients with MARD with the SIDD cluster considered as the reference (p = 0.001) (Table 3).

SIRD also had higher sHR for advanced fibrosis (≥F3) (sHR, 2.36; 95% CI: 1.21–4.60, p = 0.01) (Fig. S4). Considering the clustering classification, we analysed risk factors associated with clinically significant fibrosis in a Fine and Gray competing risks regression model. In multivariable analysis, excessive alcohol consumption (>3 units/day) (sHR, 2.60; 95% CI: 1.64–4.13, p <0.001), cluster allocation (p <0.001), and presence of hypertension (sHR, 1.47; 95% CI: 1.03–2.12, p = 0.035) at baseline were independent predictive factors of clinically significant fibrosis (Table 3). We also assessed the association between different levels of baseline alcohol consumption and the development of clinically significant fibrosis. We observed a significantly increased risk starting from three drinks/day (Fig. S5).

No association was found between glucose-lowering drugs (i.e. metformin, GLP1 receptor agonists, PPARγ agonists) and the development of clinically significant fibrosis (Table 3).

Moreover, in a competing risk regression model, we included all clinical variables used to subclassify patients with diabetes. Independent predictors of clinically significant fibrosis in multivariable models remained: excessive alcohol consumption at baseline (sHR, 2.53; 95% CI: 1.56–4.11, p <0.001), HOMA2IR score (sHR, 1.12; 95% CI: 1.04–1.21, p = 0.004), and HbA1c level (sHR, 0.98; 95% CI: 0.97–0.99, p = 0.001) (Table S2).

In addition, we compared the clinical characteristics of evaluated and non-evaluated patients for fibrosis during the follow-up (Table S3). We observed no statistically significant difference between screened and non-screened patients with SIRD concerning their clinical and biological characteristics at baseline. For the four remaining clusters, screened patients had higher resistance surrogate markers.

We also compared the prediction accuracy between insulin resistance score (HOMA2IR) and the clustering classification (Table S4). The clustering classification had a higher AUC after 5 years of follow-up compared with HOMA2IR, indicating a higher discriminatory accuracy (Fig. S6A). The clustering classification

Table 2. Factors associated with liver-related events.

Variable	Univariate analysis		Multivariable analysis	
	sHR (95% CI)	p value	sHR (95% CI)	p value
Cluster		0.001		0.009
SIDD ref.				
SIRD	8.90 (2.53–31.30)		7.62 (1.89–30.80)	
MOD	1.06 (0.27–4.25)		1.17 (0.26–5.29)	
MARD	1.34 (0.25–7.25)		1.65 (0.29–9.26)	
Sex (male)	1.11 (0.40–3.09)	0.80		
Hypertension	3.17 (1.12–9.00)	0.03	2.17 (0.69–6.86)	0.18
Hypercholesterolaemia	0.78 (0.24–2.51)	0.67		
Alcohol (>3 units/day)	12.3 (4.18–36.40)	<0.001	11.30 (3.07–41.05)	<0.001
Tobacco (active)	1.43 (0.41–5.05)	0.58		

Data are expressed as subdistribution hazard ratio (sHR) (95% CI). Univariate and multivariate regression analyses were performed using competing risk regression model. Excessive alcohol consumption was defined as >3 units/day. Values of p <0.05 were considered as significant.
MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.
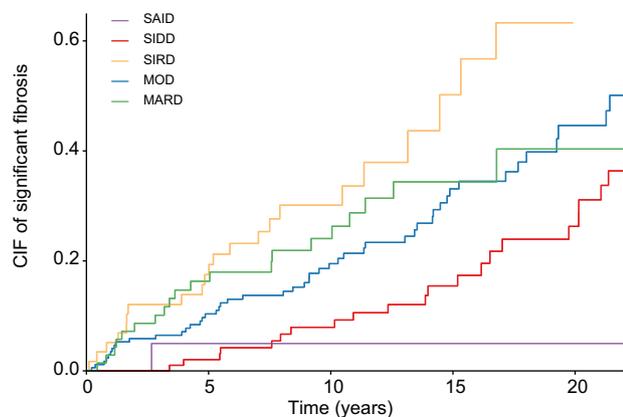
Fig. 1. Time to clinically significant fibrosis (F2–F4) by cluster. Cumulative incidence function of clinically significant fibrosis. Death was considered as competing risk factor for the diagnosis of clinically significant fibrosis. CIF, cumulative incidence function; MARD, mild age-relate diabetes; MOD, mild obesity-related diabetes; SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.

seems to have better discrimination and calibration (evaluated by the Brier score) over time compared with the HOMA2IR score (Fig. S6B).

We secondarily focused our analysis on our T1DM population. Of the 190 patients included, 23 were evaluated for liver fibrosis in the follow-up, among whom two (8.7%) developed clinically significant fibrosis. Patients without clinically significant fibrosis were younger, and had poorer metabolic control, lower BMI, and HOMA2IR/HOMA2B index as compared with patients who developed significant fibrosis (Table S5).

Association between simplified diabetes cluster assignment, liver-related events, and liver fibrosis

A total of 428 and 545 patients were assigned to the ID diabetes and the MIR diabetes clusters, respectively. The remaining 95 belonged to the SIR cluster. Their clinical and biological characteristics are summarised in Fig. S7. Using the simplified classification, the risk of developing liver-related events was, respectively, 14.3-fold (95% CI: 4.05–50.1) and 1.84-fold (95% CI: 0.52–6.57) higher for SIR and MIR as compared with patients in the ID cluster ($p < 0.001$).

At 20 years, the cumulative incidence of clinically significant fibrosis stratified by this simplified classification was 63.3% (95% CI: 41.3–85.3) for SIR, 44.3% (95% CI: 34.3–53.7) for MIR, and 23.4% (95% CI: 14.1–34.1) for ID diabetes ($p < 0.001$) (Fig. S8). The sHRs for clinically significant fibrosis were 3.30 (95% CI: 1.90–5.73, $p < 0.001$) and 1.98 (95% CI: 1.30–2.99, $p = 0.001$) for SIR and MIR, respectively, as compared to ID group (Table S6).

Liver steatosis according to diabetes cluster
Liver steatosis evaluation by imaging at baseline was available in 236 patients and revealed a higher proportion of steatosis in patients with SIRD (89%) and MOD (86%) as compared with MARD (75%), SIDD (64%), and SAID (64%) ($p = 0.013$) (Table 4). Liver steatosis evaluation by imaging at the last follow-up was available in 497 patients and also highlighted a higher proportion of steatosis in MOD (77%) and SIRD (76%) clusters ($p < 0.001$). In patients with an available CAP (n=316), SIRD and MOD also showed a higher CAP value than the other clusters. When considering the simplified classification, patients with SIR had a higher prevalence of steatosis at baseline (89%) than MIR (82%) and ID (64%) patients ($p = 0.01$) (Table S7). The prevalence of steatosis at last follow-up imaging and CAP values were also higher in SIR and were lowest in patients with an ID profile.

NASH according to diabetes cluster
We also evaluated the prevalence of biopsy-proven NASH in 124 patients. The highest prevalence of NASH was found in patients with MOD (31%) but was not significantly different from the other clusters (Table 4). Of note, none of the three patients who underwent biopsy in the SAID cluster had an NAS score >2.

Table 3. Factors associated with clinically significant fibrosis.

Variable	Univariate analysis		Multivariable analysis	
	sHR (95% CI)	*p* value	sHR (95% CI)	*p* value
Cluster		0.001		<0.001
SIDD ref.				
SIRD	2.89 (1.67–5.01)		3.14 (1.76–5.61)	
MOD	1.60 (1.03–2.48)		1.83 (1.17–2.86)	
MARD	2.05 (1.19–3.53)		2.45 (1.40–4.30)	
Sex (male)	1.17 (0.82–1.66)	0.40		
Hypertension	1.51 (1.06–2.14)	0.021	1.47 (1.03–2.12)	0.035
Hypercholesterolaemia	0.98 (0.68–1.40)	0.9		
Alcohol (>3 units/day)	1.98 (1.25–3.14)	0.004	2.60 (1.64–4.13)	<0.001
Tobacco (active)	0.77 (0.47–1.26)	0.3		
Diabetes medication				
Metformin alone	1.39 (0.86–2.24)	0.2		
PPARγ agonists	1.32 (0.38–4.58)	0.7		
GLP1R agonists	0.87 (0.50–1.50)	0.6		
GLP1R agonist–metformin	0.86 (0.49–1.52)	0.6		

Data are expressed as subdistribution hazard ratio (sHR) (95% CI). Univariate and multivariate regression analyses were performed using a competing risk regression model. Excessive alcohol consumption was defined as >3 units/day. Values of $p < 0.05$ were considered as significant.
GLP1R, glucagon-like peptide receptor; MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; PPARγ, peroxisome proliferator-activated receptor; SIDD, severe insulin-deficiency diabetes; SIRD, severe insulin-resistant diabetes.

Table 4. Distribution of steatosis and histological features across diabetes clusters.

Characteristics	n	SAID[†]	SIDD[†]	SIRD[†]	MOD[†]	MARD[†]	p value[*]
Median CAP	316	327 (277,360)	297 (256,343)	339 (310,374)	330 (293, 365)	305 (244,354)	0.005
Steatosis at baseline	236	9/14 (64%)	27/42 (64%)	32/36 (89%)	79/92 (86%)	39/52 (75%)	0.013
Steatosis at last follow-up	497	21/47 (45%)	77/124 (62%)	39/52 (76%)	145/188 (77%)	49/86 (57%)	<0.001
Definite NASH	124	0/3 (0%)	7/31 (23%)	3/18 (17%)	15/49 (31%)	2/23 (9%)	0.3

Definite NASH was defined, on liver biopsy, by a NAFLD activity score ≥5 with the presence of predominantly macrovesicular steatosis, and hepatocyte ballooning, and lobular/ or portal inflammation. Group comparisons were made using the Kruskal–Wallis rank sum test, Fisher's exact test, or Pearson's X^2 test. Values of p <0.05 were considered as significant.
CAP, controlled attenuation parameter; MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.
[*] Kruskal-Wallis rank sum test; Fisher's exact test; Pearson's X^2 test.
[†] Median (IQR); n/n_{tot} (%).

Discussion

This study showed that a novel machine learning-based classification adequately stratifies the risk of liver-related events in patients with diabetes. Patients at the highest risk of liver-related events and liver fibrosis were assigned to clusters with evidence of more pronounced insulin resistance. We also confirmed that alcohol consumption remains the most potent prognostic factor of the occurrence of liver-related events in this population.

The global prevalence of diabetes has increased over the past few decades, reaching nearly 6–10% of the general population in Western countries.[25,26] Patients with diabetes remain at high risk for liver fibrosis and liver-related outcomes. The estimated prevalence of significant fibrosis (≥F2) in patients with T2D has been evaluated across several studies and ranges between 9% and 20%.[5,27–29] Considering the worldwide increase in diabetes patients at-risk for liver fibrosis and liver-related events, identifying subtypes with different risk profiles is urgently needed.

We used a novel clustering method for adult-onset diabetes, developed by Ahlqvist et al.[11] to stratify the risk of liver-related events and liver fibrosis in patients with diabetes. Two studies have evaluated hepatic fibrosis features among clusters using serum-based markers in a limited number of patients.[12,30] Here, we used liver biopsy and elastography to show that patients with SIRD had the highest risk for clinically significant fibrosis (i.e. F2–F4). Patients with SIRD also had the highest risk for liver-related events. Conversely, none of the patients with SAID (glutamic acid decarboxylase antibody-positive) developed advanced fibrosis during follow-up and thus should be considered at very low risk. MARD and MOD had a 2.5- and twofold greater risk of developing significant fibrosis compared with patients with SIDD and should be regarded as at intermediate risk. Based on the similar metabolic profile of MARD with MOD (mild forms of insulin resistance) and SAID with SIDD (insulin deficiency), we secondarily evaluated the risk for liver fibrosis using a simplified classification into SIR (the SIRD cluster), MIR (i.e. MARD and MOD), and ID (i.e. SAID and SIDD) diabetes. Overall, patients assigned to the SIR and MIR groups had nearly three- and twofold greater risk of developing significant fibrosis, respectively, than the ID group. Here, we made use of HOMA estimates based on fasting glucose and C-peptide levels, which are less sensitive than dynamic tests; it will hence be of interest in future studies to assess insulin resistance and secretion using meal- or glucose-stimulated tests. It should be also noted that our simplified classification has not been validated so far. Therefore, it will require further evaluation in other studies, especially those assessing liver-related outcomes in patients with diabetes. Our results confirm that the risk of developing liver fibrosis is highly correlated with the degree of insulin resistance in people with diabetes and underlines the critical role of insulin resistance in the occurrence of hepatic fibrosis. There is evidence that insulin resistance plays a key role in lipid metabolism dysfunction by promoting de novo lipolysis, excessive free fatty acids accumulation in hepatocytes, and contributing to the development of lipotoxicity.[31] Liver lipotoxicity impairs insulin signalling, induces oxidative stress, and apoptotic pathways resulting in hepatocellular inflammation and fibrosis.[32]

We also confirmed that features of metabolic syndrome, including hypertension, confer an additive risk for liver fibrosis even in the population with diabetes. Moreover, we observed that excessive alcohol consumption at baseline (>3 units/day) is the most critical risk factor for liver-related outcomes even after correction for insulin-resistance level. Our data align with Mallet et al.[10] who recently showed that alcohol was the leading risk factor associated with liver-related complications in patients with T2D. These repeated observations suggest that degrees of insulin resistance and alcohol consumption are important prognostic factors of liver-related complications that should be carefully evaluated in patients with diabetes at baseline.

It should be noted that although patients with SIDD are at lower risk for liver fibrosis, the cumulative incidence of clinically significant fibrosis reached 26.3% at 20 years. Even if patients with SIDD could be considered at lower risk, this population should also be carefully monitored for liver fibrosis, especially if other features of metabolic syndrome or alcohol consumption are superimposed.

In this study, we assessed the cumulative incidence of liver-related events in patients with diabetes without advanced fibrosis at diagnosis. The observed cumulative incidence at 10 years was low, 0.34%, but consistent with the literature. The 10-year cumulative incidence of the liver-related event reported in a recent study enrolling 7,028 patients with NAFLD and T2D (including 6.4% of patients with cirrhosis at baseline) was 0.6%.[29] In another study including 80,224 patients who had obesity and 242,822 patients who did not have obesity, the cumulative incidence of severe liver disease at 10 years was <0.15%, <0.2%, and 0.4% in patients who were non-obese, overweight, and obese, respectively.[30] Despite a limited number of events, our results are in line with those of previous studies and suggest that SIRD patients should be carefully followed for liver-related complications. However, this interesting observation should be confirmed in other prospective cohorts with a longer follow-up. However, must be confirmed in a prospective manner, with a larger cohort followed for a more than 10-year period.

Age remains an important parameter to stratify the risk of liver-related events in the diabetes population. In our study the median age of patients with diabetes who developed liver-related events was 68 (IQR: 63–75), and only one patient encountered liver-related events under 50 years (a patient with

SIRD with excessive alcohol consumption). Moreover, the MARD cluster remains the second more at-risk cluster for liver fibrosis. These data are in line with previous studies that recommend screening patients with diabetes for liver-related complications starting at 50 years of age.[33]

The limitations of this study include its monocentric nature, the lack of systematic screening of liver fibrosis in the overall population, and the heterogeneity in the liver fibrosis assessment (*e.g.* liver biopsy or FibroScan). Indeed, because only patients with a higher probability of fibrosis progression were assessed by transient elastography or biopsy, this may lead to overestimating the incidence of this outcome. This applies particularly to non-SIRD clusters, where screened patients had a more pronounced level of insulin resistance. However, there was no significant difference in record clinical or biological variables between screened and unscreened patients with SIRD. Therefore, we are confident that we did not overestimate this subgroup's relative risk for liver fibrosis. We also acknowledge that by using a strict inclusion criterion for the diagnosis of NASH (*i.e.* NAS >5), our study may have inadvertently excluded a subset of patients with less severe NASH that are still considered, for example, in clinical trials.[34] This limitation could impact the generalisability of our results to the broader NASH population. Future studies evaluating this clustering approach may benefit from considering a more inclusive diagnostic criterion that captures a broader spectrum of NASH severity.

In conclusion, we showed that a novel clustering adequately stratifies the risk of liver-related complications in a population with diabetes. Patients in the SIRD cluster had the highest risk of liver-related outcomes and fibrosis progression. Furthermore, our results underline the impact of the severity of insulin resistance and alcohol consumption as key prognostic risk factors for liver-related complications.

Abbreviations

CAP, Controlled attenuation parameter; CIF, Cumulative incidence function; FIB4, Fibrosis-4; HOMA, Homeostasis model assessment; HOMA2B, HOMA of beta-cell function; HOMA2IR, HOMA of insulin resistance; ID, Insulin-deficient; MARD, Mild age-related diabetes; MIR, Mild insulin-resistant; MOD, Mild obesity-related diabetes; NAS, NAFLD activity score; SAID, Severe autoimmune diabetes; SIDD, Severe insulin-deficient diabetes; SIRD, Severe insulin-resistant diabetes; sHR, Subdistribution hazard ratio; T1D, Type 1 diabetes; T2D, Type 2 diabetes.

Conflicts of interest

TG received consulting fees from Promethera Biosciences, Martin Pharmaceuticals, Goliver therapeutics, and AbbVie and received research support from Gilead. ET received research support from Gilead. CM was paid as consultant by AbbVie, Novartis, Surrozen, and Gilead sciences; CM is a consultant for Julius clinical. JD has received research support for IRK approved studies from Fractyl In and is a shareholder of Endotools SA.

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

Authors' contributions

Designed the study: ET, MC, LOS. Collected data: LOS, CY, LC, GE, AP, DD, NB, TG, PD, AM, JD, CM, MC. Accessed, curated, and verified the data: LOS, CY. Performed the statistical analysis: LOS, HN. Directed the study: ET. All authors critically revised the manuscript and approved the final version to be published.

Data availability statement

The data shown in this article are available from the corresponding authors upon reasonable request.

Supplementary data

Supplementary data to this article can be found online at https://10.1016/j.jhepr.2023.100791.

References

Author names in bold designate shared co-first authorship

[1] Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018;15:11–20.

[2] Stefan N, Cusi K. A global view of the interplay between non-alcoholic fatty liver disease and diabetes. Lancet Diabetes Endocrinol 2022;10:284–296.

[3] Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2022;7:851–861.

[4] Kanwal F, Kramer JR, Li L, Dai J, Natarajan Y, Yu X, et al. Effect of metabolic traits on the risk of cirrhosis and hepatocellular cancer in nonalcoholic fatty liver disease. Hepatology 2020;71:808–819.

[5] Younossi ZM, Golabi P, de Avila L, Paik JM, Srishord M, Fukui N, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. J Hepatol 2019;71:793–801.

[6] European association for the study of the liver (EASL), European association for the study of diabetes (EASD), European association for the study of obesity (EASO). EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016;64:1388–1402.

[7] Cusi K, Isaacs S, Barb D, Basu R, Caprio S, Garvey WT, et al. American Association of Clinical Endocrinology Clinical Practice Guideline for the diagnosis and management of nonalcoholic fatty liver disease in primary care and endocrinology clinical settings. Endocr Pract 2022;28:528–562.

[8] Khan RS, Bril F, Cusi K, Newsome PN. Modulation of insulin resistance in nonalcoholic fatty liver disease. Hepatology 2019;70:711–724.

[9] de Vries M, Westerink J, Kaasjager KHAH, de Valk HW. Prevalence of nonalcoholic fatty liver disease (NAFLD) in patients with type 1 diabetes mellitus: a systematic review and meta-analysis. J Clin Endocrinol Metab 2020;105:3842–3853.

[10] Mallet V, Parlati L, Martinino A, Pereira JPS, Jimenez CN, Sakka M, et al. Burden of liver disease progression in hospitalized patients with type 2 diabetes mellitus. J Hepatol 2022;76:265–274.

[11] Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 2018;6:361–369.

[12] Zaharia OP, Strassburger K, Strom A, Bönhof GJ, Karusheva Y, Antoniou S, et al. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol 2019;7:684–694.

[13] Wagner R, Heni M, Tabák AG, Machann J, Schick F, Randrianarisoa E, et al. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 2021;27:49–57.

[14] Tanabe H, Saito H, Kudo A, Machii N, Hirai H, Maimaituxun G, et al. Factors associated with risk of diabetic complications in novel cluster-based diabetes subgroups: a Japanese retrospective cohort study. J Clin Med 2020;9:2083.

[15] Seitz HK, Bataller R, Cortez-Pinto H, Gao B, Gual A, Lackner C, et al. Alcoholic liver disease. Nat Rev Dis Primer 2018;4:16.

[16] Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care 1998;21:2191–2192.

[17] Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. Diabetes Care 2004;27:1487–1495.

[18] Mari A, Pacini G, Brazzale AR, Ahrén B. Comparative evaluation of simple insulin sensitivity methods based on the oral glucose tolerance test. Diabetologia 2005;48:748–751.

[19] Azzi AS, Cosentino C, Kibanda J, Féry F, Cnop M. OGTT is recommended for glucose homeostasis assessments in Friedreich ataxia. Ann Clin Transl Neurol 2019;6:161–166.

[20] Lucidarme D, Foucher J, Le Bail B, Vergniol J, Castera L, Duburque C, et al. Factors of accuracy of transient elastography (fibroscan) for the diagnosis of liver fibrosis in chronic hepatitis C. Hepatology 2009;49:1083–1089.

[21] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313–1321.

[22] Nguyen-Khac E, Thiele M, Voican C, Nahon P, Moreno C, Boursier J, et al. Non-invasive diagnosis of liver fibrosis in patients with alcohol-related liver disease by transient elastography: an individual patient data meta-analysis. Lancet Gastroenterol Hepatol 2018;3:614–625.

[23] Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Ann Intern Med 1993;118:201–210.

[24] Zhang Z, Cortese G, Combescure C, Marshall R, Lee M, Lim HJ, et al. Written on behalf of AME Big-Data Clinical Trial Collaborative Group. Overview of model validation for survival regression model with competing risks using melanoma study data. Ann Transl Med 2018;6:325.

[25] Our World in Data. Diabetes prevalence, 2019. https://ourworldindata.org/grapher/diabetes-prevalence. [Accessed 11 June 2022].

[26] International Diabetes Foundation. IDF diabetes atlas 2021 (10th ed). https://diabetesatlas.org/atlas/tenth-edition/.

[27] Kwok R, Choi KC, Wong GLH, Zhang Y, Chan HLY, Luk AOY, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut 2016;65:1359–1368.

[28] Lee HW, Wong GLH, Kwok R, Choi KC, Chan CKM, Shu SST, et al. Serial transient elastography examinations to monitor patients with type 2 diabetes: a prospective cohort study. Hepatology 2020;72:1230–1241.

[29] Roulot D, Roudot-Thoraval F, Nkontchou G, Kouacou N, Costes JL, Elourimi G, et al. Concomitant screening for liver fibrosis and steatosis in French type 2 diabetic patients using Fibroscan. Liver Int 2017;37:1897–1906.

[30] **Wang F, Zheng R**, Li L, Xu M, Lu J, Zhao Z, et al. Novel subgroups and chronic complications of diabetes in middle-aged and elderly Chinese: a prospective cohort study. Front Endocrinol 2022;12:802114.

[31] Marra F, Svegliati-Baroni G. Lipotoxicity and the gut-liver axis in NASH pathogenesis. J Hepatol 2018;68:280–295.

[32] Finck BN. Targeting metabolism, insulin resistance, and diabetes to treat nonalcoholic steatohepatitis. Diabetes 2018;67:2485–2493.

[33] Zhang X, Wong GLH, Yip TCF, Cheung JTK, Tse YK, Hui VWK, et al. Risk of liver-related events by age and diabetes duration in patients with diabetes and nonalcoholic fatty liver disease. Hepatology 2022;76:1409–1422.

[34] Francque SM, Bedossa P, Ratziu V, Anstee QM, Bugianesi E, Sanyal AJ, et al. A randomized, controlled trial of the pan-PPAR agonist lanifibranor in NASH. N Engl J Med 2021;385:1547–1558.