Transient elastography for screening of liver fibrosis: cost-effectiveness analysis from six prospective cohorts in Europe and Asia

Miquel Serra-Burriel*[1], Isabel Graupera*[2], Pere Torán[3], Maja Thiele[4], Dominique Roulot[5], Vincent Wai-Sun Wong[6], Indra Neil Guha[7], Núria Fabrellas[8], Anita Arslanow[9,10], Carmen Expósito[3], Rosario Hernández[11], Grace Lai-Hung Wong[6] , David Harman[7] Sarwa Darwish Murad[12], Aleksander Krag[4], Guillem Pera[3], Paolo Angeli[13], Peter Galle[9], Aithal Guruprasad[7] , Llorenç Caballeria[3], Laurent Castera[14], Pere Ginès[2], Frank Lammert**[10].

1. *Centre de Recerca en Economia I Salut - UPF; Universitat Pompeu Fabra*
2. *Liver Unit Hospital Clínic. Institut D'investigacions Biomèdiques August Pi I Sunyer (IDIBAPS). Centro de Investigación En Red de Enfermedades Hepáticas Y Digestivas (Ciberehd). Barcelona; Faculty of Medicine and Health Sciences, University of Barcelona*
3. *USR Metropolitana Nord. IDIAP Jordi Gol; ICS Institut Català de la Salut*
4. *Odense University Hospital, University of Southern Denmark and Odense; Patient data Exploratory Network (OPEN); Department of Gastroenterology and Hepatology, Odense, Denmark*
5. *Department of Hepatology, AP-HP, Hopital Avicenne, Bobigny, France; Université Paris 13, Sorbonne Paris Cité.*
6. *The Chinese University of Hong Kong; Dept. of Medicine and Therapeutics*
7. *NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK*
8. *Faculty of Medicine and Health Sciences; School of Nursing; University of Barcelona; Institut D'investigacions Biomèdiques August Pi I Sunyer (IDIBAPS). Centro de Investigación En Red de Enfermedades Hepáticas Y Digestivas (Ciberehd). Barcelona;*
9. *Department of Internal Medicine I, University Medical Centre of the Johannes Gutenberg-University Mainz, Langenbeckstrasse 1, 55131 Mainz, Germany; Department of Internal Medicine, University Medical Centre of the Johannes Gutenberg-University Mainz, Langenbeckstrasse 1, 55131 Mainz, Germany*
10. *Department of Medicine II, Saarland University Medical Center, Saarland University, Homburg, Germany*
11. *CAP La Marina, Institut Català de la Salut de Barcelona, Barcelona, Spain; CAP La Marina, Institut Català de la Salut de Barcelona, Barcelona, Spain*
12. *Department of Gastroenterology and Hepatology, Erasmus MC University Medical Centre, Rotterdam, the Netherlands*
13. *Of Internal Medicine and Hepatology; of Medicine (Dimed); University and General Hospital of Padova*
14. *Hôpital Beaujon; Department of Hepatology, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, Clichy, France; Université Paris VII, Inserm Umr 1149, Centre de Recherche Sur L'inflammation, Paris, France*

*First- co-authors: M. Serra and I. Graupera

**Corresponding author: Frank Lammert: frank.lammert@uks.eu

Keywords: Alcoholic Liver Disease; Liver Fibrosis; Non-Alcoholic Fatty Liver Disease; Stratified Screening; Transient Elastography

Disclosure of interests: MS, IG, LC, MT, DR, WS, NG, NF, RH, GW, SM, AK, PA, AA, PT, LC and FL have no conflicts of interests. IG has received lecture fees from Gilead. PGinès reports grants and personal fees from Grifols, grants and personal fees from Gilead, grants from Mallinckrodt, personal fees from Promethera, personal fees from Martin Pharmaceuticals, grants from Ferring Pharmaceuticals, grants and personal fees from Sequana. V. Wong has served as a consultant or advisory board member for AbbVie, Allergan, Echosens, Gilead Sciences, Janssen, Perspectum Diagnostics,

Pfizer and Terns. He has also received lecture fees from Bristol-Myers Squibb, Echosens, Gilead Sciences and Merck.

Contributions: *LiverScreen consortium*

Grant support: *EIT Health project 2018, project number EIT 18258; BMBF Liver Systems Medicine, project number LiSyM 031L005; The Danish study was funded by Innovation Fund Denmark, the European Union's Horizon 2020 Research and Innovation Program (grant agreement number 668031). This study was funded by a grant awarded to PG (PI16/00043), integrated in the Plan Nacional I+D+I and co-funded by ISCIII-Subdirección General de Evaluación and European Regional Development Fund FEDER.*

Acknowledgments: P. Ginès is a recipient of an ICREA Academia award.

Abbreviations used in this paper: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; TE, transient elastography; LSM, liver stiffness measurement; CAP, controlled attenuation parameter; QALYs, quality-adjusted life-years; ICERs, incremental cost-effectiveness ratios; HCV, hepatitis C virus; ALD alcoholic liver disease; HCC, hepatocellular carcinoma; CTREE, conditional inference tree. AIC, Akaike information criteria; GDP, gross domestic product.

Abstract

Background & Aims: Nonalcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD) pose an important challenge to current clinical healthcare pathways due to a large number of at-risk patients. Therefore, we aimed to explore the cost-effectiveness of transient elastography (TE) as a screening method to detect liver fibrosis in a primary care pathway.

Methods: Cost-effectiveness analysis using real-life individual patient data from six independent prospective cohorts (five from Europe and one from Asia). A diagnostic algorithm with conditional inference trees was developed to explore the relationships between liver stiffness, socio-demographics, comorbidities, and hepatic fibrosis, the latter assessed by fibrosis scores (FIB-4, NFS) and liver biopsy in a subset of 352 patients. We compared the incremental cost-effectiveness of a screening strategy against standard of care alongside the numbers needed to test to diagnose a patient with fibrosis stage ≥F2.

Results: The data set encompassed 6,295 participants (mean age 55±12 years, BMI 27±5 kg/m^2, liver stiffness 5.6±5.0 kPa). A 9.1 kPa TE cut-off provided the best accuracy for the diagnosis of significant fibrosis (≥F2) in general population settings, whereas a threshold of 9.5 kPa was optimal for populations at-risk for alcoholic liver disease. TE with the proposed cut-offs outperformed fibrosis scores in terms of accuracy. Screening with TE was cost-effective with mean incremental cost-effectiveness ratios ranging from 2,570 €/QALY (95% CI 2,456 - 2,683) for a population at-risk for alcoholic liver disease (age ≥45 years) to 6,217 €/QALY (95% CI 5,832 - 6,601) in the general population. Overall, there was a 12% chance of TE screening being cost-saving across countries and populations.

Conclusions: Screening for liver fibrosis with transient elastography in primary care is a cost-effective intervention for European and Asian populations and may even be cost-saving.

Lay summary

The lack of optimized public health screening strategies for the detection of liver fibrosis in adults without known liver disease presents a major healthcare challenge. Analyses from six independent international cohorts with transient elastography measurements based on economic modelling shows that a community-based risk-stratification strategy for alcoholic and non-alcoholic fatty liver diseases is cost-effective through earlier identification of patients and potentially cost-saving for our healthcare systems.

Highlights

- Optimal liver stiffness thresholds for community-based screening in populations with metabolic risk factors and alcoholic is between 9.1 and 9.5 kPa for the diagnosis of significant fibrosis (stages ≥F2)
- Transient elastography is a cost-effective intervention for identifying patients with liver fibrosis in primary care. Healthcare systems would need to invest between 2,500 (at-risk population) to 6,500 (general population) purchasing power parity-adjusted euros to gain an extra year of life, adjusted per quality of life.
- The survival effect of screening is most pronounced for the identification of significant (≥F2) fibrosis.

Introduction

Alcoholic liver disease (ALD) and non-alcoholic fatty liver disease (NAFLD) are leading causes of chronic liver diseases, hepatocellular carcinoma (HCC) and liver-related deaths worldwide[1,2]. While the causes, consequences and treatment strategies for ALD and NAFLD are being studied and developed[3–5], the majority of patients are still diagnosed at an advanced stage of disease[6]. Consequently, the course of action towards early disease detection from a public health perspective remains a grey area in hepatology[7].

Severe fibrosis and cirrhosis is the major predictor of excess mortality, liver transplantation and liver-related events in patients with NAFLD and ALD[8,9]. Consequently, early detection before progression to advanced fibrosis might be more beneficial and cost-effective, as it allows for timely lifestyle interventions, patient guidance and disease monitoring.

The high prevalence of ALD and NAFLD, in combination with a slow, asymptomatic disease progression, quickly translate into pressure on our healthcare systems, specifically on general practitioners and primary healthcare specialists. The search for optimal patient management and care pathways is thus of utmost importance, including mass preventive screening programmes[10].

The development of non-invasive diagnostic tools for the staging of liver fibrosis in ALD and NAFLD potential applications has widened. Transient elastography (TE) has been shown to be a reliable diagnostic tool for fibrosis assessment[11,12]. Use of TE could lead to highly cost-effective care pathways while minimizing costs, in comparison to magnetic resonance elastography, a

technique that either ignore or underestimate societal perspectives and limited resources[13–16]. Other alternatives, mostly based on serum biomarkers such as Fibrosis-4 (FIB-4) or NAFLD Fibrosis Score (NFS) have also been developed and validated [31].

Recently, there have been individual attempts in the literature to assess prevalence rates, diagnostic accuracy and potential algorithms to stratify patients at risk of severe fibrosis[17–22]. However, mainly due to the limited sample size and the heterogeneity of populations and healthcare systems, a broader assessment of the technology implementation encompassing both health and economic outcomes is lacking.

The aim of the present study is to assess the cost-effectiveness of a liver screening program and surveillance intervention based on individual risk factors with a TE screening algorithm confirmed by percutaneous liver biopsy into a broad spectrum of populations without known liver disease, communities, and healthcare systems.

Patients and methods

Study populations

Patients from seven previous independent prospective studies that have used TE (FibroScan©) as a screening method for liver fibrosis detection were included in the study. The final cohort includes 6,295 patients from six different countries: France (FR), Spain (ES), Denmark (DK),United Kingdom (UK), Germany (DE) and Hong Kong (HK)[23–29]. For the analysis we have considered the two studies from Spain as the same cohort, and have defined 6 cohorts, one from each country. The cohorts from ES, DE and HK include patients from the general population above 18 years, the cohort from France includes patients from the general population above 45 years, the cohort from UK includes patients above 18 years with risk factors for chronic liver disease (with hazardous alcohol use or diabetes) and the cohort from Denmark comprised only patients above 18 years at risk for hazardous alcohol consumption. Hazardous alcohol use was defined as an alcohol consumption >14 units per week for women and >21 units per week for men[5]. Data on demographics, physical exam, clinical and laboratory parameters were included alongside with comorbidities.

The cohorts from Denmark, Spain, France and UK were designed to obtain liver biopsy in patients to confirm liver fibrosis. All patients from Denmark were invited to undergo liver biopsy due to history of, either prior or ongoing, excessive drinking. The Spanish patients (general population cohort) were invited to undergo liver biopsy if LSM >6.8 kPa (n=299), but only a third of these patients accepted. French patients were referred to biopsy if LSM >8.0 kPa and

UK patients if LSM >8.2 kPa and at least one risk factor, either for alcohol or metabolic syndrome.

Finally, data from 352 patients with successful liver biopsy were included (199 from Denmark, 101 from Spain, 27 from France and 25 from UK). Liver fibrosis was classified using the Kleiner scoring system [30]. FIB-4 and NFS scores [31] were also computed for each biopsied patient in order to assess the comparative diagnostic accuracy of these screening instruments.

Figure 1 presents the flowchart of enrollment, eligibility, analysis and available liver biopsies of each cohort. Inference regarding fibrosis staging is constraint by the 5.5% of patients who underwent liver biopsy.

Statistical analysis

Databases were merged and analyzed with conditional inference trees (CTREE) [32], to explore the relationships between socio-demographics, comorbidities, LSM, and hepatic fibrosis, as assessed by liver biopsy. The CTREE was fitted to the liver biopsied group to obtain the optimal cut-off points for which the distribution of fibrosis stage classification has highest diagnostic accuracy. Model's goodness-of-fit was assessed by means of the area under the receiver-operating-characteristics curve (ROC) for binary outcome [≥F2], and 3-class [F0-F1, F2-F3, F4] accuracy was used to test the accuracy of the CTREE model[33].

Missing values in the development sample were lacking in 6% of the data fields. They were imputed with a multiple imputation random forest, obtaining a normalized root mean squared error of 0.0626 for numeric variables and a proportion of falsely classified entries of 0.0136. In order to avoid overfitting of

the diagnostic model, a 5-fold cross-validation procedure with five repeats was undertaken. Class balance across fibrosis stages was ensured using synthetic minority oversampling technique[34]. Calibration between observed and predicted probabilities was performed by visual inspection of calibration plots. Fibrosis groups F2, F3, and F4 were grouped in order to assess binary discrimination performance.

Best-subset analysis [35] was performed for the covariates other than LSM in order to achieve a theoretical reduction in unnecessary testing. The predictive models included all available predictors as potential candidates to enter the diagnostic pathway: gender, age, active alcohol consumption, active smoking, viral hepatitis, weight, height, BMI, abdominal perimeter, systolic blood pressure, diastolic blood pressure, diabetes mellitus, obesity, hypertension, glucose level, creatinine, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, AST, ALT, GGT, albumin, ferritin, leucocytes, hemoglobin, platelets, TE probe (M or XL), TE LSM value and TE IQR. Because of the population unobserved heterogeneity, fixed-effects were pre-specified in all models. The numbers needed to test in order to diagnose one case with significant fibrosis were computed for each country and risk factor, including general population, obesity, DM, and high risk alcohol consumption. All statistical analysis were performed with R 3.4.1.[36]

Economic modeling

The results of the six screening programs cohorts were used in the parameter tuning of a previously published cost-effectiveness model[37] to assess the cost-effectiveness of screening in the general population as well as populations with risk factors for chronic liver diseases such as diabetes mellitus, obesity, or

alcohol consumption. This economic model compares two different pathways of detection and risk stratification for advanced chronic liver disease (significant liver fibrosis) in adults with suspicion of NAFLD or ALD in a primary care setting. One pathway uses TE and the other pathway uses aminotransferase activities (as standard of care) to detect patients with chronic liver disease. The model considers the prevalence of each fibrosis stage in a given population diagnosed by TE and aminotransferase activities. It also considers the misclassification of diagnosis, the natural disease progression rate in each stage, the quality of life measured as quality-adjusted life years (QALYs), the annual costs of each stage of the disease, and the diagnosis costs.

The perspective of the economic model was generated with provider-direct costs only, with a 30-year time horizon and a 3% discount rate both on health outcomes and costs. Health outcomes were measured as QALYs. No assumptions regarding willingness to pay thresholds were made due to the multiplicity of healthcare systems, however, following WHO recommendations[38] for international thresholds, we define "highly cost-effective interventions" as those below one gross domestic product (GDP) per capita of the country per QALY gained. Costs are in 2017 euros, purchasing power parity (PPP) was adjusted for all six countries, and the target population was a cohort of patients with average age 45 at risk either with a history of alcohol consumption, diabetes, metabolic syndrome or a combination of above. Both failure rates for TE and liver biopsy were considered alongside their associated morbidity and mortality.[39]

Incremental cost-effectiveness ratios (ICERs) [40] was defined as the principal outcome of the economic model. ICER expresses the economic value of an

intervention with respect to a comparator, in this case screening with TE vs. standard of care. Using difference in QALYs as incremental cost-effectiveness measure allows to compare value across interventions and across pathologies.

To be able to apply the model to different healthcare systems, several assumptions had to be made, mostly regarding care and cost structure, rate of fibrosis progression and treatment effectiveness over fibrosis progression rate. Hence, probabilistic sensitivity analysis was performed for these parameters to account for the level of uncertainty associated with the estimates.

The only difference applied to the modelling setting was in the elastography testing cost structure, which is described in the Model Appendix 1; details of the assumptions, states and transition probabilities of the present study are also presented there. All modelling analyses was performed in Microsoft Excel 2016 and are available as Supplementary material for further non-commercial use.

Results

Patient demographics

Out of the 6,295 patients, 6,199 had successful LSM performed with FibroScan© devices (1.5% failure rate) and were included in the subsequent analysis. Table 1 shows the baseline characteristics of the six cohorts. Patients had a mean age of 54.7 years (±12.2), BMI of 27.1 kg/m^2 (±4.9) and abdominal perimeter 91 cm (±11.5); 52% were women, 18% smoked, 33% were obese, and 13% had type 2 diabetes. In total, 3,223 of patients (52%) were from Southern Europe, 1,984 (32%) were from Northern Europe, and 992 (16%) were from Asia.

Liver stiffness measurements

The majority of LSM were performed with an M probe (92.3%), whereas the others were performed with an XL probe (7.7%). Mean LSM was 5.6 kPa (±5.0). Figure 2 presents the non-parametric density distributions in kPa of LSM in each of the included cohorts. It can be observed that for the Asian population (HK), LSM distribution is significantly shifted to the left as compared to European populations, only the German cohort was comparable in terms of LSM distribution. Within the European populations, English (UK) and Danish (DK) patients present similarly skewed to the right LSM distributions, whereas the Spanish (ES), German (DE) and French (FR) cohorts were less skewed.

Fibrosis stages in liver biopsy and LSM cut-offs for significant fibrosis in general and at-risk populations

Out of the 299 patients from the Spanish studies who were invited to perform liver biopsy, 101 underwent successful liver biopsy. In this group, 88 had

NAFLD and 7 were diagnosed with alcoholic liver disease; the remaining patients had no major histopathological abnormalities. The distribution of fibrosis stages (according to Kleiner scores) were 57% for F0, 15% for F1, 21% for F2, 3% for F3, and 4% for F4. With respect to the Danish cohort we included 199 with successful liver biopsy. Fibrosis distribution was 6% F0, 52% F1, 18% F2, 7% F3, and 18% F4. Regarding the French cohort, 89 patients presented LSM >8.0 kPa and 27 underwent liver biopsy with a fibrosis distribution of 4% F0, 30% F1, 33% F2, 0% F3, and 33% F4. Finally, 25 biopsies were performed in the UK cohort out of 98 with LSM >8.0 kPa, with a fibrosis distribution of 32% F0, 24% F1, 12% F2, 24% F3, and 8% F4.

Figure 3 presents the LSM distribution in kPa of biopsied cohorts according to fibrosis stages. The distribution of fibrosis staging in the general population (ES and FR cohorts), risk population for both NAFLD/ALD (UK cohort) and risk population for ALD (DK cohort) differs significantly ($p<0.001$). Patients in the DK and UK cohorts showed more advanced liver fibrosis. Table 2 summarizes the clinical, demographic and analytical characteristics of the cohorts. In comparison to the DK cohort, patients from the ES and UK cohorts presented more frequently with characteristics of the metabolic syndrome and had higher prevalence of obesity (72-76% vs. 28%), arterial hypertension (42-60% vs. 28%), and diabetes (28-84% vs. 2%). On the other hand, patients from the DK cohort displayed higher serum GGT activities, probably related to alcohol consumption, and significantly lower serum albumin concentrations, corresponding to the higher prevalence of advanced fibrosis in this cohort. However, it is worth noting that the small UK biopsied cohort has a high

prevalence of concomitant risk factors for both NAFLD and ALD (76% obesity, 84% diabetes, 76% hazardous alcohol use).

Next, we evaluated the best cut-off for significant fibrosis (stage ≥F2) in the set of patients with liver biopsy (n=352). Through CTREE, an LSM cut-off greater than 19.1 kPa was shown to exhibit a conditional probability of 87.3% for fibrosis stage ≥F2 and 57.1% for stage F4 (cirrhosis) regardless of the sampled population. For LSM values below 19.1 kPa, distinct optimal cut-offs were identified, which depended upon the sampled populations: For general population sampling (ES cohort), an LSM threshold of >9.1 kPa was found to provide the best diagnostic accuracy with an average negative predictive value of 88.1% for finding fibrosis stage ≥F2 in patients below this threshold and a probability of 57.6% for finding it in patients above this cut-off. Table 3 presents the complete diagnostic yield of the CTREE model.

In cohorts with patients who had a history of hazardous alcohol use and/or metabolic syndrome (FR, DK and UK cohorts), a slightly higher optimal threshold for fibrosis staging were obtained: With LSM lower than 9.5 kPa the probability of fibrosis stage ≥F2 was 9.7% only, while higher than 9.5 kPa the probability increased to 52.4% for this at-risk population.

According to the developed predictive model, a total of 3.9% (n=238) patients of the general population samples were predicted to have ≥F2 fibrosis, whereas 28.8% (n=157) of at-risk patients were predicted to have developed at least fibrosis stage F2.

Figure 4 presents the detailed cut-off values alongside the tree structure and the distributions of fibrosis stages. The fibrosis CTREE had an average 5-fold 5

repeats cross-validated 3-class diagnostic balanced accuracy of 70.8% (95% CI 68.9 – 72.7%), i.e. compared to a random model, CTREE presented an increase of 30.7% (p<0.001, accuracy > no information rate).

The other serum surrogate fibrosis markers had the following 3-class accuracies: FIB-4 59.4% (95% CI 57.1 – 61.5%) and NFS 55.5% (95% CI 53.3 – 57.6%). Compared to the CTREE, FIB-4 had 11.4% lower accuracy (p<0.001) and NFS 15.3% lower (p<0.001). The statistical supplementary appendix comprises the distributions of these markers [Figures S18-S20] alongside the comparative diagnostic performance [Figures S21, S23-25], and the calibration of the models [Figure S22].

Table 4 presents the number needed to screen (NNS) for each cohort and population sampling method. The average NNS to detect a patient with fibrosis stage ≥F2 case by general population sampling was 34.5 individuals, for obese patients 8.5, for patients with diabetes 7.0, and for at-risk patients with alcohol consumption 12.5. However, caution must be taken in interpreting these numbers, given the population sampling heterogeneity across studies.

Cost-effectiveness results

The optimal cut-offs defined in our set of biopsied patients were used to infer the predicted fibrosis prevalence rates in each of the heterogeneous cohorts. This real data was introduced in the economic model and given the assumptions of the cost-effectiveness model, we found that the mean ICER of the risk-stratification strategy with TE ranged from 2,570 €/QALY (95% CI 2,456 - 2,683) in Spain for a population at-risk for alcoholic liver disease (age ≥ 45 years) to 6,217 €/QALY (95% CI 5,832 – 6,601) in the Hong Kong general

population setting. While there were significant differences in the results across countries and targeted populations, all results were highly cost-effective below the one gross domestic product per capita/QALY threshold.

Figure 5 presents the survival curves derived from the state-transition probabilities. The model estimates the difference in quality-adjusted survival between patients diagnosed and undiagnosed for each fibrosis stage. The area between diagnosed and undiagnosed curves is the average effectiveness in terms of QALYs of the screening program. Figure 5 illustrates that there is a benefit in QALYs of patients when they are diagnosed compared to undiagnosed in all stages, but this benefit is higher in stages ≥F2, compared to F0-1 or F4.

Figure 6 present the results of the sensitivity analyses in terms of cost-effectiveness acceptability curves depending upon patient selection either based on risk factors (obese, diabetes and alcohol) or in the general population setting by country. All targeted populations present an average probability of being cost saving of 12%, highlighting the potential savings associated with the intervention.

With respect to ICER estimates across sampling strategies, obese or diabetic populations require 93.2% (95% CI 91.2 - 95.3%) and 85.6% (95% CI 83.7 - 87.4%) of resources, respectively, compared to the general population to obtain the same value in terms of QALYs. Of note, in a population at-risk for alcoholic liver diseases, only 65.3% (95% CI 64.1 - 66.5%) of resources are needed in comparison to general population screening, highlighting the efficiency and feasibility of targeted interventions, which is highly dependent on the prevalence in the targeted population.

Discussion

The results of the present study demonstrate that non-invasive screening for liver fibrosis with TE among the general population, and among patients with risk factors for chronic liver disease, is cost-effective. We have availed of individual data from seven previous prospective studies that performed screening of liver fibrosis with TE in the general population and in patients with risk factors for chronic liver disease. In our study, data from the subset of patients who had undergone liver biopsy was used to define the diagnostic cut-offs for significant liver fibrosis. Having defined the best cut-offs, we applied them to our six different cohorts to assess the prevalence rates of significant fibrosis. This real-life data was then used to tune an economic model that compares two distinct pathways to detect significant fibrosis, i.e. the TE detection pathway versus the standard of care pathway (based on increased serum liver enzymes). The rationale behind non-invasive fibrosis detection by TE as public health intervention is earlier and more reliable patient identification, timely referral to specialist care, adequate treatment, and enrollment into surveillance programmes.

Our study shows that, irrespective of the targeted population, screening for liver fibrosis with optimized algorithms is a highly cost-effective public health intervention, with an average probability of 12% of being cost-saving. As might be expected, when we focus on patients with risk factors for chronic liver disease, including patients with diabetes, obesity or hazardous alcohol consumption, the screening program is even more cost-effective. Differences across risk factor targeting represent cost reductions around 8.3–48.9% to achieve the same value.

Liver fibrosis is, among all the histopathological changes that occur in chronic liver diseases, the major factor that predicts the long-term outcome of patients with chronic liver diseases. Fibrosis stage determines disease progression to cirrhosis and the development of liver-related complications and mortality, irrespective of the etiology of the disease[9,41]. Liver fibrosis has the potential of regression if effective treatment to control or cure the underlying disease is implemented, being more feasible at early stages[42,43]. Therefore, under the prism of public health strategies, the goal of a screening program should be to timely detect patients with advanced fibrosis or cirrhosis at high risk of liver-related decompensation or death to promote liver fibrosis regression and improve survival. In fact, our results confirm that a screening program based on the detection of liver fibrosis by TE is cost-effective, especially in early fibrosis stages ≥F2 and unsurprisingly somewhat less in patients with advanced liver disease (≥F4).

In our study we used TE as non-invasive screening method for liver fibrosis diagnosis. The efficacy of TE to detect liver fibrosis has been demonstrated during the past decade[44] however, most studies have been performed in patients with already known chronic liver diseases, such as HBV or HCV infections and NAFLD lately[45]. Data to establish the best cut-offs for the diagnosis in primary care setting is scarce[45]. In our study, the best cut-off for the diagnosis of significant fibrosis among patients from general population was 9.1 kPa. Interestingly, the subset of patients with liver biopsy from the general population had a high prevalence of components of the metabolic syndrome, the majority was obese and almost one third presented with diabetes, known risk factors for NAFLD. Moreover, the predominant etiology according to liver

biopsy in this subset of patients was NAFLD. Therefore, we can presume that the best cut-off for the diagnosis of significant fibrosis in the context of patients with metabolic risk factors for NAFLD among general population is 9.1 kPa. On the other hand, the cut-off for significant fibrosis among patients with alcohol-risk consumption was slightly higher (9.5 kPa). However, for clinical practice we can assume that any value above 9-9.5 kPa may indicate the presence of significant liver fibrosis in the setting of general population and patients with risk factors for NAFLD or ALD. Furthermore, as it has been previously shown, in our study TE is superior to serum biomarkers (FIB-4 and NFS) for fibrosis detection.

One of the main strengths of our project lays on the refinement of the parameter tuning of the economic model for target populations and diagnostic accuracy in a wide set of populations and healthcare systems. The cost-effectiveness analysis showed that the screening program targeting patients at risk for ALD and obese or diabetic populations is highly cost-effective. When implementing liver screening, we would need to invest between 2,500 (at-risk population) to 7,500 (general population) purchasing power parity-adjusted euros to gain an extra year of life, adjusted per quality of life. From a public health point of view, anything below one GDP per capita is deemed as highly cost-effective [38]. One of the sensitive issues around the cost-effectiveness estimates is the cost structure of the testing. A constant marginal cost per test is assumed, whereas in real-world settings, amortization through usage and labor imputation might result in decreasing marginal cost, hence biasing upwards our cost estimates. Our data highlights the specific diagnostic and economic consequences of targeted populations for such a public health intervention in hepatology.

Notwithstanding that the implementation in itself is not the objective of this study, in public settings, budgetary impact and difficulties in provision of public services have to be taken into account, as observed in preventive screening interventions in other fields of medicine.[46–52]. Interestingly, as compared to screening in later stages of chronic liver disease for HCC[53], screening for significant liver fibrosis (≥F2) presents 10-fold improvement in terms of efficiency, highlighting the relevance of early identification, referral and surveillance of these patients. In the present scenario of a growing epidemic of NAFLD, the implementation of screening programs to detect the patients with advanced fibrosis is essential. The implementation of TE in the primary care setting would allow involving community-based resources, including nurses and primary care physicians, to maximize value of performed interventions[24].

The current study has some limitations that should be acknowledged. First, only 5.5% of patients included in the analysis underwent liver biopsy. However, we performed analysis to evaluate the potential sample selection bias and uncertainty around parameter estimates of the economic model and both analyses were robust to exclude bias and uncertainty. Secondly, our economic model had several assumptions, mostly regarding care and cost structure, rate of fibrosis progression and treatment effectiveness over fibrosis progression rate. To account for the level of uncertainty associated with the estimates, probabilistic sensitivity analysis was performed for these parameters. Finally, it has been proposed that serum biomarkers should be used as a first step for liver fibrosis detection in general population leaving TE for a second step[10,44,54]. Our results clearly show that TE performed better than serum biomarkers for fibrosis detection. Whether a two-step approach using serum

biomarkers followed by TE it is more cost-effective and cost-saving in mass screening should be tested in future studies.

In summary, a screening program for the detection of liver fibrosis with transient elastography at primary care centers is a highly cost-effective intervention and potentially cost-saving and could represent a valuable public health strategy in the era of NAFLD epidemics.

References:

[1] Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013;10:686–90. doi:10.1038/nrgastro.2013.171.

[2] Williams R, Alexander G, Armstrong I, Baker A, Bhala N, Camps-Walsh G, et al. Disease burden and costs from excess alcohol consumption, obesity, and viral hepatitis: fourth report of the Lancet Standing Commission on Liver Disease in the UK. Lancet (London, England) 2018;391:1097–107. doi:10.1016/S0140-6736(17)32866-0.

[3] Cohen SM. Alcoholic Liver Disease. Clin Liver Dis 2016;20:xiii–xiv. doi:10.1016/j.cld.2016.05.001.

[4] Louvet A, Mathurin P. Alcoholic liver disease: mechanisms of injury and targeted treatment. Nat Rev Gastroenterol Hepatol 2015;12:231–42. doi:10.1038/nrgastro.2015.35.

[5] European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu M, European Association for the Study of the Liver A, Lackner C, Mathurin P, Moreno C, Spahr L, et al. EASL Clinical Practice Guidelines: Management of alcohol-related liver disease. J Hepatol 2018;69:154–81. doi:10.1016/j.jhep.2018.03.018.

[6] Jepsen P, Ott P, Andersen PK, Sørensen HT, Vilstrup H. Clinical course of alcoholic liver cirrhosis: A Danish population-based cohort study. Hepatology 2010;51:1675–82. doi:10.1002/hep.23500.

[7] Ginès P, Graupera I, Lammert F, Angeli P, Caballeria L, Krag A, et al. Screening for liver fibrosis in the general population: a call for action. Lancet Gastroenterol Hepatol 2016;1:256–60. doi:10.1016/S2468-1253(16)30081-4.

[8] Hagström H, Nasr P, Ekstedt M, Hammar U, Stål P, Hultcrantz R, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265–73. doi:10.1016/j.jhep.2017.07.027.

[9] 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;149:389–97.e10. doi:10.1053/j.gastro.2015.04.043.

[10] Tsochatzis EA, Newsome PN. Non-alcoholic fatty liver disease and the interface between primary and secondary care. Lancet Gastroenterol Hepatol 2018;3:509–17. doi:10.1016/S2468-1253(18)30077-3.

[11] Wong VW-S, Vergniol J, Wong GL-H, Foucher J, Chan HL-Y, Le Bail B, et al. Diagnosis

of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454–62. doi:10.1002/hep.23312.

[12] Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. J Hepatol 2008;48:835–47. doi:10.1016/j.jhep.2008.02.008.

[13] Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. Gastroenterology 2017;152:598–607.e2. doi:10.1053/j.gastro.2016.10.026.

[14] Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients with Nonalcoholic Fatty Liver Disease Than Transient Elastography. Gastroenterology 2016;150:626–637e7. doi:10.1053/j.gastro.2015.11.048.

[15] Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. Gastroenterology 2015. doi:10.1053/j.gastro.2015.11.048.

[16] Huwart L, Sempoux C, Vicaut E, Salameh N, Annet L, Danse E, et al. Magnetic Resonance Elastography for the Noninvasive Staging of Liver Fibrosis. Gastroenterology 2008;135:32–40. doi:10.1053/j.gastro.2008.03.076.

[17] Harris R, Harman DJ, Card TR, Aithal GP, Guha IN. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. Lancet Gastroenterol Hepatol 2017;2:288–97. doi:10.1016/S2468-1253(16)30205-9.

[18] Boursier J, de Ledinghen V, Leroy V, Anty R, Francque S, Salmon D, et al. A stepwise algorithm using an at-a-glance first-line test for the non-invasive diagnosis of advanced liver fibrosis and cirrhosis. J Hepatol 2017;66:1158–65. doi:10.1016/j.jhep.2017.01.003.

[19] Stål P. Liver fibrosis in non-alcoholic fatty liver disease-Diagnostic challenge with prognostic significance. World J Gastroenterol 2015;21:11077–87. doi:10.3748/wjg.v21.i39.11077.

[20] Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. J Hepatol 2016;65:570–8. doi:10.1016/j.jhep.2016.04.023.

[21] Cassinotto C, Boursier J, de Lédinghen V, Lebigot J, Lapuyade B, Cales P, et al. Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. Hepatology 2016;63:1817–27. doi:10.1002/hep.28394.

[22] Karlas T, Petroff D, Sasso M, Fan J-G, Mi Y-Q, de Lédinghen V, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. J Hepatol 2017;66:1022–30. doi:10.1016/j.jhep.2016.12.022.

[23] Roulot D, Costes J-L, Buyck J-F, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut 2011;60:977–84. doi:10.1136/gut.2010.221382.

[24] Caballería L, Pera G, Arteaga I, Rodríguez L, Alumà A, Morillas RM, et al. High Prevalence of Liver Fibrosis Among European Adults With Unknown Liver Disease: A Population-Based Study. Clin Gastroenterol Hepatol 2018. doi:10.1016/j.cgh.2017.12.048.

[25] Fabrellas N, Hernández R, Graupera I, Solà E, Ramos P, Martín N, et al. Prevalence of hepatic steatosis as assessed by controlled attenuation parameter (CAP) in subjects

with metabolic risk factors in primary care. A population-based study. PLoS One 2018;13:e0200656. doi:10.1371/journal.pone.0200656.

[26] Thiele M, Detlefsen S, Sevelsted Møller L, Madsen BS, Fuglsang Hansen J, Fialla AD, et al. Transient and 2-Dimensional Shear-Wave Elastography Provide Comparable Assessment of Alcoholic Liver Fibrosis and Cirrhosis. Gastroenterology 2016;150:123–33. doi:10.1053/j.gastro.2015.09.040.

[27] Harman DJ, Ryder SD, James MW, Jelpke M, Ottey DS, Wilkes EA, et al. Direct targeting of risk factors significantly increases the detection of liver cirrhosis in primary care: a cross-sectional diagnostic study utilising transient elastography. BMJ Open 2015;5.

[28] Arslanow A, Baum C, Lammert F, Stokes CS. Nichtinvasive Früherkennung von Lebererkrankungen im Rahmen der betrieblichen Gesundheitsförderung. Zentralblatt Fur Arbeitsmedizin, Arbeitsschutz Und Ergon 2017;67:201–10. doi:10.1007/s40664-017-0187-z.

[29] Wong VW-S, Chu WC-W, Wong GL-H, Chan RS-M, Chim AM-L, Ong A, et al. Prevalence of non-alcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. Gut 2012;61:409–15. doi:10.1136/gutjnl-2011-300342.

[30] 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;41:1313–21. doi:10.1002/hep.20701.

[31] Castera L, Vilgrain V, Angulo P. Noninvasive evaluation of NAFLD. Nat Rev Gastroenterol &Amp; Hepatol 2013;10:666.

[32] Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. J Comput Graph Stat 2006;15:651–74. doi:10.1198/106186006X133933.

[33] Holmes G, Pfahringer B, Kirkby R. Multiclass Alternating Decision Trees. Proc 13th Eur Conf Mach Learn (ECML '02) 2002;2430:161–72. doi:10.1007/3-540-36755-1_14.

[34] Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002. doi:10.1613/jair.953.

[35] Wang Y, Luangkesorn L, Shuman LJ. Best-subset selection procedure. Proc. - Winter Simul. Conf., 2011, p. 4310–8. doi:10.1109/WSC.2011.6148118.

[36] (2016), R Core Team R Foundation for Statistical Computing, Vienna A. R: A language and environment for statistical computing. n.d.

[37] Tanajewski L, Harris R, Harman DJ, Aithal GP, Card TR, Gkountouras G, et al. Economic evaluation of a community-based diagnostic pathway to stratify adults for non-alcoholic fatty liver disease: a Markov model informed by a feasibility study. BMJ Open 2017;7.

[38] Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost–effectiveness of interventions: alternative approaches. Bull World Health Organ 2015;93:118–24. doi:10.2471/BLT.14.138206.

[39] West J, Card TR. Reduced Mortality Rates Following Elective Percutaneous Liver Biopsies. Gastroenterology 2010;139:1230–7. doi:https://doi.org/10.1053/j.gastro.2010.06.015.

[40] Wright D. Incremental Cost-Effectiveness Ratio (ICER). York Heal Econ Consort 2017. doi:10.1111/j.1524-4733.2004.75003.x/full.

[41] Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;349:825–32. doi:10.1016/S0140-6736(96)07642-8.

[42] Marcellin P, Gane E, Buti M, Afdhal N, Sievert W, Jacobson IM, et al. Regression of

cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. Lancet (London, England) 2013;381:468–75. doi:10.1016/S0140-6736(12)61425-1.

[43] Poynard T, McHutchison J, Manns M, Trepo C, Lindsay K, Goodman Z, et al. Impact of pegylated interferon alfa-2b and ribavirin on liver fibrosis in patients with chronic hepatitis C. Gastroenterology 2002;122:1303–13. doi:https://doi.org/10.1053/gast.2002.33023.

[44] Wong VW-S, Adams LA, de Lédinghen V, Wong GL-H, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol 2018;15:461–78. doi:10.1038/s41575-018-0014-9.

[45] Friedrich-Rust M, Poynard T, Castera L. Critical comparison of elastography methods to assess chronic liver disease. Nat Rev Gastroenterol &Amp; Hepatol 2016;13:402.

[46] Kim JJ, Leung GM, Woo PPS, Goldie SJ. Cost-effectiveness of organized versus opportunistic cervical cytology screening in Hong Kong. J Public Health (Bangkok) 2004;26:130–7. doi:10.1093/pubmed/fdh138.

[47] Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial. N Engl J Med 2014;371:1793–802. doi:10.1056/NEJMoa1312547.

[48] Gorin SS, Gauthier J, Hay J, Miles A, Wardle J. Cancer screening and aging: Research barriers and opportunities. Cancer, vol. 113, 2008, p. 3493–504. doi:10.1002/cncr.23938.

[49] Brouwers MC, De Vito C, Bahirathan L, Carol A, Carroll JC, Cotterchio M, et al. What implementation interventions increase cancer screening rates? a systematic review. Implement Sci 2011;6:111. doi:10.1186/1748-5908-6-111.

[50] Carter-Harris L, Gould MK. Multilevel barriers to the successful implementation of lung cancer screening: Why does it have to be so hard? Ann Am Thorac Soc 2017;14:1261–5. doi:10.1513/AnnalsATS.201703-204PS.

[51] Minniear TD, Gilmore B, Arnold SR, Flynn PM, Knapp KM, Gaur AH. Implementation of and barriers to routine HIV screening for adolescents. Pediatrics 2009;124:1076–84. doi:10.1542/peds.2009-0237.

[52] Miller AB. Implementation of Colon Cancer Screening: Techniques, Costs, and Barriers. Gastroenterol Clin North Am 2008;37:83–95. doi:10.1016/j.gtc.2007.12.015.

[53] Kuo MJ, Chen HH, Chen CL, Fann JCY, Chen SLS, Chiu SYH, et al. Cost-effectiveness analysis of population-based screening of hepatocellular carcinoma: Comparing ultrasonography with two-stage screening. World J Gastroenterol 2016;22:3460–70. doi:10.3748/wjg.v22.i12.3460.

[54] Castera L. Noninvasive Evaluation of Nonalcoholic Fatty Liver Disease. Semin Liver Dis 2015;35:291–303. doi:10.1055/s-0035-1562948.

Figure Legends:

Figure 1*. Patient flowchart, combined database:* enrollment, TE and biopsy for studies in six countries in Europe and Asia.

Figure 2. *TE LSM distributions across studies:* the figure presents the non-parametric kernel distribution of liver stiffness measurements in kPa across studies. Studies ES and DE present a somewhat similar distribution, while studies DK and UK present more skewed LSM distributions in accordance with the risk factor, either NAFLD, ALD or both sampling of the population. Study FR distribution lays in between the previous studies. Study from HK, the Asian population presents a less right-shifted distribution of LSM.

Figure 3*. TE LSM distribution across fibrosis stage groups:* kernel density estimates of TE LSM by fibrosis stage and sampling strategy. Study ES (Spanish cohorts, n=101) are grouped and presented in the upper panel. UK biopsied subsample (n=25) are represented in the middle panel, while study DK (Danish cohort n=199) is presented in the lower panel. For general population (ES) the distribution of LSM is more concentrated around defined thresholds. ALD biopsied population (UK & DK) F2-F4 LSM are flattened in respect with the general population.

Figure 4*. CTREE fibrosis staging for LSM thresholds with empirical distributions:* The first split of the algorithm, LSM >19.1 kPa (node 9, p<0.001) denotes a probability of the patient having a fibrosis stage 2-3-4 of 88% independently of the sampled population. For values below 19.1 kPa the next split is the study population (p<0.001). The algorithm automatically differentiates conditional on the unobserved characteristics of studies ES and UK&DK. NAFLD and ALD populations are split. For NAFLD population an optimal LSM

kPa threshold is identified at >9.1 kPa with a probability of being F2-F3 of 57.6%. For the ALD population the optimal threshold for detecting F2-F3-F4 is identified at 9.5 kPa with a probability of 52.4%.

***Figure 5**. Cost-effectiveness model survival estimates by fibrosis group and diagnostic arm:* Survival estimates of fibrosis stage patients F0-F1, F2-F3 and F4 either diagnosed or undiagnosed. The area between diagnosed and undiagnosed curves is the average effectiveness in terms of survival of the screening program.

***Figure 6**. Cost-effectiveness acceptability curves by country and targeted population, n=1,000:* The figure presents the probability of the screening intervention being cost-effective for a given willingness to pay level. The curves are estimated from the probabilistic sensitivity analysis. It is worth noting that all models have a baseline 12% probability of being cost-saving. The upper panel shows the analysis for general population screening settings, the second for obese population, the third for diabetes type 2 and the lowest panel shows the results for an alcoholic population.

Table 1. Baseline characteristics of patients from 6 countries included in the study.

	ES N=3333	**DE** N=106	**UK** N=378	**DK** N=199	**FR** N=1357	**HK** N=922
Female n (%)	1906 (57.2%)	79 (74.5%)	123 32.5%	53 26.6%	553 40.7%	533 57.8%
Age	55.0 (12.1)	35.7(12.8)	61.8(15.0)	54.9(11.3)	57.9(8.95)	48.1(10.5)
Smoke active	0.23 (0.42)	n/a	n/a	0.59(0.49)	n/a	0.10(0.30)
Viral hepatitis	0.01 (0.12)	0.02(0.14)	0.00(0.00)	0.00(0.00)	n/a	0.00(0.00)
BMI	28.3 (4.91)	24.2(4.19)	28.5(5.25)	26.8(5.27)	26.7(4.23)	22.8(3.50)
Abd. per.	94.7 (12.8)	80.3(11.6)	n/a	n/a	90.0(11.4)	81.7(10.1)
DM	10%	4%	56%	2%	n/a	4%
Obesity	31%	10%	35%	28%	20%	3%
Arterial HT	26%	7%	46%	28%	n/a	16%
Glucose	6.05(14.0)	5.12(0.64)	n/a	6.81(1.91)	5.75(1.67)	5.14(0.92)
Creatinine	0.90(0.24)	0.85(0.16)	1.02(0.28)	0.73(0.17)	0.82(0.18)	0.75(0.16)
Cholesterol	5.53(1.03)	4.88(0.98)	4.61(1.27)	5.05(1.24)	5.41(1.02)	5.16(1.00)
Triglycerides	1.24(0.81)	1.14(0.63)	1.74(1.03)	1.54(1.02)	1.32(0.89)	1.37(1.19)
AST	23.7(9.15)	25.1(8.92)	28.0(18.9)	n/a	22.6(14.6)	21.4(12.2)
ALT	23.9(14.3)	23.5(11.8)	28.2(19.3)	40.0(31.5)	25.3(15.5)	26.0(16.0)
GGT	33.4(38.7)	23.3(17.4)	n/a	226(392)	37.4(56.3)	n/a
Bilirubin	11.3(5.08)	8.93(6.24)	12.0(5.06)	14.0(10.9)	n/a	13.5(5.96)
Albumin	44.0(3.25)	n/a	37.8(2.92)	40.2(5.27)	n/a	45.2(2.58)
Ferritin	118(119)	n/a	n/a	267(384)	165(179)	n/a
Leukocytes	9.85(176)	n/a	7.02(2.19)	n/a	6.42(1.83)	6.01(1.51)
Platelets	243(60.0)	n/a	241(62.3)	n/a	251(59.4)	244(54.3)
Probe:						
M	3233	104 (99.0%)	338 (89.4%)	186 (95.4%)	1357 (100%)	922 (100%)
XL	96	1 (0.95%)	40 (10.6%)	9 (4.62%)	0 (0.00%)	0 (0.00%)
Mean TE kPa	4.94(2.18)	4.62(1.57)	7.98(7.41)	17.2(20.6)	5.72(2.36)	4.84(2.62)

Average values and standard deviations in brackets (). All liver blood metrics are presented in mmol/L. BMI: body mass index, Abd. per.: abdominal perimeter in cm, DM: diabetes mellitus, HT: hypertension, AST: aspartate aminotransferase, ALT: alanine aminotransferase, GGT: gamma-glutamyl transpeptidase. ES= Spanish cohorts 1&2, DE= Germany cohort, UK= England cohort, DK= Danish cohort, FR= French cohort, HK= Hong Kong cohort.

Table 2. Baseline characteristics of patients from cohorts with liver biopsy available.

	ES N=101	**UK** N=25	**DK** N=199	**FR** N=27	P value
Female n (%)	33 (32.7%)	5 (20.0%)	53 (26.6%)	2 (7.41%)	0.055
Age	58.1 (8.83)	61.1 (10.2)	54.9 (11.3)	59.2 (9.33)	0.003
Hazardous alcohol use	0.16 (0.37)	0.76 (0.44)	1.00 (0.00)	0.70 (0.47)	<0.001
Active smoking	0.24 (0.43)	n/a	0.59 (0.49)	n/a	<0.001
Viral hepatitis	0.01 (0.10)	0.00 (0.00)	0.00 (0.00)	n/a	0.331
BMI	33.1 (5.30)	33.9 (6.05)	26.8 (5.27)	28.4 (5.89)	<0.001
DM	0.28 (0.45)	0.84 (0.37)	0.02 (0.14)	n/a	<0.001
Obesity	0.72 (0.45)	0.76 (0.44)	0.28 (0.45)	0.33 (0.48)	<0.001
Arterial hypertension	0.42 (0.50)	0.60 (0.50)	0.28 (0.45)	n/a	0.001
Glucose	6.30 (1.69)	n/a	6.81 (1.91)	7.31 (3.52)	0.035
Creatinine	0.95 (0.22)	0.99 (0.22)	0.73 (0.17)	0.80 (0.22)	<0.001
Cholesterol	5.51 (0.94)	4.06 (1.20)	5.05 (1.24)	5.15 (1.04)	<0.001
Cholesterol LDL	3.40 (0.95)	1.99 (1.03)	3.09 (1.52)	3.19 (0.82)	<0.001
Cholesterol HDL	1.26 (0.27)	1.18 (0.34)	1.40 (0.50)	1.28 (0.39)	0.018
Triglycerides	1.84 (1.09)	2.06 (0.94)	1.54 (1.02)	1.49 (0.98)	0.022
ALT	39.1 (25.8)	43.2 (23.3)	40.0 (31.5)	61.3 (53.0)	0.011
GGT	63.4 (87.9)	n/a	226 (392)	144 (250)	<0.001
Bilirubin	11.8 (5.24)	10.5 (4.38)	14.0 (10.9)	n/a	0.065
Albumin	44.2 (5.15)	37.2 (2.55)	40.2 (5.27)	n/a	<0.001
Ferritin	206 (221)	n/a	267 (384)	440 (678)	0.152
Mean TE kPa	10.4 (5.66)	18.0 (9.09)	17.2 (20.6)	14.0 (8.04)	0.005
IQR TE	2.20 (0.64)	3.93 (3.64)	2.22 (3.27)	2.29 (1.66)	0.025
Fibrosis Stage					<0.001
F0	58 (57.4%)	8 (32.0%)	12 (6.03%)	1 (3.70%)	
F1	15 (14.9%)	6 (24.0%)	103 (51.8%)	8 (29.6%)	
F2	21 (20.8%)	3 (12.0%)	35 (17.6%)	9 (33.3%)	
F3	3 (2.97%)	6 (24.0%)	13 (6.53%)	0 (0.00%)	
F4	4 (3.96%)	2 (8.00%)	36 (18.1%)	9 (33.3%)	

Average values and standard deviations in brackets (). All liver blood metrics are presented in mmol/L. BMI: body mass index, DM: diabetes mellitus, ALT: alanine aminotransferase, GGT: gamma-glutamyl transpeptidase, TE: transient elastography, IQR: interquartile range ES= Spanish cohorts 1&2, UK= England cohort, DK= Danish cohort.

Table 3. Diagnostic yield CTREE model, 5-fold by 5-repeat cross-validation.

F	F0-F1	F2-F3	F4
Prevalence	62.15%	24.92%	12.92%
Sensitivity	79.01%	54.54%	79.55%
Specificity	80.17%	83.56%	91.86%
PPV	85.63%	53.16%	62.50%
NPV	71.85%	84.30%	96.34%
Balanced accuracy	79.59%	69.05%	85.70%
Overall	Average	95% CI L.	95% CI U.
Accuracy	70.79%	68.86%	72.72%
P [Acc. > NIR]		$3.621*10^{-7}$	

PPV: positive predictive value, NPV: negative predictive value, Balanced accuracy: sensitivity + specificity / 2. P [Acc > NIR]: p-value for testing accuracy larger than no information rate.

Table 4. Number needed to screen by risk factor and country

	General			Obese			Diabetic			Alcohol		
	NNS	95% CI		NNS	95% CI		NNS	95% CI		NNS	95% CI	
ES	30.6	25.8	37.5	12.2	10.2	14.1	7.7	6.1	10.5	7.6	6.1	10.1
DE	52.0	21.9	139.6									
UK				2.7	2.2	3.6	3.9	3.1	5.0	5.5	4.5	7.3
DK										2.5	2.1	3.0
FR	29.1	22.6	40.8	10.7	7.6	17.7				21.9	16.4	32.9
HK	26.6	20.0	39.6	8.3	4.0	13.5	9.5	4.9	15.2	25.1	15.0	38.3
Average	34.5	22.6	64.4	8.5	6.0	12.2	7.0	4.7	10.2	12.5	8.8	18.3

ES= Spanish cohorts 1&2, DE= Germany cohort, UK= England cohort, DK= Danish cohort, FR= French cohort, HK= Hong Kong cohort.