Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study

Jean-Pierre Zarski[1,2,*], Nathalie Sturm[2,3], Jérôme Guechot[4], Adeline Paris[5], Elie-Serge Zafrani[6], Tarik Asselah[7], Renée-Claude Boisson[8], Jean-Luc Bosson[5], Dominique Guyader[9], Jean-Charles Renversez[10], Jean-Pierre Bronowicki[11], Marie-Christine Gelineau[12], Albert Tran[13], Candice Trocme[14], Victor De Ledinghen[15], Elisabeth Lasnier[4], Armelle Poujol-Robert[16], Frédéric Ziegler[17], Marc Bourliere[18], Hélène Voitot[19], Dominique Larrey[20], Maria Alessandra Rosenthal-Allieri[21], Isabelle Fouchard Hubert[22], François Bailly[23], Michel Vaubourdolle[4], The ANRS HCEP 23 Fibrostar Group

[1]*Clinique Universitaire d'Hépato-Gastroentérologie – Pôle DIGIDUNE-CHU de Grenoble, France;* [2]*Unité INSERM/UJF U823 IAPC Institut Albert Bonniot-Grenoble, France;* [3]*Département d'Anatomie et Cytologie Pathologiques-Pôle de Biologie, CHU de Grenoble, France;* [4]*Laboratoire de Biochimie A – Hôpital Saint Antoine, AP-HP, Paris, France;* [5]*Centre d'Investigation Clinique INSERM CIC003, CHU Grenoble, Grenoble F-38043, France;* [6]*Département de Pathologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Université Paris Est, Créteil, Val de Marne, France;* [7]*Service d'Hépato-Gastroentérologie, Hôpital Beaujon, AP-HP, Clichy, France;* [8]*Laboratoire de Biochimie, Centre hospitalier Lyon Sud, Hospices Civils de Lyon, France;* [9]*Service d'Hépatologie CHU de Rennes, France;* [10]*Département de Biologie Intégrée, Pôle de Biologie, CHU de Grenoble, France;* [11]*Service d'Hépato-Gastroentérologie, CHU de Nancy, France;* [12]*Laboratoire de Biochimie, Hôtel Dieu, Hospices Civils de Lyon, France;* [13]*Hépatologie, Hôpital l'Archet 2, Nice, France;* [14]*GREPI TIMC-IMAG UMR CNRS 5525, Université de Grenoble/ Laboratoire d'Enzymologie, Dépt. de Biologie et Pathologie de la Cellule, Pôle de Biologie, CHU de Grenoble, France;* [15]*Service d'Hépato-Gastroentérologie, CHU de Bordeaux, France;* [16]*Service d'Hépato-Gastroentérologie, Hôpital St Antoine, AP-HP, Paris, France;* [17]*Laboratoire de Biochimie, Institut de Biologie Clinique, EA4311, IHU, IFR23/CHU, Rouen, France;* [18]*Service d'Hépato-Gastroentérologie, Hôpital St Joseph, Marseille, France;* [19]*Laboratoire de Biochimie, Hôpital Beaujon, AP-HP, Clichy, France;* [20]*Service d'Hépato-Gastroentérologie, CHU de Montpellier, France;* [21]*Laboratoire d'Immunologie, CHU de Nice, France;* [22]*Service d'Hépato-Gastroentérologie, CHU d'Angers, France;* [23]*Service d'Hépato-Gastroentérologie, Hotel Dieu-Lyon, France*

Background & Aims: Blood tests and transient elastography (Fibroscan™) have been developed as alternatives to liver biopsy. This ANRS HCEP-23 study compared the diagnostic accuracy of nine blood tests and transient elastography (Fibroscan™) to assess liver fibrosis, *vs.* liver biopsy, in untreated patients with chronic hepatitis C (CHC).

Methods: This was a multicentre prospective independent study in 19 French University hospitals of consecutive adult patients having simultaneous liver biopsy, biochemical blood tests (performed in a centralized laboratory) and Fibroscan™. Two experienced pathologists independently reviewed the liver biopsies (mean length = 25 ± 8.4 mm). Performance was assessed using ROC curves corrected by Obuchowski's method.

Results: Fibroscan™ was not interpretable in 113 (22%) patients. In the 382 patients having both blood tests and interpretable Fibroscan™, Fibroscan™ performed similarly to the best blood tests for the diagnosis of significant fibrosis and cirrhosis. Obuchowski's measure showed Fibrometer® (0.86), Fibrotest® (0.84), Hepascore® (0.84), and interpretable Fibroscan™ (0.84) to be the most accurate tests. The combination of Fibrotest®, Fibrometer®, or Hepascore® with Fibroscan™ or Apri increases the percentage of well classified patients from 70–73% to 80–83% for significant fibrosis, but for cirrhosis a combination offers no improvement. For the 436 patients having all the blood tests, AUROC's ranged from 0.82 (Fibrometer®) to 0.75 (Hyaluronate) for significant fibrosis, and from 0.89 (Fibrometer® and Hepascore®) to 0.83 (FIB-4) for cirrhosis.

Conclusions: Contrarily to blood tests, performance of Fibroscan™ was reduced due to uninterpretable results. Fibrotest®, interpretable Fibroscan™, Fibrometer®, and Hepascore® perform best and similarly for diagnosis of significant fibrosis and cirrhosis.

Keywords: Chronic hepatitis C; Liver fibrosis; Surrogate markers; Transient elastography; Blood tests.

Received 21 October 2010; received in revised form 13 April 2011; accepted 3 May 2011; available online 23 July 2011

* Corresponding author. Address: Clinique Universitaire d'Hépato-Gastroentérologie – Pôle DIGIDUNE-CHU de Grenoble, BP 217, 38043 Grenoble Cedex 9, France. Tel.: +33 476 76 54 41; fax: +33 476 76 51 79.

E-mail address: JPZarski@chu-grenoble.fr (J.-P. Zarski).

Abbreviations: CHC, chronic hepatitis C; ROC, receiver operating characteristic curves; AUROC, area under receiver operating curve; HCV, Hepatitis C virus; LSM, liver stiffness measurement; LB, liver biopsy; BMI, body mass index; NPV, negative predictive value; PPV, positive predictive value; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma glutamyltranspeptidase.

Research Article

Introduction

Liver biopsy is the method of reference to assess the fibrosis stage in chronic hepatitis C (CHC). However, it is an invasive procedure with severe complications in about 0.5% of cases [1] and its accuracy is limited by sampling heterogeneity [2] and inter-observer and intra-observer variation [3,4]. Biopsy specimens less than 15 mm in length appear poorly reliable [3]. Semi-quantitative evaluation of fibrosis has high variability especially among non-expert pathologists [4,5]. Several blood tests with or without scores calculated from statistical models have been developed to evaluate fibrosis. Hyaluronate was proposed as a non-invasive marker [6]. Fibrotest® was the first score combining several variables proposed for patients with CHC [7]. Apri [8], Fibrometer® [9], and Hepascore® [10] were then validated in these patients. Other fibrosis scores have been recently proposed but are not often performed in practice, FIB-4 [11], Forns's score [12], MP3 [13,14], and the European Liver Fibrosis Group or ELFG score [15]. However, all these tests have limitations. Blood test results can be influenced by other associated diseases, comorbidities or different dosage techniques.

Another alternative, transient elastography (Fibroscan™; Echosens, Paris, France) is based on liver stiffness measurement. Its diagnostic performance is similar to that of serological markers [16–20]. However Fibroscan™ has some limitations (failure and unreliability) particularly in obese patients or in circumstances of limited operator experience, as recently discussed by Castera *et al.* [21].

The aim of this study was to perform a prospective independent multicentre comparative evaluation of most of the currently best evaluated non-invasive markers i.e. blood tests and transient elastography, *vs.* liver biopsy in an etiologically homogenous study group (CHC), with an appropriate number of patients, appropriate histological analysis and using well standardized biological tests.

Patients and methods

Patients

Consecutive adult patients with chronic hepatitis C were prospectively considered for inclusion if they were naïve of treatment or had no treatment during the last 6 months, interpretable liver biopsy with delay between biopsy and blood tests of <3 months. All patients had been referred for tests in order to make a decision on treatment strategy. CHC was confirmed by HCV–RNA polymerase chain reaction analysis of serum. Cirrhotic patients were compensated and asymptomatic at the time of inclusion. Patients with co-existing liver diseases attributed to alcohol, hepatitis B, auto-immune hepatitis, primary biliary cirrhosis, hemochromatosis, alpha-1-antitrypsine deficiency, or Wilson's disease were excluded by history and clinical, laboratory, imaging, and histological data. Human immunodeficiency virus co-infected and post-transplant patients were also excluded. The protocol was approved by the ethics committee "CPP Sud-Est 5". All patients gave written informed consent. Liver biopsies were performed as part of normal clinical care for staging and grading of liver disease before antiviral treatment. Demographic data were recorded at the time of the liver biopsy.

Biological scores of liver fibrosis

Fasting blood samples were collected by venipuncture. The same batches of tubes were used for all patients (BD Vacutainer®, type 9NC, K2E, and Z, Becton–Dickinson, Plymouth, UK).

Cholesterol, platelet count, and prothrombin time were immediately measured in each centre. All other biological parameters were measured in a centralized laboratory using serum samples immediately fractioned into 0.5 ml fractions in 1.5 ml screw cap micro tubes (Sarstedt, Nümbrecht, Germany), then frozen and stored at −80 °C until assayed. Samples were transported in dry-ice by a specialized transporter (AreaTime Logistics, Cergy Pontoise, France). All the tests were performed blind of clinical and histological data.

The following blood tests were evaluated: Fibrotest®, Fibrometer®, Forns score, Apri, MP3, ELFG, Hepascore®, FIB-4, Hyaluronate. Blood test scores were calculated according to the most recent published formulae [8,10–15], or patent for Fibrotest® [7] and Hepascore® [10], or by the courtesy of the manufacturer (BioLivescale) for Fibrometer® [9]. The list of variables included in each test and the measurement techniques are detailed in the Supplementary data.

Liver stiffness measurement by transient elastography (Fibroscan™)

Measurements were made on the right lobe of the liver, through the intercostal spaces with the patient lying in dorsal decubitus with the right arm in maximal abduction by the operator who performed the liver biopsy. The tip of the transducer probe was covered with coupling gel and placed on the skin, between two ribs at the level of the right lobe. Liver stiffness measurement (Fibroscan™) failure was defined as zero valid shots (after at least 10 attempts) and "unreliable examinations" were defined as fewer than 10 valid shots or an interquartile range (IQR)/LSM greater than 30% or a success rate less than 60% [16–19].

Liver biopsy

Liver biopsies (LB) were performed using Menghini's technique with a 1.6 mm needle (Hepafix, Brown, Melsungen, Germany), formalin-fixed in the centres and paraffin embedded. Sections (4 mm) were stained with hematoxylin-eosin-saffron, and picrosirius red. The liver fibrosis stage was evaluated according to the METAVIR scoring system [5], independently by two senior liver pathologists (NS, ESZ) blind to clinical and biological data. In cases of disagreement, slides were simultaneously reviewed using a multi-pipe microscope to reach a consensus. Inter-observer agreement was evaluated using the kappa index, called κ, which excludes chance-expected agreement and the weighted κ index according to a linear evolution of the METAVIR score [4]. The length of biopsy and the number of portal tracts were recorded. To be considered for scoring, LB less than 20 mm had to measure at least 15 mm and/or contain at least 11 portal tracts, except for cirrhosis.

Statistical analysis

Due to the inherent difficulty in the interpretability of Fibroscan™ we defined two populations, the first including patients with all the available blood tests (436 patients), and the second population including patients having both interpretable Fibroscan™ (excluding cases in which Fibroscan™ was not possible, failures and unreliable tests) and all blood tests (382 patients).

Descriptive results were expressed as the mean ± standard deviation or as the number (percentage) of patients. The diagnostic performance of the non-invasive methods was assessed using AUROCs, considering liver biopsy as a "gold standard", albeit imperfect, and its 95% confidence intervals. We used cut-offs corresponding to the score associated with p <0.05 in the corresponding logistic regression model. Comparison of AUROCS was performed using a Chi^2 test associated with the procedure of "ROCGOLD" (Stata™). Due to the multiple comparisons between scores, the method of Sidak was used to exclude the risk of concluding wrongly, with an alpha risk of $p_{(Sidak)} \leqslant 0.05$ for statistical significance.

Since AUROC assumes a binary gold standard while histological fibrosis staging is based on an ordinal scale we used another estimator of diagnostic test accuracy which does not require dichotomization of the gold standard. The Obuchowski measure [22], was recently recommended as a multinomial version of the AUC. With N (= 5) categories of the gold standard outcome and AUCst, it estimates the AUC of diagnostic tests differentiating between categories s and t. The Obuchowski measure is a weighted average of the $N(N-1)/2$ (= 10) different AUCst corresponding to all the pair-wise comparisons between two of the N categories. All these paired comparisons are also weighted using a penalty function proportional to the difference in METAVIR units. In our study the penalty function was 1 for each different METAVIR unit. As proposed by Lambert *et al.* [23] we thus defined a penalty function proportional to the difference in METAVIR units between stages (the penalty function was 0.25 when the difference was 1, 0.5 when the difference was 2, 0.75 when the difference was 3, and 1 when the difference was 4).

We combined the main tests pair-wise, calculating the % of concordant well classified patients given by the tests and the number of avoided biopsies (assuming biopsy to be the gold standard).