Clinics and Research in Hepatology and Gastroenterology (2011) 35, 720—730

Available online at ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

ORIGINAL ARTICLE

FibroTest® and Fibroscan® performances revisited in patients with chronic hepatitis C. Impact of the spectrum effect and the applicability rate

Thierry Poynard[a,*,1], Victor de Ledinghen[b,1], Jean-Pierre Zarski[c,1], Carol Stanciu[d,1], Mona Munteanu[e,1], Julien Vergniol[b,1], Julie France[f,1], Anca Trifan[d,1], Joseph Moussalli[a,1], Pascal Lebray[a,1], Dominique Thabut[a,1], Vlad Ratziu[a,1]

[a] *APHP UPMC Liver Center, 75013 Paris, France*
[b] *Service d'hépato-gastroentérologie, CHU de Bordeaux, 33000 Bordeaux, France*
[c] *ANRS HCEP 23 FIBROSTAR study, clinique universitaire d'hépato-gastroentérologie, 38000 Grenoble, France*
[d] *Gastroenterology and Hepatology University of Medicine and Pharmacy ''Gr. T. Popa'', Iasi, Romania*
[e] *Biopredictive, 75007 Paris, France*
[f] *Inserm CIC003, centre d'investigation clinique, CHU de Grenoble, 38000 Grenoble, France*

Available online 17 August 2011

Summary
Background: Two widely used biomarkers of fibrosis, FibroTest® and liver stiffness measurement (LSM), have been mostly validated in patients with chronic hepatitis C (CHC) using the standard area under the ROC curve (sAUROC) which is not the most appropriate method due to the risk of fibrosis spectrum effect. Furthermore the performance of these biomarkers have not been assessed in ''intention to diagnose'' which takes into account the failures and non-reliable results.
Aim: The aim was to compare the accuracy of FibroTest® and LSM for the diagnosis of fibrosis using sAUROC, the pairwise comparison of fibrosis stages by Obuchowski measure (wAUROC), and these AUROCs reassessed after taking into account the applicability rates.
Methods: One thousand two hundred and eighty-nine patients with CHC and 604 healthy volunteers were analyzed. The performances of biomarkers assessed were compared in a patients-only group (P1: $n = 1289$), in a population combining both patients and healthy volunteers (P2: $n = 1893$) and in a simulated population (P3: $n = 1893$) with the prevalence of stages observed in a reference population, to demonstrate the impact of spectrum effect. Applicability rates were estimated prospectively in 24,872 consecutive FibroTest® and in 13,669 consecutive LSM examinations.

* Corresponding author. Service d'hépatologie, hôpital La-Pitié-Salpétrière, 47-83, boulevard de l'Hôpital, 75651 Paris cedex 13, France. Tel.: +33 1 42 16 10 22.
E-mail address: tpoynard@teaser.fr (T. Poynard).

[1] **Authors' involvements.** TP: study concept and design, analysis and interpretation of data, drafting; statistical analysis; study supervision. VdL, JPZ, CS, JV, JF, AT, VR: acquisition of data; VR, MM, JM, PL, DT: critical revision of the manuscript.

2210-7401/$ – see front matter
doi:10.1016/j.clinre.2011.07.003

Results: Using wAUROC, the conclusions of studies with reliable results in P1 were different than in those of P2 and in P3. There was a lower performance of FibroTest® versus LSM in P1 (0.864 [0.855—0.873] vs. 0.883 [0.874—0.892]; $P = 0.002$) which was not found in P2 (0.893 [0.887—0.900] vs. 0.894 [0.887—0.901]; $P = 0.86$) and in P3 (0.899 [0.893—0.905] vs 0.902 [0.895—0.909]; $P = 0.60$). Using the sAUROC, in P1, P2 and P3, there was no significant difference between FibroTest® and LSM performance for advanced fibrosis and a lower performance of FibroTest® versus LSM for cirrhosis. In intention to diagnose, using wAUROCs performances were higher for FibroTest® vs. LSM in P1 (0.857 [0.848—0.866] vs. 0.814 [0.807—0.821]; $P < 0.0001$) and P2 (0.885 [0.879—0.892] vs. 0.743 [0.737—0.749]; $P < 0.0001$), without difference in P3 (0.891 [0.885—0.897] vs. 0.894 [0.887—0.901]; $P = 0.90$). Using sAUROC, the significant differences in favor of FibroTest® vs LSM persisted also for the diagnosis of advanced fibrosis, both in P1 and P2 ($P < 0.0001$) and for the diagnosis of cirrhosis in P1 ($P < 0.001$).
Conclusion: When the spectrum effects and applicability rates were taken into account, LSM had lower performance results than FibroTest® for the diagnosis of fibrosis stages.

Background

There is a major debate surrounding the efforts to find the best means of evaluating and managing the increasing numbers of patients with chronic liver disease [1—4]. Liver biopsy, due to its risks and limitations, is no longer considered mandatory as the first-line indicator of liver injury, and several markers have been developed as non-invasive alternatives [1—4]. Among patients with chronic viral hepatitis, the assessment of liver fibrosis by two validated noninvasive techniques, biomarkers by FibroTest® (FT) (Biopredictive, Paris, France [5,6]) and liver stiffness measurements (LSM) by Fibroscan® (Echosens, Paris, France [6,7]), is now widely done in countries where these techniques are available and approved [8,9].

Together with the development of new biomarkers, several advances have also been made in the methods for a better comparison of their performances. Besides the absence of a true gold (reference) standard [10], two main sources of variability have been identified which must be taken into account: the spectrum effect [11—13] and the applicability rate [11,14].

Biomarkers of fibrosis have been mostly validated using the standard area under the ROC curve (sAUROC) [5,7], which is no longer the most appropriate method [11—13]. Fibrosis staging accuracy measure designed for ordinal gold standards (such as Obuchowski measure [wAUROC]) is now recommended for assessing the diagnostic accuracy of non-invasive biomarkers of fibrosis to prevent the spectrum effect [12,13].

The applicability of biomarkers, which combines the failure rate and the reliability rate, directly impacted their performances [11—14], but has never been systematically taken into account in an ''intention-to-diagnose'' analysis as was done in the ''intention to treat'' for therapeutic trials. Therefore an ideal overview of FT and Fibroscan® performances would combine individual participant data from all studies after the exclusion of duplicate data [15—17], using appropriate statistical methods and taking into account the applicability rates.

While awaiting this ideal exhaustive overview, we analyzed individual participant data from 3 populations of patients with chronic hepatitis C (CHC), which came from three groups that were independent of the inventor of biomarkers. Performances that took into account the spectrum effect (wAUROC) and applicability rate were compared to performances that used sAUROCs.

Methods

Endpoints

The main goal of the study was to measure the impact of the spectrum effect and the applicability rate on the estimates of biomarker performance.

The ''standard'' method is to assess two sAUROCs with binary gold standards: one with stages of advanced fibrosis that are usually aggregated (stages F2, F3 and F4) versus non-advanced fibrosis stages (stages F0 and F1), and the other cirrhosis (stage F4) versus non-cirrhosis stages (F0, F1, F2 and F3) in patients with reliable results. This method does not take into account the spectrum effect and the applicability rate.

Two methods have been used for consideration of the spectrum effect: the Obuchowski measure (wAUROC) for the diagnosis of all pairwise stage comparisons, and the standard sAUROC between each pairwise adjacent stage. These estimates were multiplied by the applicability rate for each biomarker for the intention-to-diagnose analysis.

Alanine aminotransferase serum activity (ALT) was used as the ''control'' biomarker for the first-line function liver test without specificity for liver fibrosis staging.

Patients

The database included 1893 subjects: three groups of patients with CHC ($n = 1289$, population P1) that were collected prospectively, and one population of apparently healthy volunteers ($n = 604$). HCV patients belonged to one