



HAL
open science

MÉTHODOLOGIE POUR L'AMÉLIORATION DE LA PERFORMANCE DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HÉPATIQUE AU COURS DE L'HÉPATITE CHRONIQUE C

Jérôme Boursier

► **To cite this version:**

Jérôme Boursier. MÉTHODOLOGIE POUR L'AMÉLIORATION DE LA PERFORMANCE DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HÉPATIQUE AU COURS DE L'HÉPATITE CHRONIQUE C. Biochimie, Biologie Moléculaire. Université d'Angers, 2011. Français. NNT: . tel-00982618

HAL Id: tel-00982618

<https://theses.hal.science/tel-00982618v1>

Submitted on 24 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**MÉTHODOLOGIE POUR L'AMÉLIORATION DE LA
PERFORMANCE DU DIAGNOSTIC NON INVASIF DE LA FIBROSE
HÉPATIQUE AU COURS DE L'HÉPATITE CHRONIQUE C**

THÈSE DE DOCTORAT

**Spécialité : Biochimie
ÉCOLE DOCTORALE BIOLOGIE et SANTE**

Présentée et soutenue publiquement

le 17 novembre 2011

à Angers

par **Jérôme BOURSIER**

Devant le jury ci-dessous :

Dominique GUYADER (Président), Professeur, Université de Rennes

Victor de LEDINGHEN (Rapporteur), Professeur, Université de Bordeaux

Vincent LEROY (Rapporteur), Professeur, Université de Grenoble

Fabrice CARRAT (Examineur), Professeur, Université Pierre et Marie Curie, Paris VI

Directeur de Thèse :

Paul CALÈS, Professeur, Université d'Angers

Nom et coordonnées du Laboratoire :

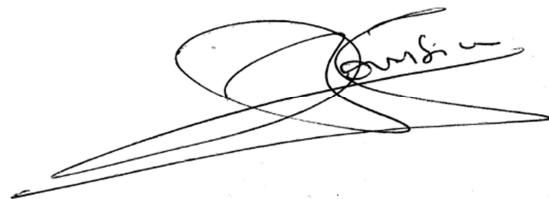
Laboratoire HIFIH, UPRES 3859, IFR 132, Université d'Angers et PRES UNAM
UFR Sciences Médicales, Pavillon Ollivier, rue Haute de Reculée, 49045 Angers Cedex
Tel : 02 41 35 34 10 ; Fax : 02 41 35 41 19

ENGAGEMENT DE NON PLAGIAT

Je, soussigné (e) Jérôme BOURSIER,

déclare être pleinement conscient(e) que le plagiat de documents ou d'une partie d'un document publiés sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

Signature :



COMPOSITION DU JURY

PRESIDENT :

Monsieur le Professeur Dominique GUYADER

DIRECTEUR DE THESE :

Monsieur le Professeur Paul CALÈS

RAPPORTEURS :

Monsieur le Professeur Victor de LEDINGHEN

Monsieur le Professeur Vincent LEROY

EXAMINATEURS :

Monsieur le Professeur Dominique GUYADER

Monsieur le Professeur Fabrice CARRAT

REMERCIEMENTS

à **MONSIEUR LE PROFESSEUR CALES**

Professeur d'Hépatogastroentérologie à la Faculté de Médecine d'Angers

**Directeur du Laboratoire HIFIH, UPRES 3859, IFR 132, Université d'Angers et
PRES UNAM**

Vous avez accepté de diriger ce travail de Thèse et m'avez accueilli au sein de votre laboratoire HIFIH.

Vous avez su dès le début de mon internat susciter puis aiguïser mon intérêt pour l'hépatologie et pour la recherche dans ce domaine. Au cours de ces 10 dernières années, j'ai pu apprécier et profiter de vos larges connaissances ainsi que de votre rigueur scientifique. Vous m'avez donné l'occasion et les moyens de m'épanouir au sein de votre équipe.

Soyez assuré de toute de ma reconnaissance ainsi que de mon profond respect.

à **MONSIEUR LE PROFESSEUR GUYADER**

Professeur d'Hépatogastroentérologie à la Faculté de Médecine de Rennes

Vous m'avez fait l'honneur d'accepter de président le jury de ma Thèse.

Au cours de nos divers échanges, en congrès ou réunions scientifiques, j'ai apprécié votre disponibilité et la pertinence de votre réflexion scientifique.

Soyez assuré de tout mon respect.

à **MONSIEUR LE PROFESSEUR DE LEDINGHEN**

Professeur d'Hépatogastroentérologie à la Faculté de Médecine de Bordeaux

Vous m'avez fait l'honneur d'accepter de siéger au jury de ma Thèse et de juger ce travail.

Nos diverses rencontres m'ont permis d'apprécier votre gentillesse et vos qualités humaines. La qualité de votre réflexion scientifique, les compétences de votre équipe, et votre volonté d'initier des travaux collaboratifs nous ont permis déjà de produire des travaux de qualité.

Soyez assuré de tout mon respect.

à **MONSIEUR LE PROFESSEUR LEROY**

Professeur d'Hépatogastroentérologie à la Faculté de Médecine de Grenoble

Vous m'avez fait l'honneur d'accepter de siéger au jury de ma Thèse et de juger ce travail.

J'ai apprécié lors des congrès votre disponibilité ainsi que la qualité de vos travaux et de vos présentations scientifiques. Les compétences de votre centre et de votre équipe nous ont permis de produire des travaux communs de qualité.

Soyez assuré de tout mon respect.

à **MONSIEUR LE PROFESSEUR CARRAT**

Professeur de Santé Publique à l'Université Pierre et Marie Curie, Paris VI

Vous m'avez fait l'honneur d'accepter de siéger au jury de ma Thèse.

J'ai apprécié lors de diverses réunions la pertinence de votre réflexion scientifique et de vos présentations.

Soyez assuré ici de tout mon respect.

à **Arnaud et Emmanuel**

Nous avons traversé ensemble les deux premiers cycles des études médicales à Lille et nos chemins se sont séparés lors de l'internat. A notre amitié, malgré les kilomètres qui nous séparent.

à **Frédéric et Isabelle, médecins hépatologues au CHU d'Angers.**

Vos grandes compétences cliniques, votre soutien quotidien, et votre amitié m'ont été d'une aide la plus précieuse. Soyez assurés de mon respect, de ma reconnaissance, et de toute mon amitié.

à **Nina, Elodie, Bruno, Dominique, François-Xavier, hépato-gastroentérologues au CHU d'Angers.**

Je mesure toute la chance de pouvoir travailler avec une équipe soudée, compétente et enthousiaste. Soyez assurés de toute mon amitié.

à **Gwenaëlle, Sandra, Sandrine, Pascal, Marc, attachés de recherche clinique au CHU d'Angers**

Travailler quotidiennement avec vous est des plus enthousiasmants. Votre disponibilité, vos compétences, et surtout votre bonne humeur nous sont précieux à tous dans le service.

à **tous mes amis** : Matthieu et Caroline, Nina, Olivier, Nadia et Benoit, Marie, Benjamin et Maïlin, Hugues et Dorothée, Jean et Caroline, Cyril et Séverine, Gérald et Raïssa, Guillaume et Anne-Florence.

A nos mémorables souvenirs, et à ceux à venir...

à **mes parents et toute ma famille**

Vous m'avez donné les moyens d'atteindre mes objectifs, qu'ils soient professionnels ou personnels. Je m'efforcerai de suivre votre exemple pour mes propres enfants.

à **ma belle-famille**

Vous avez accueilli le ch'ti qui a débarqué en Anjou avec chaleur et convivialité. Comme mes propres parents, vous serez un exemple pour mes enfants.

à **Hugo et Jules,**

Mes deux p'tits bouts qui font de moi le plus heureux et le plus fier des papas.

à **Anne-Hélène,**

Mention spéciale : sans toi rien n'aurait été pareil...

ABBREVIATIONS

AUROC : aire sous la courbe ROC

CHC : carcinome hépatocellulaire

HAS : Haute Autorité de Santé française

PBH : ponction biopsie hépatique

RLB : régression logistique binaire

VHC : virus de l'hépatite C

VPN : valeur prédictive négative

VPP : valeur prédictive positive

PLAN

1. INTRODUCTION	12
1.1 EPIDEMIOLOGIE DE L'HEPATITE CHRONIQUE C	12
1.2 HISTOIRE NATURELLE DE L'HEPATITE CHRONIQUE C.....	12
<i>1.2.1 Progression de la fibrose hépatique au cours de l'hépatite chronique C</i>	<i>13</i>
<i>1.2.2 Incidence de la cirrhose au cours de l'hépatite chronique C.....</i>	<i>14</i>
<i>1.2.3 Complications de l'hépatite chronique C</i>	<i>14</i>
1.3 OBJECTIFS DE LA PRISE EN CHARGE DE L'HEPATITE CHRONIQUE C – RATIONNEL DE LA THESE D'UNIVERSITE.....	15
2. METHODES D'EVALUATION DE LA FIBROSE HEPATIQUE	17
2.1 PONCTION-BIOPSIE HEPATIQUE	17
<i>2.1.1 Détermination du stade histologique de fibrose.....</i>	<i>17</i>
<i>2.1.2 Limites des scores histologiques de la fibrose hépatique.....</i>	<i>17</i>
2.2 TESTS SANGUINS DE FIBROSE	19
<i>2.2.1 Méthode d'élaboration des tests sanguins de fibrose.....</i>	<i>19</i>
<i>2.2.2 Performance diagnostique des tests sanguins de fibrose.....</i>	<i>21</i>
<i>2.2.3 Classement de fibrose du Fibrotest et du FibroMètre.....</i>	<i>22</i>
<i>2.2.4 Limites des tests sanguins de fibrose</i>	<i>24</i>
2.3 FIBROSCAN	25
<i>2.3.1 Utilisation en pratique du Fibroscan.....</i>	<i>26</i>
<i>2.3.2 Performance diagnostique du Fibroscan.....</i>	<i>28</i>
<i>2.3.3 Compteur de fibrose du Fibroscan</i>	<i>29</i>
<i>2.3.4 Limites du Fibroscan.....</i>	<i>30</i>
2.4 RECOMMANDATIONS DE LA HAUTE AUTORITE DE SANTE FRANÇAISE	30
2.5 NOUVELLES TECHNIQUES DE MESURE NON INVASIVE DE LA FIBROSE HEPATIQUE ..	31

3.	ASPECTS METHODOLOGIQUES DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HEPATIQUE.....	33
3.1	BIAIS DES ETUDES AYANT EVALUE LA PERFORMANCE DIAGNOSTIQUE DES TESTS NON INVASIFS DE FIBROSE	33
3.1.1	<i>Représentativité des populations où sont créés et/ou évalués les tests non invasifs de fibrose</i>	<i>33</i>
3.1.2	<i>Impact de l'absence de gold standard sur l'évaluation de la performance des tests non invasifs de fibrose.....</i>	<i>36</i>
3.1.3	<i>Discordance entre les tests non invasifs de fibrose</i>	<i>38</i>
3.2	INDICES DIAGNOSTIQUES POUR L'EVALUATION DES TESTS NON INVASIFS DE FIBROSE	38
3.2.1	<i>AUROC</i>	<i>38</i>
3.2.2	<i>Diagnostic binaire de fibrose</i>	<i>40</i>
3.2.3	<i>Profil de performance</i>	<i>46</i>
3.2.4	<i>Associations de tests non invasifs</i>	<i>47</i>
3.2.5	<i>Optimisation des valeurs prédictives.....</i>	<i>53</i>
3.2.6	<i>Intervalles de diagnostic fiable</i>	<i>55</i>
3.3	REPRODUCTIBILITE ET ROBUSTESSE DES TESTS NON INVASIFS DE FIBROSE	57
3.3.1	<i>Reproductibilité</i>	<i>58</i>
3.3.2	<i>Robustesse.....</i>	<i>59</i>
4.	TRAVAUX PERSONNELS	61
4.1	PRESENTATION DU TRAVAIL DE THESE.....	61
4.2	PATIENTS ET METHODES	62
4.2.1	<i>Patients.....</i>	<i>62</i>
4.2.2	<i>Histologie.....</i>	<i>64</i>
4.2.3	<i>Tests sanguins de fibrose</i>	<i>65</i>
4.2.4	<i>Mesure de la dureté hépatique avec le Fibroscan.....</i>	<i>65</i>
4.3	ETUDE SNIFF 17	66
4.3.1	<i>Présentation de l'étude.....</i>	<i>66</i>
4.3.2	<i>Discussion des résultats.....</i>	<i>66</i>
4.4	ETUDE VINDIAG 6	70
4.4.1	<i>Présentation de l'étude.....</i>	<i>70</i>

4.4.2	<i>Discussion des résultats</i>	70
4.5	ETUDE VINDIAG 7	73
4.5.1	<i>Présentation de l'étude</i>	73
4.5.2	<i>Discussion des résultats</i>	74
5.	DISCUSSION GENERALE	79
5.1	AMELIORATION DES RECOMMANDATIONS DE LA HAS	79
5.2	QUEL EST L'AVENIR DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HEPATIQUE ?.....	81
6.	CONCLUSION	85
7.	REFERENCES	87
8.	ANNEXES	100

1. INTRODUCTION

1.1 Epidémiologie de l'hépatite chronique C

L'hépatite C est un problème de Santé Publique majeur. On estime que 180 millions de personnes dans le monde sont des porteurs chroniques du virus de l'hépatite C (VHC) et que 3 à 4 millions de personnes sont nouvellement infectées chaque année (1).

Aux Etats-Unis, la prévalence de l'hépatite chronique C entre 1999 et 2002 était de 1,3%, soit 3,2 millions de personnes (2). L'augmentation de l'incidence du carcinome hépatocellulaire (CHC) observée aux Etats-Unis est liée à l'augmentation de l'incidence de l'hépatite chronique C (3). L'hépatite C est la principale cause de décès par maladie du foie (4) et la première indication de transplantation hépatique aux Etats-Unis (5). Enfin, certaines projections estiment que la mortalité liée à l'infection par le VHC (décès par insuffisance hépatique ou CHC) va continuer à augmenter aux Etats-Unis au cours des deux prochaines décennies pour culminer à 13 000 décès par an en 2030 (6).

En France, on estime que 232 000 personnes sont atteintes d'hépatite chronique C (7). On enregistre entre 2 700 et 4 400 nouvelles contaminations et 2 600 décès par an.

1.2 Histoire naturelle de l'hépatite chronique C

L'agression chronique du foie par le VHC provoque l'accumulation de fibrose dans le parenchyme hépatique, l'évolution vers une cirrhose, et enfin la survenue des complications impliquant le pronostic vital des patients (insuffisance hépatocellulaire, hémorragie digestive par rupture de varice, CHC, syndrome hépatorénal).

1.2.1 Progression de la fibrose hépatique au cours de l'hépatite chronique C

Une méta-analyse récente a repris les données de 111 études sur l'histoire naturelle de l'hépatite chronique C (8). La modélisation réalisée sur 33 212 patients a montré que la progression de la fibrose au cours de l'hépatite chronique C n'était pas linéaire mais exponentielle. Les facteurs influençant la progression de la fibrose sont : l'âge lors de la contamination, la durée de l'infection, le sexe, la co-infection VIH ou VHB, l'éthylisme chronique, l'insulino-résistance, l'activité d'ALAT, et le degré de fibrose hépatique (9).

Les patients ayant une fibrose hépatique minimale ont un bon pronostic. Une étude récente a ainsi montré chez 282 patients avec hépatite chronique C et des lésions de fibrose hépatique minimales (Ishak 0 ou 1) que le taux de progression vers la fibrose sévère (Ishak ≥ 4) était de seulement 5% après un suivi médian de 52,5 mois (10). Dans cette même étude, une revue de la littérature des travaux similaires montrait des taux de progression vers la fibrose sévère de 9,8%, 16,0%, et 27,4% après des suivis médians respectifs de 6,3 (11), 5,0 (12), et 7,8 années (12).

L'étude américaine multicentrique HALT-C donne l'opportunité d'évaluer l'histoire naturelle du VHC dans une large cohorte de patients ayant des lésions de fibrose significative. Cette étude américaine multicentrique a inclus 1050 patients ayant une hépatite chronique C, des lésions de fibrose hépatique significative (Ishak ≥ 2), et non-répondeurs ou rechuteurs après une bithérapie antivirale par interféron pégylé et ribavirine (13). La comparaison du groupe « traitement de maintenance » par interféron pégylé à faible dose au groupe « placebo » a montré que le traitement de maintenance ne réduisait pas la progression de la maladie hépatique. Une étude ancillaire de HALT-C a montré que le degré de fibrose hépatique était bien corrélé au pronostic des patients (14). En effet, l'incidence cumulée à 6 ans des complications hépatiques augmentait significativement avec le stade Ishak de fibrose déterminé à l'inclusion (stade 2 à l'inclusion : 5,6%, stade 3 : 16,1%, stade 4 : 19,3%, stade 5 : 37,8%, et stade 6 : 49,3%). De même, l'incidence cumulée à 6 ans du décès ou transplantation hépatique augmentait également significativement (stade 2 à l'inclusion : 4,3%, stade 3 : 8,1%, stade 4 : 10,2%, stade 5 : 21,3%, et stade 6 : 27,8%).

1.2.2 Incidence de la cirrhose au cours de l'hépatite chronique C

La prévalence de la cirrhose après 20 ans d'infection par le VHC varie de 5% à 25% selon les études (9, 15). Dans la modélisation réalisée par Thein *et al* chez plus de 33 000 patients, la prévalence de la cirrhose était estimée à 16% à 20 ans et 45% à 30 ans (8). Ces résultats étaient variables en fonction de la population étudiée et du design des études. Dans l'étude HALT-C, 30% des patients non cirrhotiques (Ishak 2-4) à l'inclusion développaient une cirrhose au cours des 3,5 ans de suivi (13).

1.2.3 Complications de l'hépatite chronique C

Deux études, dont une méta-analyse, ont montré que les patients avec une hépatite chronique C avaient 20 (16) à 24 (17) fois plus de risque de développer un CHC que les patients témoins (qui étaient, selon les études, des patients hospitalisés sans hépatopathie chronique ou la population générale). Dans l'étude HALT-C, l'incidence du CHC était de 1,0% par an et les prévalences à 3 et 5 ans respectivement de 2,0 et 5,0% (18). Dans cette étude, le CHC était plus fréquent chez les patients cirrhotiques à l'inclusion (incidence : 1,4% par an, prévalence à 3 et 5 ans : 2,6% et 7,0%) que chez les patients ayant une fibrose significative (incidence CHC : 0,8% par an, prévalence à 3 et 5 ans : 1,4% et 4,1%). Dans une récente méta-analyse incluant 13 études, Alazawi *et al* ont montré une incidence du CHC de 3,4% par an chez 2386 patients ayant une cirrhose virale C compensée (19).

Dans l'étude HALT-C, 26% des patients n'ayant pas de varice à l'inclusion en ont développé au cours des 4 ans de suivi (20). Parmi eux, 24% avaient des grosses varices à risque hémorragique relevant d'une prophylaxie primaire par bêta-bloquant ou ligature endoscopique. Enfin, après 4 ans de suivi dans cette étude, le risque de rupture de varice œsophagienne était estimé à 0,7% chez les patients sans varice à l'admission et 7,0% chez les patients ayant des varices à l'inclusion. Dans une autre étude réalisée chez 218 patients ayant une cirrhose virale C compensée, la prévalence des varices œsophagiennes était estimée à 12%, 15%, 25%, 35%, et 65% après respectivement 3, 6, 9, 12, et 15 ans de suivi (21).

Dans la méta-analyse d'Alazawi *et al*, l'incidence des décompensations des cirrhoses virales C jusqu'alors compensées était évaluée à 6,4% par an, et celle des décès ou transplantations hépatiques à 4,6% par an (19).

1.3 Objectifs de la prise en charge de l'hépatite chronique C – Rationnel de la Thèse d'Université

L'objectif principal de la prise en charge des patients ayant une hépatite chronique virale C est d'éviter l'évolution vers la cirrhose et les complications engageant le pronostic vital.

L'indication du traitement antiviral dépend du degré de fibrose hépatique (15). Chez les patients ayant une fibrose minime, le traitement antiviral n'est pas formellement indiqué car le pronostic vital du patient n'est pas modifié. Une surveillance peut donc suffire, mais le traitement peut être discuté en fonction de la motivation du patient, des comorbidités et des chances de guérison. Le traitement antiviral est, par contre, indiqué chez les patients ayant une fibrose modérée à sévère ou une cirrhose : chez les patients avec une fibrose modérée ou sévère, le but est d'éradiquer le virus pour éviter la constitution d'une cirrhose ; chez les patients cirrhotiques, l'objectif est de stabiliser la maladie hépatique et d'éviter une décompensation de la cirrhose.

Il est également important de diagnostiquer la cirrhose car à ce stade il est nécessaire de débiter un dépistage du carcinome hépatocellulaire et des grosses varices œsophagiennes.

En effet, 95% des CHC se développent dans un foie cirrhotique (22). Au stade précoce, le CHC est asymptomatique. Lorsque les symptômes cliniques apparaissent, la tumeur est généralement trop étendue pour pouvoir envisager un traitement curatif (transplantation hépatique, chirurgie, traitements percutanés). C'est pourquoi, il est recommandé de réaliser un dépistage semestriel du CHC chez le patient cirrhotique afin de diagnostiquer le CHC à un stade précoce où le traitement curatif est envisageable (22).

Les grosses varices œsophagiennes sont à risque hémorragique et nécessitent un traitement prophylactique afin d'éviter leur rupture et une hémorragie digestive (23, 24). Plusieurs études ont démontré qu'une prophylaxie primaire par propranolol ou ligature

endoscopique améliorerait la survie des patients cirrhotiques avec de grosses varices œsophagiennes (25, 26).

En conclusion, il est nécessaire d'évaluer précisément le degré de fibrose hépatique chez les patients ayant une hépatite virale C chronique car il influence le pronostic des patients et guide leur prise en charge (15).

2. METHODES D'EVALUATION DE LA FIBROSE HEPATIQUE

2.1 Ponction-biopsie hépatique

2.1.1 Détermination du stade histologique de fibrose

L'examen histologique d'une ponction-biopsie hépatique (PBH) est encore considéré comme la méthode de référence pour l'évaluation du degré de fibrose hépatique chez les patients ayant une hépatite chronique C (15). Il s'agit d'un examen invasif réalisé le plus souvent par voie percutanée transcostale, sous anesthésie locale. L'anatomopathologiste évalue le degré de fibrose du foie à l'aide de scores semi-quantitatifs qui décrivent les lésions de fibrose selon 5 à 7 stades (15). Les scores semi-quantitatifs de fibrose les plus utilisés sont : le score Metavir (27) et le score d'Ishak (28), le score de Batts-Ludwig (29), et le score IASL (30). Une étude récente a montré que le score Ishak avait une bonne valeur pronostique au cours de l'hépatite chronique C (14). En France, le score Metavir est le plus utilisé. Il comprend 5 stades : absence de fibrose (F0), fibrose portale stellaire sans septa (F1 : fibrose minimale), fibrose portale avec quelques septa (F2 : fibrose significative), fibrose portale avec nombreux septa sans cirrhose (F3 : fibrose sévère) et cirrhose (F4).

Outre la détermination du stade de fibrose, l'examen histologique d'une PBH permet également d'évaluer l'activité inflammatoire, d'identifier les lésions hépatiques qui influencent la progression de la maladie et la réponse au traitement antiviral (notamment la stéatose et la surcharge en fer hépatique), et enfin d'éliminer les diagnostics différentiels.

2.1.2 Limites des scores histologiques de la fibrose hépatique

L'évaluation histologique de la fibrose hépatique a plusieurs limites. Tout d'abord, la PBH est un examen **invasif** craint par de nombreux patients (31, 32). Le taux de

complications graves de cette procédure est d'environ 4‰ et le taux de **décès** avoisine les 0,4‰ (33).

Par ailleurs, la répartition hétérogène de la fibrose induit un **biais d'échantillonnage** car le fragment hépatique prélevé ne représente que 1/50 000^{ème} du foie total. Deux facteurs peuvent ainsi influencer la performance de l'évaluation histologique : le site de la PBH et la longueur de la PBH (34).

La reproductibilité inter-site de l'évaluation histologique de la fibrose hépatique a été évalué dans un étude de 124 patients avec hépatite chronique C (35). Dans ce travail, les patients avaient deux biopsies au cours d'une laparoscopie : une dans le lobe hépatique gauche et l'autre dans le lobe droit. Les fragments prélevés étaient examinés par un anatomopathologiste expérimenté qui déterminait le stade de fibrose selon le score de Batts-Ludwig (29). Une différence d'un stade entre les deux lobes était observée chez 38 patients (30,6%) et une différence ≥ 2 stades chez 3 patients (2,4%). Chez 12 patients (9,7%), le stade de fibrose était F0-2 dans un lobe et F3-4 dans l'autre. Chez 18 patients (14,5%), une cirrhose était diagnostiquée dans un lobe alors que le stade de fibrose était F3 dans l'autre. Au total, la concordance inter-lobe était « seulement bonne » ($\kappa = 0,57$).

L'influence de la longueur de la PBH a été quant à elle évaluée sur une large série de 10 659 biopsies virtuelles obtenues à partir de 17 prélèvements hépatiques chirurgicaux (36). Dans ce travail, l'évaluation du degré de fibrose hépatique sur les biopsies de 15 mm de long n'était correcte que dans 65% des cas lorsqu'elle était comparée à la référence qu'était l'examen de l'ensemble de la pièce opératoire. Ce taux augmentait à 75% pour les biopsies de 25 mm de long, sans amélioration significative pour les biopsies de longueurs supérieures. D'autre part, l'étude d'Everhart *et al* a montré que la valeur pronostique du score Ishak chez les patients avec une hépatite chronique C disparaissait lorsqu'on ne prenait en compte que les biopsies fragmentées (14).

Enfin, l'évaluation histologique des stades de fibrose à l'aide des scores semi-quantitatifs a une **médiocre reproductibilité inter-observateur**. Notre équipe a déterminé la reproductibilité inter-observateur de l'évaluation histologique du Metavir F chez 254 patients ayant une hépatite chronique virale. La concordance inter-observateur entre deux anatomopathologistes experts seniors était « seulement bonne » ($\kappa = 0,59$) (37). Par ailleurs, la

concordance inter-observateur entre un anatomopathologiste expert senior et 10 anatomopathologistes non-experts de première ligne était très faible ($\kappa = 0,13$).

En conclusion, et compte tenu de ses limites, l'évaluation histologique de la fibrose hépatique n'est pas considérée comme le « gold standard », mais comme le « best standard » (38).

2.2 Tests sanguins de fibrose

Des tests sanguins de fibrose ont été mis au point pour effectuer une évaluation non-invasive de la fibrose hépatique (39). Ces tests sanguins ont été principalement construits pour prédire le diagnostic binaire (présence ou absence) de fibrose hépatique significative (Metavir $F \geq 2$, indiquant le traitement antiviral) à l'aide de plusieurs marqueurs de fibrose combinés dans une formule arithmétique.

2.2.1 Méthode d'élaboration des tests sanguins de fibrose

Les tests sanguins de première génération (**Tableau 1**) utilisent des marqueurs de fibrose indirects combinés selon une formule arithmétique simple. Les tests sanguins de seconde génération (**Tableau 1**) combinent des marqueurs directs et/ou indirects de fibrose selon des formules complexes obtenues à l'aide d'une régression logistique binaire.

Tableau 1 : Tests sanguins les plus étudiés dans l'hépatite chronique C (en italique : tests sanguins de première génération; autres : tests sanguins de seconde génération)

Etude princeps	Test sanguin	Patients (n)	Marqueurs
<i>Wai (40)</i>	<i>APRI</i>	192	<i>ASAT, plaquettes</i>
<i>Sterling (41)</i>	<i>Fib4</i>	555	<i>Age, ASAT, ALAT, plaquettes</i>
Imbert-Bismuth (42)	Fibrotest	205	Age, sexe, GGT, bilirubine, apolipoprotéine 1, haptoglobine, $\alpha 2$ macroglobuline
Forns (43)	Forns index	351	Age, plaquettes, GGT, cholestérol
Patel (44)	Fibrospect	294	Acide hyaluronique, TIMP-1, $\alpha 2$ macroglobuline
Rosenberg (45)	ELFG	1021	Age, acide hyaluronique, PIIIIP, TIMP-1
Calès (46)	FibroMètre	383	Age, sexe, ASAT, urée, TP, plaquettes, acide hyaluronique, $\alpha 2$ macroglobuline
Adams (47)	Hepascore	117	Age, sexe, GGT, bilirubine, $\alpha 2$ macroglobuline, acide hyaluronique

La régression logistique binaire (RLB) est une méthode statistique qui crée un modèle mathématique dans une population permettant de discriminer les patients ayant l'événement étudié de ceux ne l'ayant pas. En pratique, la RLB identifie parmi toutes les variables testées celles qui sont indépendamment liées à l'événement étudié, et leur attribue un « poids » selon leur capacité de discrimination. On obtient donc un score de régression $R = ax + by + cz + (\dots) + \text{constante}$ où x, y, z, \dots sont les variables indépendantes sélectionnées par la RLB ; et a, b, c, \dots les coefficients β reflétant « le poids » de chaque variable. Dans les études princeps de chaque test sanguin de 2^e génération, les auteurs ont ainsi utilisé la RLB pour identifier les marqueurs de fibrose permettant de discriminer les patients ayant une fibrose absente ou minimale (Metavir F0/1) de ceux ayant une fibrose significative (Metavir $F \geq 2$). Le test sanguin de fibrose est ensuite créé selon la formule $1 / [1 + \exp(-R)]$ où R correspond au score de régression de la RLB. La valeur du test sanguin s'échelonne de 0 à 1 et son interprétation est la suivante : plus la valeur du test sanguin est proche de 0, plus le patient a de chance de ne pas avoir la cible diagnostique et donc d'avoir une fibrose absente ou minimale (Metavir F0/1) ; inversement, plus la valeur est proche de 1, plus le patient a de risque d'avoir une fibrose significative (Metavir $F \geq 2$). Le diagnostic reste incertain pour les valeurs intermédiaires autour de 0,5 (**Figure 1**).

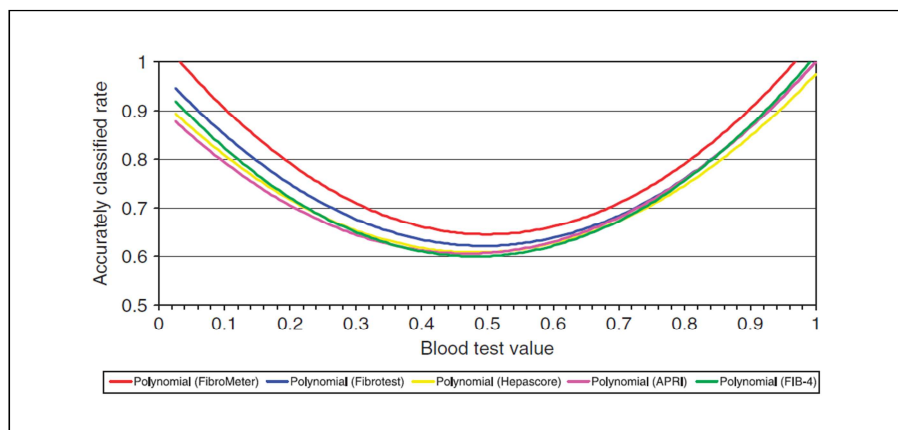


Figure 1 : Performance de cinq tests sanguins pour le diagnostic de fibrose significative chez 1056 patients atteints d'hépatite C. Les courbes sont lissées par régression polynomiale, d'après Calès et al (48).

Un bon modèle sera donc celui qui permet de répartir le plus grand nombre de patients vers les valeurs extrêmes du test sanguin, là où le diagnostic est le plus fiable. La **Figure 2** montre l'exemple de deux tests sanguins évalués dans une même population pour un

diagnostic binaire de fibrose significative ($F0/1$ vs $F\geq 2$). Le test sanguin A est le meilleur car il écarte un plus grand nombre de patients vers les valeurs extrêmes du test, là où la performance pour le diagnostic de fibrose significative est excellente.

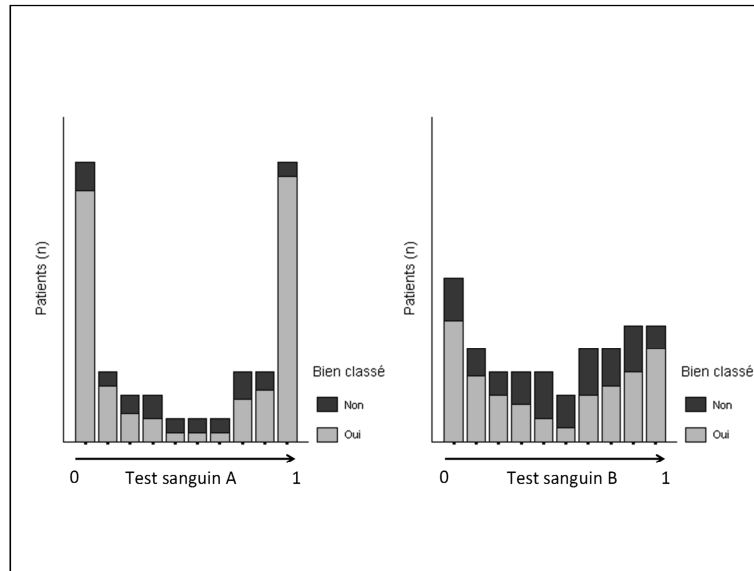


Figure 2 : Répartition des patients selon les valeurs de deux tests sanguins évalués dans une même population pour le diagnostic binaire de fibrose significative ($F0/1$ vs $F\geq 2$). Le test sanguin A est le meilleur : il classe plus de patients vers les valeurs extrêmes du test, là où la probabilité diagnostique est la plus élevée.

2.2.2 Performance diagnostique des tests sanguins de fibrose

Le Fibrotest est le test sanguin de fibrose le plus validé dans l'hépatite chronique C. Une récente méta-analyse de 19 études incluant 3 501 patients avec une hépatite chronique C a montré que son aire sous la courbe ROC (AUROC) pour le diagnostic de fibrose hépatique significative était égale à 0,79 (49).

Notre équipe a mis au point le FibroMètre. Dans une étude récente incluant 1056 patients avec une hépatite chronique C dans 9 centres français (secondaires et tertiaires, publics et privés), la performance du FibroMètre pour le diagnostic de fibrose hépatique significative, de fibrose sévère, et de cirrhose était significativement supérieure à celles des autres tests sanguins : Fibrotest, Hepascore, APRI et Fib4 (48).

Récemment, une grande étude indépendante promue par l'Agence Nationale de Recherche sur le SIDA et les hépatites a comparé les performances de 8 tests sanguins de fibrose (Fibrotest, FibroMètre, Forns index, APRI, ELF, Hepascore, MP3, et hyaluronate) chez 590 patients avec une hépatite chronique C (50). La comparaison des courbes ROC des différents tests sanguins a montré que les performances du Fibrotest, du FibroMètre, et de l'Hepascore n'étaient pas significativement différentes, mais supérieures à celles des autres tests, pour le diagnostic de fibrose significative (Metavir $F \geq 2$), de fibrose sévère (Metavir $F \geq 3$), et de cirrhose (Metavir F4).

La reproductibilité inter-laboratoire des tests sanguins est excellente (51). Par ailleurs, au contraire de la PBH, les tests sanguins sont non-invasifs et mieux acceptés par les patients, et peuvent ainsi être répétés dans le temps. Une étude préliminaire récente a ainsi montré que la réalisation annuelle d'un FibroMètre permettait de suivre l'évolution de la fibrose hépatique de façon fiable avec une bonne reproductibilité (52).

2.2.3 Classement de fibrose du Fibrotest et du FibroMètre

Bien que les tests sanguins de fibrose aient été construits pour un diagnostic binaire de fibrose significative ($F0/1$ vs $F \geq 2$), leur valeur est bien corrélée aux stades de fibrose de la classification Metavir (**Figure 3**).

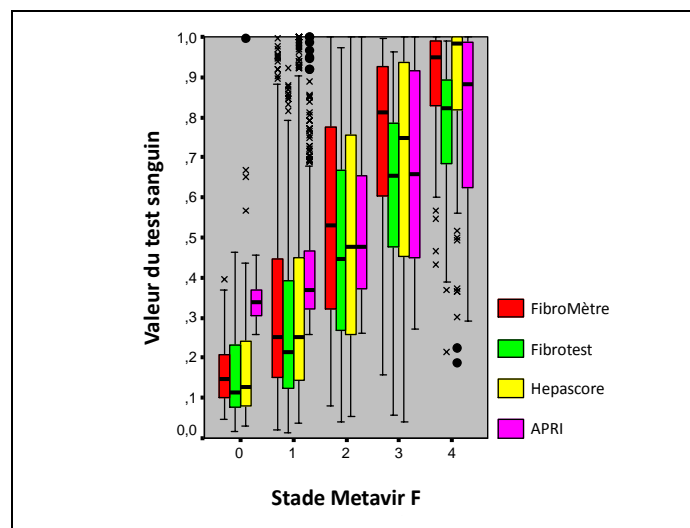


Figure 3 : Box plots (médiane, quartile, extrêmes) des tests sanguins de fibrose selon le stade Metavir F chez 1056 patients atteints d'hépatite chronique C (48). L'APRI a été standardisé par régression logistique binaire. x et ● : valeurs éloignées et extrêmes. D'après Calès et al (53).

Les concepteurs du Fibrotest et du FibroMètre ont donc créé des classifications, ou « compteurs de fibrose », permettant de déterminer le stade Metavir F à partir de la valeur de leur test sanguin. La méthodologie de mise au point du compteur du Fibrotest n'a jamais été clairement décrite, mais il semble qu'elle ait utilisé les bornes des box-plots du Fibrotest pour chaque stade de fibrose (**Figures 4a à 4c**) (54). La méthode d'élaboration du compteur du FibroMètre a été publiée (55) : une population de 1056 patients atteints d'hépatite chronique C a été divisée en 25 groupes d'effectifs égaux selon les 2,5% percentiles du FibroMètre (**Figure 4d**). Dans chacun des groupes, le(s) stade(s) de fibrose dont la prévalence était $\geq 75\%$ indiquai(en)t le diagnostic histologique pour l'intervalle de valeur du FibroMètre correspondant, permettant ensuite la création du compteur (**Figure 4e**).

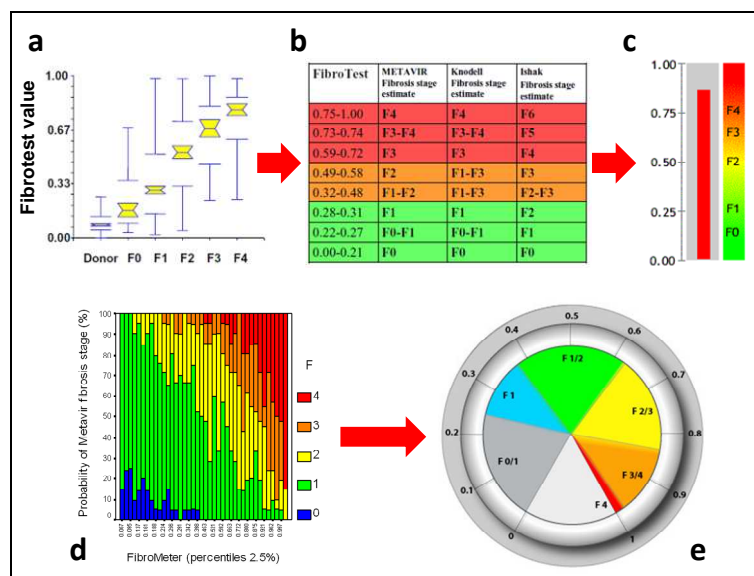


Figure 4 : Méthodologies utilisées pour la création des compteurs de fibrose du Fibrotest et du FibroMètre. Box-plots du Fibrotest pour chaque stade de fibrose (figure 4a) permettant d'élaborer le compteur du Fibrotest (figures 4b et 4c); d'après Poynard et al (54). Pourcentage de stades de fibrose Metavir (F) en fonction des valeurs du FibroMètre exprimées en 2,5% percentiles (figure 4d), permettant la création du compteur du FibroMètre (figure 4e); d'après Leroy et al (55).

L'évaluation de la performance diagnostique des compteurs du Fibrotest et du FibroMètre dans une large population d'hépatite chronique C n'a encore jamais été publiée. Seules sont disponibles les données d'une petite étude (56) et d'un abstract (57). Dans l'étude de Gressner *et al* (56), le compteur du Fibrotest a été testé sur 24 prélèvements sanguins de patients ayant une hépatite chronique C. Le résultat de ce compteur était en accord avec le

stade histologique de fibrose dans seulement 25% des cas. L'étude de Calès *et al* publiée sous forme d'abstract (57) incluait 1298 patients atteints d'hépatite chronique C issus de deux précédentes études diagnostiques. Dans ce travail, la performance diagnostique du compteur du FibroMètre était significativement supérieure à celle du compteur du Fibrotest, avec respectivement 74,7% vs 38,4% de bien classés dans la population 1 ($p < 10^{-4}$) et 68,2% vs 33,2% dans la population 2 ($p < 10^{-3}$). Dans la population 1, le résultat du compteur du FibroMètre avait un écart ≥ 2 stades de fibrose avec le Metavir F histologique chez 4,6% des patients versus 18,2% avec le compteur du Fibrotest ($p < 10^{-3}$). Dans la population 2, ces taux étaient respectivement de 6,0% et 22,2% ($p < 10^{-3}$).

2.2.4 Limites des tests sanguins de fibrose

Les tests sanguins de fibrose sont issus d'une modélisation dans une population donnée, un peu comme un costume taillé « sur mesure » : la sélection des variables indépendantes et le calcul du poids attribué à chaque variable par la RLB dépend des caractéristiques de la population où a été créé le test. La performance des tests sanguins de fibrose dans les études de validation externe est ainsi souvent inférieure à celle décrite dans la population exploratoire. C'est un peu comme un vêtement créé « sur-mesure » sur un mannequin qui ne convient pas à toutes les personnes qui l'essaient en magasin de « prêt-à-porter ». Pour une utilisation en pratique clinique, les tests sanguins de fibrose doivent donc être élaborés et/ou validés dans de larges populations dont les caractéristiques sont représentatives de la population générale des patients avec hépatite chronique C vus en pratique clinique.

L'existence d'une pathologie extra-hépatique peut modifier la valeur d'un ou de plusieurs marqueurs de fibrose inclus dans le test sanguin et ainsi induire un risque de résultat faux-positif ou faux-négatif. Une hémolyse (élévation de la bilirubine et diminution de l'haptoglobine) et un syndrome inflammatoire (élévation de l' $\alpha 2$ macroglobuline sans diminution de l'haptoglobine) peuvent ainsi être responsables d'un résultat faux-positif du Fibrotest (58). Au contraire, un syndrome inflammatoire avec une élévation isolée de l'haptoglobine peut être responsable d'un résultat faux-négatif du Fibrotest.

Enfin, les tests sanguins de fibrose ont été construits pour le diagnostic de fibrose significative et non pas pour estimer le stade de fibrose histologique. Une étude récente dans

une large cohorte de patients ayant une hépatite chronique C a ainsi montré que la performance des compteurs du FibroMètre et du Fibrotest était médiocre pour le diagnostic de cirrhose « F4 » avec des sensibilités respectives de 32,2% et 61,8% ($p < 10^{-3}$), et des valeurs prédictives positives (VPP) de 68,5% et 37,1% (59). L'utilisation de ces compteurs à partir du stade « F3/4 » améliorerait la sensibilité pour le diagnostic de cirrhose (FibroMètre : 74,8% vs Fibrotest : 64,7%; $p=0.078$), mais au prix d'une diminution de la VPP (FibroMètre : 39,8%; Fibrotest : 35,1%).

2.3 Fibroscan

Le Fibroscan® (Echosens, Paris, France) est un procédé de mesure non-invasive de la dureté hépatique par élastométrie impulsionnelle commercialisé en 2003 (**Figure 5**). Outre sa simplicité et son caractère indolore, l'examen avec le Fibroscan présente l'avantage de pouvoir être pratiqué rapidement (moins de 5 minutes) au lit du malade ou en consultation, avec un résultat immédiat (60).



Figure 5 : Appareil Fibroscan.

2.3.1 Utilisation en pratique du Fibroscan

L'examen s'effectue chez un malade en position allongée, dos à plat, le bras droit sous la tête pour dégager l'hypochondre droit. Après application de gel sur la peau, la sonde du Fibroscan est positionnée par l'opérateur au niveau d'un espace intercostal en regard du lobe droit du foie (en pleine matité hépatique, **Figure 6**).



Figure 6 : *Utilisation du Fibroscan en pratique. Le patient est en position allongée, le bras droit sous la tête pour dégager l'hypochondre droit. Après application de gel sur la peau, la sonde du Fibroscan est positionnée par l'opérateur au niveau d'un espace intercostal en regard du lobe droit du foie.*

La sonde produit une impulsion mécanique sur la peau et génère ainsi une onde élastique de cisaillement de basse fréquence qui traverse le tissu sous-cutané puis le foie (**Figure 7**). A l'aide d'acquisitions ultrasonores, la sonde mesure ensuite la vitesse de propagation de l'onde élastique dans une zone d'intérêt du parenchyme hépatique. Cette zone d'intérêt correspond à un cylindre de 1 cm de diamètre sur 4 cm de longueur, situé entre 25 et 65 mm de profondeur sous la peau, ce qui représente un volume au moins 100 fois supérieur à celui d'un fragment obtenu par biopsie hépatique.

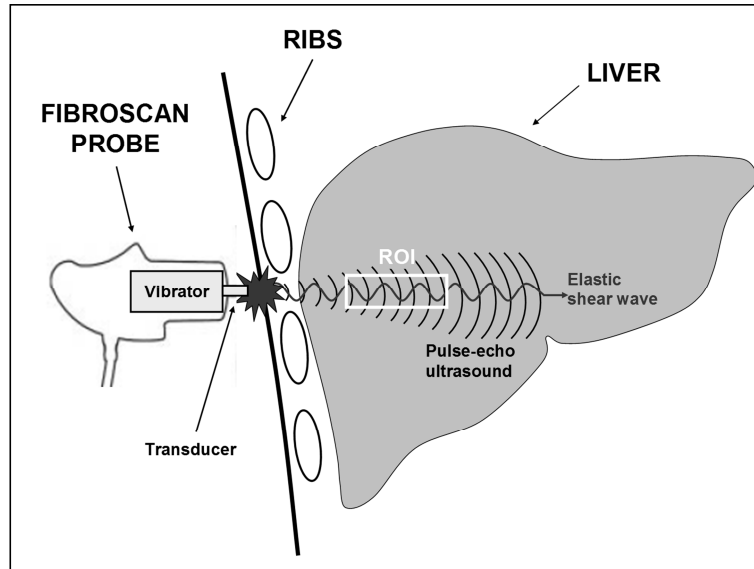


Figure 7 : Mesure de la dureté hépatique par le Fibroscan. La sonde, appliquée sur la peau en regard du foie, produit une impulsion mécanique indolore sur la peau. Cette impulsion mécanique génère une onde élastique de cisaillement de basse fréquence qui traverse le foie. A l'aide d'acquisitions ultrasonores, la sonde mesure ensuite la vitesse de propagation de l'onde élastique dans une région d'intérêt du foie (ROI) et en déduit la dureté hépatique : plus l'onde va vite, plus le foie est dur (61).

La vitesse de l'onde de cisaillement est d'autant plus importante que le milieu dans lequel elle se propage est dur (61). De la vitesse de propagation de l'onde, le Fibroscan déduit donc la dureté du foie exprimée en kilopascals (kPa) à l'aide d'une équation simple (module de Young) : $E = 3\rho V_s^2$ ou E représente l'élasticité, ρ la densité (considérée comme une constante pour les tissus) et V_s la vitesse de propagation de l'onde. Le Fibroscan est donc un test indirect de dureté hépatique.

Les résultats sont compris entre 2,5 et 75,0 kPa. Il est recommandé d'obtenir 10 mesures valides, et le résultat final correspond à la valeur médiane de ces 10 mesures. Une étude récente a suggéré que le rapport IQR/médiane devait être $<0,21$ pour que le résultat du Fibroscan soit considéré comme fiable (62). Nous avons montré que le rapport IQR/médiane définissait trois catégories de mesures de la dureté hépatique par le Fibroscan : celles avec un rapport IQR/médiane $\leq 0,10$ sont « très fiables », surtout chez les patients ayant un résultat >13 kPa ; celles avec rapport IQR/médiane entre 0,10 et 0,30 sont « fiables » ; et celles avec un rapport IQR/médiane $>0,30$ sont « peu fiables » quand le résultat est >7.6 kPa (63).

2.3.2 Performance diagnostique du Fibroscan

Quatre méta-analyses ont évalué la performance du Fibroscan pour le diagnostic non invasif de la fibrose hépatique, toutes incluant des études avec diverses causes d'hépatopathies chroniques (64-67) (**Annexe 1**). Dans ces travaux, l'AUROC du Fibroscan variait de 0,83 à 0,87 pour le diagnostic de fibrose significative, et de 0,94 à 0,96 pour le diagnostic de cirrhose (64-66). Seule la méta-analyse de Friedrich-Rust *et al* (8478 patients) a évalué la performance du Fibroscan pour le diagnostic de fibrose sévère avec une AUROC calculée à 0,89 (66). Dans ce travail, les auteurs ont également montré que l'hétérogénéité des résultats entre les études pour le diagnostic de fibrose significative était significativement réduite lorsque les causes d'hépatopathies chroniques étaient séparées. L'AUROC pour le diagnostic de fibrose significative avait ainsi été calculé à 0,85 pour les hépatites chroniques C. Par contre, l'étiologie n'impliquait pas d'hétérogénéité significative pour le diagnostic de cirrhose. Les méta-analyses de Friedrich-Rust *et al* (66) et de Stebbing *et al* (67) ont également déterminé les meilleurs seuils pour le diagnostic binaire de fibrose significative (respectivement 7,65 kPa et 7,81 kPa) et pour celui de cirrhose (respectivement 13,01 kPa et 15,56 kPa). Dans l'étude de Stebbing *et al* (4430 patients), ces seuils ont été calculés pour chaque étiologie d'hépatopathie chronique ; pour l'hépatite chronique C, ils étaient de 8,44 kPa pour le diagnostic de fibrose significative et 16,14 kPa pour la cirrhose (67).

Deux grandes études indépendantes ont comparé la performance du Fibroscan à celle des tests sanguins de fibrose pour le diagnostic non invasif de la fibrose hépatique au cours des hépatites chroniques virales (50, 68). La première est celle de Zarski *et al*, promue par l'Agence Nationale de Recherche sur le SIDA et les hépatites et ayant inclus 590 patients avec une hépatite chronique C (50). Dans ce travail, les AUROC du Fibroscan pour le diagnostic de fibrose significative et de fibrose sévère n'étaient pas significativement différentes de celles du Fibrotest, du FibroMètre, et de l'Hepascore. Par contre, le Fibroscan était le meilleur test non invasif pour le diagnostic de cirrhose. La seconde étude a été promue par l'Assistance Publique des Hôpitaux de Paris et a inclus 1993 patients avec une hépatite chronique virale (68). Dans ce travail, l'AUROC du Fibroscan pour le diagnostic de fibrose significative n'était pas significativement différente de celles du Fibrotest, de l'Hepascore et de l'APRI, mais était significativement inférieure à celle du FibroMètre. Par contre, comme pour l'étude précédente, le Fibroscan était le meilleur test pour le diagnostic non invasif de cirrhose.

2.3.3 Compteur de fibrose du Fibroscan

Plusieurs études ont montré que la dureté hépatique estimée par le Fibroscan était bien corrélée aux stades histologiques de fibrose (60). Un compteur de fibrose a ainsi été proposé pour estimer les stades histologiques de fibrose à partir du résultat du Fibroscan chez les patients avec une hépatite chronique C (**Figure 8**) (69).

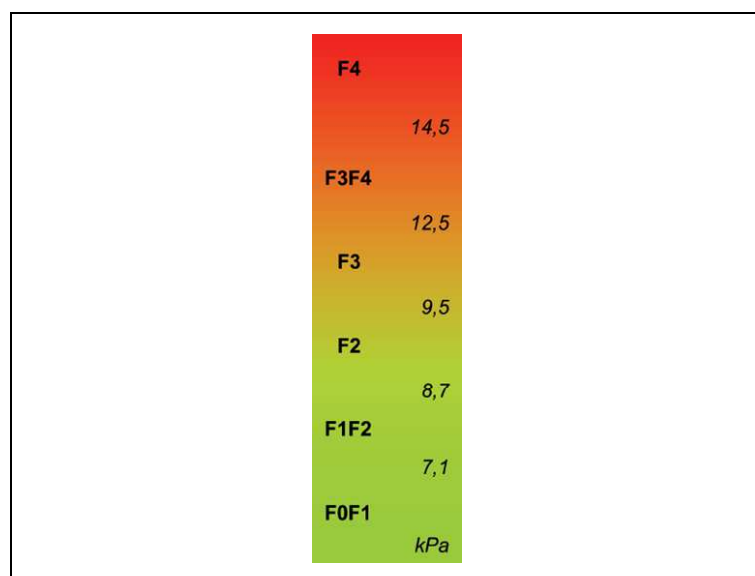


Figure 8 : Compteur de fibrose du Fibroscan permettant d'estimer le stade histologique de fibrose (Metavir F) à partir du résultat du Fibroscan (kPa) chez les patients avec une hépatite chronique C; d'après de Ledinghen et al (69).

La méthode d'élaboration de ce compteur de fibrose n'a pas été détaillée, mais il semble qu'il ait été créé à partir de seuils diagnostiques de deux précédentes études (70, 71). L'évaluation de la performance diagnostique du compteur du Fibroscan dans une large population d'hépatite chronique C n'a encore jamais été publiée. Seules sont disponibles les données de l'abstract de Calès *et al* (57) : chez 343 patients avec une hépatite chronique C, ce compteur avait une performance diagnostique de 54,8%, significativement supérieure à celle du compteur du Fibrotest (33,2%, $p < 10^{-3}$) mais significativement inférieure à celle du compteur du FibroMètre (68,2%, $p < 10^{-3}$). Le résultat du compteur du Fibroscan avait un écart ≥ 2 stades de fibrose Metavir F histologique chez 12,3% des patients versus 22,2% avec le compteur du Fibrotest et 6,0% avec celui du FibroMètre ($p < 10^{-3}$).

2.3.4 Limites du Fibroscan

Le Fibroscan présente cependant plusieurs limites :

- Le taux d'échec de mesure de la dureté hépatique est relativement élevé : il varie entre 2,4% et 9,4% (60) et est essentiellement lié à l'obésité des patients (72, 73).
- Notre équipe a montré que la reproductibilité de la mesure de la dureté hépatique par le Fibroscan était médiocre pour les faibles valeurs de dureté hépatique, celles correspondant aux stades intermédiaires et faibles de fibrose (74). Dans notre étude, l'écart de mesure entre les deux opérateurs était en moyenne de $\pm 30\%$. Ainsi, pour une dureté hépatique mesurée à 10 kPa, le résultat du second opérateur pouvait aller de 7 kPa (en faveur d'une fibrose significative) à 13 kPa (en faveur d'une fibrose sévère, voire d'une cirrhose).
- Il existe un risque de surestimation du degré de fibrose hépatique lorsque la mesure de la dureté hépatique est effectuée en cas de cytolyse aiguë (75-78), de cholestase (79), ou d'insuffisance cardiaque (80).
- Le Fibroscan est un dispositif coûteux (environ 80 000 euros dans son modèle « hospitalier ») disponible uniquement dans les centres spécialisés (hôpitaux et quelques cabinets d'hépatogastroentérologues libéraux), limitant ainsi son accès aux patients.

2.4 Recommandations de la Haute Autorité de Santé française

L'ensemble des résultats des études ayant évalué la performance des tests sanguins de fibrose et du Fibroscan a amené la Haute Autorité de Santé française (HAS) à effectuer les recommandations suivantes en décembre 2008 (81) :

- « Dans le cas d'une hépatite chronique C non traitée et sans comorbidité, chez l'adulte (hors diagnostic évident de cirrhose), la mesure de la fibrose/cirrhose hépatique peut donc se faire avec cinq techniques (la ponction-biopsie hépatique, le Fibroscan®, le score Fibrotest®, le score FibroMètreV® et le score Hepascore). Il est proposé en première intention la réalisation d'une des quatre techniques non invasives. En seconde intention (s'il n'y a pas de concordance entre la clinique et le résultat de la technique réalisée en première intention, ou si ce résultat est non interprétable), une des trois autres techniques non invasives ou la PBH peut être alors réalisée. »

- « Dans le cas d'une hépatite chronique C non traitée avec co-infection VIH, chez l'adulte (hors diagnostic évident de cirrhose), la mesure de la cirrhose hépatique peut donc se faire avec la PBH ou le Fibroscan®. Il est proposé en première intention la réalisation du Fibroscan®. En seconde intention (s'il n'y a pas de concordance entre la clinique et le résultat du Fibroscan®, ou si il y a eu un échec technique dans la réalisation de cette technique), la PBH peut être alors réalisée. »

Bien qu'elles constituent une avancée, ces recommandations restent discutables sur certains points. En effet, en pratique clinique, les résultats du Fibrotest ou FibroMètre sont interprétés à l'aide de leurs compteurs. Cependant, la performance de ces compteurs n'a pas été publiée et ils nécessitent donc d'être validés. Par ailleurs, l'interprétation du résultat et donc l'utilisation en pratique clinique de l'Hepascore ne sont pas connues. Enfin, les recommandations ne précisent pas les valeurs des seuils diagnostiques à utiliser pour interpréter le résultat du Fibroscan en pratique.

2.5 Nouvelles techniques de mesure non invasive de la fibrose hépatique

Des travaux récents ont cherché à identifier de nouveaux marqueurs de fibrose hépatique en utilisant la « glycomique clinique ». La « glycomique clinique » permet de caractériser les profils de glycosylation des protéines sériques. En utilisant cette technique, Callewaert *et al* ont identifié 3 nouveaux biomarqueurs (NGA2FB, NA3, et NA2FB) permettant de calculer le GlycoFibroTest et le GlycoCirrhoTest (82, 83). Une étude récente incluant 376 patients avec hépatite chronique C a montré que le GlycoFibroTest était bien corrélé aux stades de fibrose du Metavir (83).

De nouvelles techniques d'élastométrie ont été mise au point. *L'élastométrie par résonance magnétique* semble performante pour le diagnostic non invasif de la fibrose hépatique (84-86), mais cette technique n'est pas encore commercialisée, nécessite encore d'être validée, et sera probablement peu disponible.

L'élastométrie acoustique (Acoustic Radiation Force Impulse : ARFI) est commercialisée sous la forme d'un module additionnel d'un appareil d'échographie-Doppler

standard. Le principe consiste à mesurer la vitesse de propagation dans le foie d'une onde de cisaillement créée par un faisceau ultrasonore de haute énergie (87), et non pas par une impulsion mécanique comme pour le Fibroscan. Plusieurs études ont montré que l'élastométrie acoustique avait une bonne performance pour le diagnostic non invasif de la fibrose hépatique dans les hépatopathies chroniques, mais les populations étudiées étaient de faibles effectifs (88-93). Une étude réalisée chez 86 patients avec hépatite chronique virale a montré que les performances de l'ARFI et du Fibroscan n'étaient pas significativement différentes pour le diagnostic non invasif de la fibrose hépatique (88). Cependant, un travail récent a montré que l'ARFI avait une meilleure faisabilité que le Fibroscan (94), et plusieurs études ont montré que son taux d'échec de mesure était de 0% (88, 92, 94, 95).

Enfin, l'*imagerie d'élasticité* (Supersonic Shear Imaging) couple une imagerie échographique ultrarapide (supérieure à 5000 images par seconde) à une palpation ultrasonore à distance à l'aide d'une vibration basse fréquence générée par un faisceau ultrasonore focalisé, permettant ainsi de produire des images en 2 dimensions de l'élasticité des tissus (96). Dans une étude publiée sous forme d'abstract et incluant 104 patients avec hépatite chronique virale, le Supersonic Shear Imaging avait une AUROC significativement supérieure à celle du Fibroscan pour le diagnostic de la fibrose significative (Metavir F \geq 2), respectivement 0,95 vs 0,89 (p=0,04) (97). Dans ce travail, les AUROC des 2 techniques pour le diagnostic de fibrose sévère (Metavir F \geq 3) et de cirrhose n'étaient pas significativement différentes : respectivement 0,96 vs 0,94 (p=0,25) et 0,97 vs 0,94 (p=0,16).

3. ASPECTS METHODOLOGIQUES DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HEPATIQUE

3.1 Biais des études ayant évalué la performance diagnostique des tests non invasifs de fibrose

Pour mettre au point et évaluer la performance d'un test non invasif de fibrose, il faut contourner essentiellement deux biais méthodologiques :

- **Le biais de sélection.** Pour l'éviter, le test diagnostique doit être créé/évalué dans une population représentative de la population générale des patients vus en pratique clinique.
- **Le biais de classement.** Pour l'éviter, la performance du test diagnostique doit être évaluée versus les résultats d'un gold standard.

Nous discutons ici l'implication de ces biais méthodologiques dans l'évaluation de la performance des tests non invasifs de fibrose.

3.1.1 Représentativité des populations où sont créés et/ou évalués les tests non invasifs de fibrose

Pour que les travaux évaluant la performance des tests non invasifs de fibrose produisent des résultats exportables à l'exercice clinique, il est nécessaire qu'ils étudient des populations représentatives de la population générale des patients vus en pratique clinique. Ceci est encore plus vrai pour les tests sanguins de fibrose : ils sont issus d'une modélisation qui a cherché à établir le modèle « sur-mesure » le plus performant dans la population de l'étude princeps (cf § 2.2.1). Ainsi, la sélection des variables par l'analyse multivariée, et le calcul des coefficients β pour les tests de 2^e génération, dépendait des caractéristiques des populations où ont été créés les tests sanguins.

La méta-analyse de Thein donne une référence pour la prévalence des stades de fibrose dans la population générale des patients avec hépatite chronique C (8). Ce travail a inclus 33 212 patients ayant une hépatite chronique C non sélectionnés par les tests non invasifs pour la réalisation de la PBH. Dans cette étude, la prévalence des stades de fibrose était : F0: 17%, F1: 35%, F2: 22%, F3: 14%, F4: 12%.

3.1.1.1 Evolution des caractéristiques des patients inclus dans les études diagnostiques

Depuis les recommandations de la HAS (81), les tests non invasifs de fibrose sont de plus en plus utilisés en pratique clinique. Désormais, les praticiens ne proposent plus de PBH lorsque l'évaluation non invasive suggère l'absence de fibrose significative. Par conséquent, la PBH n'est surtout réalisée qu'en cas de discordance entre les tests non invasifs ou pour confirmer l'existence d'une fibrose significative suggérée par le bilan non invasif.

Ceci pose un problème dans les études de recherche clinique qui évaluent la performance diagnostique des tests non invasifs de fibrose. En effet, ce type d'étude nécessite d'inclure des patients avec PBH car c'est le résultat de cette dernière qui constitue la référence de fibrose. Les indications de la PBH devenant réservées à un sous-groupe de patients sélectionnés par les tests non invasifs de fibrose, les études diagnostiques ont désormais tendance à étudier des populations dont les caractéristiques sont différentes de celles de la population générale des patients suivis en pratique clinique. Dans les faits, dans les études diagnostiques avec la PBH comme référence, cela se traduit par une diminution de la prévalence des faibles stades de fibrose (Metavir F0/1), et par conséquent par une augmentation "artificielle" de la prévalence des stades de fibrose élevés (Metavir $F \geq 2$).

3.1.1.2 Influence de la prévalence des stades de fibrose sur la performance des tests non invasifs de fibrose

Une étude a récemment évalué la relation entre la prévalence des stades de fibrose et l'AUROC du Fibrotest pour le diagnostic de fibrose significative (98). Dans ce travail, un nouvel index était proposé : l'index DANA. Dans une population donnée, cet index reflète la fréquence des stades de fibrose dans chacun des deux groupes de la cible diagnostique "fibrose significative" (groupes Metavir F0/1 et $F \geq 2$). L'index DANA est calculé ainsi :

$$\text{Index DANA} = \frac{(2 \times \text{prévalence F2} + 3 \times \text{prévalence F3} + 4 \times \text{prévalence F4})}{(\text{prévalence F2} + \text{prévalence F3} + \text{prévalence F4})} - \frac{(0 \times \text{prévalence F0} + 1 \times \text{prévalence F1})}{(\text{prévalence F0} + \text{prévalence F1})}$$

Les auteurs ont tout d'abord montré dans une large population de 1392 patients avec hépatite chronique C qu'il existait une très forte corrélation positive entre l'index DANA et l'AUROC du Fibrotest pour le diagnostic de fibrose significative (coefficient de Spearman = 0,95; $p < 10^{-4}$). La formule de régression de la corrélation était : $\text{AUROC} = 0,582 + (0,1056 \times \text{DANA})$. Un index DANA de référence a ensuite été calculé dans une large population de référence (99). La prévalence des stades de fibrose dans cette population (F0 : 10%, F1 : 36%, F2 : 21%, F3 : 15%, et F4 : 18%) permettait de calculer le DANA de référence à 2,16. Les auteurs ont ainsi estimé que l'AUROC du Fibrotest pour le diagnostic de fibrose significative dans une population idéale (DANA à 2,16) serait égale à 0,85. Enfin, une méthode de correction de l'AUROC en fonction de la prévalence des stades de fibrose dans la population étudiée a été proposée : $\text{AUROC}_{\text{corrigée}} = \text{AUROC}_{\text{calculée}} + 0,1056 \times (2,16 - \text{DANA})$.

Ce travail apporte deux messages importants : 1/ la prévalence des stades de fibrose dans une population influence la performance diagnostique des tests non invasifs de fibrose évalués dans cette population, et 2/ une forte prévalence de stades de fibrose élevés (Metavir $F \geq 2$) va dans le sens d'une surestimation de la performance des tests non invasifs pour le diagnostic de fibrose significative. Ainsi, l'enrichissement en stades de fibrose Metavir $F \geq 2$ dans les études récentes, lié au fait que la PBH n'est plus réalisée aux patients dont le bilan non invasif suggère l'absence de fibrose significative, implique probablement une surévaluation de la performance diagnostique des tests non invasifs dans ces études.

L'AUROC permet d'évaluer la performance globale d'un test pour une cible diagnostique binaire ce qui impose de regrouper les stades histologiques de fibrose: fibrose minimale (F0/1) versus fibrose significative ($F \geq 2$), ou encore : absence de cirrhose ($F \leq 3$) versus cirrhose (F4). En plus de la prévalence des stades de fibrose dans la population étudiée, le résultat de l'AUROC sera influencé par les modalités de regroupement des stades de fibrose. La méthode d'Obuchowski permet de comparer deux tests définis sur un même échantillon pour un gold-standard ordinal (notamment la classification Metavir en 5 stades) et requiert

comme paramètre une matrice de pénalités des discordances entre modalités de prédiction (100). Cette même méthode peut aussi s'appliquer à des gold-standard continus et nominaux. Brièvement, les AUROCs des tests sont calculées pour le diagnostic entre chaque paire de stades de fibrose individuel (F0 vs F1, F0 vs F2, F0 vs F3, F0 vs F4, F1 vs F4...) et la mesure d'Obuchowki correspond à la moyenne pondérée des 10 AUROCS obtenues. La pondération s'effectue à l'aide de matrice des pénalités et de la prévalence des stades de fibrose dans une population de référence. La matrice des pénalités peut être proportionnelle à la différence entre les stades Metavir : 1 quand la différence est de 1 stade, 2 pour 2 stades etc. lorsqu'on veut tenir compte de la magnitude des erreurs de classement ou uniforme (1 pour toute différence) si on ne veut quantifier que l'erreur de classement (101). Cette méthode permet donc d'évaluer la performance réelle des tests de fibrose, même dans les populations où les stades élevés de fibrose sont sur représentés du fait des critères de sélection des patients pour la réalisation de la PBH (cf § 3.1.1.1).

3.1.2 Impact de l'absence de gold standard sur l'évaluation de la performance des tests non invasifs de fibrose

La performance d'un test diagnostique doit être évaluée en prenant comme référence un gold standard. Pour le diagnostic de la fibrose hépatique, le gold standard théorique est l'examen histologique de la totalité du foie, ce qui n'est pas réalisable en pratique. Les études diagnostiques ont donc pris la PBH comme référence de fibrose. Cependant, compte tenu de ses limites (biais d'échantillonnage, variabilité inter-observateur des scores semi-quantitatifs de fibrose), il s'agit d'une référence imparfaite (cf § 2.1.2). Dès lors se pose la question : quel est le bon diagnostic lorsque la PBH et le test non invasif ont des résultats discordants ?

3.1.2.1 Discordances entre tests non invasifs de fibrose et ponction-biopsie hépatique

Quatre études ont étudié les discordances entre les résultats de la PBH et ceux du Fibrotest (58, 102, 103) ou du Fibroscan (62) chez les patients avec une hépatite chronique C. L'écart entre le Metavir F histologique et le stade de fibrose estimé par le compteur du Fibrotest était ≥ 2 stades dans 16-20% des cas (**Annexe 2**). Dans l'étude de Lucidarne *et al*, un écart ≥ 2 stades entre le Metavir F histologique et le stade de fibrose estimé par le Fibroscan était observé chez 11 à 16% des patients selon les seuils diagnostiques utilisés pour le Fibroscan (62).

Les 3 études ayant étudié les discordances entre la PBH et le Fibrotest ont cherché à attribuer l'imputabilité de la discordance à l'une des deux méthodes (58, 62, 103). L'imputabilité de la discordance était attribuée à la PBH dans 20 à 50% des cas, au Fibrotest dans 15 à 45% des cas, et restait indéterminée dans 30 à 50% des cas (**Annexe 2**). Il faut cependant préciser que les règles d'attribution des discordances étaient sensiblement différentes entre les 3 études. De plus, en l'absence de gold standard, aucune règle d'attribution des discordances entre deux tests de fibrose n'est parfaite.

Dans leur étude, Poynard *et al* ont identifié les causes des erreurs de la PBH et du Fibrotest (58). Les résultats faux-négatifs de la PBH étaient expliqués par une petite longueur de biopsie et la présence d'une stéatose, alors que les résultats faux-positifs étaient liés à la fragmentation de la biopsie. Une récente mise au point a ainsi suggéré que la PBH devait faire au moins 20 mm de longueur et inclure au moins 11 espaces portes complets pour que l'évaluation du stade de fibrose hépatique soit fiable (34).

3.1.2.2 Performance diagnostique maximale des tests non invasifs de fibrose

Dans les nombreuses études ayant évalué la performance des tests non invasifs de fibrose, la référence était l'examen histologique d'une PBH et en cas de discordance le résultat du test non invasif était considéré comme faux. Cependant, les performances d'un test diagnostique sont sous-évaluées lorsque les erreurs du test diagnostique et de la référence sont indépendantes (104-106).

Mehta *et al* ont récemment évalué l'impact des erreurs de la PBH sur la performance diagnostique des tests non invasifs de fibrose (107). Dans ce travail, les auteurs ont montré quelles seraient les variations d'AUROC des tests non invasifs en fonction, d'une part, de la sensibilité et de la spécificité supposées de la PBH pour le diagnostic de la fibrose significative et, d'autre part, de la prévalence de la fibrose significative dans la population étudiée. Dans le meilleur scénario pour la PBH (sensibilité et spécificité pour le diagnostic de fibrose significative estimée à 90%) et pour une prévalence de la fibrose significative à 40%, un test non invasif parfait verrait son AUROC diminuer de 0,99 à 0,90. Inversement, dans l'hypothèse où la sensibilité et la spécificité de la PBH sont à 80% et la prévalence entre 30 et 50%, l'AUROC d'un test de fibrose calculée à 0,76 dans une étude versus PBH devrait "réellement" être entre 0,93 et 0,99.

3.1.3 Discordance entre les tests non invasifs de fibrose

Les tests sanguins et le Fibroscan sont de plus en plus à disposition des praticiens qui peuvent désormais effectuer facilement ces deux examens. Depuis les recommandations de la HAS, les PBH ayant pour but d'évaluer le degré de fibrose hépatique chez les patients ayant une hépatite chronique C sont de plus en plus souvent réalisées en seconde ligne, en raison d'une discordance entre les tests non invasifs de fibrose. On peut donc supposer que les études diagnostiques évaluant la performance des tests non invasifs de fibrose versus la PBH incluent de plus en plus de patients pour lesquels il existe une discordance entre les tests non invasifs. Cependant, inclure ce profil de patient induit inévitablement une diminution de la performance diagnostique des tests non invasifs dans la population étudiée car une discordance entre deux tests implique d'emblée qu'un des tests donne un faux résultat.

En conclusion, plusieurs facteurs influencent l'évaluation de la performance des tests non invasifs de fibrose dans les études scientifiques :

- L' "**effet sélection**" (enrichissement des populations étudiées en stade de fibrose élevé) qui va dans le sens d'une surévaluation de la performance des tests non invasifs.
- L' "**effet PBH**" (référence imparfaite de fibrose) et l' "**effet discordance**" (inclusion de patients avec discordances entre les tests non invasifs de fibrose) qui vont dans le sens d'une sous-évaluation de la performance des tests non invasifs.

Ainsi, il est fort probable que la performance diagnostique réelle des tests non invasifs de fibrose ne soit actuellement pas connue.

3.2 Indices diagnostiques pour l'évaluation des tests non invasifs de fibrose

3.2.1 AUROC

L'AUROC (Area Under the Receiver-Operating Characteristic) correspond à l'aire sous la courbe ROC. La courbe ROC est la représentation graphique de la sensibilité et de la spécificité globales d'un test diagnostique : ces deux indices sont calculés pour chacune des

valeurs du test diagnostic et le résultat est reporté sur un graphique représentant la sensibilité (axe Y) en fonction de 1 – spécificité (axe X, **Figure 9**).

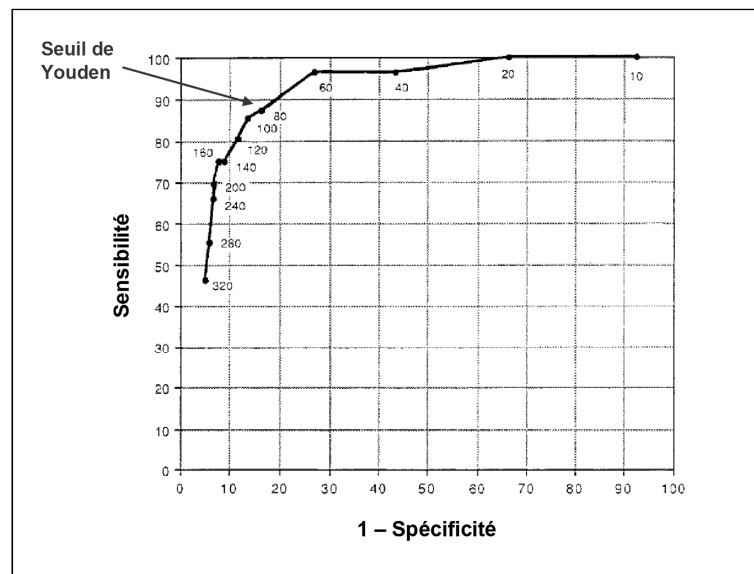


Figure 9 : Courbe ROC de l'acide hyaluronique pour le diagnostic de cirrhose chez 191 patients avec une hépatopathie chronique compensée; d'après Oberti et al (108).

L'AUROC a donc l'avantage d'être un indice qui résume en un seul chiffre la performance globale d'un test diagnostic, indépendamment du seuil diagnostic. Cet indice est informatif dans les études visant à évaluer et à comparer les performances globales des tests non invasifs de fibrose.

Cependant, comme nous l'avons précisé plus haut, les résultats de ces études doivent être interprétés avec précaution, en prenant en considération les effets "sélection", "PBH", et "discordance" qui influencent la valeur des AUROCs des tests (cf § 3.1). Par ailleurs, l'AUROC est un indice peu informatif pour la pratique clinique. Rappelons qu'elle évalue la performance intrinsèque d'un test. En effet, le clinicien a besoin de connaître les valeurs seuils du test qui affirment ou infirment le diagnostic chez son patient, afin d'orienter sa prise en charge médicale.

3.2.2 Diagnostic binaire de fibrose

Lorsqu'il prend en charge un patient avec une hépatite chronique C, le clinicien doit répondre à deux questions "binaires" (réponse : oui/non) : 1/ Faut-il débiter un traitement antiviral ? et 2/ Faut-il dépister les complications telles que le CHC ou les grosses varices œsophagiennes? Avant que les tests non invasifs de fibrose ne soient recommandés par la HAS, c'est le résultat de la PBH qui permettait de répondre à ces questions : le traitement antiviral était indiqué à partir du stade Metavir F2 (fibrose significative) et le dépistage des complications à partir du stade F4 (cirrhose). La première ambition des tests non invasifs de fibrose a donc été de répondre également à ces deux questions. Les différentes études ont ainsi cherché à déterminer les meilleurs seuils diagnostiques des tests non invasifs de fibrose permettant de prédire avec la meilleure performance la fibrose significative et la cirrhose.

3.2.2.1 Différents types de seuils diagnostiques

Plusieurs types de seuils diagnostiques peuvent être utilisés pour les tests non invasifs de fibrose :

- Le seuil "**a priori**". Il s'agit du seuil fixé arbitrairement lors de la construction du test sanguin par la RLB pour discriminer dans le modèle final les patients ayant l'événement étudié de ceux ne l'ayant pas. Ce seuil s'applique donc par définition uniquement aux tests sanguins de 2^e génération (créés par RLB), et pour le diagnostic de fibrose significative qui était la cible utilisée pour élaborer les tests sanguins. En pratique, le seuil *a priori* est fixé à 0,5 : lorsqu'elle modélise la population étudiée, la RLB identifie les variables indépendamment liées à la fibrose significative (Metavir $F \geq 2$) et calcule leur coefficient β de sorte que, dans le modèle final (test sanguin allant de 0 à 1), les patients sans fibrose significative aient une valeur $< 0,5$ et ceux avec une fibrose significative aient une valeur $\geq 0,5$.

- Le seuil "**a posteriori**". Les tests sanguins ont souvent été élaborés dans des échantillons de faible effectif (**Tableau 1**), peu représentatifs de la population générale des patients vus en pratique clinique. Dans ces conditions, le seuil *a priori* de 0,5 de la RLB est peu robuste et non exportable à la pratique clinique. Il est donc nécessaire dans un second temps (*a posteriori*) de calculer un meilleur seuil diagnostique dans une population plus représentative et de plus large effectif. Ceci est bien entendu possible pour les tests sanguins, qu'ils soient de 1^{ère} et de 2^e génération, comme pour le Fibroscan. Le seuil *a posteriori* peut

être calculé de plusieurs façons. Une technique courante est celle de *l'indice de Youden*. Elle détermine le seuil diagnostique du test pour lequel l'indice de Youden (sensibilité + spécificité - 1) est le plus élevé, offrant ainsi le meilleur compromis entre sensibilité et spécificité pour la cible diagnostique choisie (**Figure 10**). Il correspond en général au point d'inflexion de la courbe ROC du test étudié (**Figure 9**). Dans la littérature, les études ayant déterminé les seuils diagnostiques des tests non invasifs de fibrose ont majoritairement utilisé la méthode du seuil de Youden.

- Autres types de seuils diagnostiques. D'autres seuils diagnostiques peuvent être calculés (**Figure 10**) : seuil de meilleure performance diagnostique; seuil de sensibilité, spécificité, VPP, ou valeur prédictive négative (VPN) fixés à une certaine valeur (généralement 90% ou 95%).

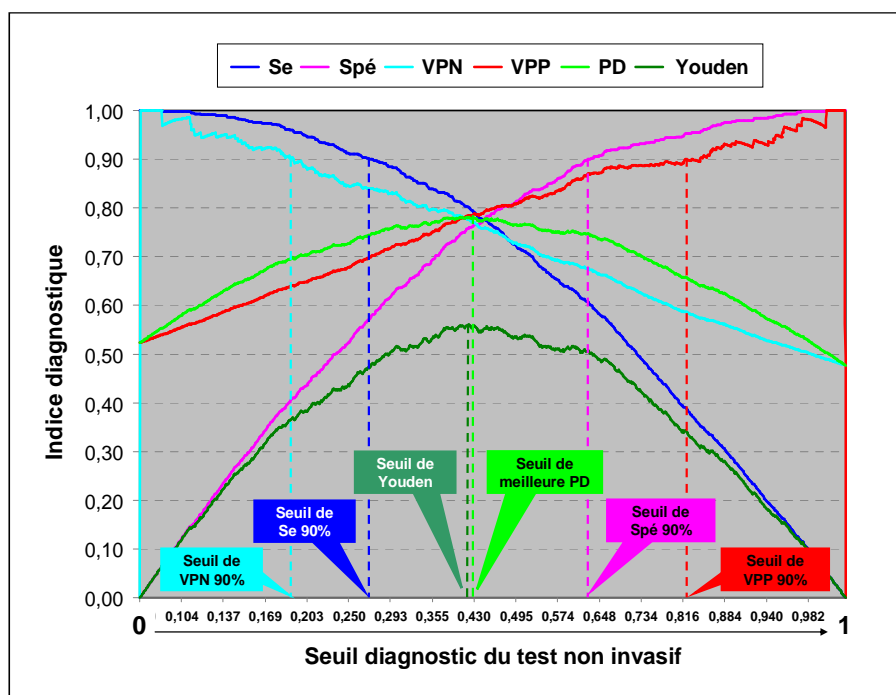


Figure 10 : Variation des indices diagnostiques en fonction de la valeur du test non invasif de fibrose (Se : sensibilité, Spé : spécificité, VPN : valeur prédictive négative, VPP : valeur prédictive positive, PD : performance diagnostique, Youden : indice de Youden).

3.2.2.2 Diagnostic binaire de fibrose significative

Par définition, le seuil *a priori* de 0,5 pour le diagnostic de fibrose significative n'est applicable qu'aux tests sanguins de 2^e génération construits par RLB. Notre équipe a calculé

les seuils *a posteriori de Youden* et ainsi comparé les performances de 5 tests sanguins (FibroMètre, Fibrotest, Hepascore, Fib4 et APRI) pour le diagnostic de fibrose significative (Metavir $F \geq 2$) dans une large cohorte de 1056 patients avec une hépatite chronique C inclus dans 9 centres français (secondaires et tertiaires, publics et privés) (48). La répartition des stades Metavir dans cette population était proche de celle décrite dans la large cohorte de référence de Thein (8) : F0 : 4,4%, F1 : 43,5%, F2 : 27,0%, F3 : 14,0%, et F4 : 11,2%. Les seuils *a posteriori de Youden* du FibroMètre, du Fibrotest, et de l'Hepascore étaient inférieurs au seuil a priori de 0,5 (**Annexe 3**). Les indices diagnostiques calculés à l'aide de ces seuils sont détaillés en **Annexe 3** : la sensibilité, la VPN, et le diagnostic odds ratio étaient les plus élevés avec le FibroMètre, alors que la spécificité et la VPP semblaient similaires entre les tests sanguins (sauf pour le Fib4). Dans cette étude, du fait d'une prévalence de la cible diagnostique proche de 50%, le calcul du seuil *a posteriori de meilleure performance diagnostique* pour la fibrose significative donnait des valeurs proches du seuil *a posteriori de Youden* pour chaque test sanguin. Il serait cependant nécessaire de valider la robustesse de ces seuils ainsi que leur performance diagnostique dans d'autres larges populations de patients avec hépatite chronique C.

De nombreuses études ont cherché à déterminer les seuils diagnostiques du Fibroscan (66, 67). Par définition, il n'existe pas de seuil diagnostique *a priori* pour le Fibroscan. Les seuils diagnostiques *a posteriori* dans des populations d'hépatite chronique C n'ont été publiés que dans 2 études (70, 71). Le seuil *a posteriori de Youden* pour le diagnostic de fibrose significative (Metavir $F \geq 2$) était calculé à 7,1 kPa dans l'étude de Castera *et al*, et à 8,8 kPa dans l'étude de Ziol *et al*. Les performances diagnostiques du Fibroscan calculées à l'aide de ces seuils semblaient similaires entre les deux travaux (**Annexe 4**). Cependant, les populations étudiées étaient d'effectif relativement faible (183 et 251 patients), et la prévalence de la fibrose significative était élevée (75,3% et 62,2%). Ainsi, il faut être prudent quant à l'utilisation de ces seuils de Fibroscan en pratique clinique où la prévalence de la fibrose significative est notamment plus faible (8). En effet, la valeur d'un seuil diagnostique calculé dans une population dépend de la prévalence de l'événement étudié dans cette population.

Dans la méta-analyse de Friedrich-Rust *et al* (incluant 8478 patients issus de 50 études ou abstracts), le seuil du Fibroscan pour le diagnostic de fibrose significative était calculé à 7,65 kPa avec une sensibilité à 80% et une spécificité à 85% (66). Dans la méta-analyse de

Stebbing *et al* (incluant 22 études et 3514 patients), le seuil pour la fibrose significative était calculé à 7,81 kPa avec une sensibilité à 72% et une spécificité à 82% (67). Bien que ces méta-analyses aient étudié de larges cohortes de patients, leurs résultats sont de nouveau à interpréter avec précaution car elles incluaient des études avec des hépatopathies de causes diverses dont les prévalences de fibrose significative étaient hétérogènes et parfois très élevées (24% à 92%). Dans leur travail, Stebbing *et al* ont calculé le seuil diagnostique de fibrose significative pour chaque étiologie d'hépatopathie, avec une valeur à 8,44 kPa pour l'hépatite chronique C. Cependant, les performances diagnostiques du Fibroscan avec ce seuil n'étaient pas précisées.

Ainsi, peu d'études ont calculé et évalué la performance de seuils *a posteriori* des tests non invasifs pour le diagnostic de la fibrose significative dans de larges cohortes de patients avec hépatite chronique C. D'après les études disponibles et avec les seuils *a posteriori* calculés (**Annexes 3 et 4**), les tests sanguins semblent avoir une meilleure sensibilité et une meilleure VPN que le Fibroscan, alors que ce dernier semble avoir une meilleure spécificité et une meilleure VPP. Néanmoins, il s'agit de résultats d'études distinctes, sans comparaison directe entre les tests sanguins et le Fibroscan. De plus, les populations étudiées n'étaient pas comparables avec notamment des prévalences de fibrose significative dans les 2 études du Fibroscan (62,2% et 75,3%) plus élevées que dans celle ayant évalué les tests sanguins (52.2%).

3.2.2.3 Diagnostic binaire de cirrhose

Par définition, il n'existe pas de seuil *a priori* pour le diagnostic de cirrhose pour les tests sanguins (du moins ceux construits pour le diagnostic de fibrose significative) et pour le Fibroscan.

Les seuils *a posteriori de Youden* du FibroMètre et du Fibrotest pour le diagnostic de cirrhose ont été calculés dans notre cohorte de 1056 patients avec hépatite chronique C (59). La valeur de ces seuils et leur performances diagnostiques sont détaillées en **Annexe 3**.

Pour le Fibroscan, seules sont disponibles les données des 2 études ayant calculé les seuils *a posteriori de Youden* dans l'hépatite chronique C (**Annexe 4**).

3.2.2.4 Limites des seuils de diagnostic binaire

D'un point de vue pratique, les données de la littérature sont insuffisantes pour une utilisation en pratique clinique des seuils de diagnostic binaire dans l'hépatite chronique C. En effet, il est encore nécessaire de : 1/ calculer les meilleurs seuils de Fibroscan pour le diagnostic de fibrose significative et de cirrhose, et 2/ valider les seuils des tests sanguins et du Fibroscan pour le diagnostic de fibrose significative et de cirrhose. Ceci nécessiterait de large cohortes de patients avec hépatite chronique C pour lesquels la réalisation de la PBH ne soit pas sélectionnée par les tests non invasifs de fibrose, afin que les populations étudiées soient représentatives des patients vus en pratique clinique. La mise en place d'études remplissant cette dernière condition semble cependant difficile en France depuis la publication des recommandations de la HAS qui autorisent l'utilisation des tests non invasifs en première ligne.

D'un point de vue méthodologique, le diagnostic binaire implique probablement une vision "trop binaire" du diagnostic non invasif de la fibrose hépatique. En effet, du fait de leur construction par RLB, la performance des tests sanguins pour une cible diagnostique est maximale aux valeurs extrêmes du test et médiocre pour les valeurs intermédiaires (cf § 2.2.1 et **Figure 1**). Ceci est également valable pour le Fibroscan. Un diagnostic binaire implique une suppression de la zone grise et donc un changement de diagnostic pour une faible variation de la valeur du test lorsqu'elle est proche du seuil diagnostique. Par exemple, si on considère la seuil du FibroMètre à 0,419 pour le diagnostic de fibrose significative (48), un patient avec un FibroMètre à 0,405 est classé F0/1 alors qu'un patient avec un FibroMètre à 0,425 est classé F \geq 2. Néanmoins, le stade histologique de fibrose n'est pas différent entre ces deux patients n'ayant qu'un écart de 0,02 pour la valeur du FibroMètre (**Figure 11**). De même, bien que Stebbing *et al* aient proposé un seuil de 8,44 kPa pour le diagnostic de fibrose significative par le Fibroscan, les lésions histologiques ne sont probablement pas différentes entre deux patients ayant respectivement des résultats à 8,2 kPa et 8,6 kPa.

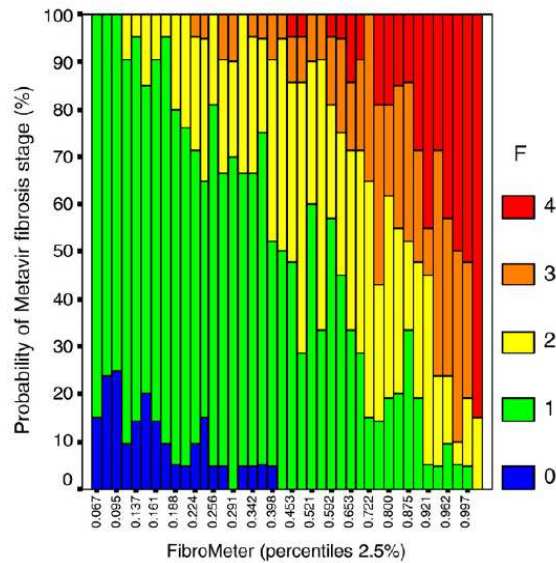


Figure 11 : Pourcentage de stades de fibrose Metavir (F) en fonction des valeurs du FibroMètre exprimées en 2,5% percentiles chez 1056 patients avec hépatite chronique C ; d'après Leroy et al (55).

De plus, le calcul d'un seuil diagnostique implique inmanquablement une imprécision liée aux calculs statistiques. C'est pourquoi il serait très informatif de préciser l'intervalle de confiance à 95% des seuils diagnostiques. En effet, la pertinence d'un seuil diagnostique pourrait être remise en cause si les performances du test sont très différentes entre les seuils situés aux bornes de l'intervalle de confiance.

Un diagnostic binaire induit également des "pseudo-discordances" entre les tests non invasifs de fibrose lorsque les résultats de ces tests sont proches des seuils diagnostiques. Par exemple, on peut parler de "vraie discordance" lorsque le résultat du FibroMètre est 0,115 (<0,419 donc F0/1) et celui du Fibroscan 25,8 kPa (>8,44 donc $F \geq 2$). Par contre, bien que les résultats du diagnostic binaire soient les mêmes que dans la situation précédente, un FibroMètre à 0,405 n'est pas réellement discordant avec un Fibroscan à 8,6 kPa.

Enfin, dans sa pratique, le clinicien souhaite établir un diagnostic pour son patient et a donc besoin d'un test ayant de bonnes valeurs prédictives. Les études ayant déterminé les seuils diagnostiques des tests non invasifs ont souvent calculé le seuil *a posteriori* de Youden qui optimise la sensibilité et la spécificité du test. Cependant, les valeurs prédictives négative et positive sont insuffisantes lorsque les tests de fibrose sont utilisés avec ce seuil *a posteriori* de Youden (**Annexes 3 et 4**) : pour le diagnostic de fibrose significative, les valeurs

prédictives des tests sanguins ne dépassent pas 80% et la VPP du Fibroscan est très faible (50-60%). Seule la VPN du Fibroscan est très bonne (90-95%). Pour le diagnostic de cirrhose, les VPN des tests sanguins et du Fibroscan sont excellentes ($\geq 95\%$). Cependant, la VPP du Fibroscan est insuffisante ($< 80\%$) et celle des tests sanguins ne dépasse pas 30%. Ces résultats correspondent cependant à des valeurs prédictives « globales » et plus le résultat du test s'approche de ses valeurs extrêmes, plus la valeur prédictive augmente.

3.2.3 Profil de performance

La performance des tests non invasifs de fibrose dépend de la prévalence des stades de fibrose dans la population étudiée. Afin de contourner cette limite, Halfon *et al* ont proposé une nouvelle méthode d'évaluation des tests non invasifs : le profil de performance (102). Le profil de performance décrit le taux de patient mal classés par le seuil diagnostique pour la fibrose significative, en fonction du stade de fibrose (**Figure 12**). La prévalence des stades de fibrose n'a pas d'influence sur cette évaluation du test non invasif car le taux de patients mal classés est calculé pour chaque stade de fibrose histologique.

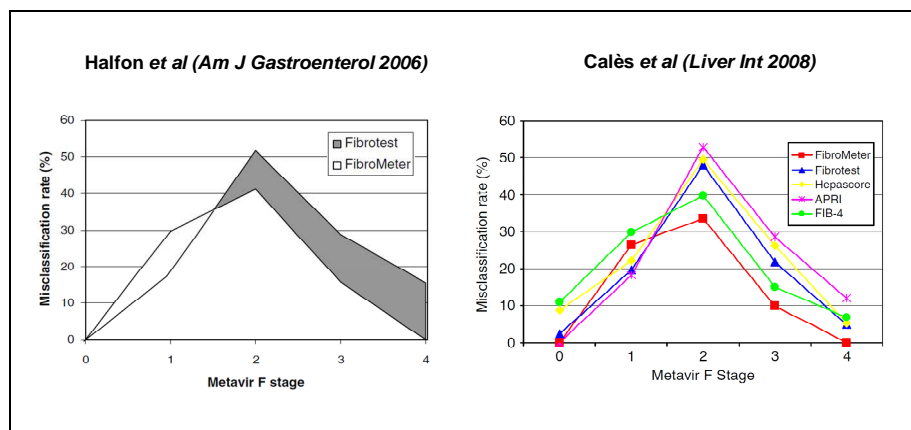


Figure 12 : Profil de performance. Taux de patients mal classés par le seuil diagnostique pour la fibrose significative de 5 tests sanguins, en fonction du stade histologique de fibrose ; d'après Halfon *et al* et Calès *et al* (48).

Trois études ont décrit le profil de performance des tests sanguins dans l'hépatite chronique C. Les résultats de ces travaux étaient similaires et montraient que les tests sanguins avaient un profil de performance en « V » : le taux de patients mal classés était faible pour les stades extrêmes de fibrose (F0 et F4) et plus élevé pour les stades

intermédiaires de fibrose avec un pic pour le stade F2 (**Figure 12**). Le FibroMètre était le test sanguin qui offrait le meilleur profil de performance (élevée et mieux répartie), notamment pour les stades de fibrose Metavir $F \geq 2$ (**Figure 12**).

Le profil de performance a donc l'avantage d'être une méthode d'évaluation de la performance des tests non invasifs de fibrose qui n'est pas influencée par la prévalence des stades de fibrose dans la population étudiée. Cependant, cette méthode se rapproche du diagnostic binaire de fibrose significative (avec ses inconvénients cités plus haut) car elle nécessite l'utilisation et donc le choix d'un seuil diagnostique pour la fibrose significative. De plus, le profil de performance emprunte le chemin inverse du raisonnement clinique : il montre la performance d'un test non invasif en fonction du stade histologique, alors que le clinicien veut connaître la performance pour le diagnostic histologique en fonction de la valeur du test. Il n'a donc qu'une valeur évaluative.

3.2.4 Associations de tests non invasifs

Plusieurs associations de tests ont été proposées pour améliorer la performance du diagnostic non invasif de la fibrose hépatique. L'objectif était de combiner les tests non invasifs dans des algorithmes décisionnels afin d'obtenir un diagnostic fiable chez le plus grand nombre de patients, et ainsi de réduire le taux de patients avec un diagnostic incertain pour lesquels une PBH est nécessaire.

Plusieurs types d'algorithmes décisionnels peuvent être individualisés, selon la manière d'utiliser les tests non invasifs et selon le type de tests utilisés.

Certains algorithmes décisionnels utilisent les tests non invasifs de fibrose de façon séquentielle : un test sanguin de 1^{ère} génération simple à calculer est utilisé dans un premier temps ; en cas de diagnostic indéterminé, un test plus complexe de 2^e génération est calculé ; enfin, le recours à la PBH s'effectue dans un 3^{ème} temps si le diagnostic reste incertain (109-111). D'autres algorithmes décisionnels sont basés sur la concordance et nécessitent donc l'utilisation d'emblée de 2 tests non invasifs : le diagnostic non invasif est accepté lorsque les résultats des tests sont concordants, et la PBH est réservée aux cas discordants (70, 103, 112-114).

D'autre part, on peut distinguer les algorithmes associant uniquement des tests sanguins (103, 109-112) de ceux associant un test sanguin avec le Fibroscan (70, 113, 114).

Actuellement deux algorithmes décisionnels ont été évalués et directement comparés dans plusieurs études : le Sequential Algorithm for Fibrosis Evaluation (SAFE) (115) et l'algorithme de Bordeaux (70).

3.2.4.1 SAFE

Le SAFE est un algorithme décisionnel qui utilise de façon séquentielle deux tests sanguins de fibrose : l'APRI en 1^{ère} intention, le Fibrotest en seconde, et la PBH en dernier recours (115). Les seuils diagnostiques utilisés pour interpréter les résultats des tests sanguins sont ceux publiés dans l'étude princeps de l'APRI (40) et ceux du compteur du Fibrotest (cf § 2.2.3). Il existe 3 algorithmes SAFE : un pour le diagnostic de fibrose significative, un pour la cirrhose, et un dernier pour les deux cibles diagnostiques (**Figure 13**).

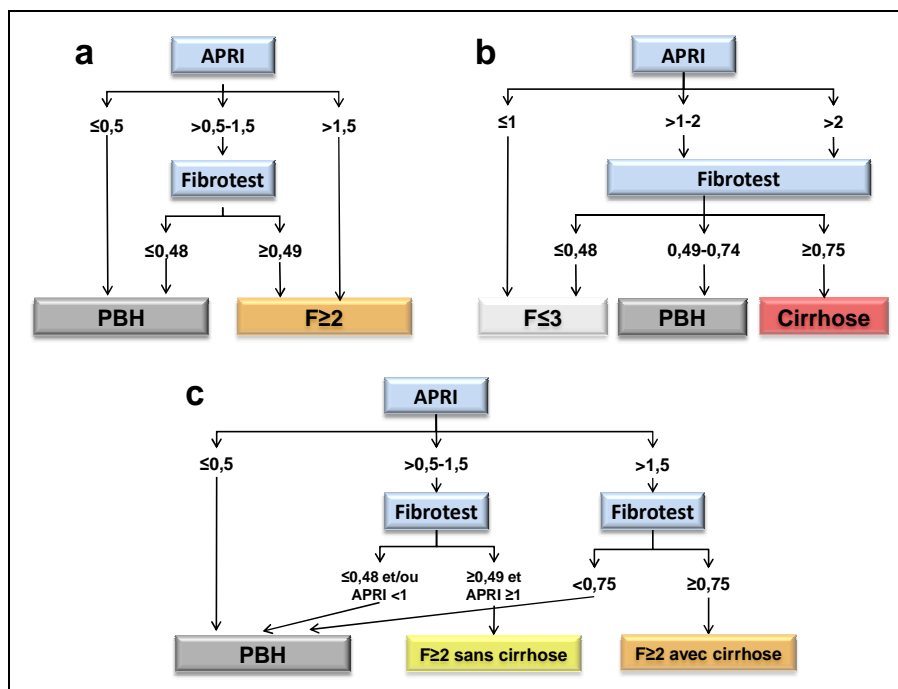


Figure 13 : Sequential Algorithm for Fibrosis Evaluation (SAFE). Algorithmes décisionnels séquentiels associant l'APRI et le Fibrotest pour le diagnostic non invasif de fibrose significative (a), de cirrhose (b), ou les deux (c) (115).

Quatre études de validation ont évalué la performance du SAFE dans l'hépatite chronique C (**Annexe 5**). Les populations de ces travaux avaient des caractéristiques différentes : la prévalence de la fibrose significative variait de 46% à 76%, et celle de la cirrhose de 9% à 25%. Le SAFE avait une excellente performance pour le diagnostic de fibrose significative avec 90% à 97% de bien classés, mais la PBH était requise chez 52% à 64% des patients. La performance du SAFE pour le diagnostic de cirrhose était très bonne avec 89 à 93% de biens classés, pour un taux de PBH nécessaire entre 6% à 29%. A noter que notre équipe, au contraire des autres études, a montré une très faible sensibilité (62%) et VPP (56%) pour le diagnostic de cirrhose de SAFE (116).

3.2.4.2 Algorithme de Bordeaux

L'algorithme de Bordeaux est un algorithme décisionnel basé sur la concordance entre le résultat d'un test sanguin de fibrose (Fibrotest) et celui du Fibroscan. Dans l'étude princeps (70), les auteurs ont calculé les meilleurs seuils de Youden du Fibroscan pour le diagnostic de fibrose significative (Metavir $F \geq 2$), de fibrose sévère (Metavir $F \geq 3$), et de cirrhose (Metavir F4). A l'aide de ces seuils diagnostiques et de ceux du compteur du Fibrotest, un algorithme où la PBH était réservée aux discordances entre les deux tests non invasifs a été proposé (**Figure 14**).

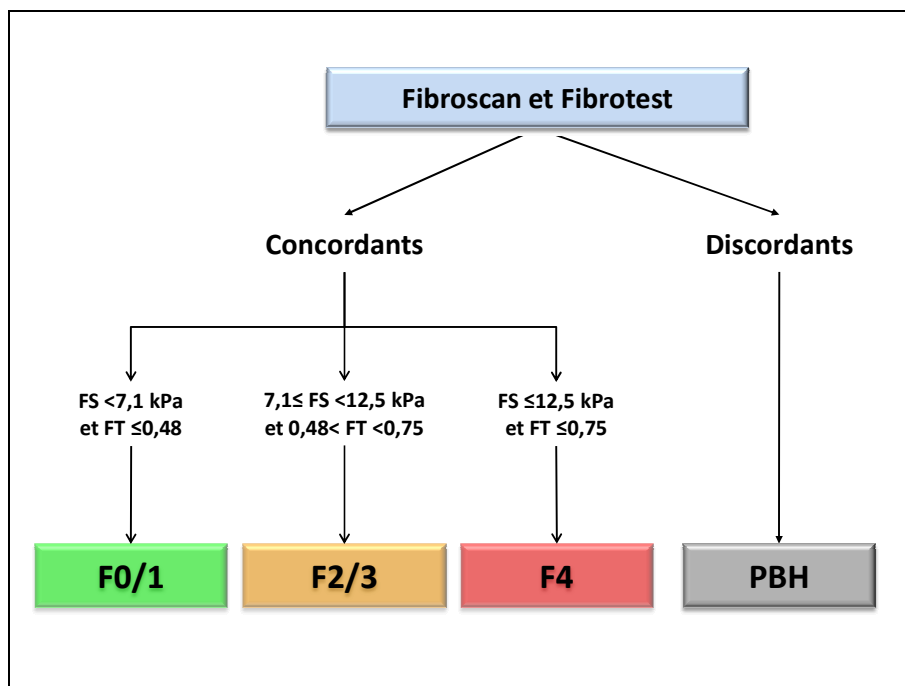


Figure 14 : Algorithme de Bordeaux. Algorithme décisionnel basé sur la concordance entre les résultats du Fibrotest et du Fibroscan (70).

Deux études de validation ont évalué la performance de l'algorithme de Bordeaux dans l'hépatite chronique C (**Annexe 5**). Les populations de ces 2 études étaient sensiblement différentes avec notamment des prévalences de fibrose significative à 58% et 76%, et de cirrhose à 15% et 25%. La performance diagnostique et le taux de PBH étaient cependant similaires dans les 2 études. La performance de l'algorithme de Bordeaux était très bonne pour le diagnostic de fibrose significative avec 88% de biens classés, et était excellente pour le diagnostic de cirrhose avec 95% de biens classés. Les taux de PBH nécessaires pour le diagnostic de fibrose significative étaient de 28 et 35%, et de 21 et 25% pour le diagnostic de cirrhose. Par contre, les résultats des valeurs prédictives étaient sensiblement différents entre les 2 études avec notamment des VPN pour la fibrose significative à 67% et 85%, et des VPP pour le diagnostic de cirrhose à 76% et 95%.

Ces deux études ont également comparé les performances de l'algorithme de Bordeaux à celles du SAFE (**Annexe 5**). Dans les deux travaux, la performance du SAFE pour le diagnostic de fibrose significative était significativement supérieure à celle de l'algorithme de Bordeaux, mais le taux de PBH avec le SAFE était également nettement significativement supérieur. Pour le diagnostic de cirrhose, l'algorithme de Bordeaux avait une performance significativement supérieure à celle du SAFE dans les deux études ; les taux de PBH n'étaient pas significativement différents entre les deux algorithmes dans l'étude de Castera *et al* alors que notre équipe montrait un taux de PBH très significativement inférieur avec le SAFE.

3.2.4.3 Limites des algorithmes décisionnels

Les algorithmes décisionnels qui combinent les tests non invasifs de fibrose ont pour principal avantage d'offrir une très bonne performance pour les deux cibles diagnostiques d'intérêt en pratique clinique (fibrose significative et cirrhose), avec des taux de bien classés globalement supérieurs à 90%.

Ces algorithmes décisionnels ont cependant plusieurs limites :

- Les algorithmes séquentiels comme le SAFE nécessitent de réaliser les tests non invasifs consécutivement, ce qui implique de **convoquer le patient plusieurs fois** (dans l'option économique) avant d'obtenir un diagnostic.

- **La réalisation d'une PBH est nécessaire pour une proportion élevée de patients** : 30 à 65% pour le diagnostic de fibrose significative, et 20 à 30% pour celui de cirrhose. Ceci n'est pas compatible avec le concept de « diagnostic non invasif » de la fibrose hépatique. Ces algorithmes décisionnels ne peuvent donc pas être proposés dans une stratégie de dépistage.

- **Les seuils diagnostiques utilisés pour l'interprétation des tests non invasifs dans les algorithmes décisionnels ne sont pas validés.** En effet, les seuils diagnostiques utilisés pour interpréter le résultat de l'APRI dans le SAFE ou du Fibroscan dans l'algorithme de Bordeaux sont ceux publiés dans les études princeps de ces tests de fibrose. Or, ces études avaient inclus des populations de faible effectif et bien souvent non représentatives des patients vu en pratique clinique. D'autre part, dans le SAFE comme dans l'algorithme de Bordeaux, le résultat du Fibrotest est interprété avec les seuils de son compteur dont la performance diagnostique n'a jamais été validée. Finalement, comme nous l'avons précisé plus haut (cf § 3.2.2.4), l'idéal serait de disposer de larges cohortes représentatives de patients avec hépatite chronique C pour recalculer les meilleurs seuils diagnostiques des tests non invasifs et ensuite y élaborer les algorithmes décisionnels

- **L'utilisation des algorithmes décisionnels en pratique clinique ne correspond pas à celle décrite dans les publications les ayant validés.**

SAFE. Il existe un SAFE pour chacune des 2 questions que se pose le clinicien : « Ce patient a-t-il une fibrose significative ? » (**Figure 13a**), et « Ce patient a-t-il une cirrhose ? » (**Figure 13b**). Cependant, ces questions ne sont pas exclusives et le clinicien a besoin de répondre aux deux pour correctement prendre en charge son patient. Il doit donc appliquer le SAFE pour le diagnostic de fibrose significative et ensuite, si le diagnostic de fibrose significative est retenu par le diagnostic non invasif, celui pour le diagnostic de cirrhose. Nous avons évalué et interprété les résultats des 2 SAFE chez 1785 patients avec hépatite chronique C (116). 303 patients avaient un diagnostic non invasif « $F \geq 2$ » par le SAFE pour la fibrose significative et « pas de cirrhose » par le SAFE pour la cirrhose (**Tableau 2**). L'interprétation de ces deux résultats donnait donc un diagnostic final « F2/3 ». De même les diagnostics non invasifs « $F \geq 2$ » et « F4 » chez 203 patients donnaient un diagnostic final « F4 ». La PBH était nécessaire chez tous les autres patients. Au total, l'interprétation des résultats des 2 SAFE permettait d'obtenir 3 classes diagnostiques : « F2/3 », « F4 », et « PBH

nécessaire » (**Tableau 2**). Cette classification avait 87,3% de performance diagnostique versus 94,6% avec le SAFE pour la fibrose significative ($p < 10^{-3}$) et 89,5% avec le SAFE pour la cirrhose ($p < 10^{-3}$). Le taux de PBH nécessaire était de 70,8% versus 64,0% avec le SAFE pour la fibrose significative ($p < 10^{-3}$) et 6,4% avec le SAFE pour la cirrhose ($p < 10^{-3}$). Enfin, il était nécessaire de réaliser un Fibrotest chez 49,2% des patients versus 35,8% avec le SAFE pour la fibrose significative ($p < 10^{-3}$) et 22,2% avec le SAFE pour la cirrhose ($p < 10^{-3}$).

Tableau 2 : *Interprétation des résultats du SAFE pour la fibrose significative (Metavir $F \geq 2$) et du SAFE pour la cirrhose (F4) chez 1785 patients avec hépatite chronique C (116). En italique : diagnostic final (en gras) et taux de patients bien classés par le diagnostic final.*

		Résultat Algorithme SAFE pour $F \geq 2$	
		PBH	$F \geq 2$
Résultat	$F \leq 3$	<i>PBH</i>	<i>F2/3</i>
algorithme		<i>(1116/1116)</i>	<i>(185/303)</i>
SAFE	F4	<i>PBH</i>	<i>F4</i>
pour F4		<i>(0/0)</i>	<i>(101/203)</i>
	PBH	<i>PBH</i>	<i>PBH</i>
		<i>(0/0)</i>	<i>(112/112)</i>

Ainsi, les algorithmes séquentiels créés pour différentes cibles diagnostiques binaires ont individuellement de très bonnes performances diagnostiques, mais il est nécessaire en pratique clinique de les appliquer successivement avec ainsi une augmentation du taux de PBH nécessaire et une diminution de la performance diagnostique.

Algorithme de Bordeaux. L'algorithme de Bordeaux donne un diagnostic en 4 classes (**Figure 14**). Cependant, il a été évalué pour le diagnostic binaire de fibrose significative et celui de cirrhose (**Annexe 5**). Evaluer la performance d'une classification pour un diagnostic binaire réduit artificiellement le taux de mal classés. Nous avons évalué la performance de l'algorithme de Bordeaux à 4 classes (F0/1, F2/3, F4, PBH) chez 729 patients avec hépatite chronique C (116). Le taux de patients bien classés par l'algorithme à 4 classes était 84,7% versus 88,3% lorsque la cible était le diagnostic binaire de fibrose significative ($p = 10^{-3}$, test de MacNemar) et 94,2% lorsque la cible était le diagnostic binaire de cirrhose ($p < 10^{-3}$). Le taux de PBH nécessaire avec l'algorithme à 4 classes était égal à 49,8% versus

34,6% pour le diagnostic binaire de fibrose significative ($p < 10^{-3}$) et 24,6% pour celui de cirrhose ($p < 10^{-3}$).

Ainsi, la performance de l'algorithme de Bordeaux à 4 classes est surestimée dans les études publiées car elles ne l'ont évaluée que pour des cibles diagnostiques binaires. En pratique clinique, la performance de cet algorithme diminue légèrement, mais le taux de PBH nécessaire augmente considérablement et avoisine les 50%.

Au total, combiner les tests non invasifs dans des algorithmes décisionnels pour augmenter la performance diagnostique est une approche séduisante. Cependant, le taux de patients pour lesquels une PBH est nécessaire reste élevé avec les algorithmes publiés dans la littérature. Par ailleurs, l'évaluation des performances diagnostiques de ces algorithmes n'a pas été conforme à leur utilisation en pratique clinique. Enfin, il est nécessaire d'optimiser les seuils diagnostiques des tests non invasifs utilisés dans les algorithmes décisionnels.

3.2.5 Optimisation des valeurs prédictives

Aucun test non invasif de fibrose ne permet d'obtenir une très bonne VPN et une très bonne VPP avec un simple seuil diagnostique (**Figure 10**). La performance des tests non invasifs étant maximale à leurs valeurs extrêmes, il est possible de définir deux intervalles de valeurs du test où le diagnostic (par exemple, le % de bien classés pour la fibrose significative) est fiable. En pratique, deux seuils sont calculés : le seuil de VPN à 90% (ou 95% selon le degré d'exigence) pour le diagnostic de fibrose significative et le seuil de VPP à 90% (**Figure 15**). Lorsque le résultat du test diagnostique est inférieur au seuil de VPN à 90% (intervalle de $VPN \geq 90\%$) du test, le patient a au moins 90% de chance de ne pas avoir de fibrose significative. Au contraire, lorsque le résultat du test est supérieur au seuil de VPP à 90% (intervalle de $VPP \geq 90\%$), le patient a au moins 90% de risque d'avoir une fibrose significative. Enfin, lorsque le résultat du test est inclus dans l'intervalle situé entre les deux seuils, le diagnostic n'est pas assez fiable et indéterminé (VPN ou $VPP < 90\%$). Cette optimisation des valeurs prédictives pour le diagnostic de fibrose significative permet donc d'obtenir 3 classes diagnostiques : F0/1, indéterminé, et $F \geq 2$. Il est également possible avec cette même méthode d'optimiser les valeurs prédictives pour le diagnostic de cirrhose et ainsi d'obtenir les 3 classes diagnostiques suivantes : $F \leq 3$, indéterminé, et F4.

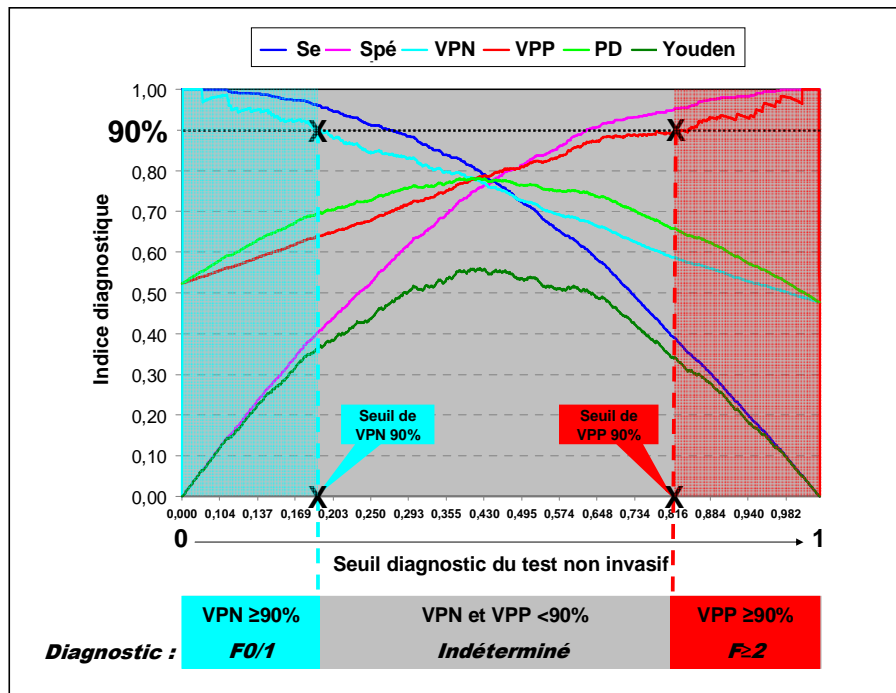


Figure 15 : Intervalles de valeurs prédictives $\geq 90\%$ pour le diagnostic de fibrose significative. Deux seuils sont calculés: le seuil de valeur prédictive négative (VPN) à 90% pour la fibrose significative et le seuil de valeur prédictive positive (VPP) à 90%. Ces deux seuils permettent d'obtenir 3 intervalles de diagnostic : F0/1, indéterminé, et F ≥ 2 .

Un test non invasif de fibrose performant est donc celui qui inclue le taux le plus élevé de patients dans les intervalles de valeurs prédictives élevées (par exemple $\geq 90\%$). Ces intervalles ont été déterminés pour 5 tests sanguins dans notre cohorte de 1056 patients avec hépatite chronique C (48, 59). Le FibroMètre était le test sanguin le plus performant car il incluait 43,5% des patients dans les intervalles de valeurs prédictives $\geq 90\%$ pour le diagnostic de fibrose significative, versus APRI : 19,6%, Fibrotest : 17,1%, Hepascore : 3,9%, et Fib4 : 1,7% ($p < 10^{-3}$ entre les tests, **Figure 16**) (48). Pour le diagnostic de cirrhose, les taux de patients inclus par le FibroMètre (standard) dans les intervalles de valeurs prédictives à 90%, 95%, et 100% étaient, respectivement : 98,0%, 85,6%, et 47,5% (59).

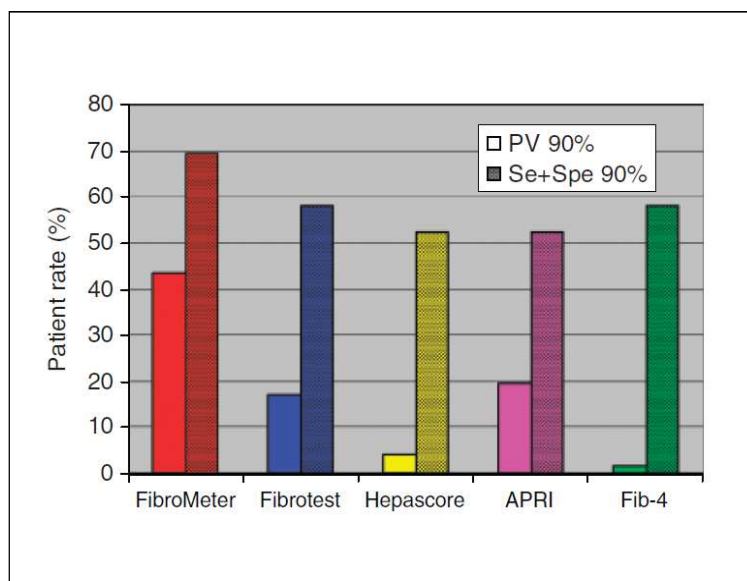


Figure 16 : Taux de patients inclus dans les intervalles de valeurs prédictives négative/positive $\geq 90\%$ ou de sensibilité/spécificité $\geq 90\%$ pour le diagnostic de fibrose significative par 5 tests sanguins de fibrose. PV : valeurs prédictives; Se : sensibilité, Spe : spécificité ; d'après Calès et al (48).

La méthode des intervalles de valeurs prédictives a donc l'avantage de préciser les modalités d'utilisation des tests non invasifs en pratique clinique. En effet, elle détermine les valeurs du test qui permettent au clinicien d'affirmer ou d'exclure de façon fiable le diagnostic chez son patient.

Cette méthode a cependant des limites. Elle laisse une « zone grise » où le diagnostic reste indéterminé. Ainsi, avec le meilleur test sanguin, il est nécessaire de réaliser une PBH chez plus de 55% des patients. D'autre part, les valeurs prédictives dépendent de la prévalence de la cible diagnostique dans la population étudiée. Les seuils de valeur prédictive positive et négative doivent donc être calculés dans une large population représentative des patients vus en pratique clinique, et ensuite être validées dans une population indépendante avant une utilisation en pratique clinique.

3.2.6 Intervalles de diagnostic fiable

Notre équipe a développé le concept des « intervalles de diagnostic fiable ». Cette méthode consiste à déterminer dans un premier temps les intervalles de valeurs prédictives pour la cible diagnostique (par exemple : intervalles de valeurs prédictives $\geq 90\%$ pour le diagnostic de fibrose significative). Dans un second temps, la zone indéterminée ou grise

entre les deux intervalles de valeurs prédictives est divisée en deux nouveaux intervalles selon le seuil diagnostique (par exemple : seuil diagnostique *a posteriori* de Youden pour la fibrose significative). Le diagnostic de stade histologique dans ces deux nouveaux intervalles est différent de la cible diagnostique initiale et correspond à l'intervalle des stades de fibrose dont la prévalence est $\geq 90\%$ (**Figure 17**).

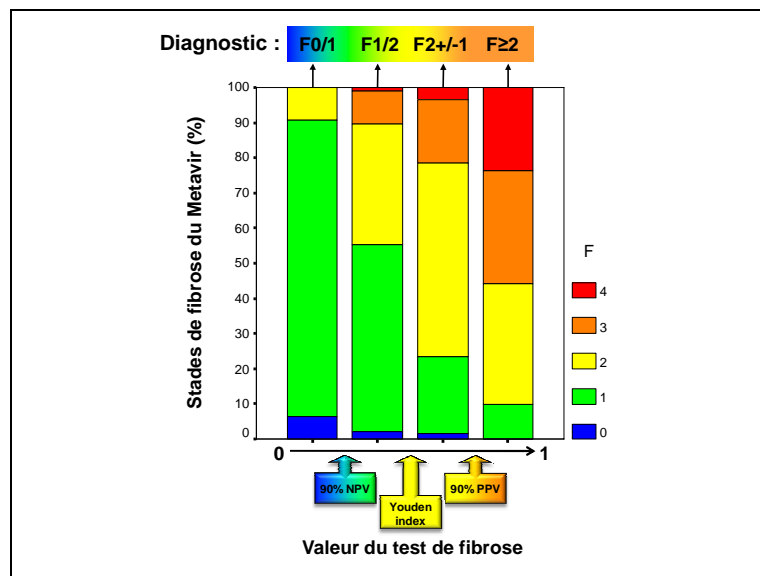


Figure 17 : Intervalles de diagnostic fiable pour la fibrose significative. La valeur du test de fibrose est divisée en 4 intervalles. Les deux intervalles extrêmes correspondent aux intervalles de valeurs prédictives (négative et positive) $\geq 90\%$. La zone grise entre ces deux intervalles est divisée en deux nouveaux intervalles selon le seuil *a posteriori* de Youden pour le diagnostic de fibrose significative. Le diagnostic pour chacun de ces deux nouveaux intervalles correspond à la somme des stades de fibrose dont la prévalence atteint $\geq 90\%$.

Nous avons déterminé les intervalles de diagnostic fiable du FibroMètre pour la fibrose significative (48) et pour la cirrhose (59) dans notre cohorte de 1056 patients avec hépatite chronique C.

Lorsque la cible diagnostique était la fibrose significative (Metavir $F \geq 2$), le FibroMètre pouvait être divisé en 4 intervalles de diagnostic fiable ($VP \geq 90\%$) : F0/1, F1/2, F2 \pm 1, et F \geq 2 (48). 89,5% des patients étaient correctement classés par cette nouvelle classification, sans aucune PBH nécessaire. La performance des intervalles de diagnostic fiable du FibroMètre pour la fibrose significative a été testée et validée dans d'autres populations de patients avec hépatite chronique C (57).

La cirrhose est un diagnostic important car c'est à ce stade que surviennent les complications qui engagent le pronostic vital des patients, et qu'il faut débiter un dépistage du carcinome hépatocellulaire et des grosses varices à risque hémorragique. Nous avons donc déterminé les intervalles de diagnostic fiable du FibroMètre pour la cirrhose, mais en fixant un seuil à 95% pour les intervalles de valeurs prédictives (59). Au final, le FibroMètre pouvait être divisé en 3 intervalles de diagnostic fiable (pas de cirrhose, $F \geq 2$, et cirrhose) qui classaient correctement 95,1% des patients sans aucune PBH nécessaire.

La méthode des intervalles de diagnostic fiable permet donc un diagnostic non invasif performant chez 100% des patients, sans aucune PBH nécessaire. L'inconvénient de cette méthode est qu'elle aboutit à une classification moins précise que celle du Metavir, le diagnostic final étant un intervalle incluant aux moins deux stades de fibrose. Ceci peut poser des problèmes décisionnels pour la prise en charge des patients, par exemple pour l'indication du traitement antiviral lorsque le diagnostic est F1/2.

En conclusion, de nombreuses méthodes ont été utilisées pour évaluer la performance diagnostique des tests non invasifs de fibrose. Elles ont toutes leurs propres avantages et inconvénients (**Annexe 6**). Les méthodes permettant d'améliorer le diagnostic individuel des patients (intervalles de valeurs prédictives ou de diagnostic fiable) sont celles d'intérêt pour la pratique clinique quotidienne.

3.3 Reproductibilité et robustesse des tests non invasifs de fibrose

Outre la performance diagnostique des tests de fibrose, il faut également évaluer leur reproductibilité et leur robustesse. La **reproductibilité** étudie la variabilité de la mesure d'un test de fibrose à un temps donné, par exemple par différents opérateurs chez un même patient pour le Fibroscan ou par différents laboratoires sur le même prélèvement pour les tests sanguins. La **robustesse** d'un test correspond à la reproductibilité de sa performance (sa stabilité), par exemple dans une large population indépendante.

3.3.1 Reproductibilité

Nous avons évalué la reproductibilité des tests sanguins entre différents laboratoires d'analyses biologiques : le laboratoire du CHU d'Angers, un laboratoire spécialisé (Mérieux / Biomnis, Lyon, France), et plusieurs laboratoires de 1^{ère} ligne (51). La reproductibilité inter-laboratoire du FibroMètre était excellente avec des coefficients de corrélation intra-classe qui variaient selon les sous-études entre 0,942 et 0,991. Une sous-étude ayant directement comparé les différents tests sanguins montrait qu'ils avaient tous une excellente reproductibilité avec les coefficients de corrélation suivants : 0,963 pour le FibroMètre, 0,984 pour le Fibrotest, et 0,949 pour l'APRI.

La reproductibilité de la mesure de la dureté hépatique par le Fibroscan s'évalue par la concordance inter-observateur. Elle est globalement excellente : selon les études, le coefficient de corrélation intra-classe varie entre 0,87 et 0,98 pour la concordance inter-observateur, et entre 0,94 et 0,98 pour la concordance intra-observateur (74, 75, 117-121). Cependant, plusieurs facteurs influencent la concordance inter-observateur : le rapport IQR/médiane (74, 119, 121), le niveau de dureté/fibrose hépatique (74, 118), le BMI (74, 118), le site de mesure (74), et la présence d'une stéatose hépatique (118). Notre équipe a ainsi montré que la reproductibilité inter-observateur était médiocre pour les faibles valeurs de dureté hépatique, celles correspondant aux stades intermédiaires et faibles de fibrose (74). Dans notre étude, l'écart de mesure entre les deux opérateurs était en moyenne de $\pm 30\%$. Ainsi, pour une dureté hépatique mesurée à 10 kPa, le résultat du second opérateur pouvait aller de 7 kPa (en faveur d'une fibrose significative) à 13 kPa (en faveur d'une fibrose sévère, voire d'une cirrhose). Nos résultats montraient également que la concordance inter-observateur diminuait lorsque le BMI était ≥ 25 , le rapport IQR/médiane $> 25\%$, et lorsque la mesure de la dureté hépatique n'était pas effectuée sur le site de référence (au croisement entre l'espace intercostal situé sous la limite supérieure de la matité hépatique et la ligne axillaire médiane).

Finalement, nous avons comparé la reproductibilité des résultats de la PBH, du Fibroscan et du FibroMètre dans plusieurs études (122). Après avoir converti les résultats du FibroMètre et du Fibroscan en stades Metavir, les indices kappa étaient les suivants : 0,336 pour la PBH lue par un pathologiste de 1^{ère} ligne (faible reproductibilité), 0,640 pour le Fibroscan (bonne reproductibilité), et 0,874 pour le FibroMètre (excellente reproductibilité).

Ces résultats permettent de classer les tests de fibrose selon leur reproductibilité « dans la vraie vie » : tests sanguins > Fibroscan >> PBH.

3.3.2 Robustesse

Nous avons évalué la robustesse des tests sanguins dans une large population de 825 patients avec hépatite chronique C et une répartition représentative des stades de fibrose (F0 : 5%, F1 : 47%, F2 : 25%, F3 : 12%, F4 : 11%) (55). Une RLB incluant toutes les variables du test a été réalisée pour chaque test sanguin. La RLB pour l'Hepascore ne sélectionnait plus la bilirubine et celle pour le Fibrotest ne sélectionnait ni l'âge ni la bilirubine. Ces variables n'avaient donc pas de valeur indépendante et n'apportaient pas d'information supplémentaire pour le diagnostic non invasif de fibrose dans la population étudiée, montrant ainsi que les tests sanguins correspondants n'étaient pas robustes. Pour le FibroMètre, toutes les variables composites étaient de nouveau sélectionnées par la RLB. Ces résultats ont été confirmés avec une technique bootstrap dans notre cohorte de 1056 patients avec hépatite chronique C (123). La RLB a été répétée sur 1000 échantillons déterminés par tirage au sort avec remise à partir de la population initiale. Chaque variable du FibroMètre était sélectionnée par la RLB dans au moins 700 échantillons, alors que la bilirubine du Fibrotest et de l'Hepascore n'était sélectionnée respectivement que dans 25 et 96 échantillons (**Figure 18**).

Figure 18 :

Number of significant independent predictions for significant fibrosis of markers included in blood tests among 1000 samples according to bootstrap resampling in population #1. The most robust markers (selected in >94% of the bootstrap models) are in roman characters and the least robust (selection <77%) in italic characters.

Marker	FibroMeter		Fibrotest	Hepascore
	2G	3G		
Alpha 2 macroglobulin	1000	1000	1000	1000
AST	999	997	-	-
Age	991	999	1000	990
Platelets	980	989	-	-
Prothrombin index	969	997	-	-
Sex	968	943	891	980
Haptoglobin	-	-	965	-
Apolipoprotein A1	-	-	989	-
GGT	-	723	1000	992
Urea	768	630	-	-
Hyaluronic acid	698	-	-	999
Bilirubin	-	-	25	96

Calès et al, Clin Chem 2010

Le seuil diagnostique de chaque test sanguin pour le diagnostic binaire de fibrose significative a également été recalculé et, comparativement aux seuils diagnostiques initialement publiés, le taux de patients pour lesquels le diagnostic était modifié avec le nouveau seuil était de 18,2% pour l'APRI, 5,8% pour le Fibrotest, 2,4% pour l'Hepascore, et 0,8% pour le FibroMètre. La reproductibilité du seuil diagnostique était donc variable selon les tests.

L'ensemble de ces résultats suggéraient que le FibroMètre était le test sanguin offrant la plus grande robustesse.

4. TRAVAUX PERSONNELS

4.1 Présentation du travail de thèse

Depuis décembre 2008, La HAS recommande en première intention l'utilisation d'un test non invasif de fibrose (Fibroscan ou Fibrotest ou FibroMètre ou Hepascore) pour le diagnostic de fibrose hépatique ou de cirrhose chez l'adulte avec une hépatite chronique C non traitée et sans comorbidité. Si le résultat du test non invasif utilisé est discordant avec la clinique ou si le résultat n'est pas interprétable, un des trois autres tests ou la PBH peut être réalisé en seconde intention. Ces recommandations constituent une réelle avancée dans la prise en charge des patients avec une hépatite chronique C car il est désormais possible de se passer de l'examen invasif qu'est la PBH pour évaluer le degré de fibrose hépatique.

Cependant ces recommandations restent assez théoriques car elles ne détaillent pas les modalités d'utilisation des tests non invasifs de fibrose en pratique clinique. En effet, il n'est pas précisé si les tests de fibrose doivent être utilisés à l'aide de seuils diagnostiques qui permettent de répondre à une question binaire (présence/absence de fibrose significative, présence/absence de cirrhose) ou avec leur compteur de fibrose qui permet d'estimer le stade Metavir. Concernant les seuils de fibrose, il n'est pas précisé lesquels sont à utiliser. Pour les raisons que nous avons détaillées plus haut (cf 3.1.1), les seuils diagnostiques calculés dans les études publiées varient d'un travail à l'autre et ne sont donc actuellement pas clairement définis. Enfin, les performances diagnostiques des compteurs de fibrose du Fibroscan, du Fibrotest, et du FibroMètre (non disponible pour l'Hepascore) ont été insuffisamment évaluées pour une utilisation en pratique clinique.

L'objectif de notre travail de Thèse était donc d'améliorer l'utilisation et la performance du diagnostic non invasif de fibrose hépatique en pratique clinique au cours de l'hépatite chronique C en développant plusieurs concepts méthodologiques :

1. Création d'un test sanguin spécifiquement dédié pour le diagnostic de cirrhose (**étude SNIFF 17**)
2. Evaluation de la performance diagnostique d'associations synchrones de tests non invasifs de fibrose (**étude VINDIAG 6**)
3. Utilisation d'associations synchrones de tests non invasifs de fibrose adaptées à la cible diagnostique pour la création d'une nouvelle classification non invasive de fibrose permettant une estimation performante du stade histologique de fibrose (**étude VINDIAG 7**).

4.2 Patients et méthodes

4.2.1 Patients

4.2.1.1 Etude SNIFF 17

Cette étude a regroupé les données individuelles de 3 études indépendantes précédemment publiées (46, 124, 125) et de 2 populations non publiées (Angers et Bordeaux). Les critères d'inclusion et de non inclusion étaient similaires entre les 5 populations. Les patients avec une hépatite chronique C (sérologie et ARN du VHC positifs dans le sérum) et pour lesquels étaient disponibles une PBH ainsi que les marqueurs sanguins de fibrose ont été prospectivement inclus entre 1994 et 2007 dans 9 centres : Angers (46), Tours et 5 centres de Provence-Alpes-Côte d'Azur (124), Grenoble (125), et Bordeaux. Les 5 centres de Provence-Alpes-Côte d'Azur incluaient 3 centres de soins secondaires (dont 2 publics et 1 privé) et 2 centres de soins tertiaires (124).

Les critères de non inclusion étaient : autres causes d'hépatopathie chronique (notamment la co-infection hépatite B et/ou VIH, et une consommation de boissons alcoolisées >30 g/j dans les 5 ans précédant l'inclusion), cirrhose décompensée, et traitement anti-fibrosant dans les 6 mois précédant l'inclusion. A Tours et dans les 5 centres de Provence-Alpes-Côte d'Azur, les patients avec une PBH <15mm étaient également exclus.

1056 patients avec une hépatite chronique C, une PBH, et les tests sanguins de fibrose ont été inclus dans cette méta-analyse (Angers : 300+135 patients, Provence-Alpes-Côte d'Azur : 217, Tours : 159, Grenoble : 149, et Bordeaux : 96).

4.2.1.2 Etude VINDIAG 6

Les critères d'inclusion de cette étude étaient : patients avec une hépatopathie chronique, quelque soit l'étiologie, pour lesquels étaient disponibles une PBH, les tests sanguins de fibrose, et une mesure de la dureté hépatique par le Fibroscan. Les patients avec une complication de cirrhose (ascite, rupture de varice, infection systémique, ou carcinome hépatocellulaire) étaient exclus de l'étude.

390 patients avec une hépatopathie chronique, une PBH, les tests sanguins de fibrose, et une mesure de la dureté hépatique ont été inclus dans l'étude VINDIAG 6 : 194 à Angers (avril 2004 – juin 2007 : population exploratoire), et 196 à Bordeaux et Rabat (septembre 2003 – avril 2007 : population de validation).

4.2.1.3 Etude VINDIAG 7

L'étude VINDIAG 7 a regroupé deux populations multicentriques, une exploratoire et une de validation.

3 centres (Angers, Bordeaux, et Grenoble) ont inclus les patients dans la population exploratoire. Les critères d'inclusion étaient : patients avec une hépatite chronique C (sérologie et ARN du VHC positifs dans le sérum) pour lesquels une PBH, les tests sanguins de fibrose, et une mesure de la dureté hépatique par le Fibroscan étaient disponibles. Les patients avec une complication de cirrhose (ascite, rupture de varice, infection systémique, ou carcinome hépatocellulaire) étaient exclus de l'étude.

La population de validation était celle de l'étude multicentrique (19 centres) ANRS HC EP 23 FIBROSTAR (50). Les critères d'inclusion dans cette étude étaient : patient avec une hépatite chronique C, naïf de tout traitement ou n'ayant pas reçu de traitement permettant l'éradication virale depuis 6 mois, intervalle d'au moins 6 mois entre la PBH et le traitement antérieur pour les patients non naïfs, et PBH >20mm. Les critères de non inclusion étaient : autre cause d'hépatopathie chronique associée (co-infection hépatite B et/ou VIH, consommation chronique d'alcool ≥ 30 g/j chez l'homme et ≥ 20 g/j chez la femme, maladie métabolique du foie, hémochromatose, hépatite auto-immune, exposition à des toxiques), antécédent de maladie systémique sévère, traitement anti-VHC soit à visée anti-virale soit à visée anti-fibrosante depuis moins de 6 mois. 512 patients ont été inclus dans l'étude FIBROSTAR. Les 132 patients de l'étude FIBROSTAR déjà inclus dans la population

exploratoire de l'étude VINDIAG 7 ont été exclus de la population de validation de VINDIAG 7.

729 patients avec une hépatite chronique C, une PBH, les tests sanguins de fibrose, et une mesure de la dureté hépatique ont été inclus dans l'étude VINDIAG 7 : 349 dans la population exploratoire (Angers : 99 patients, Bordeaux : 133, Grenoble : 117), et 380 dans la population de validation.

Au total, ce travail de Thèse a regroupé 3 études incluant un total de 2175 patients (**Tableau 3**). Tous les patients ont donné par écrit leur consentement libre et éclairé avant de participer à ces études. Toutes les études étaient conformes aux règles éthiques de la déclaration d'Helsinki en vigueur, et avaient été toutes approuvées par un comité local d'éthique.

Tableau 3 : patients inclus dans les études SNIFF 17, VINDIAG 6, et VINDIAG 7

Etude	Etiologie	Population	Centres	Patients (n)	PBH	Test sanguins	Fibroscan
SNIFF 17	VHC	-	9	1056	√	√	-
		Exploratoire	Angers	194	√	√	√
VINDIAG 6	Toutes	Validation	Bordeaux Rabat	196	√	√	√
		Exploratoire	Angers Bordeaux Grenoble	349	√	√	√
VINDIAG 7	VHC	Validation	19	380	√	√	√

4.2.2 Histologie

Dans chaque étude, les PBH étaient lues par des anatomo-pathologistes expérimentés et spécialisés en hépatologie, en aveugle des données cliniques, biologiques et élastométriques des patients. Le degré de fibrose hépatique était évalué avec le score semi-quantitatif de

fibrose Metavir. La fibrose significative était définie par un stade de fibrose Metavir $F \geq 2$, la fibrose sévère par $F \geq 3$, et la cirrhose par F4.

La lecture des PBH était effectuée par un seul anatomo-pathologiste à Grenoble et à Bordeaux. A Angers, Tours, et en Provence-Alpes-Côte d'Azur, la lecture était effectuée par deux anatomo-pathologistes avec une lecture consensus en cas de discordance. Dans l'étude FIBROSTAR, la lecture des lames histologiques était centralisée et réalisée par deux anatomo-pathologistes indépendants référents, avec une relecture consensus en cas de discordance.

4.2.3 Tests sanguins de fibrose

L'analyse des prélèvements sanguins et le dosage des marqueurs de fibrose étaient réalisés dans les laboratoires de chaque centre. Nous avons montré que la reproductibilité inter-centre des tests sanguins de fibrose était excellente (51). Pour l'étude FIBROSTAR, les dosages de l'acide hyaluronique, de l' $\alpha 2$ macroglobuline, de l'apolipoprotéine A1, et de l'haptoglobine étaient centralisés dans des laboratoires de référence.

Dans chaque étude, les tests sanguins de fibrose ont été calculés selon les formules publiées pour le FibroMètre (55), le Fibrotest (70), l'Hepascore (47), le Fib4 (41), et l'APRI (40).

4.2.4 Mesure de la dureté hépatique avec le Fibroscan

La dureté hépatique était mesurée avec le Fibroscan par un opérateur expérimenté (au moins 50 examens) en aveugle des données cliniques, biologiques, et histologiques du patient. L'examen s'effectuait selon les recommandations du fabricant. L'objectif était d'obtenir 10 mesures valides, le résultat final donné par le Fibroscan correspondant à la médiane de ces 10 mesures valides.

4.3 ETUDE SNIFF 17

L'article de l'étude SNIFF 17 est présenté en **Annexe 7**

4.3.1 Présentation de l'étude

Les tests sanguins de fibrose ont été majoritairement conçus pour prédire le diagnostic binaire de fibrose significative (cf § 2.2). Leur performance pour le diagnostic de fibrose sévère ou de cirrhose, une cible clinique d'intérêt en pratique clinique, a été peu évaluée.

L'objectif de l'étude SNIFF 17 était d'évaluer la performance des tests sanguins (initialement conçus pour le diagnostic de fibrose significative) pour le diagnostic de fibrose sévère (Metavir $F \geq 3$) et celui de cirrhose (Metavir F4), et d'élaborer un test sanguin spécifiquement dédié pour le diagnostic de cirrhose.

L'étude SNIFF 17 a inclus 1056 patients avec hépatite chronique C pour lesquels étaient disponibles une PBH et les tests sanguins. Les patients ont été inclus dans 9 centres français : Angers, Provence-Alpes-Côte d'Azur (5 centres), Tours, Grenoble, et Bordeaux.

4.3.2 Discussion des résultats

Les 9 centres ayant inclus les patients dans l'étude SNIFF 17 incluaient des centres de soins secondaires ou tertiaires, publiques ou privés. De plus, la sélection des patients, le dosage des marqueurs de fibrose, et l'analyse histologique des PBH étaient réalisés de façon indépendante dans chaque centre. Enfin, la prévalence des stades Metavir de fibrose dans notre population (F0 : 4,4%, F1 : 43,5%, F2 : 27,0%, F3 : 14,0%, F4 : 11,2%) était proche de la prévalence de référence de la méta-analyse de Thein (8). Ainsi, compte tenu de son caractère multicentrique, des différents types de centres investigateurs, du large nombre de patients inclus, et de la répartition adéquate des stades de fibrose, notre population était

estimée représentative de la population générale des patients avec une hépatite chronique C vus en pratique clinique.

Nos résultats montrent que les tests sanguins construits pour le diagnostic de fibrose significative avaient de très bonnes AUROC pour le diagnostic de fibrose sévère (0,822 – 0,885) et de cirrhose (0,841 – 0,907). Les AUROC du FibroMètre pour le diagnostic de fibrose sévère et de cirrhose étaient significativement plus élevées que celles des autres tests, sauf celle de l’Hepascore pour le diagnostic de cirrhose. Néanmoins, l’AUROC est un indice de performance global qui ne montre pas la performance des tests sanguins en pratique clinique.

Nous avons donc calculé 2 types de seuils diagnostiques : celui correspondant au meilleur indice de Youden (sensibilité + spécificité - 1) et celui correspondant à la meilleure performance diagnostique. Comme attendu, le taux de patients bien classés pour le diagnostic de fibrose sévère et surtout celui de cirrhose était plus élevé avec le seuil de meilleure performance diagnostique qu’avec le seuil de Youden. Le seuil de meilleure performance diagnostique semblait donc le plus adapté à la pratique quotidienne des cliniciens qui cherchent à établir le diagnostic individuel de leur patient.

Néanmoins, le calcul du seuil de meilleure performance diagnostique était fortement influencé par la faible prévalence de la cible diagnostique (fibrose sévère : 25%, cirrhose : 11%). En effet, augmenter la valeur du seuil diagnostique permettait d’augmenter considérablement **le nombre** de patient « non malades » bien classés (car leur prévalence était élevée), alors que **le nombre** de patients « malades » bien classés ne diminuait que faiblement (car leur prévalence était faible). Ainsi, dans notre population où la prévalence de la cirrhose était faible, on améliorait la performance diagnostique en augmentant la valeur du seuil diagnostique, mais cette amélioration était surtout liée à un meilleur taux de patients non cirrhotiques bien classés. Parallèlement, comparé au meilleur seuil de Youden, le seuil de meilleure performance diagnostique permettait d’augmenter la VPP pour le diagnostic de cirrhose (+20% à +40% selon le test sanguin), mais au prix d’une diminution considérable de la sensibilité (-40% à -60%). Ainsi, optimiser la performance du diagnostic de cirrhose en calculant le seuil de meilleure performance diagnostique amenait à une situation paradoxale : avec le meilleur test, le taux de patients bien classés pour la cirrhose était excellent dans la

population globale (environ 90,9%), mais était seulement de 35,7% dans le sous-groupe des patients cirrhotiques (cible diagnostique).

Nous avons donc déterminé les intervalles de VPN $\geq 95\%$ et à 100% permettant d'exclure de façon fiable le diagnostic de cirrhose. Il n'était pas possible de déterminer les intervalles de VPP $\geq 95\%$ et à 100% car les VPP des tests sanguins n'atteignaient pas ces valeurs quelque soit le seuil diagnostique choisi. Le FibroMètre était le test sanguin qui permettait d'exclure le diagnostic de cirrhose avec 100% d'exactitude chez le plus grand nombre de patients : 47,5% versus 31,9% avec le Fibrotest ($p < 10^{-3}$), 24,7% avec l'Hepascore ($p < 10^{-3}$), et 5,4% avec l'APRI ($p < 10^{-3}$). Ainsi, malgré leur très bonne performance globale, les tests sanguins construits pour le diagnostic de fibrose significative ne permettaient pas d'affirmer le diagnostic de cirrhose mais seulement de l'exclure.

Nous avons finalement tenté d'optimiser le diagnostic de cirrhose en élaborant un test sanguin de fibrose spécifiquement dédié pour le diagnostic de cirrhose. La méthode d'élaboration était la même que pour le FibroMètre dédié pour la fibrose significative : les 7 variables du FibroMètre ont été introduites comme variables indépendantes dans une régression logistique binaire dont la variable dépendante était le diagnostic de cirrhose (F0-3 vs F4). De façon intéressante, toutes les variables du FibroMètre spécifiquement dédié pour la fibrose significative étaient indépendamment liées au diagnostic de cirrhose. Le « *FibroMètre cirrhose* » comprenait donc les mêmes marqueurs de fibrose que le « *FibroMètre standard* » mais avec des coefficients β différents. Un des avantages de ce nouveau test sanguin de fibrose est donc qu'il peut être calculé sans surcoût et dans le même temps que le test sanguin original. Le *FibroMètre cirrhose* avait plusieurs autres avantages :

- L'AUROC du *FibroMètre cirrhose* pour le diagnostic de cirrhose était supérieure à celle du *FibroMètre standard*.

- Le taux de patients bien classés pour le diagnostic de cirrhose avec le seuil de meilleure performance diagnostique était supérieur avec le *FibroMètre cirrhose* (93,0% vs 90,9% avec le *FibroMètre standard* ; $p=0,005$). Ceci était dû à une amélioration du taux de patients bien classés parmi les cirrhotiques (54,8% vs 35,7% ; $p < 10^{-3}$) sans modification significative du taux de patients bien classés parmi les non cirrhotiques (97,9% vs 98,0% ; $p=1$).

- Le *FibroMètre cirrhose* permettait d'affirmer le diagnostic de cirrhose avec une VPP à 100% chez 1,3% des patients (11,3% des patients cirrhotiques), alors que le *FibroMètre standard* ne dépassait pas une VPP de 88,2% quelque soit la valeur du seuil diagnostique testée.

En conclusion, l'étude SNIFF 17 a montré que les tests sanguins de fibrose construits pour le diagnostic de fibrose significative avaient une très bonne performance globale pour le diagnostic de cirrhose (AUROC, taux de biens classés par le seuil de meilleure performance diagnostique). Cependant, ces tests permettaient surtout d'exclure le diagnostic de cirrhose alors que leur performance pour affirmer ce diagnostic était médiocre. Un test sanguin spécifiquement construit pour le diagnostic de cirrhose et incluant les mêmes marqueurs de fibrose que le *FibroMètre* permettait d'améliorer la performance du diagnostic non invasif de cirrhose et notamment la performance pour affirmer ce diagnostic.

4.4 ETUDE VINDIAG 6

L'article de l'étude VINDIAG 6 est présenté en **Annexe 8**

4.4.1 Présentation de l'étude

Outre la conception de tests sanguins adaptés à la cible diagnostique, une autre manière d'améliorer le diagnostic non invasif de la fibrose hépatique est d'associer les tests de fibrose. Plusieurs équipes ont ainsi proposé des algorithmes associant divers tests sanguins de fibrose et/ou le Fibroscan (70, 103, 109-111, 114). Les deux algorithmes les mieux validés sont le SAFE et l'algorithme de Bordeaux (cf §3.2.4). Ces algorithmes permettent d'obtenir une excellente performance pour le diagnostic de fibrose significative ou de cirrhose. Ils ont cependant plusieurs inconvénients (cf § 3.2.4.3).

L'objectif de l'étude VINDIAG 6 était d'améliorer la performance du diagnostic non invasif de fibrose en mettant au point une combinaison de tests non invasifs de fibrose qui contourne les limites des algorithmes déjà publiés.

L'étude VINDIAG 7 a inclus 390 patients avec une hépatopathie chronique pour lesquels étaient disponibles une PBH, les tests sanguins, et une mesure de la dureté hépatique par le Fibroscan. 194 patients inclus à Angers constituaient la population exploratoire et 196 patients inclus à Bordeaux et Rabat la population de validation.

4.4.2 Discussion des résultats

Nous avons choisi de combiner les tests de fibrose par régression logistique binaire. Ceci permettait de s'affranchir des limites des algorithmes publiés : la combinaison était synchrone et non séquentielle avec plusieurs étapes comme pour le SAFE, et il n'y avait plus nécessité d'utiliser de seuils diagnostiques pour les tests de fibrose comme avec l'algorithme de Bordeaux. La régression logistique binaire permettait ainsi d'élaborer un nouveau test de

fibrose de « 3^e génération » qui combinait les tests sanguins et/ou le Fibroscan. Par ailleurs, compte tenu des résultats de l'étude SNIFF 17, nous avons choisi d'adapter nos nouveaux tests de fibrose à la cible diagnostique en en créant un pour le diagnostic de fibrose significative et un pour celui de cirrhose.

Que ce soit pour le diagnostic de fibrose significative ou celui de cirrhose, c'est le FibroMètre et le Fibroscan qui étaient les tests sélectionnés par la régression logistique binaire. Nous avons donc élaboré le *significant fibrosis index* (SF-index) et le *cirrhosis index* (C-index) pour les diagnostics respectifs de fibrose significative et de cirrhose, selon la méthode déjà utilisée pour les tests sanguins de 2^e génération.

Le SF-index avait plusieurs avantages. Premièrement, ce nouveau test avait une AUROC pour le diagnostic de fibrose significative significativement supérieure à celles du FibroMètre et du Fibroscan, ainsi qu'un meilleur profil de performance. Deuxièmement, le SF-index résolvait deux tiers des discordances entre le FibroMètre et le Fibroscan pour le diagnostic de fibrose significative. Ainsi, le SF-index corrigeait une grande partie des erreurs diagnostiques individuelles du FibroMètre ou du Fibroscan. Troisièmement, le SF-index incluait significativement plus de patients dans les intervalles de VP $\geq 90\%$ pour le diagnostic de fibrose significative que le FibroMètre ou le Fibroscan. Utilisé avec ces intervalles de VP $\geq 90\%$ et une PBH pour les patients dans l'intervalle intermédiaire, le SF-index permettait d'obtenir une excellente performance diagnostique (91,9%) et un taux de PBH nécessaires significativement inférieur à ceux du SAFE ou de l'algorithme de Bordeaux. Enfin, utilisé avec la méthode des intervalles de diagnostic fiable (cf §3.2.6), le SF-index permettait de classer correctement 90,6% des patients sans aucune PBH nécessaire.

Le C-index offrait pour le diagnostic de cirrhose les mêmes avantages que le SF-index pour le diagnostic de fibrose significative. L'AUROC du C-index était supérieure à celles du FibroMètre ou du Fibroscan, mais la différence n'était significative qu'avec le FibroMètre. Utilisé avec les intervalles de VP $\geq 90\%$ pour la cirrhose et une PBH pour les patients dans l'intervalle intermédiaire, le C-index permettait d'obtenir une excellente performance diagnostique (91,0%, significativement supérieure à celle du SAFE) et un taux de PBH nécessaires très faible (9,3%, significativement inférieur à celui de l'algorithme de Bordeaux). Enfin, utilisé avec la méthode des intervalles de diagnostic fiable, le SF-index permettait de classer correctement 90,3% des patients sans aucune PBH nécessaire.

En conclusion, la combinaison synchrone du FibroMètre et du Fibroscan permettait d'améliorer la performance du diagnostic non invasif de fibrose significative, la fiabilité du diagnostic de fibrose significative et de celui de cirrhose, et donc de réduire le recours à la PBH. Au contraire des combinaisons de tests de fibrose précédemment publiées, la combinaison synchrone s'effectuait en une seule étape et s'affranchissait du problème des discordances entre les tests. Enfin, avec la méthode des intervalles de diagnostic fiable, la combinaison synchrone du FibroMètre et du Fibroscan permettait de classer correctement plus de 90% des patients sans aucune PBH nécessaire.

Néanmoins l'étude VINDIAG 7 avait quelques limites. Premièrement, elle incluait des patients ayant des hépatopathies de cause diverses. Deuxièmement, la prévalence des cibles diagnostiques était élevée dans les populations exploratoires et de validation : respectivement 76,3% et 72,4% pour la fibrose significative, et 22,7% et 39,3% pour la cirrhose. Ces populations n'étaient donc pas représentatives des patients avec hépatopathie chronique vus en pratique clinique, ce qui limite l'exportabilité des résultats à la pratique quotidienne. Troisièmement, deux nouveaux tests de fibrose ont été créés : un pour la fibrose significative et un pour la cirrhose. Ainsi, comme pour le SAFE, le clinicien doit d'abord appliquer le test pour la fibrose significative puis éventuellement, en cas de diagnostic de fibrose significative, celui pour la cirrhose. Enfin, bien que très réduit par rapport aux algorithmes précédents, il était toujours nécessaire d'effectuer une PBH chez certains patients.

4.5 ETUDE VINDIAG 7

L'article de l'étude VINDIAG 7 est présenté en **Annexe 9**

4.5.1 Présentation de l'étude

L'étude VINDIAG 6 a montré les différents avantages d'une combinaison synchrone de tests non invasifs de fibrose. Cependant, ce travail était limité par le fait qu'il avait inclus des patients ayant des hépatopathies chroniques de causes diverses, et que les populations étudiées n'étaient pas représentatives des patients vus en pratique clinique car la prévalence des cibles diagnostiques y était plus élevée. Il était donc nécessaire de confronter la méthode et les résultats de l'étude VINDIAG 6 dans une population homogène de patients ayant une hépatite chronique C, représentative des patients vus en pratique clinique.

Dans l'étude VINDIAG 6, les combinaisons synchrones de tests de fibrose ont été adaptées à la cible diagnostique, car l'étude SNIFF17 avait démontré les avantages à élaborer un test spécifiquement dédié pour le diagnostic de cirrhose. Cependant, comme pour l'algorithme SAFE, ceci impose d'appliquer dans un premier temps la combinaison synchrone pour le diagnostic de fibrose significative puis, si le diagnostic de fibrose significative est retenu, d'appliquer dans un second temps la combinaison synchrone pour le diagnostic de cirrhose. Néanmoins, nos combinaisons synchrones pour la fibrose significative et pour la cirrhose utilisent les mêmes tests non invasifs de fibrose et il est donc possible de les calculer en un seul et même temps.

L'objectif de l'étude VINDIAG 7 était donc de confirmer les avantages de la combinaison synchrone de tests non invasifs de fibrose dans une population homogène et représentative de patients ayant une hépatite chronique C ; et d'utiliser les combinaisons synchrones de tests adaptées à la cible diagnostique pour élaborer une nouvelle classification non invasive de fibrose plus performante que celles des compteurs des tests non invasifs de fibrose recommandés par la HAS.

4.5.2 Discussion des résultats

Tous les aspects méthodologiques développés dans nos précédentes études ont été utilisés dans l'étude VINDIAG 7 pour créer une nouvelle classification non invasive de fibrose : 1/ combinaison synchrone de tests non invasifs de fibrose par régression logistique binaire (126, 127), 2/ tests adaptés à la cible diagnostique (127, 128), et 3/ intervalles de diagnostic fiable (48, 59). Finalement, l'association des intervalles de diagnostic fiable de deux nouveaux index, le CSF-index et le SF-index, chacun combinant le FibroMètre et le Fibroscan par RLB et adaptés respectivement à la fibrose significative et à la fibrose sévère, permettait d'obtenir une nouvelle classification (**Figure 19**) totalement non invasive (0% de PBH), précise (6 classes diagnostiques), performante (87% de bien classés), et robuste.

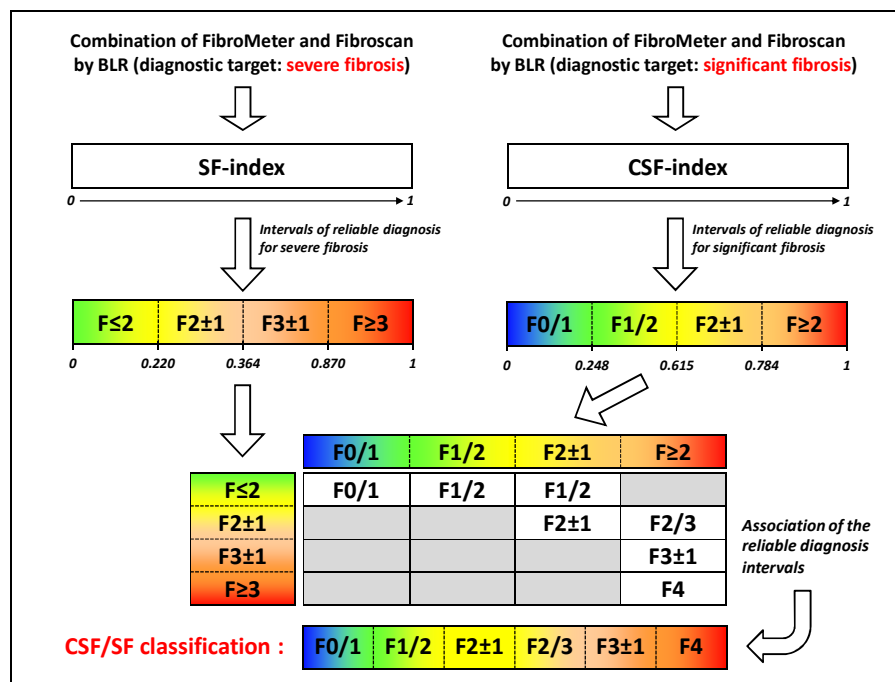


Figure 19 : Nouvelle classification CSF/SF. Tout d'abord, deux nouveaux index sont calculés : le CSF-index qui combine par RLB le FibroMètre et le Fibroscan pour le diagnostic de fibrose significative (Metavir $F \geq 2$), et le SF-index qui combine le FibroMètre et le Fibroscan pour le diagnostic de fibrose sévère (Metavir $F \geq 3$). Dans un second temps, les intervalles de diagnostic fiable du CSF-index sont déterminés pour le diagnostic de fibrose significative ainsi que ceux du SF-index pour la fibrose sévère. Finalement, l'association des intervalles de diagnostic fiable du CSF-index et du SF-index permet de déterminer la classification CSF/SF.

L'étude VINDIAG 7 a confirmé dans l'hépatite chronique C les résultats obtenus dans l'étude pilote VINDIAG 6. En effet, parmi les 6 tests de fibrose étudiés, le FibroMètre et le Fibroscan étaient bien les tests de fibrose indépendamment liés à chaque cible diagnostique étudiée. Il est à noter qu'une simple régression logistique binaire était utilisée dans VINDIAG 6, alors que l'étude VINDIAG 7 a utilisé la méthode des échantillons bootstrap pour répéter l'analyse multivariée sur 1000 échantillons. Au final, le FibroMètre et le Fibroscan étaient de loin les tests de fibrose les plus souvent sélectionnés comme indépendamment liés à la fibrose significative, la fibrose sévère, ou la cirrhose. Nous avons donc construit 3 nouveaux tests non invasifs de fibrose (combinaisons synchrones), spécifiquement dédiés à l'évaluation non invasive de la fibrose hépatique au cours de l'hépatite chronique C et chacun adapté à la cible diagnostique (CSF-index pour la fibrose significative, SF-index pour la fibrose sévère, et C-index pour la cirrhose). Comme dans l'étude VINDIAG 6, les combinaisons synchrones amélioraient significativement l'AUROC pour le diagnostic non invasif de fibrose significative (et de fibrose sévère) par rapport aux tests seuls. De la même manière, l'AUROC du C-index était supérieure à celle du FibroMètre et du Fibroscan mais la différence n'était significative qu'avec le FibroMètre. Enfin, les combinaisons synchrones amélioraient le diagnostic fiable de chaque cible diagnostique.

Utilisés avec la méthode des intervalles de diagnostic fiable, les 3 combinaisons synchrones permettaient de bien classer plus de 90% des patients pour le diagnostic de fibrose significative ou de fibrose sévère, et plus de 95% pour celui de cirrhose. Comme nos 3 combinaisons synchrones incluaient les mêmes tests de fibrose, il était possible de les calculer en un seul et même temps et donc de produire 3 diagnostics fiables (issus des intervalles de diagnostic fiable de chaque combinaison synchrone). Premier résultat marquant : il n'existait jamais de discordance entre les 3 résultats. Nous avons donc cherché à évaluer l'association de ces résultats de diagnostic fiable. L'association du diagnostic fiable du CSF-index avec celui du SF-index permettait d'élaborer une nouvelle classification non invasive de fibrose incluant 6 classes (classification CSF/SF : F0/1, F1/2, F2±1, F2/3, F3±1, F4). L'association du diagnostic fiable du CSF-index avec celui du C-index permettait d'obtenir une autre classification incluant les 6 mêmes classes (classification CSF/C).

Nous avons également déterminé les intervalles de diagnostic fiable du FibroMètre pour la fibrose significative et ceux du Fibroscan pour la fibrose sévère. Nos résultats montraient

que leur performance diagnostique n'était pas significativement différente de celles des combinaisons synchrones. Une question se posait alors : l'association des intervalles de diagnostic fiable de simples tests de fibrose (FibroMètre et Fibroscan) ne serait-elle pas aussi performante que celle des combinaisons synchrones, plus complexes ? Nos résultats ont clairement montré les inconvénients d'une combinaison des diagnostics fiables du FibroMètre et du Fibroscan. Tout d'abord, il existait des discordances entre les diagnostics fiables du FibroMètre et ceux du Fibroscan, avec par conséquent un diagnostic indéterminé et la nécessité de réaliser une PBH. La performance diagnostique de l'association des intervalles de diagnostic fiable du FibroMètre et du Fibroscan n'était pas robuste : le taux de bien classés était de 82,8% dans la population exploratoire mais diminuait très significativement à 69,4% dans la population de validation. Enfin, que ce soit dans la population exploratoire ou dans celle de validation, l'association des intervalles de diagnostic fiable du FibroMètre et du Fibroscan avait une performance diagnostique significativement inférieure à celle de la classification CSF/SF.

Ayant confirmé l'intérêt des combinaisons synchrones de tests de fibrose dans l'hépatite chronique C et la possibilité d'élaborer une nouvelle classification non invasive de fibrose, la seconde question était : quelle est la meilleure classification entre la classification CSF/SF et la classification CSF/C ? Nos résultats ont montré que la classification CSF/SF avait une performance diagnostique significativement supérieure à celle de la classification CSF/C dans la population de validation, et surtout qu'elle permettait de très significativement mieux classer les patients cirrhotiques, le sous-groupe de patients ayant le plus haut risque de complication.

Enfin, la dernière question était : nos résultats sont-ils exportables à la pratique clinique quotidienne ? Comme dans l'étude VINDIAG 6, la prévalence des cibles diagnostiques dans la population exploratoire était plus élevée que celle de la population de référence de la méta-analyse de Thein (8). Néanmoins, et de façon très intéressante, la performance diagnostique de la classification CSF/SF n'était pas significativement différente entre la population exploratoire et la population de validation dans laquelle la prévalence des stades de fibrose était beaucoup plus proche de celle de référence. Ainsi, un autre avantage de la classification CSF-SF est la robustesse de ses résultats qui permet leur exportation à la pratique clinique. Ceci est très probablement dû à la robustesse des intervalles de diagnostic fiable que nous

avons démontré dans 4 populations regroupant 2068 patients avec hépatite chronique C (données non publiées).

La classification CSF/SF avait une performance diagnostique très significativement supérieure à celle des compteurs de fibrose des tests de fibrose recommandés par la HAS. En effet, la performance diagnostique de ces compteurs ne dépassait pas 70%, et même pas plus de 45% pour le Fibrotest qui a été le premier test non invasif de fibrose à être recommandé pour une utilisation en pratique clinique. La classification CSF/SF permettait de classer correctement 87% des patients, amenant par la même une amélioration très significative du diagnostic non invasif de fibrose en pratique clinique. Il est à souligner que notre référence de fibrose était la PBH qui n'est pas un gold-standard. La performance de la classification CSF/SF est donc très probablement encore supérieure en réalité, et donc excellente.

Bien qu'elle permet une amélioration très significative du diagnostic non invasif de fibrose en pratique clinique, la classification CSF/SF peut donner l'impression d'être très complexe à utiliser car elle nécessite plusieurs étapes : 1/ calcul du FibroMètre et mesure de la dureté hépatique, 2/ calcul du CSF-index et du SF-index, 3/ détermination des intervalles de diagnostic fiable du CSF-index et du SF-index, et enfin 4/ association des intervalles de diagnostic fiable. Cependant, toutes ces étapes peuvent être informatisées, le clinicien n'ayant finalement qu'une simple manipulation supplémentaire à faire : indiquer le résultat du Fibroscan lors du calcul du FibroMètre. Par ailleurs, l'apparente limite de la classification CSF/SF liée au fait qu'elle nécessite d'effectuer deux méthodes différentes est en fait un atout car elle aborde la fibrose sous deux aspects différents : marqueurs biologiques et dureté physique du foie. Oberti *et al* ont montré que les paramètres biologiques mais également la dureté physique (foie ferme) étaient des paramètres indépendamment liés au diagnostic de cirrhose (108). La classification CSF/SF prend en compte cette dureté physique par le résultat du Fibroscan et regroupe ainsi un ensemble clinico-biologique et physique dont le résultat du Fibroscan serait la 9^e variable. En outre, en pratique pour le patient, tous les examens (bilan biologique et mesure de la dureté hépatique par le Fibroscan) peuvent être réalisés en un seul temps.

Ainsi, la classification CSF/SF permet de contourner toutes les limites des associations de tests de fibrose précédemment publiées (SAFE, algorithme de Bordeaux) : l'association synchrone des tests non invasifs de fibrose s'affranchi de l'utilisation de seuils diagnostiques

et donc des discordances entre les tests, le résultat de la classification CSF/SF est plus détaillé (6 classes) qu'un simple diagnostic binaire, il est obtenu en une seule étape, et enfin aucune PBH n'est nécessaire.

Une limite de la classification CSF/SF pourrait être la relative imprécision de son diagnostic car chacune de ses classes diagnostiques correspond à 2 ou 3 stades histologiques de fibrose. Néanmoins, cette relative imprécision permet très probablement de réduire le taux de mal classés. Par ailleurs, bien que moins précise, la classification CSF/SF comprend plus de classes diagnostiques que la classification Metavir (respectivement 6 versus 5 classes). La valeur pronostique de cette classification à « 6 stades non invasifs » de fibrose devra être évaluée dans des études longitudinales qui prendront comme critère de jugement principal les événements cliniques. De tels travaux permettront de déterminer une prise en charge des patients basée sur le résultat des tests non invasifs sans aucune référence à la classification histologique du Metavir. Par ailleurs, d'autres études évaluant le rapport coût-efficacité d'une prise en charge des patients basée sur les résultats soit de la PBH soit de la classification CSF/SF permettront de déterminer la pertinence de l'utilisation de la classification CSF/SF en pratique clinique.

En conclusion, l'étude VINDIAG 7 a confirmé les résultats de l'étude VIDDIAG 6 dans une population homogène et représentative de patients ayant une hépatite chronique C, et y a démontré l'intérêt des combinaisons synchrones de tests non invasifs de fibrose hépatique. L'association des nouvelles combinaisons synchrones de tests de fibrose, basée sur intervalles de diagnostic fiable, permet d'élaborer une nouvelle classification de fibrose à 6 classes dont la performance diagnostique est robuste et très significativement supérieure à celles des compteurs des tests de fibrose recommandés par la HAS (Fibrotest, FibroMètre, et Fibroscan). Cette nouvelle classification permet ainsi d'améliorer très significativement la performance du diagnostic non invasif de fibrose en pratique clinique, sa précision, et d'être réellement exclusivement non invasif.

5. DISCUSSION GENERALE

5.1 Amélioration des recommandations de la HAS

Les recommandations de la HAS « Méthodes non invasives de mesure de la fibrose/cirrhose hépatique » de décembre 2006 puis de décembre 2008 (81) constituent une réelle avancée dans la prise en charge des patients avec une hépatite chronique C. En effet, pour évaluer le degré de fibrose hépatique chez les patients adultes avec une hépatite chronique C non traitée et sans comorbidité, il est désormais possible d'utiliser en première intention un test non invasif de fibrose parmi le Fibrotest, le FibroMètre, l'Hepascore, ou le Fibroscan. En cas de discordance entre le résultat et la clinique ou lorsque le test est ininterprétable, un des 3 autres tests non invasifs ou la PBH peut être réalisé en seconde intention. La PBH est donc devenu un examen de seconde voire de troisième ligne.

Bien qu'elles affirment la place du diagnostic non invasif de fibrose dans la prise en charge des patients avec une hépatite chronique C, ces recommandations ont des limites (**Annexe 10**). En effet, elles ne donnent aucune précision sur les modalités d'utilisation des tests non invasifs en pratique clinique. Ainsi, quels sont les seuils diagnostiques à utiliser pour les tests non invasifs de fibrose ? En pratique clinique, les tests sont utilisés avec leur compteur de fibrose (54, 55, 69) mais nous avons montré dans l'étude VINDIAG 7 que leur performance diagnostique était médiocre voire très mauvaise. Il est ainsi impossible de savoir aujourd'hui si les recommandations de la HAS ont amélioré la prise en charge des patients avec une hépatite chronique C. Le nombre de PBH a certes diminué mais combien de patients n'ont pas eu de traitement antiviral ou de programme de surveillance de cirrhose en raison d'une erreur diagnostique des tests de fibrose ?

Les algorithmes combinant les tests de fibrose, tels que le SAFE (115) ou l'algorithme de Bordeaux (70), permettent d'améliorer le diagnostic non invasif de fibrose (**Annexe 10**). Tout d'abord, ils précisent les modalités d'utilisation des tests de fibrose : quel test utiliser, quel

seuil diagnostique appliquer, et quel résultat en attendre. Ensuite, ils ont une excellente performance pour le diagnostic de fibrose significative ou de cirrhose (**Annexe 5**). Cependant, comme nous l'avons déjà précisé, ils ont plusieurs limites : 1/ l'excellente performance s'obtient au prix d'un taux de PBH nécessaire qui reste élevé (et dont le principe même recèle quelque part une certaine antinomie avec plusieurs applications d'un diagnostic non invasif), 2/ les algorithmes SAFE sont séquentiels, en plusieurs étapes, ce qui peut imposer de reconvoquer le patient plusieurs fois, et enfin 3/ les algorithmes SAFE ne donnent que des diagnostics binaires de fibrose significative ou de cirrhose ce qui impose d'appliquer les deux algorithmes avec une augmentation du risque d'erreur diagnostique et également du taux de PBH nécessaire. Il faut également rappeler que l'algorithme de Bordeaux, à 3 classes dans la publication princeps (70), n'a été évalué que pour des diagnostics binaires avec par conséquent une surestimation de sa performance diagnostique.

Nous avons développé plusieurs méthodes pour améliorer la performance du diagnostic non invasif de la fibrose hépatique au cours de l'hépatite chronique C (**Figure 20**): tests sanguins adaptés à la cause de l'hépatopathie (46), tests sanguins adaptés à la cible diagnostique (étude SNIFF 17), intervalles de valeurs prédictives $\geq 90\%$ (55), intervalles de diagnostic fiable (55), combinaison synchrone de tests (études VINDIAG 6 et 7), et enfin association des intervalles de diagnostic fiable des combinaisons synchrones de tests (étude VINDIAG 7). Tous ces concepts méthodologiques ont été utilisés dans l'étude VINDIAG 7 pour aboutir à la conception d'une nouvelle classification non invasive de fibrose : la classification CSF/SF. Cette classification a une excellente performance pour le diagnostic non invasif du stade histologique Metavir de fibrose (87% de patients bien classés), très significativement supérieure à celles des compteurs de fibrose du FibroMètre, du Fibroscan, et surtout du Fibrotest. Par ailleurs, la classification CSF/SF contourne toutes les limites des recommandations de l'HAS ainsi que celle des algorithmes combinant les tests non invasifs de fibrose (**Annexe 10**). En effet, cette classification est déterminée en une seule étape, ne nécessite pas d'utilisation de seuils diagnostiques pour les tests non invasifs de fibrose et s'affranchi donc des problèmes de discordances entre les tests, le diagnostic est plus précis (6 classes) qu'un simple diagnostic binaire (2 classes), la performance diagnostique est robuste, et enfin aucune PBH n'est nécessaire. Nous avons donc mis au point une nouvelle classification de fibrose permettant un diagnostic totalement non invasif (0% de PBH) des stades histologiques de fibrose avec une excellente performance diagnostique. En pratique, le

clinicien n'a qu'à préciser le résultat du Fibroscan lors de la saisie des variables du FibroMètre.

Etude	Etiologie	Méthode	Classes diagnostiques	PD (%)	
SNIFF 19	VHC	Compteur du FibroMètre (FM)	F0/1 F1 F1/2 F2/3 F3/4 F4	68-75	
-	VHC	Compteur du Fibroscan (FS)	F0/1 F1/2 F2 F3 F3/4 F4	55-65	
SNIFF 17	VHC	Test construit pour la cirrhose CirroMètre	F≤3 F4		
SNIFF 32	VHC	Intervalles de diagnostic fiable FibroMètre	F0/1 F1/2 F2±1 F≥2		
SNIFF 34	VHC	Intervalles de diagnostic fiable CirroMètre	F≤3 F≥2 F4		
VINDIAG 6	Toutes	Combinaison FibroMètre + Fibroscan Pour la fibrose significative	F0/1 PBH F≥2		
		Combinaison FibroMètre + Fibroscan Pour la cirrhose	F≤3 PBH F4		
VINDIAG 7	VHC	Combinaison de tests adaptées à la cible + intervalles de diagnostic fiable	F0/1 F1/2 F2±1 F2/3 F3±1 F4		86,7

Classification CSF/SF

Figure 20 : Méthodes développées pour améliorer la performance du diagnostic non invasif de la fibrose hépatique dans l'hépatite chronique C. L'utilisation de toutes ces méthodes dans la même procédure a permis de créer la classification CSF/SF dont la performance diagnostique est significativement supérieure à celles du FibroMètre et du Fibroscan recommandés par la Haute Autorité de Santé (HAS) française. Les études présentées dans ce travail de Thèse sont surlignées en jaune. PD : performance diagnostique (taux de patients bien classés) ; VHC : virus de l'hépatite C ; PBH : ponction-biopsie hépatique.

5.2 Quel est l'avenir du diagnostic non invasif de la fibrose hépatique ?

Nous disposons de plusieurs outils utilisables en pratique clinique qui permettent d'estimer le stade histologique de fibrose hépatique : les tests sanguins de fibrose, le Fibroscan, et maintenant la nouvelle classification CSF/SF. Ces outils ont été élaborés et évalués pour le **diagnostic** du degré de fibrose hépatique car il est bien établi que c'est un déterminant majeur du **pronostic** des patients ayant une hépatite chronique C. Cependant, les

tests non invasifs de fibrose et plus particulièrement la classification CSF/SF sont bien corrélés aux stades histologiques et donc au degré de fibrose hépatique : n'ont-ils donc pas eux aussi une valeur pronostique dans l'hépatite chronique C ? Plusieurs études ont suggéré que les tests non invasifs de fibrose ont une valeur pronostique au cours des principales causes d'hépatopathies chroniques : hépatite C (129-131), co-infection hépatite C et VIH (132), hépatite B (133, 134), alcool (135), cirrhose biliaire primitive (136), ou causes diverses (137). Dans l'étude de Naveau *et al* (135), l'AUROC des tests sanguins pour prédire le décès n'était pas significativement différente de celle de la classification histologique. Ainsi, plutôt que de continuer à évaluer la performance diagnostique des tests et des classifications non invasives de fibrose en prenant comme référence la PBH (une référence imparfaite avec pour conséquence une évaluation diagnostique biaisée), il semble désormais nécessaire d'évaluer leur valeur pronostique en prenant comme référence les événements cliniques (138). Ceci permettra de déterminer les sous-groupes de patients à risque et de proposer une prise en charge basée uniquement sur les résultats du diagnostic non invasif de fibrose, permettant ainsi de s'affranchir complètement du degré histologique de fibrose et donc de la PBH pour la validation de ces tests.

Une modélisation de l'évolution de l'hépatite C en France a prédit que, dans les conditions actuelles de dépistage, la mortalité imputable au VHC augmentera jusqu'en 2010 avec 1100 décès par CHC et 2000 par insuffisance hépatocellulaire, puis diminuera ensuite (139). Par ailleurs, les anti-protéases Boceprevir et Telaprevir vont enfin être disponibles en pratique clinique en 2011 par le biais d'Autorisations Temporaires d'Utilisation. Actuellement, le stade histologique de fibrose n'est pas un critère d'indication de traitement antiviral dans les hépatites C chroniques de génotype 2 ou 3 : compte tenu des excellents taux de guérison (80-90%) le traitement peut être proposé à tous les patients afin d'obtenir l'éradication virale. Compte tenu des très bons taux de réponse virale soutenue avec les trithérapies associant interféron pégylé + ribavirine + anti-protéase dans l'hépatite chronique C de génotype 1, on peut supposer que le critère stade histologique de fibrose prenne de moins en moins d'importance dans la décision de débiter un traitement anti-viral. Cependant, l'hépatite chronique C reste une maladie endémique dans le monde et il est fort probable que de nombreux pays continueront, pour des raisons économiques, à sélectionner les patients pour un traitement antiviral sur le degré de fibrose hépatique. Par ailleurs, les tests non

invasifs de fibrose garderont toujours leur intérêt pour le diagnostic de cirrhose dans l'hépatite chronique C.

L'ensemble des travaux dans l'hépatite chronique C a permis de mettre au point des outils (tests sanguins, Fibroscan) et des concepts méthodologiques qui doivent maintenant être transposés aux autres causes d'hépatopathie chronique dans lesquelles le diagnostic non invasif de fibrose a encore été peu évalué. Il est désormais nécessaire d'évaluer les tests non invasifs de fibrose dans les hépatopathies chroniques alcooliques car elles sont fréquentes en France, mais aussi et surtout dans les hépatopathies dysmétaboliques qui deviennent un problème de santé publique avec l'augmentation croissante de l'obésité dans les pays développés comme le notre. Les futurs travaux de recherche devront y évaluer : la performance des tests de fibrose existants, les nouvelles méthodes de diagnostic non invasif de la fibrose hépatique (nouveaux marqueurs biologiques de fibrose, élastométrie par résonance magnétique, élastométrie acoustique, imagerie d'élasticité ... cf §2.5), l'intérêt des combinaisons de tests de fibrose pour augmenter la performance diagnostique, et enfin la valeur pronostique des tests de fibrose afin de proposer une prise en charge des patients qui ne nécessite pas le recours à la PBH.

Les concepts développés pour le diagnostic non invasif de la fibrose hépatique pourront également être transposés aux autres lésions hépatiques, notamment la stéatose hépatique et la NASH, car il est fort probable que le diagnostic et la quantification de ces lésions seront des éléments importants pour guider la prise en charge des patients avec un syndrome dysmétabolique.

Enfin, les outils de diagnostic non invasifs ouvrent le champ du dépistage de la fibrose hépatique dans la population générale, ce qui n'était bien entendu pas possible avec la PBH. Les critères établis par l'OMS en 1970 pour la mise en œuvre du dépistage d'une maladie sont les suivants : « 1/ la maladie dont on recherche les cas constitue une menace grave pour la santé publique, 2/ un traitement d'efficacité démontrée peut être administré aux sujets chez lesquels la maladie a été décelé, 3/ les moyens appropriés de diagnostic et de traitement sont disponibles, 4/ la maladie est décelable pendant une phase de latence ou au début de la phase clinique, 5/ une épreuve ou un examen de dépistage efficace existe, 6/ l'épreuve utilisée est acceptable pour la population, 7/ l'histoire naturelle de la maladie est connue et notamment

son évolution de la phase de latence à la phase symptomatique, 8/ le choix des sujets qui recevront un traitement est opéré selon les critères préétablis, 9/ le coût de la recherche des cas n'est pas disproportionné par rapport au coût global des soins médicaux, 10/ la recherche des cas est continue et elle n'est pas considérée comme une opération exécutée une fois pour toute. » Les hépatopathies chroniques et les tests non invasifs de fibrose répondent à tous ces critères. Il est désormais pertinent d'envisager l'évaluation et la réalisation de programmes de dépistage de la fibrose hépatique dans la population générale à l'aide des tests non invasifs de fibrose.

6. CONCLUSION

Le virus de l'hépatite C est présent de façon épidémique dans le monde. La majeure partie des patients atteints d'hépatite chronique C ne parviennent pas à détruire le virus et souffrent d'hépatite virale chronique dont l'histoire naturelle est l'évolution vers la cirrhose et ses complications. Le pronostic de ces patients est déterminé par le degré de fibrose hépatique initialement évalué par l'examen histologique d'une ponction-biopsie hépatique. Cependant, la ponction-biopsie hépatique est une procédure invasive avec un risque de complication grave voire de décès, il existe un biais d'échantillonnage, et la reproductibilité de l'évaluation histologique est médiocre. Depuis une dizaine d'années, plusieurs méthodes non invasives d'évaluation de la fibrose hépatique ont été mises au point, notamment les tests sanguins de fibrose et le Fibroscan. De nombreuses publications ayant montré leur performance diagnostique, la Haute Autorité de Santé française a recommandé depuis décembre 2008 leur utilisation en première ligne pour l'évaluation du degré de fibrose hépatique chez les patients adultes avec une hépatite chronique C non traitée et sans comorbidité.

Ces recommandations restent cependant limitées car la performance des tests non invasifs de fibrose utilisés en pratique clinique n'est pas bien connue. Ce travail de Thèse a permis de développer plusieurs concepts méthodologiques qui améliorent la performance du diagnostic non invasif de fibrose hépatique : tests spécifiquement adaptés à la cible diagnostique, combinaisons synchrone de tests non invasifs de fibrose, association des intervalles de diagnostic fiable des tests de fibrose. Ces nouvelles méthodes ont permis d'élaborer une nouvelle classification non invasive de fibrose très significativement plus performante que celles utilisées jusqu'alors en pratique clinique. Cette nouvelle classification, très simple d'utilisation, permettra d'améliorer la prise en charge clinique des patients atteints d'hépatite chronique C. Il est désormais nécessaire d'évaluer la valeur pronostique de cette classification ainsi que celle des tests non invasifs de fibrose en général dans l'hépatite chronique C. Par ailleurs, nos concepts méthodologiques pourront être appliqués dans les autres moyens

diagnostiques et causes d'hépatopathie chronique afin d'y développer le diagnostic non invasif de la fibrose qui y est jusqu'alors peu évalué.

7. REFERENCES

1. www.who.int/immunization/topics/hepatitis_c/en/.
2. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006;144:705-714.
3. Blonski W, Reddy KR. Hepatitis C virus infection and hepatocellular carcinoma. *Clin Liver Dis* 2008;12:661-674.
4. Kim WR. The burden of hepatitis C in the United States. *Hepatology* 2002;36:S30-34.
5. Thuluvath PJ, Guidinger MK, Fung JJ, Johnson LB, Rayhill SC, Pelletier SJ. Liver transplantation in the United States, 1999-2008. *Am J Transplant* 2010;10:1003-1019.
6. Deuffic-Burban S, Poynard T, Sulkowski MS, Wong JB. Estimating the future health burden of chronic hepatitis C and human immunodeficiency virus infections in the United States. *J Viral Hepat* 2007;14:107-115.
7. www.invs.sante.fr/beh/2009/20_21/beh_20_21_2009.pdf.
8. Thein HH, Yi Q, Dore GJ, Krahn MD. Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression. *Hepatology* 2008;48:418-431.
9. Bialek SR, Terrault NA. The changing epidemiology and natural history of hepatitis C virus infection. *Clin Liver Dis* 2006;10:697-715.
10. Williams MJ, Lang-Lenton M. Progression of initially mild hepatic fibrosis in patients with chronic hepatitis C infection. *J Viral Hepat* 2011;18:17-22.
11. Hui CK, Belaye T, Montegrando K, Wright TL. A comparison in the progression of liver fibrosis in chronic hepatitis C between persistently normal and elevated transaminase. *J Hepatol* 2003;38:511-517.
12. Fartoux L, Chazouilleres O, Wendum D, Poupon R, Serfaty L. Impact of steatosis on progression of fibrosis in patients with mild hepatitis C. *Hepatology* 2005;41:82-87.

13. Di Bisceglie AM, Shiffman ML, Everson GT, Lindsay KL, Everhart JE, Wright EC, Lee WM, et al. Prolonged therapy of advanced chronic hepatitis C with low-dose peginterferon. *N Engl J Med* 2008;359:2429-2441.
14. Everhart JE, Wright EC, Goodman ZD, Dienstag JL, Hoefs JC, Kleiner DE, Ghany MG, et al. Prognostic value of Ishak fibrosis stage: findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. *Hepatology* 2010;51:585-594.
15. Ghany MG, Strader DB, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009;49:1335-1374.
16. Sun CA, Wu DM, Lin CC, Lu SN, You SL, Wang LY, Wu MH, et al. Incidence and cofactors of hepatitis C virus-related hepatocellular carcinoma: a prospective study of 12,008 men in Taiwan. *Am J Epidemiol* 2003;157:674-682.
17. Donato F, Boffetta P, Puoti M. A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. *Int J Cancer* 1998;75:347-354.
18. Lok AS, Seeff LB, Morgan TR, di Bisceglie AM, Sterling RK, Curto TM, Everson GT, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. *Gastroenterology* 2009;136:138-148.
19. Alazawi W, Cunningham M, Dearden J, Foster GR. Systematic review: outcome of compensated cirrhosis due to chronic hepatitis C infection. *Aliment Pharmacol Ther* 2010;32:344-355.
20. Fontana RJ, Sanyal AJ, Ghany MG, Lee WM, Reid AE, Naishadham D, Kahn JA, et al. Factors that determine the development and progression of gastroesophageal varices in patients with chronic hepatitis C. *Gastroenterology* 2010;138:2321-2331.
21. Bruno S, Crosignani A, Facciotto C, Rossi S, Roffi L, Redaelli A, de Franchis R, et al. Sustained virologic response prevents the development of esophageal varices in compensated, Child-Pugh class A hepatitis C virus-induced cirrhosis. A 12-year prospective follow-up study. *Hepatology* 2010;51:2069-2076.
22. Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology* 2005;42:1208-1236.
23. Consensus conference: complications of portal hypertension in adults (Paris, December 4-5, 2003). Long text. *Gastroenterol Clin Biol* 2004;28 Spec No 2:B324-334.

24. De Franchis R. Revising consensus in portal hypertension: Report of the Baveno V consensus workshop on methodology of diagnosis and therapy in portal hypertension *J Hepatol* 2010;53:762-768.
25. Poynard T, Cales P, Pasta L, Ideo G, Pascal JP, Pagliaro L, Lebrech D. Beta-adrenergic-antagonist drugs in the prevention of gastrointestinal bleeding in patients with cirrhosis and esophageal varices. An analysis of data and prognostic factors in 589 patients from four randomized clinical trials. Franco-Italian Multicenter Study Group. *N Engl J Med* 1991;324:1532-1538.
26. Tripathi D, Graham C, Hayes PC. Variceal band ligation versus beta-blockers for primary prevention of variceal bleeding: a meta-analysis. *Eur J Gastroenterol Hepatol* 2007;19:835-845.
27. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289-293.
28. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, Denk H, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995;22:696-699.
29. Batts KP, Ludwig J. Chronic hepatitis. An update on terminology and reporting. *Am J Surg Pathol* 1995;19:1409-1417.
30. Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology* 1994;19:1513-1520.
31. Castera L, Negre I, Samii K, Buffet C. Pain experienced during percutaneous liver biopsy. *Hepatology* 1999;30:1529-1530.
32. Poynard T, Lebrech D. The inconvenience of investigations used in hepatology: patients' and hepatologists' opinions. *Liver* 1982;2:369-375.
33. Nousbaum JB. The role of liver biopsy in the management of chronic hepatitis C. *Gastroenterol Clin Biol* 2002;26 (suppl 2):B168-179.
34. Guido M, Rugge M. Liver biopsy sampling in chronic viral hepatitis. *Semin Liver Dis* 2004;24:89-97.
35. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, Feng ZZ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002;97:2614-2618.
36. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;38:1449-1457.

37. Rousselet MC, Michalak S, Dupre F, Croue A, Bedossa P, Saint-Andre JP, Cales P. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005;41:257-264.
38. Bedossa P, Carrat F. Liver biopsy: the best, not the gold standard. *J Hepatol* 2009;50:1-3.
39. Boursier J, Dib N, Oberti F, Cales P. Characteristics and interpretation of blood tests for liver fibrosis. *Gastroenterol Clin Biol* 2007;31:511-523; quiz 500, 531-512.
40. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, Lok AS. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003;38:518-526.
41. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, M SS, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;43:1317-1325.
42. Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poinard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001;357:1069-1075.
43. Forn X, Ampurdanes S, Llovet JM, Aponte J, Quinto L, Martinez-Bauer E, Bruguera M, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 2002;36:986-992.
44. Patel K, Gordon SC, Jacobson I, Hezode C, Oh E, Smith KM, Pawlotsky JM, et al. Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients. *J Hepatol* 2004;41:935-942.
45. Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, Hubscher S, et al. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004;127:1704-1713.
46. Calès P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A, Gallois Y, et al. A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005;42:1373-1381.
47. Adams LA, Bulsara M, Rossi E, DeBoer B, Speers D, George J, Kench J, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* 2005;51:1867-1873.

48. Cales P, De Ledinghen V, Halfon P, Bacq Y, Leroy V, Boursier J, Foucher J, et al. Evaluating accuracy and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C. *Liver Int* 2008;28:1352-1362.
49. Poynard T, Morra R, Halfon P, Castera L, Ratziu V, Imbert-Bismut F, Naveau S, et al. Meta-analyses of FibroTest diagnostic value in chronic liver disease. *BMC Gastroenterol* 2007;7:40.
50. Zarski H, Sturm N, Guechot J, Paris A. Comparison of 9 blood tests and transient elastography for liver fibrosis in chronic hepatitis C: the ANRS HCEP-23 study. *J Hepatol* 2011; in press 2009.
51. Cales P, Veillon P, Konate A, Mathieu E, Ternisien C, Chevaller A, Godon A, et al. Reproducibility of blood tests of liver fibrosis in clinical practice. *Clin Biochem* 2008;41:10-18.
52. Cales P, Boursier J, Juttin I, Mahuet J, Gicquel S, Morin D, Maryheck G, et al. Repeatability of non invasive tests for liver fibrosis. *AASLD* 2010.
53. Cales P, Boursier J, Oberti F, Hubert I, Gallois Y, Rousselet MC, Dib N, et al. FibroMeters: a family of blood tests for liver fibrosis. *Gastroenterol Clin Biol* 2008;32:40-51.
54. Poynard T, Imbert-Bismut F, Munteanu M, Messous D, Myers RP, Thabut D, Ratziu V, et al. Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004;3:8.
55. Leroy V, Halfon P, Bacq Y, Boursier J, Rousselet MC, Bourliere M, de Muret A, et al. Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008;41:1368-1376.
56. Gressner OA, Beer N, Jodlowski A, Gressner AM. Impact of quality control accepted inter-laboratory variations on calculated Fibrotest/Actitest scores for the non-invasive biochemical assessment of liver fibrosis. *Clin Chim Acta* 2009;409:90-95.
57. Calès P, Boursier J, Bertrais S, Oberti F, Gallois Y, Fouchard-Hubert I, Dib N, et al. Accuracy of liver fibrosis classifications provided by non-invasive tests. *J Hepatol* 2010;52 (Suppl 1):S406.
58. Poynard T, Munteanu M, Imbert-Bismut F, Charlotte F, Thabut D, Le Calvez S, Messous D, et al. Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. *Clin Chem* 2004;50:1344-1355.

59. Cales P, Boursier J, de Ledinghen V, Halfon P, Bacq Y, Leroy V, Dib N, et al. Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests. *Gastroenterol Clin Biol* 2008;32:1050-1060.
60. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008;48:835-847.
61. Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, Christidis C, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol* 2003;29:1705-1713.
62. Lucidarme D, Foucher J, Le Bail B, Vergniol J, Castera L, Duburque C, Forzy G, et al. Factors of accuracy of transient elastography (fibroscan) for the diagnosis of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1083-1089.
63. Boursier J, Zarski JP, De Ledinghen V, Bacq Y, Rousselet MC, Fouchard-Hubert I, Gallois Y, et al. Determination of reliability criteria for liver stiffness evaluation in a cohort of 1165 patients. *J Hepatol* 2011;54:S130.
64. Shaheen AA, Wan AF, Myers RP. FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy. *Am J Gastroenterol* 2007;102:2589-2600.
65. Talwalkar JA, Kurtz DM, Schoenleber SJ, West CP, Montori VM. Ultrasound-based transient elastography for the detection of hepatic fibrosis: systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2007;5:1214-1220.
66. Friedrich-Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, Zeuzem S, Herrmann E. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. *Gastroenterology* 2008;134:960-974.
67. Stebbing J, Farouk L, Panos G, Anderson M, Jiao LR, Mandalia S, Bower M, et al. A meta-analysis of transient elastography for the detection of hepatic fibrosis. *J Clin Gastroenterol* 2010;44:214-219.
68. Degos F, Perez P, Roche B, Mahmoudi A, Asselineau J, Voitot H, Bedossa P, et al. Diagnostic accuracy of Fibroscan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study). *J Hepatol* 2010;53:1013-1021.
69. de Ledinghen V, Vergniol J. Transient elastography (FibroScan). *Gastroenterol Clin Biol* 2008;32:58-67.

70. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343-350.
71. Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, de Ledinghen V, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology* 2005;41:48-54.
72. Foucher J, Castera L, Bernard PH, Adhoute X, Laharie D, Bertet J, Couzigou P, et al. Prevalence and factors associated with failure of liver stiffness measurement using FibroScan in a prospective study of 2114 examinations. *Eur J Gastroenterol Hepatol* 2006;18:411-412.
73. Castera L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, Couzigou P, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010;51:828-835.
74. Boursier J, Konate A, Gorea G, Reaud S, Quemener E, Oberti F, Hubert-Fouchard I, et al. Reproducibility of liver stiffness measurement by ultrasonographic elastometry. *Clin Gastroenterol Hepatol* 2008;6:1263-1269.
75. Coco B, Oliveri F, Maina AM, Ciccorossi P, Sacco R, Colombatto P, Bonino F, et al. Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat* 2007;14:360-369.
76. Arena U, Vizzutti F, Corti G, Ambu S, Stasi C, Bresci S, Moscarella S, et al. Acute viral hepatitis increases liver stiffness values measured by transient elastography. *Hepatology* 2008;47:380-384.
77. Sagir A, Erhardt A, Schmitt M, Haussinger D. Transient elastography is unreliable for detection of cirrhosis in patients with acute liver damage. *Hepatology* 2008;47:592-595.
78. Seo YS, Lee KG, Jung ES, An H, Park S, Keum B, Yim HJ, et al. Dynamic changes in liver stiffness during the course of acute hepatitis A. *Scand J Gastroenterol* 2010;45:449-456.
79. Millonig G, Reimann FM, Friedrich S, Fonouni H, Mehrabi A, Buchler MW, Seitz HK, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *Hepatology* 2008;48:1718-1723.
80. Millonig G, Friedrich S, Adolf S, Fonouni H, Golriz M, Mehrabi A, Stiefel P, et al. Liver stiffness is directly influenced by central venous pressure. *J Hepatol* 2010;52:206-210.
81. http://www.has-sante.fr/portail/jcms/c_476401/methodes-non-invasives-devaluation-de-la-fibrose/cirrhose-hepatique.

82. Callewaert N, Van Vlierberghe H, Van Hecke A, Laroy W, Delanghe J, Contreras R. Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics. *Nat Med* 2004;10:429-434.
83. Vanderschaeghe D, Laroy W, Sablon E, Halfon P, Van Hecke A, Delanghe J, Callewaert N. GlycoFibroTest is a highly performant liver fibrosis biomarker derived from DNA sequencer-based serum protein glycomics. *Mol Cell Proteomics* 2009;8:986-994.
84. Huwart L, Sempoux C, Salameh N, Jamart J, Annet L, Sinkus R, Peeters F, et al. Liver fibrosis: noninvasive assessment with MR elastography versus aspartate aminotransferase-to-platelet ratio index. *Radiology* 2007;245:458-466.
85. Huwart L, Sempoux C, Vicaut E, Salameh N, Annet L, Danse E, Peeters F, et al. Magnetic resonance elastography for the noninvasive staging of liver fibrosis. *Gastroenterology* 2008;135:32-40.
86. Talwalkar JA, Gross JB, Venkatesh SK, Yin M, Glockner J, Takahashi N, Charlton MR, et al. Magnetic resonance elastography for the detection of moderate to severe hepatic fibrosis. *Hepatology* 2008;48 (Suppl 1):1110A.
87. Palmeri ML, Wang MH, Dahl JJ, Frinkley KD, Nightingale KR. Quantifying hepatic shear modulus in vivo using acoustic radiation force. *Ultrasound Med Biol* 2008;34:546-558.
88. Friedrich-Rust M, Wunder K, Kriener S, Sotoudeh F, Richter S, Bojunga J, Herrmann E, et al. Liver fibrosis in viral hepatitis: noninvasive assessment with acoustic radiation force impulse imaging versus transient elastography. *Radiology* 2009;252:595-604.
89. Yoneda M, Suzuki K, Kato S, Fujita K, Nozaki Y, Hosono K, Saito S, et al. Nonalcoholic Fatty Liver Disease: US-based Acoustic Radiation Force Impulse Elastography. *Radiology* 2010;256:640-647.
90. Fierbinteanu-Braticevici C, Andronescu D, Usvat R, Cretoiu D, Baicus C, Marinoschi G. Acoustic radiation force imaging sonoelastography for noninvasive staging of liver fibrosis. *World J Gastroenterol* 2009;15:5525-5532.
91. Lupsor M, Badea R, Stefanescu H, Sparchez Z, Branda H, Serban A, Maniu A. Performance of a new elastographic method (ARFI technology) compared to unidimensional transient elastography in the noninvasive assessment of chronic hepatitis C. Preliminary results. *J Gastrointest Liver Dis* 2009;18:303-310.
92. Takahashi H, Ono N, Eguchi Y, Eguchi T, Kitajima Y, Kawaguchi Y, Nakachita S, et al. Evaluation of acoustic radiation force impulse elastography for fibrosis staging of liver disease: a pilot study. *Liver Int* 2010;30:538-545.

93. Goertz RS, Zopf Y, Jugl V, Heide R, Janson C, Strobel D, Bernatik T, et al. Measurement of liver elasticity with acoustic radiation force impulse (ARFI) technology: an alternative noninvasive method for staging liver fibrosis in viral hepatitis. *Ultraschall Med* 2010;31:151-155.
94. Boursier J, Isselin G, Fouchard-Hubert I, Oberti F, Dib N, Bertrais S, Gallois Y, et al. Acoustic Radiation Force Impulse (ARFI): a new ultrasonographic technology for the widespread non-invasive diagnosis of liver fibrosis. *Eur J Gastroenterol Hepatol* 2010;22:1074-1084.
95. Rifai K, Bahr MJ, Mederacke I, Bantel H, Bayer D, Boozari B, Wedemeyer H, et al. Acoustic radiation force impulse (ARFI) as a new method of ultrasonographic elastography allows accurate and flexible assessment of liver stiffness. *J Hepatol* 2009;50 (Suppl 1):S88.
96. Muller M, Gennisson JL, Deffieux T, Tanter M, Fink M. Quantitative viscoelasticity mapping of human liver using supersonic shear imaging: preliminary in vivo feasibility study. *Ultrasound Med Biol* 2009;35:219-229.
97. Bavu E, Gennisson JL, Mallet V, Osmanski BF, Couade M, Bercoff J, Fink M, et al. Supersonic Shear Imaging is a new potent morphological non-invasive technique to assess of liver fibrosis. Part II: Comparison with Fibroscan. *J Hepatol* 2010;52 (Suppl 1):S166.
98. Poynard T, Halfon P, Castera L, Munteanu M, Imbert-Bismut F, Ratziu V, Benhamou Y, et al. Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007;53:1615-1622.
99. 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;349:825-832.
100. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol* 2005;12:1198-1204.
101. Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008;54:1372-1378.
102. Halfon P, Bourliere M, Deydier R, Botta-Fridlund D, Renou C, Tran A, Portal I, et al. Independent prospective multicenter validation of biochemical markers (fibrotest-actitest) for the prediction of liver fibrosis and activity in patients with chronic hepatitis C: the fibropaca study. *Am J Gastroenterol* 2006;101:547-555.

103. Bourliere M, Penaranda G, Renou C, Botta-Fridlund D, Tran A, Portal I, Lecomte L, et al. Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies. *J Viral Hepat* 2006;13:659-670.
104. Phelps CE, Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard". *Med Decis Making* 1995;15:44-57.
105. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol* 1990;93:252-258.
106. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol* 1999;52:943-951.
107. Mehta SH, Lau B, Afdhal NH, Thomas DL. Exceeding the limits of liver histology markers. *J Hepatol* 2009;50:36-41.
108. Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, Aube C, Gallois Y, et al. Noninvasive diagnosis of hepatic fibrosis or cirrhosis. *Gastroenterology* 1997;113:1609-1616.
109. Sebastiani G, Vario A, Guido M, Noventa F, Plebani M, Pistis R, Ferrari A, et al. Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol* 2006;44:686-693.
110. Bourliere M, Penaranda G, Ouzan D, Renou C, Botta-Fridlund D, Tran A, Rosenthal E, et al. Optimized stepwise combination algorithms of non-invasive liver fibrosis scores including Hepascore in hepatitis C virus patients. *Aliment Pharmacol Ther* 2008;28:458-467.
111. Becker L, Salameh W, Sferruzza A, Zhang K, Chen R, Malik R, Reitz R, et al. Validation of Hepascore, compared to simple indices of fibrosis, in US patients with chronic hepatitis C virus infection. *Clin Gastroenterol Hepatol* 2009.
112. Maor Y, Cales P, Bashari D, Kenet G, Lubetsky A, Luboshitz J, Schapiro JM, et al. Improving estimation of liver fibrosis using combination and newer noninvasive biomarker scoring systems in hepatitis C-infected haemophilia patients. *Haemophilia* 2007;13:722-729.
113. Kim SU, Ahn SH, Park JY, Kang W, Kim do Y, Park YN, Chon CY, et al. Liver stiffness measurement in combination with noninvasive markers for the improved diagnosis of B-viral liver cirrhosis. *J Clin Gastroenterol* 2009;43:267-271.
114. Wong GL, Wong VW, Choi PC, Chan AW, Chan HL. Development of a non-invasive algorithm with transient elastography (Fibroscan) and serum test formula for advanced liver fibrosis in chronic hepatitis B. *Aliment Pharmacol Ther* 2010;31:1095-1103.

115. Sebastiani G, Halfon P, Castera L, Pol S, Thomas DL, Mangia A, Di Marco V, et al. SAFE biopsy: a validated method for large-scale staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1821-1827.
116. Boursier J, De Ledinghen V, Zarski JP, Hubert I, Rousselet MC, Gallois Y, Oberti F, et al. New synchronous combinations of fibrosis tests provide a very accurate and precise diagnosis of liver fibrosis stages without liver biopsy. *AASLD* 2010.
117. Boursier J, Konate A, Guilluy M, Gorea G, Sawadogo A, Quemener E, Oberti F, et al. Learning curve and interobserver reproducibility evaluation of liver stiffness measurement by transient elastography. *Eur J Gastroenterol Hepatol* 2008;20:693-701.
118. Fraquelli M, Rigamonti C, Casazza G, Conte D, Donato MF, Ronchi G, Colombo M. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* 2007;56:968-973.
119. Neukam K, Macias J, Ferrete C, Palomares JC, Fernandez A, Mira JA, Pineda JA. Impact of observer experience on the reproducibility of transient elastometry in HIV/HCV co-infected patients. *HIV Clin Trials* 2009;10:276-281.
120. Nobili V, Vizzutti F, Arena U, Abraldes JG, Marra F, Pietrobattista A, Fruhwirth R, et al. Accuracy and reproducibility of transient elastography for the diagnosis of fibrosis in pediatric nonalcoholic steatohepatitis. *Hepatology* 2008;48:442-448.
121. Lucidarme D, Forzy G, Gremaux V, Filoche B. Interobserver Reproducibility of Liver Stiffness Measurement by Transient Elastography (FIBROSCAN®). *Hepatology* 2007;46 (Suppl 1):836A.
122. Cales P, Boursier J, Rousselet MC, Michalak S, Oberti F, Gallois Y, De Ledinghen V, et al. Comparison of reproducibility of histology, blood tests and Fibroscan for liver fibrosis. *Hepatology* 2007;46 (Suppl 1):834A.
123. Cales P, Boursier J, Bertrais S, Oberti F, Gallois Y, Fouchard-Hubert I, Dib N, et al. Optimization and robustness of blood tests for liver fibrosis and cirrhosis. *Clin Biochem* 2010;43:1315-1322.
124. Halfon P, Bacq Y, De Muret A, Penaranda G, Bourliere M, Ouzan D, Tran A, et al. Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C. *J Hepatol* 2007;46:395-402.
125. Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P, Morel F, et al. Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. *J Hepatol* 2007;46:775-782.

126. Boursier J, Cales P. Combination of fibrosis tests: sequential or synchronous? *Hepatology* 2009;50:656-657; author reply 657.
127. Boursier J, Vergniol J, Sawadogo A, Dakka T, Michalak S, Gallois Y, Le Tallec V, et al. The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. *Liver Int* 2009;29:1507-1515.
128. Boursier J, Bacq Y, Halfon P, Leroy V, De Ledinghen V, De Muret A, Bourliere M, et al. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009;21:28-38.
129. Masuzaki R, Tateishi R, Yoshida H, Goto E, Sato T, Ohki T, Imamura J, et al. Prospective risk assessment for hepatocellular carcinoma development in patients with chronic hepatitis C by transient elastography. *Hepatology* 2009;49:1954-1961.
130. Ngo Y, Munteanu M, Messous D, Charlotte F, Imbert-Bismut F, Thabut D, Lebray P, et al. A prospective analysis of the prognostic value of biomarkers (FibroTest) in patients with chronic hepatitis C. *Clin Chem* 2006;52:1887-1896.
131. Yu ML, Lin SM, Lee CM, Dai CY, Chang WY, Chen SC, Lee LP, et al. A simple noninvasive index for predicting long-term outcome of chronic hepatitis C after interferon-based therapy. *Hepatology* 2006;44:1086-1097.
132. Nunes D, Fleming C, Offner G, Craven D, Fix O, Heeren T, Koziel MJ, et al. Noninvasive markers of liver fibrosis are highly predictive of liver-related death in a cohort of HCV-infected individuals with and without HIV infection. *Am J Gastroenterol* 2010;105:1346-1353.
133. Fung J, Lai CL, Seto WK, Wong DK, Yuen MF. Prognostic significance of liver stiffness for hepatocellular carcinoma and mortality in HBeAg-negative chronic hepatitis B. *J Viral Hepat* 2010.
134. Ngo Y, Benhamou Y, Thibault V, Ingiliz P, Munteanu M, Lebray P, Thabut D, et al. An accurate definition of the status of inactive hepatitis B virus carrier by a combination of biomarkers (FibroTest-ActiTest) and viral load. *PLoS One* 2008;3:e2573.
135. Naveau S, Gaude G, Asnacios A, Agostini H, Abella A, Barri-Ova N, Dauvois B, et al. Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. *Hepatology* 2009;49:97-105.
136. Mayo MJ, Parkes J, Adams-Huet B, Combes B, Mills AS, Markin RS, Rubin R, et al. Prediction of clinical outcomes in primary biliary cirrhosis by serum enhanced liver fibrosis assay. *Hepatology* 2008;48:1549-1557.

137. Forestier J, Dumortier J, Guillaud O, Ecochard M, Roman S, Boillot O, Lutringer D, et al. Noninvasive diagnosis and prognosis of liver cirrhosis: a comparison of biological scores, elastometry, and metabolic liver function tests. *Eur J Gastroenterol Hepatol* 2010;22:532-540.
138. Boursier J, Cales P. Clinical interpretation of Fibroscan(R) results: a real challenge. *Liver Int* 2010;30:1400-1402.
139. Deuffic-Burban S, Deltenre P, Louvet A, Canva V, Dharancy S, Hollebecque A, Boitard J, et al. Impact of viral eradication on mortality related to hepatitis C: a modeling approach in France. *J Hepatol* 2008;49:175-183.
140. Castera L, Sebastiani G, Le Bail B, de Ledinghen V, Couzigou P, Alberti A. Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. *J Hepatol* 2010;52:191-198.

8. ANNEXES

Annexe 1 : Résultats des méta-analyses ayant évalué la performance du Fibroscan pour le diagnostic non invasif de la fibrose hépatique dans les hépatopathies chroniques.
(p 102)

Annexe 2 : Analyse des discordances entre le résultat du Fibrotest ou du Fibroscan et celui de l'examen histologique de la ponction-biopsie hépatique chez les patients atteints d'hépatite chronique C.
(p 103)

Annexe 3 : Performance des seuils diagnostiques *a posteriori de Youden* de 5 tests sanguins, calculés pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (F4) dans une large population de 1056 patients avec hépatite chronique C inclus dans 9 centres français (secondaires et tertiaires, publics et privés).
(p 104)

Annexe 4 : Performance des valeurs seuils du Fibroscan pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (Metavir F4) dans 2 populations de patients avec hépatite chronique C.
(p 105)

Annexe 5 : Performances de deux algorithmes décisionnels (cellules blanches : SAFE, cellules grises : algorithme de Bordeaux) combinant les tests non invasifs de fibrose pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (Metavir F4).
(p 106)

Annexe 6 : Avantages et inconvénients des divers indices diagnostiques utilisés pour l'évaluation de la performance des tests non invasifs de fibrose.
(p 107)

Annexe 7 : **Etude SNIFF 17**
(p 108) *Boursier Jérôme, Bacq Yannick, Halfon Philippe, Leroy Vincent, de Ledinghen Victor, de Muret Anne, Bourlière Marc, Sturm Nathalie, Foucher Juliette, Oberti Frédéric, Rousselet Marie-Christine, Calès Paul. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. Eur J Gastroenterol Hepatol 2009;21:28-38.*

Annexe 8 : **Etude VINDIAG 6**
(p 120) *Boursier Jérôme, Vergniol Julien, Sawadogo Apollinaire, Dakka Taoufiq, Michalak Sophie, Gallois Yves, Le Tallec Véronique, Oberti Frédéric, Fouchard-Hubert Isabelle, Dib Nina, Rousselet Marie-Christine, Konate Anselme, Amrani Naïma, de Ledinghen Victor, Calès Paul. The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. Liver Int 2009;29:1507-1515*

(p 130) Article Vindiag 6 : Annexes

Annexe 9 : Etude VINDIAG 7

(p 136)

Boursier Jérôme, de Ledinghen Victor, Zarski Jean-Pierre, Rousselet Marie-Christine, Sturm Nathalie, Foucher Juliette, Leroy Vincent, Fouchard-Hubert Isabelle, Bertrais Sandrine, Gallois Yves, Oberti Frédéric, Dib Nina, Cales Paul, and multicentric group ANRS HC EP 23 Fibrostar. A new combination of blood test and Fibroscan for accurate non-invasive diagnosis of liver fibrosis stages in chronic hepatitis C. Am J Gastroenterol, in press.

(p 146)

Article Vindiag 7 : Annexes

Annexe 10 :

(p 157)

Avantages (cases vertes) et inconvénients (cases rouges) des recommandations de la Haute Autorité de Santé française, et des combinaisons de tests non invasifs de fibrose pour le diagnostic non invasif de fibrose hépatique au cours de l'hépatite chronique C.

Annexe 1 : Résultats des méta-analyses ayant évalué la performance du Fibroscan pour le diagnostic non invasif de la fibrose hépatique dans les hépatopathies chroniques.

Meta-analyse	Etudes	Patients	Toutes étiologies					Hépatite chronique C			
			AUROC			Seuil diagnostique		AUROC		Seuil diagnostique	
			F \geq 2	F \geq 3	F4	F \geq 2	F4	F \geq 2	F4	F \geq 2	F4
Shaheen 2007 (64)	4	546	0.83	-	0.95	-	-	-	-	-	-
Talwalkar 2007 (65)	9	2083	0.87	-	0.96	-	-	-	-	-	-
Friedrich-Rust 2008 (66)	50	8478	0.84	0.89	0.94	7.65	13.01	0.85	-	-	-
Stebbing 2010 (67)	22	4430	-	-	-	7.81	15.56	-	-	8.44	16.14

Annexe 2 : Analyse des discordances entre le résultat du Fibrotest ou du Fibroscan et celui de l'examen histologique de la ponction-biopsie hépatique chez les patients atteints d'hépatite chronique C.

Auteur	Patients (n)	Test non invasif évalué	Discordances ^a n (%)	Attribution des discordances				
				Test non invasif		PBH	Indéterminé	
				n (%)	Causes	n (%)	Causes	n (%)
Poynard 2004 (58)	537	Fibrotest	87 (16,2)	13 (14,9)	<i>Faux positif :</i> • Hémolysé • Inflammation <i>Faux négatif :</i> • Inflammation	42 (48,3)	<i>Faux positif :</i> • Longueur • Stéatose <i>Faux négatif :</i> • PBH fragmentée	32 (36,8)
Halfon 2006 (102)	504	Fibrotest	92 (18,3)	27 (29,3)	• Hémolysé • Inflammation	19 (20,7)	-	46 (50,0)
Bourlière 2006 (103)	235	Fibrotest	46 (19,6)	21 (45,7)	-	11 (23,9)	-	14 (30,4)
Lucidarme 2009 (62)	254	Fibroscan	28 à 40 ^b (11,0 à 15,8) ^c	-	-	-	-	-

PBH : ponction-biopsie hépatique

^a Entre le test non invasif de fibrose évalué et la PBH

^b Selon le type de seuil diagnostique utilisé pour le Fibroscan.

^c Variables indépendamment associées à la présence d'une discordance : stade histologique de fibrose (F0-2 vs F3-4) et rapport IQR/médiane (%)

Annexe 3 : Performance des seuils diagnostiques *a posteriori de Youden* de 5 tests sanguins, calculés pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (F4) dans une large population de 1056 patients avec hépatite chronique C inclus dans 9 centres français (secondaires et tertiaires, publics et privés) ; d'après Calès *et al* (48, 59).

Cible	Test	Seuil	Se	Spe	VPP	VPN	RV+	RV-	DOR	PD	AUROC
F \geq 2	FibroMètre	0,419	80,0	76,0	78,5	77,6	3,33	0,26	12,65	78,1	0,853
	Fibrotest	0,435	67,7	81,9	80,1	70,2	3,74	0,39	9,48	74,5	0,811
	Hepascore	0,465	66,2	79,1	77,5	68,3	3,17	0,43	7,42	72,4	0,784
	APRI	0,548	62,4	83,5	80,5	67,0	3,78	0,45	8,38	72,5	0,786
	Fib4	1,116	73,9	71,9	74,2	71,6	2,63	0,36	7,25	73,0	0,799
F4	FibroMètre	0,628	95,7	71,1	29,7	99,2	-	-	-	-	-
	Fibrotest	0,660	82,4	77,2	29,2	97,5	-	-	-	-	-

Se : sensibilité; Spé : spécificité; VPP : valeur prédictive positive; VPN : valeur prédictive négative; RV+ : rapport de vraisemblance positif; RV- : rapport de vraisemblance négatif; DOR : diagnostic odds ratio; PD : performance diagnostique (taux de bien classés)

Annexe 4 : Performance des valeurs seuils du Fibroscan pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (Metavir F4) dans 2 populations de patients avec hépatite chronique C.

Cible	Etude	Patient (n)	Prévalence ^a (%)	AUROC	Seuil		Se	Spé	VPP	VPN	RV+	RV-	DOR
					Type	Valeur							
F \geq 2	Castera (70)	183	74,3	0,83	Youden	7,1	67	89	95	48	6,09	0,37	16,46
	Ziol (71)	251	65,0	0,79	Youden	8,8	56	91	88	56	6,63	0,48	13,81
F4	Castera (70)	183	25,1	0,95	Youden	12,5	87	91	77	95	9,66	0,14	69,00
	Ziol (71)	251	19,5	0,97	Youden	14,6	86	96	78	97	23,05	0,14	164,64

Se : sensibilité; Spé : spécificité; VPP : valeur prédictive positive; VPN : valeur prédictive négative; RV+ : rapport de vraisemblance positif; RV- : rapport de vraisemblance négatif; DOR : diagnostic odds ratio

^a Prévalence de la fibrose significative

Annexe 5 : Performances de deux algorithmes décisionnels (cellules blanches : SAFE, cellules grises : algorithme de Bordeaux) combinant les tests non invasifs de fibrose pour le diagnostic de fibrose significative (Metavir F \geq 2) ou de cirrhose (Metavir F4).

Cible	Etude	Algorithme	Patients (n)	Cible (%)	PD	Se	Spe	VPN	VPP	RV+	RV-	PBH
F \geq 2	Sebastiani (109) ^a	SAFE	100	65.0	94.2	-	-	100.0	92.7	5.7	0.0	57.0
	Sebastiani (115)	SAFE	2035	45.7	90.1	100.0	77.0	100.0	83.7	4.3	0.0	53.5
	Castera (140)	SAFE	302	76.0	97.0 ^b	100.0	87.3	100.0	96.3	7.9	0.0	51.7
	Boursier (116)	SAFE	1785	54.7	94.6	100.0	88.2	100.0	91.0	8.5	0.0	64.0
	Castera (140)	Bordeaux	302	76.0	87.7 ^b	85.1	97.2	66.9	98.9	30.4	0.15	28.1
	Boursier (116)	Bordeaux	729	58.3	88.3 ^c	88.0	88.7	84.5	91.4	7.8	0.14	34.6 ^d
F4	Sebastiani (109) ^a	SAFE	100	21.0	92.7	-	-	98.6	58.3	12.9	0.13	29.0
	Sebastiani (115)	SAFE	2035	9.4	92.5	90.4	92.7	99.0	55.7	16.5	0.11	18.5
	Castera (140)	SAFE	302	25.0	88.7 ^b	86.4	89.7	94.1	77.6	8.4	0.15	25.2
	Boursier (116)	SAFE	1785	12.7	89.5	61.5	93.4	94.6	56.2	9.3	0.41	6.4
	Castera (140)	Bordeaux	302	25.0	95.7 ^b	89.4	98.2	95.9	95.0	49.6	0.10	21.2
	Boursier (116)	Bordeaux	729	15.0	94.2 ^d	86.5	95.4	97.7	76.1	19.0	0.14	24.6 ^d

Abréviations : voir annexes précédentes.

^a Résultats dans la population de validation

^b p<10⁻⁴ entre le SAFE et l'algorithme de Bordeaux dans l'étude de Castera *et al*

^c p=0.010 entre le SAFE et l'algorithme de Bordeaux dans l'étude de Boursier *et al*

^d p<10⁻³ entre le SAFE et l'algorithme de Bordeaux dans l'étude de Boursier *et al*

Annexe 6 : Avantages et inconvénients des divers indices diagnostiques utilisés pour l'évaluation de la performance des tests non invasifs de fibrose.

Indice	Cible diagnostique	Classes diagnostiques	Diagnostic	Avantages	Inconvénients
<i>AUROC</i>	F \geq 2 cirrhose	-	-	<ul style="list-style-type: none"> • Evaluation globale de la performance du test 	<ul style="list-style-type: none"> • Non informatif pour l'utilisation du test en pratique clinique
<i>Diagnostic binaire</i>	F \geq 2	2	F0/1, F \geq 2	<ul style="list-style-type: none"> • Permet de poser un diagnostic en pratique clinique 	<ul style="list-style-type: none"> • Seuils diagnostiques encore à calculer et/ou valider dans de larges cohortes de patients. • "Pseudo-discordances" entre les tests ayant des valeurs proches des seuils diagnostiques.
	Cirrhose	2	F \leq 3, F4		
<i>Profil de performance</i>	F \geq 2	-	-	<ul style="list-style-type: none"> • Non influencé par la prévalence des stades de fibrose dans la population étudiée 	<ul style="list-style-type: none"> • Nécessite l'utilisation d'un seuil de diagnostic binaire pour la fibrose significative • Peu informatif pour l'utilisation du test en pratique clinique
<i>Algorithmes décisionnels</i>	F \geq 2	3	F0/1, F \geq 2, PBH	<ul style="list-style-type: none"> • Très bonne performance diagnostique 	<ul style="list-style-type: none"> • Taux de PBH nécessaire élevé • Nécessité de convoquer plusieurs fois le patient pour les algorithmes séquentiels (attitude économique) • Seuils diagnostiques des tests non invasifs non optimisés • Evaluation des performances diagnostiques dans les études non conforme à l'utilisation en pratique clinique
	Cirrhose	3	Pas de cirrhose, cirrhose, PBH		
<i>Intervalles de valeurs prédictives \geq90%</i>	F \geq 2	3	F0/1, indéterminé, F \geq 2	<ul style="list-style-type: none"> • Permet d'exclure ou d'affirmer le diagnostic de façon fiable en pratique clinique 	<ul style="list-style-type: none"> • Zone grise où le diagnostic est indéterminé (\geq55% des patients avec les tests sanguins) • Le calcul des seuils de valeurs prédictives dépend de la prévalence de l'événement dans la population étudiée
	Cirrhose	3	F \leq 3, indéterminé, F4		
<i>Intervalles de diagnostic fiable</i>	F \geq 2	4	F0/1, F1/2, F2/3, F \geq 2	<ul style="list-style-type: none"> • Diagnostic non invasif de fibrose fiable chez 100% des patients, sans aucune PBH nécessaire 	<ul style="list-style-type: none"> • Classification moins précise que celle du Metavir : le diagnostic final inclue aux moins deux stades de fibrose
	Cirrhose	3	Pas de cirrhose, F \geq 2, cirrhose		

F \geq 2 : fibrose cliniquement significative (classification Metavir)

Annexe 7 : Etude SNIFF 17.

Boursier Jérôme, Bacq Yannick, Halfon Philippe, Leroy Vincent, de Ledinghen Victor, de Muret Anne, Bourlière Marc, Sturm Nathalie, Foucher Juliette, Oberti Frédéric, Rousselet Marie-Christine, Calès Paul. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. Eur J Gastroenterol Hepatol 2009;21:28-38.

Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C

Jérôme Boursier^a, Yannick Bacq^c, Philippe Halfon^e, Vincent Leroy^g, Victor de Ledinghenⁱ, Anne de Muret^d, Marc Bourlière^f, Nathalie Sturm^h, Juliette Foucher^l, Frédéric Oberti^a, Marie Christine Rousselet^b and Paul Calès^a

Objective Blood tests are usually designed to identify significant fibrosis. We evaluated their diagnostic accuracy, and how to increase it, for the clinically important targets of severe fibrosis and cirrhosis.

Methods The accuracy for severe fibrosis or cirrhosis of four blood tests was evaluated based on Metavir staging in 1056 patients with chronic hepatitis C recruited in five independent hospitals.

Results Using original scores, an original diagnostic target (significant fibrosis) and best diagnostic cutoff, the correct classification rates in severe fibrosis and cirrhosis stages were, respectively: FibroMeter: 90.1, 100%, Fibrotest: 78.2, 95.1%, Hepascore: 73.8, 94.9%, aspartate aminotransferase to platelet ratio index (APRI): 71.4, 88.0% ($P < 0.003$, $P = 0.004$, respectively, between tests). The corresponding area under the receiver operating characteristics were FibroMeter: 0.885, 0.907, Fibrotest: 0.837, 0.882, Hepascore: 0.834, 0.896, APRI: 0.822, 0.841 ($P < 0.003$, respectively). Observed 100% negative predictive values for severe fibrosis and cirrhosis were, respectively, FibroMeter: 15.4, 47.5%, Fibrotest: 3.6, 31.9%, Hepascore: 0.3, 24.6%, APRI: 1.4, 5.3% of patients ($P < 0.003$, respectively, between tests). By calculating a specific test for cirrhosis, including the FibroMeter markers, the correct classification (93.0%) was significantly higher for the cirrhosis diagnosis compared with the original FibroMeter (90.9%, $P = 0.005$).

Introduction

Blood tests for liver fibrosis are based on an algorithm (or score) that takes into account several blood or clinical markers [1]. They are usually designed to diagnose significant fibrosis, which includes all stages with bridging fibrosis [2]. The result is a binary or qualitative diagnosis, as distinct from a quantitative diagnosis obtained by measuring parameters such as liver stiffness [3] or the area of fibrosis by a blood test [4]. Significant fibrosis has been adopted as a target because it can be identified using semiquantitative histological staging in patients with chronic hepatitis C, and because it is an indicator for treatment. Severe fibrosis and cirrhosis, however, are also clinically relevant targets, which warrant screening

This specific test provided a 100% positive predictive value for cirrhosis diagnosis versus 88% for original FibroMeter.

Conclusion Using the most accurate original test, cirrhosis can be excluded in 47.5% of patients and is correctly diagnosed, as significant fibrosis, in 100% of patients. A specific test for cirrhosis provides a significant gain in diagnostic accuracy to 93% and in positive predictive value to 100% compared with the original test. *Eur J Gastroenterol Hepatol* 21:28–38 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.

European Journal of Gastroenterology & Hepatology 2009, 21:28–38

Keywords: blood markers, cirrhosis, diagnostic accuracy, liver biopsy, liver fibrosis, Metavir staging, noninvasive diagnosis, sensitivity, specificity, viral hepatitis C

^aHepatogastroenterology Department, ^bPathology Department and HIFIH Laboratory, CH&U, Angers, ^cHepatogastroenterology Department, ^dPathology Department, Trousseau Hospital, CHRU, Tours, ^eAlphabio Laboratory, Marseille, ^fHepatogastroenterology Department, Saint-Joseph Hospital, Marseille, ^gHepatogastroenterology Department, INSERM/UJF U823, IAPC, IAB, ^hPathology Department, CHU, Grenoble and ⁱHepatogastroenterology Department, Haut Leveque Hospital, Pessac, CHU, Bordeaux, France

Correspondence to Professor Paul Calès, MD, Hepatogastroenterology Department, CHU, 49933 Angers Cedex 09, France
Tel: +33 2 41 35 34 10; fax: +33 2 41 35 41 19;
e-mail: PaCales@chu-angers.fr

Received 16 March 2008 Accepted 16 June 2008

for liver complications including esophageal varices and hepatocellular carcinoma. It is therefore important to be aware of, and optimize, the performance of blood tests in identifying these alternative diagnostic targets.

This objective requires that the accuracy of blood tests be evaluated in the stages of severe fibrosis and cirrhosis. In addition, we must determine how accurate blood tests, originally constructed to detect significant fibrosis ('original' tests), are at identifying the alternative diagnostic targets. This implies that the determination of optimal diagnostic cutoffs of blood test values that distinguish patients with or without these alternative targets is necessary. Finally, we have to find out whether

there is any value in replacing the original score of a blood test constructed for significant fibrosis with another score using the same markers but specifically constructed for the alternative diagnostic targets.

Thus, the main aims of this study were to comprehensively describe the overall accuracy of original blood tests in identifying alternative diagnostic targets, including severe fibrosis and cirrhosis, and to determine whether it was possible to improve overall accuracy by implementing a specific score for them. To achieve those objectives, we needed a large population of patients for whom liver biopsy and numerous blood markers were available for evaluation and comparison. We decided to include four different tests: aspartate aminotransferase (AST) to platelet ratio index (APRI) [5], Fibrotest [6], Hepascore [7], and FibroMeter [4].

Methods

Data source

Using the Medline database and a manual search, we systematically reviewed the literature from 1997 to June 2007 for studies comparing FibroMeter and Fibrotest in patients with chronic viral hepatitis C (HCV), for whom liver biopsy data were available. Information provided by the tests enabled Hepascore and APRI to be calculated. Three independent publications were retrieved. The first study, involving one center (Angers), was the original publication that described the FibroMeter test [4]. The second study included two independent centers, Provence Alpes-Côte d'Azur (PACA) and Tours [8]. The third study was from Grenoble [9]. The PACA center incorporated three secondary and two tertiary care settings. Two additional unpublished populations were provided by Angers and Bordeaux centers [3]. Thus, in this meta-analysis, individual patient data were available from five centers, independent for study design, patient recruitment, blood marker analysis, and interpretation of liver histology.

Patients

Inclusion and exclusion criteria were very similar at all five centers. Patients with chronic HCV were prospectively included from 1994 to 2007 if they had anti-HCV antibodies, HCV RNA in serum, and available liver biopsy and blood markers. Fasting blood samples were collected immediately before or no more than 3 months after the liver biopsy was performed. Patients in the Tours and PACA centers were excluded if their liver specimen was smaller than 15 mm. Other exclusion criteria were additional causes of liver disease, particularly HIV or HBV coinfection, complicated cirrhosis, antifibrotic treatment in the earlier 6 months, and alcohol consumption of more than 30 g/day in the 5 years before inclusion. Overall, the five centers provided 1535 patients, of whom 479 were excluded because of missing criteria or data. This left a

core population of 1056, distributed over the centers as follows: Angers ($n = 300 + 135$), PACA ($n = 217$), Tours ($n = 159$), Grenoble ($n = 149$), and Bordeaux ($n = 96$). The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and was approved by local ethics committees.

Blood measurements

Blood samples were processed independently at each center. Variables determined were platelet count, urea, bilirubin, γ -glutamyl transpeptidase, AST and alanine aminotransferase, prothrombin index, apolipoprotein A1, haptoglobin, hyaluronic acid, and α 2-macroglobulin. Direct markers were measured in either fresh blood or a frozen sample of serum stored at less than or equal to -20°C . Indirect markers were usually measured in fresh blood. Analytical methods for blood tests were recorded to check that they were in agreement with recommendations. They were centralized in the five centers, in particular in the PACA area the laboratory participated to several quality studies for Fibrotest [10]. We have observed an excellent interlaboratory reproducibility of blood tests with various methods [11]. Original blood tests were calculated according to the most recent published [4,5,7,12] formulae and according to WO 02/16949 patent for Fibrotest. The area of liver fibrosis was estimated using the FibroMeter designed for that target and expressed as the percentage of whole liver area [4]. AST used in APRI was divided by a common upper limit of normal, as several studies performed in numerous laboratories in France have shown high interlaboratory reproducibility [11].

Liver biopsy

Liver biopsies were performed using Menghini's technique with a 1.4–1.6-mm diameter needle. Biopsy specimens were fixed in a formalin–alcohol–acetic solution and embedded in paraffin; 5- μm thick sections were then cut and stained with hematoxylin–eosin–safran. Liver fibrosis was staged from F0 to F4 according to the Metavir staging system [13]. Three diagnostic targets were defined as follows: significant fibrosis, F2 + F3 + F4; severe fibrosis, F3 + F4; and cirrhosis, F4. Readings were performed by independent, blinded, senior pathologists specialized in hepatology. Histological assessments were made twice by the same pathologist in Grenoble, once in Bordeaux and once by each of two pathologists in Angers, Tours, and PACA, with a common final reading in cases of disagreement. Therefore, a centralized reading was performed in three centers.

Outcomes

The main objective of this study was to evaluate and compare the diagnostic accuracy of original blood tests, initially constructed for the diagnosis of significant

fibrosis, in the diagnosis of severe fibrosis and cirrhosis (Fig. 1, ②), here called ‘alternative diagnostic targets’.

Secondary objectives were to evaluate:

1. the best diagnostic cutoff values of original blood tests to diagnose alternative diagnostic targets (Fig. 1, ②),
2. the rates of correct classification of the original blood tests in the stages of severe fibrosis or cirrhosis (Fig. 1, ①),
3. the rates of reliable diagnosis, that is, with high predictive values (Fig. 1, ②), and
4. a new test specific for the diagnosis of severe fibrosis or cirrhosis compared with the original test (Fig. 1, ③ or ④, respectively); this endpoint was restricted to the most accurate blood test.

Detailed definitions are listed in the Appendix.

Statistical analysis

The diagnostic performance of each test was expressed as the area under the receiver operating characteristic (AUROC), the overall accuracy and detailed indices such as likelihood ratio and diagnostic odds ratio [14–16]. AUROC findings were compared using the DeLong test [17]. Data were reported according to Standards for Reporting of Diagnostic Accuracy statements [18]. The size of the population was that necessary to detect a significant difference between the two most accurate tests according to a preliminary study [19] in the diagnosis of cirrhosis. With an α risk of 0.05, β risk of

0.2, cirrhosis prevalence of 0.11, AUROC correlation of 0.75 and bilateral testing, the required sample size was 910 patients for the following AUROC values: FibroMeter: 0.93, Fibrotest: 0.89.

Results

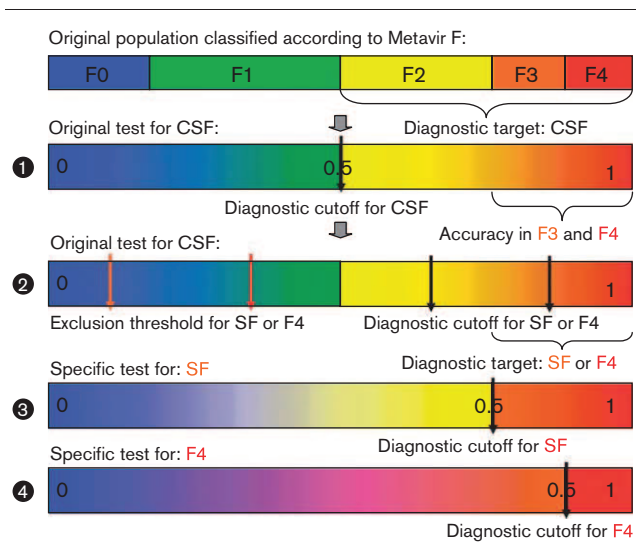
Characteristics of patients

The principal characteristics of the 1056 patients were as follows: sex: 59.5% male; mean age: 45.6 ± 12.5 years; Metavir fibrosis stage: F0: 4.4%, F1: 43.5%, F2: 27.0%, F3: 14.0%, F4: 11.2%. Thus, the prevalence of diagnostic targets was significant fibrosis: 52.0%, severe fibrosis: 25.0%, and cirrhosis: 11.0%. Mean liver specimen length was 21 ± 8 mm; 84.0% of liver specimens were greater than or equal to 15 mm and 58.2% were greater than or equal to 20 mm. No correlation between Metavir fibrosis stage and liver specimen length was observed (r_s : 0.00, $P = 0.872$).

Classification rates

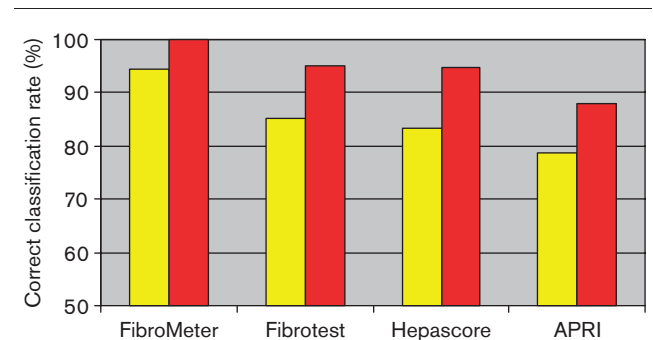
The correct classification rates of blood tests for the diagnosis of significant fibrosis, based on liver biopsy, are presented as a function of Metavir fibrosis stage in Fig. 2 (Table A). A rate of 100% in F4 means that 100% of patients with cirrhosis were correctly classified as having significant fibrosis. In brief, the correct classification rate for significant fibrosis was higher in cirrhosis than in severe fibrosis for all blood tests. This rate was significantly higher with FibroMeter than with other blood tests in F3 or severe fibrosis (F3 + F4) or F4 (borderline significance with Hepascore) stages. No significant difference between Fibrotest, Hepascore, and APRI was observed (with Youden diagnostic cutoff; except in severe fibrosis). These results are explained by the distribution of blood test values as a function of the Metavir fibrosis stages (Fig. A). For example, the

Fig. 1



Design of methodology. Numbers from ① to ④ refer to Methods section. CSF, clinically significant fibrosis; SF, severe fibrosis; F4, cirrhosis.

Fig. 2



Correct classification rate for the diagnosis of significant fibrosis by blood tests in the stages of severe fibrosis ($P < 0.003$ between all tests) and cirrhosis ($P = 0.004$). Diagnostic cutoff according to maximum Youden index (see Table A). APRI, aspartate aminotransferase to platelet ratio index. ■ Severe fibrosis, ■ Cirrhosis.

higher and narrower values in cirrhosis explain the high performance of Hepascore and FibroMeter in the cirrhosis stage and their suitability for cirrhosis diagnosis compared with other blood tests.

Overall performance

Original blood tests

We present here the original blood tests constructed for significant fibrosis. Main indices of diagnostic performance of tests for the diagnosis of severe fibrosis (F3 + F4) are presented in Table 1 and those for the diagnosis of cirrhosis (F4) in Table 2. AUROC data are presented in Table 3.

Comparison between tests: The AUROC of FibroMeter was significantly higher than that of other blood tests for severe fibrosis and cirrhosis (with the exception of Hepascore in cirrhosis). With the exception of Hepascore being superior to APRI in cirrhosis, the AUROCs of other blood tests for severe fibrosis and cirrhosis were not significantly different.

Comparison between diagnostic targets: Considering all blood tests, overall accuracy and AUROC values for cirrhosis were better than those for severe fibrosis.

Impact on the diagnostic target: We detail here the original FibroMeter (constructed for significant fibrosis) using the best performing diagnostic cutoff. In Table 4, the overall accuracy shows a significant increase in severe fibrosis or cirrhosis compared with significant fibrosis as the diagnostic target. Paradoxically, this was associated with a significant decrease in the rate of correct classification in the fibrosis stages corresponding to the diagnostic

Table 3 AUROCs (95% CI) of original blood tests and their comparisons (P value of Delong test) as a function of diagnostic target

	Severe fibrosis	Cirrhosis
FibroMeter	0.885 (0.863–0.906)	0.907 (0.885–0.929)
Fibrotest	0.837 (0.809–0.865)	0.882 (0.855–0.910)
Hepascore	0.834 (0.806–0.862)	0.896 (0.868–0.924)
APRI	0.822 (0.792–0.852)	0.841 (0.803–0.880)
All blood tests	< 0.003	< 0.003
FibroMeter vs. Fibrotest	< 0.003	0.041
FibroMeter vs. Hepascore	< 0.003	0.203
FibroMeter vs. APRI	< 0.003	< 0.003
Fibrotest vs. Hepascore	0.307	0.592
Fibrotest vs. APRI	0.324	0.131
Hepascore vs. APRI	0.479	0.013

APRI, aspartate aminotransferase to platelet ratio index; AUROC, area under the receiving operating characteristic; CI, confidence interval. Significant differences are indicated in bold.

Table 1 Diagnostic indices of original blood tests for severe fibrosis

Test	Cutoff ^a	κ^b	Sensitivity	Specificity	+ PV	– PV	+ LR	– LR	DOR	OA	AUROC ^c
FibroMeter	0.628	0.549	83.6	79.4	57.8	93.4	4.06	0.21	19.63	80.5	0.885
	0.830	0.534	59.8	91.2	69.6	87.0	6.75	0.44	15.30	83.2	
Fibrotest	0.448	0.447	84.4	71.1	48.7	93.5	2.93	0.21	13.76	74.5	0.837
	0.631	0.479	66.8	83.8	57.2	88.6	4.13	0.40	10.43	79.7	
Hepascore	0.497	0.428	82.1	70.7	48.3	92.2	2.80	0.25	11.04	73.5	0.834
	0.904	0.458	48.1	93.0	69.6	84.3	6.86	0.56	12.30	81.8	
APRI	0.581	0.443	77.5	74.6	50.6	90.9	3.05	0.30	10.18	75.3	0.822
	1.159	0.463	50.8	91.7	67.3	84.7	6.14	0.54	11.43	81.4	

APRI, aspartate aminotransferase to platelet ratio index; AUROC, area under the receiver operating characteristic; DOR, diagnostic odds ratio; LR, likelihood ratio; OA, overall accuracy; PV, predictive value.

^aDiagnostic cutoff determined *a posteriori* according to maximum Youden index (first) or the maximum overall accuracy (second).

^bKappa index measuring the agreement between blood tests and liver biopsies for the diagnosis of severe fibrosis.

^cAUROC is independent of cutoff.

Table 2 Diagnostic indices of original blood tests for cirrhosis

Test	Cutoff ^a	κ^b	Sensitivity	Specificity	+ PV	– PV	+ LR	– LR	DOR	OA	AUROC ^c
FibroMeter	0.628	0.339	95.7	71.1	29.7	99.2	3.30	0.06	53.99	73.8	0.907
	0.979	0.427	35.7	98.0	69.5	92.2	17.79	0.66	27.09	90.9	
Specific test ^d	0.089	0.428	87.0	81.0	37.0	97.9	4.59	0.16	28.55	81.8	0.919
	0.442	0.602	54.8	97.9	76.8	94.4	25.90	0.46	56.05	93.0	
Fibrotest	0.660	0.330	82.4	77.2	29.2	97.5	3.62	0.23	15.83	77.7	0.882
	0.862	0.411	42.2	95.5	51.8	93.6	9.44	0.61	15.60	90.1	
Hepascore	0.801	0.412	80.4	82.5	36.6	97.1	4.58	0.24	19.23	82.2	0.896
	0.999	0.472	38.5	98.5	76.3	92.7	25.55	0.62	40.89	91.8	
APRI	0.652	0.291	84.6	71.5	27.1	97.4	2.97	0.22	13.79	72.3	0.841
	2.532	0.333	27.4	97.7	61.5	91.5	12.76	0.74	17.19	90.0	

One specific test is also depicted.

APRI, aspartate aminotransferase to platelet ratio index; AUROC, area under the receiving operating characteristic; DOR, diagnostic odds ratio; LR, likelihood ratio; OA, overall accuracy; PV, predictive value.

^aDiagnostic cutoff determined *a posteriori* according to maximum Youden index (first) or the maximum overall accuracy (second).

^bKappa index measuring the agreement between blood tests and liver biopsies for the diagnosis of cirrhosis.

^cAUROC is independent of cutoff.

^dDerived from FibroMeter (same blood markers but constructed for the diagnosis of cirrhosis).

Table 4 Correct classification rate (%) for the original FibroMeter test as a function of diagnostic target and fibrosis stage (F) according to different diagnostic cutoffs

Cutoff	Value	Diagnostic target															
		Significant fibrosis					Severe fibrosis					Cirrhosis					
		≥ F3	F3	F4	F0–F4	Value	≥ F3	F3	F4	F0–F4	<i>P</i> ^a	Value	≥ F3 ^b	F3 ^b	F4	F0–F4	<i>P</i> ^a
Youden	0.41	94.5	90.1	100	78.1	0.628	83.6	73.8	95.7	80.5	0.181	0.628	57.4	26.2	95.7	73.8	0.034
Accuracy	– ^c	–	–	–	–	0.830	59.8	46.8	75.7	83.2	0.004	0.979	67.6	93.6	35.7	90.9	<0.003
<i>P</i> ^d	–	–	–	–	–	–	<0.003	<0.003	<0.003	0.027	–	–	0.051	<0.003	<0.003	<0.003	–

^aComparison of overall rate (F0–F4) for the alternative target vs. corresponding overall rate for significant fibrosis by McNemar test.

^bStrong decrease in rates because of antagonism between specific and original diagnostic targets.

^cDiagnostic cutoff with maximum Youden and overall accuracy were identical because of the prevalence (0.52) of diagnostic target.

^dComparison of classification rates according to cutoffs by McNemar test.

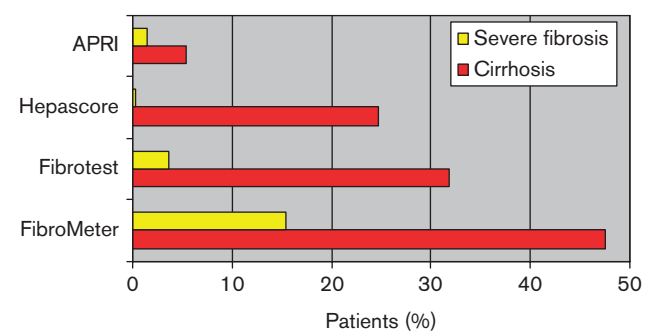
target (i.e. F3 + F4 in severe fibrosis and F4 in cirrhosis). For example, the overall accuracy was 78.1% for the diagnosis of significant fibrosis versus 90.9% for the diagnosis of cirrhosis, whereas the correct classification rate in patients with cirrhosis decreased from 100 to 35.7%, respectively.

Specific blood test

We compare here the original FibroMeter and a derived test including the same markers but specifically constructed for the diagnosis of severe fibrosis or cirrhosis (its diagnostic indices are described in Table 2). We described the results with the best-performing diagnostic cutoff (Table B). In brief, the specific test significantly increased the overall accuracy for the diagnosis of cirrhosis compared with the original FibroMeter: 93.0 versus 90.9% ($P = 0.005$), respectively (AUROCs: 0.919 vs. 0.907). In addition, the specific test significantly increased the correct classification rate in patients with cirrhosis compared with the original FibroMeter (54.8 vs. 35.7%, respectively, $P < 0.003$) without decreasing it in patients without cirrhosis (97.9 vs. 98.0% respectively, $P = 1$) contrary to the original FibroMeter as described previously (Table 4). The specific test did not significantly increase the overall accuracy for the diagnosis of severe fibrosis compared with the original FibroMeter: 84.4 versus 83.2% ($P = 0.213$), respectively (AUROCs: 0.888 vs. 0.885).

Reliable diagnosis: predictive values

As diagnosis of cirrhosis or severe fibrosis is clinically important, we looked at 95 and 100% predictive values. With original tests, these thresholds were available only for the negative predictive values (NPV) that are presented in Fig. 3 (Table C). The proportion of patients with exclusion (100% NPV) of severe fibrosis or cirrhosis was in the following order of significance: FibroMeter > Fibrotest > Hepascore > APRI (the order of the last two was inverted for severe fibrosis). For example, with the two most accurate tests, cirrhosis or severe fibrosis could be excluded in, respectively, 47.5 and 15.4% of patients with FibroMeter versus 31.9 and 3.6% with Fibrotest ($P < 0.003$ between tests, respectively).

Fig. 3

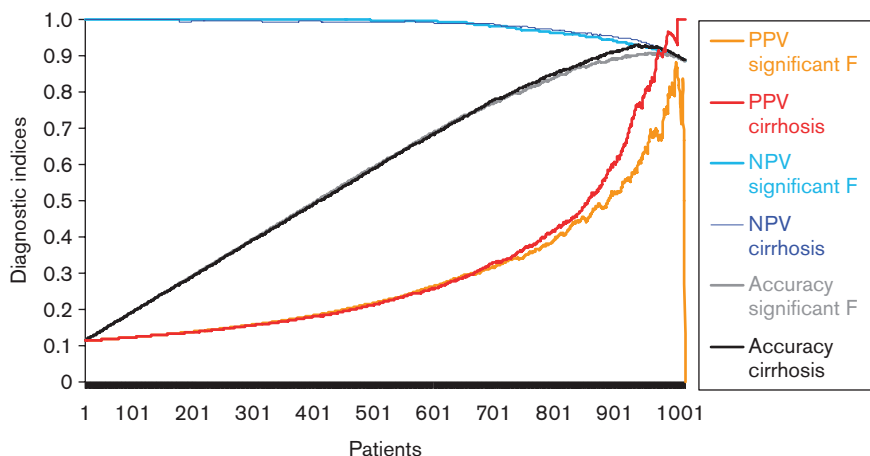
Patient rate (%) where the diagnosis of cirrhosis or severe fibrosis is excluded (100% NPV) by original blood tests. $P < 0.005$ for all pairs of tests within each diagnostic target. APRI, aspartate aminotransferase to platelet ratio index; NPV, negative predictive value.

In contrast, with the original FibroMeter the maximum positive predictive value (PPV) for cirrhosis was only 88.2%. The corresponding test specific for cirrhosis, however, was able to provide a 100% PPV for cirrhosis in 1.3% of patients (Fig. 4), that is, 11.3% of patients with cirrhosis. The patients with 100% PPV were characterized by a much higher estimated area of fibrosis than in other patients with cirrhosis: 29.8 ± 10.8 versus $13.0 \pm 6.3\%$, respectively ($P < 0.003$).

Comparison of original and specific tests

The different tests, original or specific, described here are both available in practice, as the constitutive blood variables are the same. The two options, however, imply that we have to make a choice for clinical application. Figure 4 depicts several diagnostic indices for the diagnosis of cirrhosis as a function of the original FibroMeter and the specific test. The specific test for cirrhosis produces a close NPV and, as already mentioned, a higher overall accuracy and, above all, a higher PPV compared with the original FibroMeter. In addition, the peak PPV (100%) of the specific test is well shaped, following a progressive curve. In contrast, the peak PPV

Fig. 4



Value of diagnostic indices for cirrhosis (y-axis) as a function of patients sorted by ascending blood test values (x-axis) according to original FibroMeter test constructed for significant fibrosis (F, lighter color) or specific test constructed for cirrhosis (darker color). PPV, positive predictive values (red and orange lines); NPV, negative predictive value (blue lines); accuracy, overall accuracy (black and gray lines).

(88%) of original FibroMeter is followed by a marked decrease that precludes reliable clinical use. The 100% NPV of the specific test, however, includes only 15.6% of patients versus 47.5% ($P < 0.003$) for the original FibroMeter. Thus, the main advantage of a specific test is to provide a high (100%) and robust PPV, whereas the original FibroMeter provides 100% NPV in a larger population. In addition, the specific test better depicts the degree of fibrosis than does the original FibroMeter, as reflected by a higher correlation with the estimated area of fibrosis: r_p of 0.72 and 0.48, respectively, $P < 0.003$ (Fig. 5).

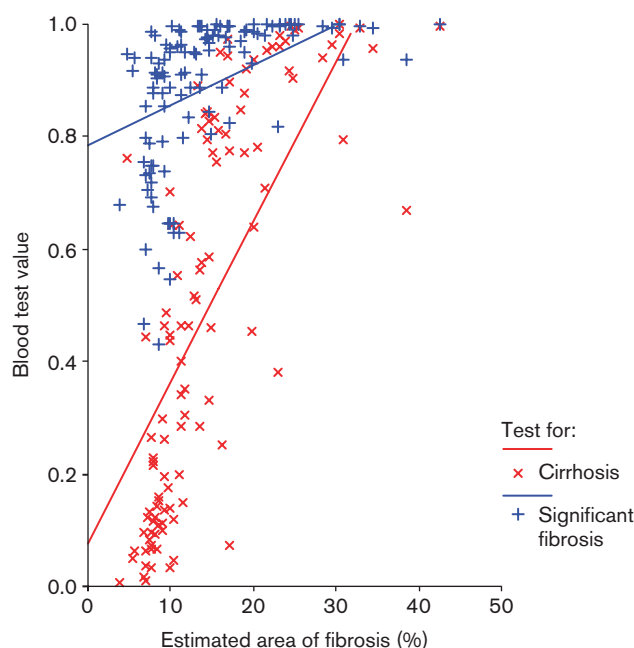
Discussion

Methods

The expression of the results of a blood test is a function of four linked factors: the diagnostic target, the fibrosis stage of special interest (e.g. cirrhosis when diagnostic target is severe fibrosis), the score used (original for significant fibrosis or specific for other targets), and the diagnostic cutoff. It is not necessarily a matter of choice over what to calculate, as software can provide a range of different results from the same data, but the presentation has to be simplified to be clinically applicable.

With the most accurate test, only 7% of patients were misclassified as having cirrhosis. This suggests that high overall accuracy can be attained despite the misclassification rate attributed to the reference, that is, liver biopsy [20]. Nevertheless, interobserver variability is higher in intermediate fibrosis stages, especially Metavir F2 and F3, than in F4 [20], and therefore could be largely responsible for the apparent decrease of the correct classification rate of blood tests in intermediate fibrosis

Fig. 5



Correlation between the estimated area of fibrosis and original FibroMeter for clinically significant fibrosis (in blue) (r_p : 0.48) or derived specific test for cirrhosis (in red) (r_p : 0.72) in patients with cirrhosis.

stages and consequently for the diagnosis of significant fibrosis or severe fibrosis compared with that of cirrhosis.

Biopsy specimens were not reexamined centrally in two centers, as we have shown in an earlier study (involving most of the same pathologists) that interobserver

agreement on the Metavir staging system is excellent among senior hepatopathologists [20]. In addition, the lack of correlation between Metavir fibrosis stage and liver specimen length indicates that the specimen length had no influence on staging. This in turn suggests an excellent pathologist expertise in the present series, contrary to other series [20].

The center effect was evaluated in another study issued from the same cohort: the coefficient of variation in diagnostic cutoff of FibroMeter was low, 4%, and this percentage was significantly lower than that of other blood tests [21].

Overall accuracy

Overall accuracy of the original tests for the diagnosis of severe fibrosis was in the following significant order: FibroMeter > Fibrotest \approx Hepascore \approx APRI, and for cirrhosis: FibroMeter > Fibrotest \approx APRI with an intermediate rank for Hepascore \approx FibroMeter or Fibrotest, and Hepascore > APRI (Table 3). The performance of FibroMeter for virus has been validated in other populations such as coinfections [22] or HBV infection [23]. Overall accuracy and the AUROC clearly show that performance increases as a function of the diagnostic target (higher in cirrhosis).

Classification rate

The diagnostic indices are usually global (overall accuracy and AUROC). The correct classification rate has the advantage of describing the performance of a test in individual stages of fibrosis in addition to all or several stages combined [8] and thus, is independent of the prevalence of the diagnostic target. Therefore, in severe fibrosis and cirrhosis, as well as in all stages combined, the correct classification rate of FibroMeter was significantly higher than that of other blood tests (Fig. 2). Finally, with the best diagnostic cutoff, FibroMeter was able to identify 100% of patients with cirrhosis in the group correctly classified as having significant fibrosis. Another advantage of the correct classification rate is to allow comparison of performance between different diagnostic targets. Indeed, unpaired comparisons (between different targets or patient groups) of AUROCs are unusual. Thus, Table 4 clearly indicates that the overall accuracy of FibroMeter was significantly higher for the diagnosis of severe fibrosis (with the best performing cutoff), and even more so for that of cirrhosis, compared with the diagnosis of significant fibrosis. All these comparisons are, however, based on original tests constructed for the diagnosis of significant fibrosis.

Specific test for alternative diagnostic targets

In this study, we have evaluated a specific test constructed for the diagnosis of severe fibrosis or cirrhosis including the same markers as in the original FibroMeter.

The gain in overall accuracy for cirrhosis was significant when the test was specific for cirrhosis compared with the original test. In contrast, the overall accuracy of cirrhosis was significantly higher when the blood test, including the same markers, was constructed for the diagnosis of cirrhosis rather than for that of significant fibrosis. That may sound like a truism, but it was not the case for the diagnosis of severe fibrosis. Adapting a test with a specific score to a diagnostic target, however, induces an apparent paradox. Indeed, overall accuracy increases, but the correct classification rate in the fibrosis stage(s) corresponding to the diagnostic target decreases. Nevertheless, this can be partially circumvented by careful selection of the diagnostic cutoff (Appendix). In brief, in terms of cirrhosis diagnosis, the best option is to use the specific test with an accuracy cutoff. The issue of choice regarding diagnostic cutoff, with the implication of binary diagnosis (yes/no) of the diagnostic target, can also be easily circumvented. Indeed, there is another statistical technique that determines the most probable fibrosis stage(s) for a given range of blood test values [21]. Combining tests represents another possibility to increase rates of reliable binary diagnosis. Combinations can be synchronous algorithms, for example, blood test and Fibroscan [3] or sequential algorithms of blood tests followed by liver biopsy in the gray zone of blood tests with unreliable diagnosis [24].

Original and/or specific test?

The choice between the original and specific tests concerns several aspects of the diagnosis of cirrhosis. Besides diagnostic cutoff, the other thresholds of interest are those concerning a reliable diagnosis. A patient-based approach is usually taken in this area, with thresholds corresponding to 90% PPV and NPV. Outside this range, liver biopsy is considered unnecessary for fibrosis staging [6,9]. These predictive value thresholds, however, depend on the prevalence of the diagnostic target and are therefore not easily exportable. Here, we circumvented this difficulty by considering 100% predictive values. Indeed, 100% NPV and PPV also correspond to 100% sensitivity and specificity, respectively, which are independent of diagnostic target prevalence. Thus, 100% NPV and PPV can be applied everywhere contrary to 90% predictive values. In addition, it is especially important to avoid any false diagnosis for cirrhosis.

With the original tests, a 100% NPV threshold was available in a significantly larger population than with specific test. Interestingly, this exclusion threshold included 47.5% of the present population with original FibroMeter, a much higher proportion than with other blood tests (Fig. 3). The only useful test for the exclusion of severe fibrosis was the original FibroMeter, but the proportion of patients was lower, at 15.3%.

It is noteworthy that the specific test for cirrhosis (derived from FibroMeter) provided a 100% PPV, unlike the original test (Fig. 4). This involved 1.3% of all patients or 11.3% of patients with (compensated) cirrhosis. Indeed, it is quite unusual to reach a 100% rate with PPV in clinical diagnostic tests, contrary to NPV, especially for the diagnosis of cirrhosis or severe fibrosis in liver diseases [25–27].

Considering that a positive likelihood ratio greater than 5 and a negative likelihood ratio less than 0.2 favor strong diagnostic evidence, and a diagnostic odds ratio greater than 30 gives reasonable test performance [15,16], there should be no suitable blood test according to this study (Table 2). This interpretation, however, should take into account the well recognized misclassification rate of liver biopsy [20], which would lead to underestimation of the accuracy of the blood tests. Meanwhile, the specific test with the accurate diagnostic cutoff offers the best compromise between diagnostic indices, especially in terms of equilibrium between PPV and NPV. In addition, the specific test better depicts the degree of fibrosis than does the original test (Fig. 5). For other uses, the optimization of specific tests by sequential algorithms [24] would be useful. Finally, both original and specific blood tests, based on the same markers, can be simultaneously used in clinical practice, the original test for NPV and the specific test for PPV and overall accuracy, which is facilitated by available software [21].

In conclusion, this study shows that by using the most accurate original test for the diagnosis of significant fibrosis with an optimized diagnostic cutoff, 100% of patients with cirrhosis can be correctly classified. Using the same original test, the diagnosis of cirrhosis can be excluded, that is a 100% NPV, in almost half of patients. Moreover, a higher overall accuracy (93%) is attained with a specific test for cirrhosis, including the same markers; it also provides a 100% PPV, which is its most robust clinical application. Finally, this study shows that the diagnostic use of blood tests is optimized by using appropriate diagnostic cutoffs and additional specific score for the diagnosis of cirrhosis, which implies the use of accessible software for facilitated clinical application.

Acknowledgements

The authors thank other investigators from Angers: Frédéric Oberti, Sophie Michalak, Isabelle Hubert-Fouchard, Anselme Konaté, Catherine Ternisien, Alain Chevailler, Françoise Lunel, Dermot O'Toole, and Gilles Hunault; PACA: Denis Ouzan, Albert Tran, Danielle Botta, Christophe Renou, Christophe Sattonnet, Marie-Christine Saint-Paul, Thierry Benderitter, Stéphane Garcia, Henri Pierre Bonneau, and Guillaume Penaranda; Tours: Marie-Claude Bréchet; Grenoble: Marie-Noëlle Hilleret, Jean-Pierre Zarski, Patrice Faure, Jean-Charles

Renversez, Françoise Morel, and Candice Trocme; Bordeaux: Laurent Castéra, Patrice Couzigou, Pierre-Henri Bernard, Wassil Merrouche, Paulette Bioulac-Sage, and Brigitte Le Bail; and Clichy: Claude Degott and Valérie Paradis. PHRC grants in 1994 and 2002 from the French Department of Health. The sponsor had no role in study design, in data collection, analysis, and interpretation, in the writing of the report, or in the decision to submit the paper for publication. Two coauthors (Paul Calès and Frédéric Oberti) have mentioned potential conflict of interest with stock ownership in a society (BioLiveScale) created under the auspices of University of Angers. Steering committee: Jérôme Boursier, Paul Calès, Philippe Halfon, Yannick Bacq, Vincent Leroy, Victor de Ledinghen. Review committee: Jérôme Boursier, Frédéric Oberti; for statistics: Thomas Moyon, Gilles Hunault, Guillaume Penaranda; and Kevin L. Erwin (for English proofreading). Data analysis: Thomas Moyon, Gilles Hunault, Paul Calès.

References

- Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, Aube C, *et al.* Noninvasive diagnosis of hepatic fibrosis or cirrhosis. *Gastroenterology* 1997; **113**:1609–1616.
- Sebastiani G, Alberti A. Noninvasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. *World J Gastroenterol* 2006; **12**:3682–3694.
- Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, *et al.* Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005; **128**:343–350.
- Cales P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A, *et al.* A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005; **42**:1373–1381.
- Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, Lok AS. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003; **38**:518–526.
- Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poinard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001; **357**:1069–1075.
- Adams LA, Bulsara M, Rossi E, DeBoer B, Speers D, George J, *et al.* Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* 2005; **51**:1867–1873.
- Halfon P, Bacq Y, De Muret A, Penaranda G, Bourliere M, Ouzan D, *et al.* Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C. *J Hepatol* 2007; **46**:395–402.
- Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P, *et al.* Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. *J Hepatol* 2007; **46**:775–782.
- Halfon P, Imbert-Bismut F, Messous D, Antoniotti G, Benchetrit D, Cart-Lamy P, *et al.* A prospective assessment of the inter-laboratory variability of biochemical markers of fibrosis (FibroTest) and activity (ActiTest) in patients with chronic liver disease. *Comp Hepatol* 2002; **1**:3.
- Cales P, Veillon P, Konate A, Mathieu E, Ternisien C, Chevailler A, *et al.* Reproducibility of blood tests of liver fibrosis in clinical practice. *Clin Biochem* 2008; **41**:10–18.
- Calès P, De Ledinghen V, Halfon P, Bacq Y, Leroy V, Boursier J, *et al.* Evaluating accuracy and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C. *Liver Int* 2008; (in press).
- Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology* 1994; **20**:15–20.
- Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *Br Med J* 1997; **315**:540–543.
- Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; **56**:1129–1135.

- 16 Parkes J, Guha IN, Roderick P, Rosenberg W. Performance of serum marker panels for liver fibrosis in chronic hepatitis C. *J Hepatol* 2006; **44**:462–474.
- 17 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**:837–845.
- 18 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003; **49**:7–18.
- 19 Calès P, Halfon P, Bacq Y, Leroy V, Rousselet MC, Bourliere M, et al. Meta-analysis of blood scores for liver fibrosis in chronic hepatitis C. *Hepatology* 2006; **44**:279A.
- 20 Rousselet MC, Michalak S, Dupre F, Croue A, Bedossa P, Saint-Andre JP, Cales P. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005; **41**:257–264.
- 21 Leroy V, Halfon P, Bacq Y, Boursier J, Rousselet M, Bourliere M, et al. Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008; (in press).
- 22 Cacoub P, Carrat F, Bedossa P, Lambert J, Penaranda G, Perronne C, et al. Comparison of non-invasive liver fibrosis biomarkers in HIV/HCV co-infected patients: the fibrovic study-ANRS HC02. *J Hepatol* 2008; **48**:765–773.
- 23 Leroy V, Sturm N, Hilleret M, Renversez P, Trocme C, Faure P, Zarski J. Diagnostic accuracy of blood tests of liver fibrosis in chronic hepatitis B: comparison with hepatitis C. *Hepatology* 2007; **46**:900–901.
- 24 Sebastiani G, Vario A, Guido M, Noventa F, Plebani M, Pistis R, et al. Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol* 2006; **44**:686–693.
- 25 Shaheen AA, Myers RP. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C-related fibrosis: a systematic review. *Hepatology* 2007; **46**:912–921.
- 26 Kim BK, Kim SA, Park YN, Cheong JY, Kim HS, Park JY, et al. Noninvasive models to predict liver cirrhosis in patients with chronic hepatitis B. *Liver Int* 2007; **27**:969–976.
- 27 Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. *Hepatology* 2007; **46**:32–36.
- 28 Colletta C, Smirne C, Fabris C, Toniutto P, Rapetti R, Minisini R, Pirisi M. Value of two noninvasive methods to detect progression of fibrosis among HCV carriers with normal aminotransferases. *Hepatology* 2005; **42**:838–845.
- 29 Poynard T, Munteanu M, Ngo Y, Torres M, Benhamou Y, Thabut D, Ratziu V. Diagnostic value of FibroTest with normal serum aminotransferases. *Hepatology* 2006; **43**:374–375.

Appendix

Methods: definitions

Diagnostic target: The main objective of a blood test, usually defined by two ranges of all stages of fibrosis, with a diagnostic cutoff between them, for example, significant fibrosis (yes/no).

Original and specific test: Multivariate analysis (usually forward stepwise binary logistic regression) provides the regression score probability (from 0 to 1), which corresponds to the blood test value. The original test is that obtained with the original diagnostic target, usually significant fibrosis, in the pivotal population, characterized by a particular β coefficient of regression for each marker. A specific test is one obtained with the same markers as in the original test, but under different conditions; for example, specific tests were designed here for the alternative diagnostic targets, providing different β coefficients.

Overall accuracy: Sum of true positives and negatives as a proportion of the total. Diagnostic accuracy is a synonym.

Classification/misclassification rate: The misclassification rate of a blood test for significant fibrosis, also called performance profile [8], was calculated using Metavir staging as a reference. Thus, a patient with Metavir F0 or F1 classified in the significant fibrosis group by blood tests was considered as misclassified and vice versa. The misclassification rate was calculated in each or in possible combined Metavir fibrosis stage(s) determined by histological staging. Finally, misclassification rates were compared between pairs of blood tests using the McNemar test. Here, to simplify the presentation, results are expressed as the converse proportion of correctly classified patients. The correct classification rate in all stages combined corresponds to the overall accuracy.

Diagnostic cutoff: The diagnostic cutoff of a blood test value distinguishes between patients with and without the diagnostic target. Here, it was fixed in two ways:

Table A Correct classification rate (%) for significant fibrosis by the different original blood tests as a function of fibrosis stage according to different diagnostic cutoffs

	Diagnostic cutoff ^a	Fibrosis stage		
		≥ 3	3	4
Correct classification				
FibroMeter	0.50	91.0	85.1	98.3
	0.41	94.5	90.1	100
	<i>P</i> ^b	0.004	0.016	NA
Fibrotest	0.50	79.1	71.1	90.2
	0.43	85.3	78.2	95.1
	<i>P</i> ^b	<0.003	0.002	0.063
Hepascore	0.50	81.3	71.7	93.2
	0.47	83.2	73.8	94.9
	<i>P</i> ^b	0.063	0.250	0.500
APRI	1	54.2	44.9	65.8
	0.57	78.8	71.4	88.0
	<i>P</i> ^b	<0.003	<0.003	<0.003
Comparison with fixed cutoffs ^c				
All tests		<0.003	<0.003	<0.003
FibroMeter vs. Fibrotest		<0.003	<0.003	0.011
FibroMeter vs. Hepascore		<0.003	<0.003	0.034
FibroMeter vs. APRI		<0.003	<0.003	<0.003
Fibrotest vs. Hepascore		0.739	1	0.527
Fibrotest vs. APRI		<0.003	<0.003	<0.003
Hepascore vs. APRI		<0.003	<0.003	<0.003
Comparison with a posteriori cutoffs ^c				
All tests		<0.003	<0.003	0.004
FibroMeter vs. Fibrotest		<0.003	0.001	NA
FibroMeter vs. Hepascore		<0.003	<0.003	NA
FibroMeter vs. APRI		<0.003	<0.003	NA
Fibrotest vs. Hepascore		0.131	0.127	0.706
Fibrotest vs. APRI		0.018	0.086	0.083
Hepascore vs. APRI		0.096	0.446	0.059

APRI, aspartate aminotransferase to platelet ratio index; NA, not available.

^aDiagnostic cutoff: the first one is *a priori* (fixed at 0.5 except for APRI: 1) and the second was determined *a posteriori* according to the maximum Youden index.

^bComparison of correctly classified patients between cutoffs by McNemar test.

^cComparison of correctly classified patients between blood tests by Cochran test (all) or McNemar test (pair).

Significant differences are indicated in bold.

Table B Correct classification rate (%) as a function of diagnostic target (cirrhosis or severe fibrosis) and test (original FibroMeter or derived specific test) according to fibrosis stage and diagnostic cutoff

Cirrhosis											
Score	Original (constructed for significant fibrosis)				Specific (constructed for cirrhosis)						
	0-3			4	0-3		<i>P</i> ^a	4	<i>P</i> ^a	0-4	
Fibrosis stage											
Cutoff	Value			Value	Value				Value		
Youden	0.628	71.0	95.7	73.8	0.089	81.1	<0.003	87.0	0.013	81.8	<0.003
Accuracy	0.979	98.0	35.7	90.9	0.442	97.9	1	54.8	<0.003	93.0	0.005
<i>P</i> ^c	-	-	-	<0.003	-	-	-	-	-	<0.003	-

Severe fibrosis											
Score	Original (constructed for significant fibrosis)				Specific (constructed for severe fibrosis)						
	0-2		3+4	0-4	0-2		<i>P</i> ^b	3+4	<i>P</i> ^b	0-4	
Fibrosis stage											
Cutoff	Value		Value	Value					Value		
Youden	0.628	79.4	83.6	80.5	0.232	80.8	0.135	81.3	0.307	80.9	0.630
Accuracy	0.830	91.1	59.8	83.2	0.557	95.6	<0.003	51.2	<0.003	84.4	0.213
<i>P</i> ^c	-	-	-	0.027	-	-	-	-	-	0.013	-

^aComparison of same fibrosis stage(s) between different scores by McNemar test: cirrhosis vs. significant fibrosis.
^bComparison of same fibrosis stage(s) between different scores by McNemar test: severe fibrosis vs. significant fibrosis.
^cComparison of all fibrosis stage(s) between different diagnostic cutoffs by McNemar test: maximum Youden index vs. maximum overall accuracy.

Table C Rates of patients (%) with exclusion (95 and 100% negative predictive values) of severe fibrosis or cirrhosis according to the original blood tests

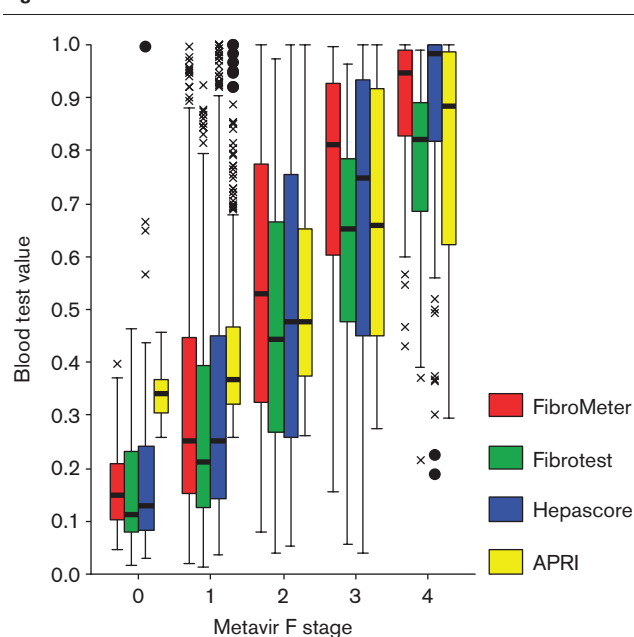
Negative predictive value (%)	Severe fibrosis		Cirrhosis	
	95	100	95	100
Patient rate				
FibroMeter	57.3	15.4	85.7	47.5
Fibrotest	41.3	3.6	87.5	31.9
Hepascore	0.3	0.3	85.6	24.7
APRI	1.4	1.4	79.7	5.4
Comparison ^a				
FibroMeter vs. Fibrotest	<0.003	<0.003	0.278	<0.003
FibroMeter vs. Hepascore	<0.003	<0.003	1	<0.003
FibroMeter vs. APRI	<0.003	<0.003	<0.003	<0.003
Fibrotest vs. Hepascore	<0.003	<0.003	0.141	<0.003
Fibrotest vs. APRI	<0.003	0.002	<0.003	<0.003
Hepascore vs. APRI	0.005	0.005	<0.003	<0.003

APRI, aspartate aminotransferase to platelet ratio index.
^aComparison between blood tests by McNemar test.
 Significant differences are indicated in bold.

a priori to 0.5 according to statistical convention (with the exception of APRI), and *a posteriori* according to the highest Youden index (sensitivity + specificity - 1) or the maximum overall accuracy to optimize test performance.

Reliable diagnosis: Corresponds to the intervals of blood test values where the accuracy is considered sufficiently reliable for clinical practice. Considering the diagnostic target of blood test in clinical practice, this means that liver biopsy can be avoided. Thus, with the traditional

Fig. A



Box plots (median, quartiles, extremes) of blood tests for the diagnosis of significant fibrosis against Metavir fibrosis stages in the whole population. APRI was standardized by binary logistic regression. x and •, distant and extreme values.

definition based on thresholds of greater than or equal to 90% predictive values, a reliable diagnosis for a patient means 90% chance to be greater than or equal to F2 in the highest interval and 90% chance to be F0/F1 in the lower interval of blood test values.

Results

Discussion: choice of diagnostic cutoff

Several cutoffs can be distinguished within a diagnostic blood test. Bear in mind that most blood tests are derived from binary logistic regression, which provides the probability, from 0 to 1, of the diagnostic between target being present. The diagnostic cutoff distinguishes between patients with or without the diagnostic target. As a convention, it is fixed *a priori* (before running the regression) to 0.5 but *a posteriori* it seems that this fixed cutoff does not always provide the highest performance, particularly when it is applied in populations larger than the original one. Although diagnostic cutoffs are rarely reported, they are important and can be a source of discrepancy and/or controversy [28,29]. The best diagnostic cutoff must thus be specified.

Whether using an original (Table 4) or specific (Table B) test, again the best performing *a posteriori* cutoff (accuracy) increased the overall accuracy compared with the reference cutoff (Youden), while decreasing correct classification rate within the diagnostic target. Finally, for the diagnosis of cirrhosis, when the correct classification rate of specific versus original test in the cirrhosis

stage is compared, the Youden cutoff resulted in a significant decrease, whereas the accuracy cutoff resulted in a significant increase (Table B). Thus, in the diagnosis of cirrhosis, the best option is to use the specific test with an accuracy cutoff. Nevertheless, there is no ideal cutoff as the Youden cutoff offers a higher correct classification rate than accuracy cutoff in cirrhosis stage at the cost of a marked significant decrease in overall accuracy.

Thus, the choice of the diagnostic cutoff depends on the overall accuracy as well as the correct classification rate for important fibrosis stage such as cirrhosis. In addition, this cutoff value has to be fixed for further clinical use as noninvasive markers (without liver biopsy as reference) or further comparisons of blood tests between different populations (with liver biopsy as reference). Here, the most accurate diagnostic cutoffs were also recalculated for the diagnosis of significant fibrosis as this was a multicentric population larger than the original population, thus providing a more robust cutoff. For example, the best cutoff for FibroMeter was the Youden/accuracy cutoff providing a maximization of overall accuracy and correct classification rate in cirrhosis stage.

Annexe 8 : Etude VINDIAG 6

Boursier Jérôme, Vergniol Julien, Sawadogo Apollinaire, Dakka Taoufiq, Michalak Sophie, Gallois Yves, Le Tallec Véronique, Oberti Frédéric, Fouchard-Hubert Isabelle, Dib Nina, Rousselet Marie-Christine, Konate Anselme, Amrani Naïma, de Ledinghen Victor, Calès Paul. The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. Liver Int 2009;29:1507-1515

CLINICAL STUDIES

The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis

Jérôme Boursier^{1,2}, Julien Vergniol^{3,4}, Apollinaire Sawadogo^{1,2}, Taoufiq Dakka⁵, Sophie Michalak^{1,6}, Yves Gallois^{1,7}, Véronique Le Tallec², Frédéric Oberti^{1,2}, Isabelle Fouchard-Hubert^{1,2}, Nina Dib^{1,2}, Marie Christine Rousselet^{1,6}, Anselme Konaté^{1,2}, Naïma Amrani⁵, Victor de Ledinghen^{3,4} and Paul Calès^{1,2}

1 HIFIH Laboratory, IFR 132, University, PRES UNAM, Angers, France

2 Hepato-Gastroenterology Department, CHU, Angers, France

3 Hepato-Gastroenterology Department, Haut Leveque Hospital, Pessac, CHU, Bordeaux, France

4 INSERM U889, Victor Segalen University, Bordeaux, France

5 Hepato-Gastroenterology Exploration Unit, Ibn Sina Hospital, CHU, Rabat, Morocco

6 Pathology Department, CHU, Angers, France

7 Biochemistry Department, CHU, Angers, France

Keywords

blood fibrosis test – cirrhosis – FibroMeter – liver fibrosis – liver stiffness – non-invasive diagnosis – sequential diagnostic algorithm – ultrasonographic elastometry

Correspondence

Paul Calès, Service d'Hépatogastroentérologie, CHU, 49933 Angers Cedex 09, France
Tel: +33 2 41 35 34 10
Fax: +33 2 41 35 41 19
e-mail: paul.cales@univ-angers.fr

Received 3 June 2009

Accepted 10 July 2009

DOI:10.1111/j.1478-3223.2009.02101.x

Abstract

Background and aims: Blood tests and liver stiffness evaluation (LSE) by ultrasonographic elastometry are accurate tools for diagnosing liver fibrosis. We evaluated whether their synchronous combination in new scores could improve the diagnostic accuracy and reduce liver biopsy requirement in algorithm. **Methods:** Three hundred and ninety patients with chronic liver disease of miscellaneous causes were included. Five blood fibrosis tests were evaluated: APRI, FIB-4, Hepascore, Fibrotest and FibroMeter. The reference was fibrosis Metavir staging. **Results:** *Diagnosis of significant fibrosis (Metavir $F \geq 2$).* The most accurate synchronous combination was FibroMeter+LSE, which provided a significantly higher area under the receiver operating characteristic curve (0.892) than LSE alone (0.867, $P=0.011$) or Fibrometer (0.834, $P < 10^{-3}$). An algorithm using the FibroMeter+LSE combination and then a liver biopsy in indeterminate cases had 91.9% diagnostic accuracy and required significantly fewer biopsies (20.2%) than previously published Bordeaux algorithm (28.6%, $P=0.02$) or sequential algorithm for fibrosis evaluation (SAFE) (55.7%, $P < 10^{-3}$). The Angers algorithm performance was not significantly different between viral hepatitis and other causes. *Diagnosis of cirrhosis.* The most accurate synchronous combination was LSE+FibroMeter, which provided $\geq 90\%$ predictive values for cirrhosis in 90.6% of patients vs 87.4% for LSE ($P=0.02$) and 57.9% for FibroMeter ($P < 10^{-3}$). An algorithm including the LSE+FibroMeter combination, and then a liver biopsy in indeterminate cases, had a significantly higher diagnostic accuracy than the SAFE algorithm (91.0 vs 79.8%, $P < 10^{-3}$), and required significantly fewer biopsies than the Bordeaux algorithm (9.3 vs 25.3%, $P < 10^{-3}$). **Conclusion:** The synchronous combination of a blood test plus LSE improves the accuracy of the non-invasive diagnosis of liver fibrosis and, consequently, markedly decreases the biopsy requirement in the diagnostic algorithm, notably to $< 10\%$ in cirrhosis diagnosis.

The non-invasive diagnosis of liver fibrosis has gained considerable attention over the last 10 years. Initially, several studies demonstrated that some blood markers correlated well with the fibrosis stages measured on liver specimens (1–3). This was followed by the development of blood tests combining several fibrosis markers to improve the diagnostic accuracy. The first generation of simple blood tests combined common indirect fibrosis

markers into a simple ratio, like APRI (4) and FIB-4 (5). The second generation of calculated blood tests combined fibrosis markers by logistic regression, either indirect and direct markers like Fibrotest (6), FibroMeter (7) and Hepascore (8), or direct markers like the ELF score (9) and Fibrospect (10). The third step was liver stiffness evaluation (LSE) by ultrasonographic elastometry (11).

Finally, blood fibrosis tests were combined into *sequential algorithms* in order to increase the diagnostic accuracy and limit the rate of liver biopsy (12–16). These sequential algorithms are usually based on a stepwise diagnosis including intervals of reliable diagnosis by blood tests in a first step, followed by a biopsy for the indeterminate cases. However, clinical applicability is somewhat difficult and a biopsy is still required in 20–50% of patients.

The association of non-invasive fibrosis tests with a *synchronous combination* circumvents these limitations while improving diagnostic accuracy and resolving the discordances between tests (17). In this setting, Castera *et al.* (18) have suggested that an association of LSE and blood fibrosis tests could improve the accuracy in patients with chronic hepatitis C. However, the statistical differences between the area under the receiver operating characteristic curves (AUROCs) – the only diagnostic index used – of the fibrosis tests and their synchronous combination were not evaluated. In another recent study, Castera *et al.* (19) have suggested that an association of LSE and blood test does not improve the cirrhosis diagnosis but only a few non-recent blood tests were evaluated.

Our primary aim was to improve the diagnostic accuracy of liver fibrosis by developing synchronous combinations of blood fibrosis tests and LSE. The ensuing secondary aim was to reduce liver biopsy requirement in diagnostic algorithms in comparison with previously published algorithms.

Patients and methods

Patients

We prospectively enrolled 390 patients with chronic liver disease (CLD) hospitalized for a percutaneous liver biopsy in the University Hospitals of Angers, Bordeaux (France) and Rabat (Morocco): 194 patients from April 2004 to June 2007 at Angers (group A, derivation set), and 196 from September 2003 to April 2007 at Bordeaux and Rabat (group B, validation set). Patients with cirrhosis complications (ascites, variceal bleeding, systemic infection and hepatocellular carcinoma) were not included. Blood fibrosis tests and LSE were performed within 1 week before biopsy. All patients gave their informed consent. The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and was approved by the local Ethics committee.

Methods

Histological liver fibrosis assessment

A percutaneous liver biopsy was performed using Menghini's technique with a 1.4–1.6-mm-diameter needle. Liver fibrosis was evaluated according to Metavir staging and significant fibrosis was defined by Metavir stages $F \geq 2$. Although initially designed for chronic viral hepatitis, Metavir staging was also validated in alcoholic liver diseases (20) and used in non-alcoholic fatty liver disease (21). Liver fibrosis evaluation was performed by two

senior experts with a consensus reading in group A and by a senior expert pathologist in group B. Fibrosis staging was considered as reliable when the liver specimen length was ≥ 15 mm and/or the portal tract number ≥ 8 (22).

Fibrosis blood tests

The following blood tests were performed according to published formulas or patents: APRI, FIB-4, Fibrotest, Hepascore and FibroMeter, where cause-specific scores were used (7, 23, 24). All blood assays were performed in the same laboratories of each centre. The interlaboratory reproducibility was excellent for these tests (25).

Liver stiffness evaluation

Liver stiffness evaluation (FibroScan[®], EchoSens[™], Paris, France) was performed by an experienced observer (> 50 LSE before this study), blinded to patient data. LSE conditions were those recommended by the manufacturer (26). LSE was stopped when 10 valid measurements were recorded. The LSE result was expressed as the median and the interquartile range (kPa) of all valid measurements performed (26).

Statistical analysis

Quantitative variables were expressed as mean \pm standard deviation, unless otherwise specified. The diagnostic cut-off of fibrosis tests was calculated according to the highest Youden index (sensitivity + specificity – 1).

Accuracy of fibrosis tests

The performance of fibrosis tests was mainly expressed as AUROC which were compared with the Delong test (27). The reliable individual diagnosis was determined by the classical intervals of $\geq 90\%$ negative (NPV) and positive (PPV) predictive values (precise definitions in supporting information) as well as an intermediate diagnosis, with respect to the diagnostic target, in the indeterminate zone between these classical intervals (23).

Synchronous combination of fibrosis tests

To identify the best combination of fibrosis tests for the diagnosis of significant fibrosis, we performed a forward binary logistic regression in the derivation set (group A) using blood fibrosis tests and LSE results as independent variables. Using the regression score combining the independent tests, we constructed a new diagnostic test for significant fibrosis called the *significant fibrosis index* (SF-index). This test was validated in groups B and A+B. We constructed the *cirrhosis index* (C-index) for the diagnosis of cirrhosis in the same way.

New algorithms

We determined the thresholds of 90% NPV and 90% PPV of the SF-index (precise definitions in supporting

information). A patient was considered as having no or mild fibrosis (Metavir F0/F1) when the SF-index values corresponded to $NPV \geq 90\%$, and as having significant fibrosis (Metavir $F \geq 2$) when the SF-index values corresponded to $PPV \geq 90\%$ (precise definitions in supporting information). A liver biopsy was required when the SF-index value was in the indeterminate zone between the two previous thresholds. This Angers algorithm was compared with the Bordeaux algorithm (18) and with the sequential algorithm for fibrosis evaluation (SAFE) (16). We constructed another Angers algorithm designed for the diagnosis of cirrhosis by using the C-index in the same way.

Sample size

It was found to show a significant difference for the diagnosis of significant fibrosis between FibroMeter and the synchronous combination of fibrosis tests in the validation population. With α risk: 0.05, β risk: 0.20, significant fibrosis prevalence: 0.70, AUROC correlation: 0.70, and in a bilateral test, the sample size was 159 patients for the following hypothesis of AUROC: FibroMeter: 0.84, synchronous combination: 0.90. The softwares were SPSS, version 11.5.1 (SPSS Inc., Chicago, IL, USA) and SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

Results

Patients

The mean age of the patients was 52.4 years; 67.9% were males, and 74.4% had significant fibrosis (Table 1). 89.5% of the patients had a liver biopsy considered as reliable. LSE failure occurred in 12 patients (3.1%). All five blood tests and LSE were available in 332 patients.

Diagnosis of significant fibrosis

Synchronous combination of fibrosis tests

Combination of non-invasive tests. The LSE AUROC for the diagnosis of significant fibrosis was significantly higher than that of Hepascore, FIB-4 and APRI, and was not significantly different from FibroMeter and Fibrotest AUROCs (Table S1). Significant fibrosis was independently diagnosed by FibroMeter at the first step and by LSE at the second step in the derivation set as well as in the validation set (Table S2). The regression score of the SF-index was: $3.6224\text{FibroMeter} + 0.4408\text{LSE}$ result $- 3.9850$ (group A).

Performance of the significant fibrosis index. The SF-index AUROCs were not significantly different between groups A and B. SF-index AUROC was significantly higher than that of FibroMeter or LSE in the whole population (Table 2). The SF-index improved the classification rate among fibrosis stages (Fig. 1): its misclassification rate was significantly lower than LSE in Metavir $F \leq 1$ stages and significantly lower than FibroMeter in Metavir $F \geq 2$ stages (details not shown).

Discordances between liver stiffness evaluation and FibroMeter. Discordances were calculated in the whole population. FibroMeter and LSE were discordant in 103 (27.0%) patients, of whom 68 (66.0%) were correctly classified by the SF-index (Table S3). Finally, the SF-index correctly classified 316 (82.7%) patients and improved the correct classification (i.e. discordances between FibroMeter and LSE resolved by the SF-index) in 33 (8.6%) patients.

Table 1. Patient characteristics at inclusion

	Group			P*
	All (n = 390)	A (n = 194)	B (n = 196)	
Age (years)	52.4 ± 13.4	50.8 ± 12.7	53.9 ± 14.0	0.03
Male sex (%)	67.9	68.0	67.9	0.97
Cause of liver disease (%)				< 10 ⁻³
Virus	48.7	54.1	43.4	
Alcohol	27.2	26.3	28.1	
NAFLD	4.9	9.8	0.0	
Other	19.2	9.8	28.6	
Metavir fibrosis stage (%)				< 10 ⁻³
F0	7.2	4.1	10.2	
F1	18.5	19.6	17.3	
F2	23.1	26.3	19.9	
F3	20.3	27.3	13.3	
F4	31.0	22.7	39.3	< 10 ⁻³
Significant fibrosis (%)	74.4	76.3	72.4	0.39
Reliable biopsy (%)	89.5	95.3	82.6	< 10 ⁻³
IQR/LSE result < 0.21 (%)	59.4	58.5	60.3	0.73

*By t-test or χ^2 between the groups A and B.

IQR, interquartile range (kPa); LSE, liver stiffness evaluation; NAFLD, non-alcoholic fatty liver disease.

Table 2. The area under the receiver operating characteristic curve comparisons between synchronous combinations (FM+LSE indexes), FibroMeter and liver stiffness evaluation, as a function of diagnostic target and patient group

Patient group	Significant fibrosis			Cirrhosis		
	All	A	B	All	A	B
AUROC						
FM	0.834	0.839	0.843	0.835	0.822	0.839
LSE	0.867	0.889	0.850	0.923	0.931	0.922
FM+LSE index*	0.892	0.917	0.874	0.917	0.923	0.913
Comparison (<i>P</i>)†						
FM vs LSE	0.162	0.150	0.839	< 10 ⁻³	10 ⁻³	0.004
FM vs FM+LSE index*	< 10 ⁻³	< 10 ⁻³	0.210	< 10 ⁻³	< 10 ⁻³	< 10 ⁻³
LSE vs FM+LSE index*	0.011	0.081	0.042	0.458	0.463	0.445

*SF-index for significant fibrosis, C-index for cirrhosis.

†By Delong test.

A, Angers; AUROC, area under the receiver operating characteristic curve; B, Bordeaux and Rabat; FM, FibroMeter; LSE, liver stiffness evaluation.

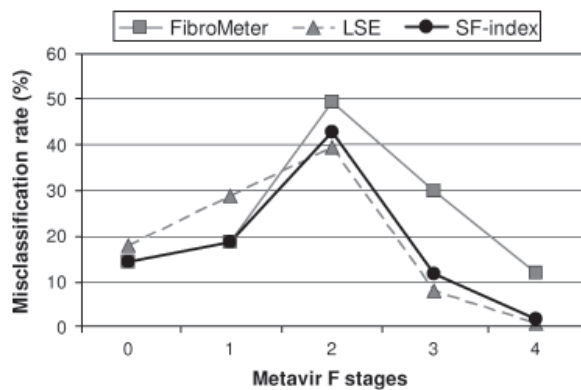


Fig. 1. Misclassification rate (%) for significant fibrosis of FibroMeter, liver stiffness evaluation (LSE) and their synchronous combination (SF-index) as a function of Metavir fibrosis stages. According to the highest Youden index, the diagnostic cut-offs used for significant fibrosis were: FibroMeter, 0.538; LSE, 6.9 kPa; SF-index, 0.753.

Methods reliably classifying 100% of patients

The Angers algorithm. This algorithm based on predictive values of the SF-index is described in Figure 2a. The SF-index included significantly more patients than FibroMeter or LSE in the intervals of $\geq 90\%$ predictive values, especially in the NPV interval (Table 3). By using the SF-index with $\geq 90\%$ predictive values and liver biopsy required in the remaining 18.3% of the patients, the Angers algorithm provided a correct diagnosis of significant fibrosis in 91.9% of the patients.

Algorithm comparison. Sequential algorithm for fibrosis evaluation had a significantly higher accuracy (97.0%) at the cost of a significantly higher biopsy rate (Table 4). The Angers algorithm required a significantly lower biopsy rate. Thus, the Angers algorithm had the best compromise with a high correct classification and a low biopsy requirement, reflected by a much lower biopsy/accuracy ratio. The accuracies of SAFE, Bordeaux and

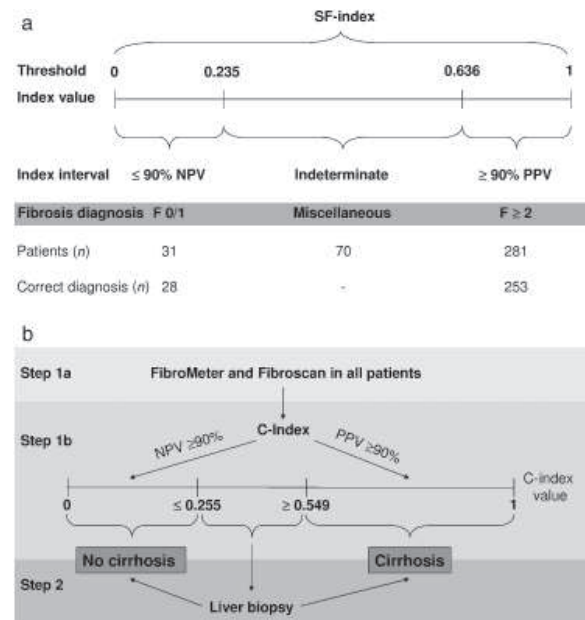


Fig. 2. The Angers algorithms designed for the diagnosis of significant fibrosis (a) or cirrhosis (b). A specific score combining FibroMeter and liver stiffness evaluation (LSE) is initially used (the SF-index for significant fibrosis or the C-index for cirrhosis), and a liver biopsy is subsequently required in the case of an indeterminate diagnosis. C-index, cirrhosis index; NPV, negative predictive value; PPV, positive predictive value; SF, significant fibrosis.

the Angers algorithms were not significantly different between patients with chronic viral hepatitis and those with other causes of CLD (detailed data not shown).

Diagnosis without liver biopsy (23). In the indeterminate interval between the thresholds of 90% predictive values of the SF-index (Fig. 2a), 92.9% of the patients had no severe fibrosis (F0: 20.0%, F1: 40.0%, and F2: 32.9%) (Fig. S1). Thus, it was possible to obtain three intervals of reliable diagnosis: F0/F1 in the $\leq 90\%$ NPV interval, F1 \pm 1 in the intermediate interval

Table 3. The Angers algorithms

Diagnostic target	Fibrosis test	Rate (%) of patients included in the intervals defined by 90% predictive values			Accuracy (%)	
		≥ 90% NPV	Indeterminate*	≥ 90% PPV	Fibrosis test(s)	Algorithm†
Significant fibrosis ($F \geq 2$)	FibroMeter	0.3	36.4	63.4	57.3	93.7
	LSE	0.5	28.8	70.7	64.1	92.9
	SF-index	8.1	18.3	73.6	73.6	91.9
Cirrhosis (F4)	FibroMeter	44.2	42.1	13.6	52.1	94.2
	LSE	68.3	12.6	19.1	78.8	91.4
	C-index	70.4	9.4	20.2	81.7	91.1

Rates of patients included and correctly classified by fibrosis tests in the intervals of $\geq 90\%$ predictive values for the diagnosis of significant fibrosis or cirrhosis in the whole population, as a function of fibrosis test(s).

*Proportion of patients for whom the diagnosis remains uncertain (NPV and PPV < 90%), thus requiring a liver biopsy. Comparison of patient rates by the McNemar test: significant fibrosis: LSE vs FibroMeter: $P = 0.006$, SF-index vs FibroMeter or LSE: $P < 10^{-3}$; cirrhosis: FibroMeter vs C-index or LSE: $P < 10^{-3}$, C-index vs LSE: $P = 0.02$.

†Based on the intervals of $\geq 90\%$ predictive values of the fibrosis test(s) and liver biopsy in the subsequent interval of indeterminate diagnosis. Thus, algorithm accuracy is calculated as the sum of patients correctly classified by the fibrosis test(s) in the whole population (4th result column) and biopsy requirement (2nd result column) where accuracy is 100% by definition. Comparison of rates by the McNemar test between FibroMeter and the C-index for cirrhosis: $P = 0.04$, others: P : NS.

LSE, liver stiffness evaluation; NPV, negative predictive value; PPV, positive predictive value.

Table 4. Diagnostic indices of sequential algorithm for fibrosis evaluation (16), Bordeaux (18) and the Angers algorithms for the diagnosis of significant fibrosis or cirrhosis

	Significant fibrosis ($F \geq 2$)			Cirrhosis (F4)		
	SAFE	Bordeaux	Angers	SAFE	Bordeaux	Angers
DA (%)*	97.0	90.4	91.9	79.8	94.3	91.0
Se (%)	100.0	90.6	98.8	44.0	89.0	74.7
Spe (%)	88.6	89.8	72.7	93.4	96.3	97.1
NPV (%)	100.0	77.5	95.5	81.5	95.9	91.1
PPV (%)	96.1	96.1	90.9	71.4	90.0	90.7
+LR	8.80	8.86	3.62	6.62	23.84	25.73
-LR	0.0	0.11	0.02	0.60	0.11	0.26
DOR	NA	84.3	214.3	11.0	208.8	98.8
Biopsy (%)†	55.7	28.6	20.2	6.6	25.3	9.3
Biopsy/DA	0.57	0.32	0.22	0.08	0.27	0.10

Three hundred and thirty-two patients having Fibrotest, FibroMeter, APRI and LSE were available.

The prevalence of significant fibrosis was 73.5% and that of cirrhosis was 27.4%.

*Comparison of algorithm diagnostic accuracies in the whole population, by the McNemar test: significant fibrosis: SAFE vs Bordeaux or Angers, $P < 10^{-3}$; Bordeaux vs Angers, $P = 0.50$; cirrhosis: SAFE vs Bordeaux or Angers, $P < 10^{-3}$; Bordeaux vs Angers: $P = 0.04$.

†Comparison of rates of biopsy required by the algorithms in the whole population, by the McNemar test: significant fibrosis: SAFE vs Bordeaux or Angers, $P < 10^{-3}$; Bordeaux vs Angers, $P = 0.02$; cirrhosis: Bordeaux vs SAFE or Angers, $P < 10^{-3}$; SAFE vs Angers, $P = 0.26$.

Biopsy, rate of liver biopsy required in the algorithm; biopsy/DA, ratio (rate of liver biopsy required)/(diagnostic accuracy); DA, diagnostic accuracy; DOR, diagnostic odds ratio; -LR: negative likelihood ratio; +LR: positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; SAFE, sequential algorithm for fibrosis evaluation; Se, sensitivity; Spe, specificity.

(correct classification: 92.9%) and $F \geq 2$ ($F3 \pm 1$) in the $\geq 90\%$ PPV interval. This classification correctly classified 90.6% of the patients without any biopsy performed.

Sensitivity analysis

By multivariate analysis, the accuracy of the SF-index was not associated with the biopsy length, the ratio interquartile range/LSE result or body mass index. In biopsy ≥ 25 mm, the SF-index still had a significantly higher AUROC (0.895) than FibroMeter (0.816, $P < 10^{-3}$) or LSE (0.860, $P = 0.03$). The SF-index was not dependent on the putative reliability of LSE: in interquartile range/LSE result ratio < 0.21 and ≥ 0.21 (28), AUROC were 0.881 and 0.906 ($P = 0.51$) respectively.

Diagnosis of cirrhosis

Synchronous combination of fibrosis tests

Combination of non-invasive tests. Liver stiffness evaluation had a significantly higher AUROC than the blood tests for the diagnosis of cirrhosis (Table S1). The most accurate combination of fibrosis tests was LSE + FibroMeter (Table S2). The regression score of the C-index was: $0.1162\text{LSE result} + 1.9714\text{FibroMeter} - 4.6616$ (group A). In each group tested, the C-index had a significantly higher AUROC than FibroMeter, but the difference with LSE was not significant (Table 2).

Discordances between liver stiffness evaluation and FibroMeter. FibroMeter and LSE were discordant in

79 patients, of whom 54 (68.4%) were correctly classified by the C-index (Table S3). Finally, the C-index correctly classified 329 (86.1%) patients and improved correct classification in 29 (7.6%) patients.

Methods reliably classifying 100% of the patients

The Angers algorithm. This algorithm, based on predictive values of the C-index, is described in Figure 2b. The C-index included significantly more patients (90.6%) than LSE or FibroMeter in the intervals of $\geq 90\%$ predictive values (Table 3). By using a liver biopsy for the remaining 9.4% of the patients, the Angers algorithm provided a correct diagnosis of cirrhosis in 91.1% of the patients.

Algorithm comparison. The Bordeaux algorithm had a significantly higher accuracy (Table 4). However, the Angers algorithm had a significantly lower biopsy rate. Thus, again, the Angers algorithm had the best compromise with a high correct classification and a low biopsy requirement.

Diagnosis without liver biopsy (23). In the indeterminate interval between the thresholds of 90% predictive values of the C-index (Fig. 2b), 91.6% had significant fibrosis ($F \geq 2$) (Fig. S1). Thus, it was possible to obtain three intervals of a reliable diagnosis: no cirrhosis ($F \leq 3$), significant fibrosis ($F3 \pm 1$) and cirrhosis (F4) that correctly classified 90.3% of patients without any biopsy performed.

Discussion

Study design

This study compared for the first time the diagnostic accuracy of LSE and five blood tests, and their synchronous combination, in a large population of patients with various causes of CLD. The liver fibrosis reference in our study was the Metavir staging on liver biopsy. This reference is hampered by inter-observer variability (29) and sampling variability (30). Nevertheless, histological reading was performed by expert pathologists who have been shown to provide the best inter-observer reproducibility (29). Moreover, SF-index accuracy was not independently influenced by liver specimen length. In addition, in liver specimens ≥ 25 mm displaying the highest reliability for liver fibrosis evaluation (30), the SF-index still had a significantly higher accuracy than FibroMeter and LSE.

A recent study has suggested that LSE is accurate when the ratio interquartile range/LSE result is < 0.21 (28). Although 59.4% of our patients had this ratio < 0.21 , all LSE were included in the statistical analysis according to the 'intention-to-diagnose' principle (31). In addition,

discarding the so-called inaccurate LSE would not have been within the scope of a combination study. Finally, SF-index accuracy was not independently influenced by this ratio.

Accuracy of blood tests and liver stiffness evaluation

The AUROC of LSE, FibroMeter and Fibrotest were not statistically different for the diagnosis of significant fibrosis. However, Figure 1 shows that the best-performing blood test had the lowest misclassification rate in F0–F1 stages, whereas LSE had the lowest misclassification rate in $F \geq 2$ stages, especially in F3 and F4. This suggests that these tests are complementary. We thus hypothesized that their synchronous combination could cumulate their advantages, erase their drawbacks and thus significantly increase their single diagnostic accuracy for liver fibrosis.

Accuracy gain by the synchronous combination

Our results show that a synchronous combination blood test+LSE increases the AUROC for the diagnosis of significant fibrosis (Table 2). However, AUROC is a global index of performance that is not meaningful in clinical practice. Indeed, physicians want to know whether a fibrosis test result is reliable in a patient. Therefore, we have developed the comparison of patient rates with high predictive values (23). The present study clearly shows that the synchronous combination blood test+LSE (SF-index) significantly improves the patient rate with $\geq 90\%$ predictive value for significant fibrosis (Table 3) in a higher proportion than could have suggested the AUROC comparison. The indeterminate zone of fibrosis tests was thus reduced: the SF-index provided a highly significantly lower rate of biopsy than the blood test or LSE. Finally, the synchronous combination provided two other advantages: the SF-index offered the lowest misclassification rate provided by each single test in each fibrosis stage (the blood test in F0/F1 and LSE in $F \geq 2$, see Fig. 1), and the SF-index resolved 66% of discordant cases between the blood test and LSE. Alternatively, by identifying an intermediate diagnosis ($F1 \pm 1$) in the indeterminate zone of the SF-index, this fully non-invasive diagnosis classified 100% of the population with 90.6% accuracy.

The C-index produced no significant increase in AUROC for the diagnosis of cirrhosis compared with LSE alone. However, the combination LSE+blood test also provided two advantages for the diagnosis of cirrhosis: the patient rate with $\geq 90\%$ predictive values was significantly higher than with LSE alone (Table 3), thus resulting in a very low rate of biopsy required (9%), and it resolved 68.4% of discordant cases between LSE and blood test. Alternatively, by identifying an intermediate diagnosis ($F \geq 2$) in the indeterminate zone of the C-index, this fully non-invasive diagnosis classified 100% of patients with 90.3% accuracy.

Algorithms for liver fibrosis diagnosis

Significant fibrosis

In our study, the accuracies of Bordeaux and SAFE algorithms were similar to those published previously (16, 18, 32), thus providing an independent external validation of these algorithms. Moreover, the accuracies of the SAFE, Bordeaux and Angers algorithms were not significantly different between patients with chronic viral hepatitis and those with other CLD. Because SAFE and Bordeaux algorithms were elaborated in chronic viral C hepatitis, this suggests that these algorithms could be extended to other causes of CLD.

Our results show that SAFE had a significantly higher diagnostic accuracy for significant fibrosis than the Bordeaux and Angers algorithms. However, a sequential combination of fibrosis tests has the major disadvantage of leaving a large final indeterminate zone responsible for a high rate of biopsy required, which is unsuitable in a screening setting (13–16). In our study, SAFE selected two and three times more patients for a biopsy than the Bordeaux and Angers algorithms respectively. In this respect, the Angers algorithm offered the best compromise between a high diagnostic accuracy (91.9%) and the lowest biopsy rate (20.2%). Finally, it should be noticed that a part of the apparent misclassifications of patients provided by an algorithm are in fact attributable to the misclassification of liver biopsy used as the reference. The fully non-invasive diagnosis based on three reliable intervals (23) might compensate for this previous uncertainty.

At first glance, it seems that SAFE would be the less expensive strategy: APRI is used as the first-line test with virtually no cost, and Fibrotest is only performed at the second step in a subgroup of patients. However, this apparently cost-effective strategy is dramatically counteracted by the high biopsy rate required by SAFE, liver biopsy being an expensive procedure in some countries (approximately 1000 euros for direct costs in France).

Cirrhosis

The Angers algorithm also provided the best compromise between a high diagnostic accuracy for cirrhosis (91.0%) and a very low rate of biopsy (9.3%). SAFE also had a low biopsy rate but an unacceptable significantly lower diagnostic accuracy (79.8%) than the other two algorithms, which was because of causes other than viral hepatitis (data not shown).

How is this integrated into clinical practice?

In a first step, the biological and LSE values are recorded and computerized (on a website for example) to obtain the combined index. A second step, based on a liver biopsy, may be required in a minority of patients. In cirrhosis, this simple two-step diagnosis had an accuracy of 91% and a biopsy was required in less than 10% of the patients (Fig. 2b). Finally, the practitioner has to evaluate

the clinical reliability of the results, especially those of fibrosis test results.

Limitations of this study

The present study has two limitations. First, the high prevalence of significant fibrosis (72%) as it influences diagnostic indices like accuracy (33) and predictive values. LSE was thus favoured in this study because it had a higher accuracy in $F \geq 2$ stages. However, this limitation can be easily circumvented with the misclassification profile, which shows the rate of misclassified patients in each fibrosis stage (Fig. 1).

The second limitation is the inclusion of patients with various CLD causes. This was mainly because of the sample size required. However, there was no significant difference in the diagnostic accuracy of the Angers algorithm between chronic viral hepatitis and other causes. Nevertheless, further studies should explore the possibility of increased accuracy via CLD cause-specific combination indexes.

In conclusion, the synchronous combination blood test+LSE significantly increases the diagnostic accuracy for significant fibrosis compared with a single test, resolves a majority of discordant results between non-invasive tests and improves the reliable individual diagnosis for significant fibrosis and cirrhosis while reducing liver biopsy requirement. Alternatively, one might use a fully non-invasive diagnosis, where the liver biopsy is replaced by an intermediate diagnosis with respect to the diagnostic target, at the cost of only a 1% decrease in accuracy.

Acknowledgements

French and Morocco research departments for grant (PAI Volubilis).

Potential conflict of interest: Paul Calès, Frederic Oberti and Isabelle Fouchard-Hubert have stock ownership in BioLiveScale Inc. that has a license for FibroMeters from Angers University.

References

1. Afdhal NH, Nunes D. Evaluation of liver fibrosis: a concise review. *Am J Gastroenterol* 2004; **99**: 1160–74.
2. Croquet V, Vuillemin E, Ternisien C, *et al.* Prothrombin index is an indirect marker of severe liver fibrosis. *Eur J Gastroenterol Hepatol* 2002; **14**: 1133–41.
3. Oberti F, Valsesia E, Pilette C, *et al.* Noninvasive diagnosis of hepatic fibrosis or cirrhosis. *Gastroenterology* 1997; **113**: 1609–16.
4. Wai CT, Greenson JK, Fontana RJ, *et al.* A simple non-invasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003; **38**: 518–26.
5. Sterling RK, Lissen E, Clumeck N, *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006; **43**: 1317–25.

6. Imbert-Bismut F, Ratziu V, Pieroni L, *et al.* Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001; **357**: 1069–75.
7. Cales P, Oberti F, Michalak S, *et al.* A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005; **42**: 1373–81.
8. Adams LA, Bulsara M, Rossi E, *et al.* Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* 2005; **51**: 1867–73.
9. Rosenberg WM, Voelker M, Thiel R, *et al.* Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004; **127**: 1704–13.
10. Patel K, Gordon SC, Jacobson I, *et al.* Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients. *J Hepatol* 2004; **41**: 935–42.
11. Sandrin L, Fourquet B, Hasquenoph JM, *et al.* Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol* 2003; **29**: 1705–13.
12. Sebastiani G, Vario A, Guido M, *et al.* Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol* 2006; **44**: 686–93.
13. Bourliere M, Penaranda G, Renou C, *et al.* Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies. *J Viral Hepat* 2006; **13**: 659–70.
14. Sebastiani G, Vario A, Guido M, *et al.* Sequential algorithms combining non-invasive markers and biopsy for the assessment of liver fibrosis in chronic hepatitis B. *World J Gastroenterol* 2007; **13**: 525–31.
15. Bourliere M, Penaranda G, Ouzan D, *et al.* Optimized stepwise combination algorithms of non-invasive liver fibrosis scores including Hepascore in hepatitis C virus patients. *Aliment Pharmacol Ther* 2008; **28**: 458–67.
16. Sebastiani G, Halfon P, Castera L, *et al.* SAFE biopsy: a validated method for large-scale staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2009; **49**: 1821–7.
17. Boursier J, Cales P. Combination of fibrosis tests: sequential or synchronous? *Hepatology* 2009; **50**: 656–7.
18. Castera L, Vergniol J, Foucher J, *et al.* Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005; **128**: 343–50.
19. Castera L, Le Bail B, Roudot-Thoraval F, *et al.* Early detection in routine clinical practice of cirrhosis and oesophageal varices in chronic hepatitis C: comparison of transient elastography (FibroScan) with standard laboratory tests and non-invasive scores. *J Hepatol* 2009; **50**: 59–68.
20. Michalak S, Rousselet M C, Bedossa P, *et al.* Respective roles of porto-septal fibrosis and centrilobular fibrosis in alcoholic liver disease. *J Pathol* 2003; **201**: 55–62.
21. Laine F, Bendavid C, Moirand R, *et al.* Prediction of liver fibrosis in patients with features of the metabolic syndrome regardless of alcohol consumption. *Hepatology* 2004; **39**: 1639–46.
22. Noursbaum JB, Cadranet JF, Bonnemaïson G, *et al.* Clinical practice guidelines on the use of liver biopsy. *Gastroenterol Clin Biol* 2002; **26**: 848–78.
23. Cales P, De Ledinghen V, Halfon P, *et al.* Evaluating accuracy and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C. *Liver Int* 2008; **28**: 1352–62.
24. Cales P, Laine F, Boursier J, *et al.* Comparison of blood tests for liver fibrosis specific or not to NAFLD. *J Hepatol* 2009; **50**: 165–73.
25. Cales P, Veillon P, Konate A, *et al.* Reproducibility of blood tests of liver fibrosis in clinical practice. *Clin Biochem* 2008; **41**: 10–8.
26. Boursier J, Konate A, Gorea G, *et al.* Reproducibility of liver stiffness measurement by ultrasonographic elastometry. *Clin Gastroenterol Hepatol* 2008; **6**: 1263–9.
27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45.
28. Lucidarme D, Foucher J, Le Bail B, *et al.* Factors of accuracy of transient elastography (Fibroscan) for the diagnosis of liver fibrosis in chronic hepatitis C. *Hepatology* 2009; **49**: 1083–9.
29. Rousselet MC, Michalak S, Dupre F, *et al.* Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005; **41**: 257–64.
30. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003; **38**: 1449–57.
31. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003; **49**: 7–18.
32. Castera L, Sebastiani G, Le Bail B, *et al.* Prospective comparison of two algorithms combining non invasive tests for staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2007; **46**: 320A.
33. Poynard T, Halfon P, Castera L, *et al.* Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; **53**: 1615–22.

Supporting information

Additional supporting information may be found in the online version of this article:

Figure S1. Sequential algorithm.

Figure S2. Intervals of $\geq 90\%$ negative (NPV) and positive (PPV) predictive values of SF-score, for the diagnosis of significant fibrosis (Metavir $F \geq 2$). NPV is $\geq 90\%$ in the 8.1% of patient with a score ≤ 0.235 ; PPV is $\geq 90\%$ in the 73.6% of patients with a score ≥ 0.636 .

Figure S3. Diagnosis based on high predictive values for significant fibrosis (3a) or cirrhosis (3b) and intermediate fibrosis stages in the remaining patients.

Table S1. AUROCs of blood tests and liver stiffness evaluation (LSE) as a function of diagnostic target.

Table S2. Combination by logistic regression. Fibrosis tests independently associated with significant fibrosis or cirrhosis, as a function of patient group.

Table S3. Discordances. Impact of FM+LSE index on discordances between FibroMeter (FM) and liver stiffness evaluation (LSE) for the diagnosis of significant fibrosis ($F \geq 2$) or cirrhosis (F4) in the whole population.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

SUPPLEMENTARY MATERIAL

GLOSSARY

Score: Regression formula provided by the binary logistic regression including the fibrosis markers. This score is included in a logit function to obtain the result of the fibrosis test.

Synchronous combination of fibrosis tests: Combination of fibrosis tests by binary logistic regression providing a new score and thus a new fibrosis test.

Sequential algorithm: Sequential ordered procedure including two fibrosis tests and liver biopsy if needed. A simple fibrosis test is firstly used to identify a group of patients with or without the diagnostic target. Then a second test is applied to reduce the indeterminate zone following the first test. Finally, biopsy is used for patients remaining in the indeterminate zone (Figure S1).

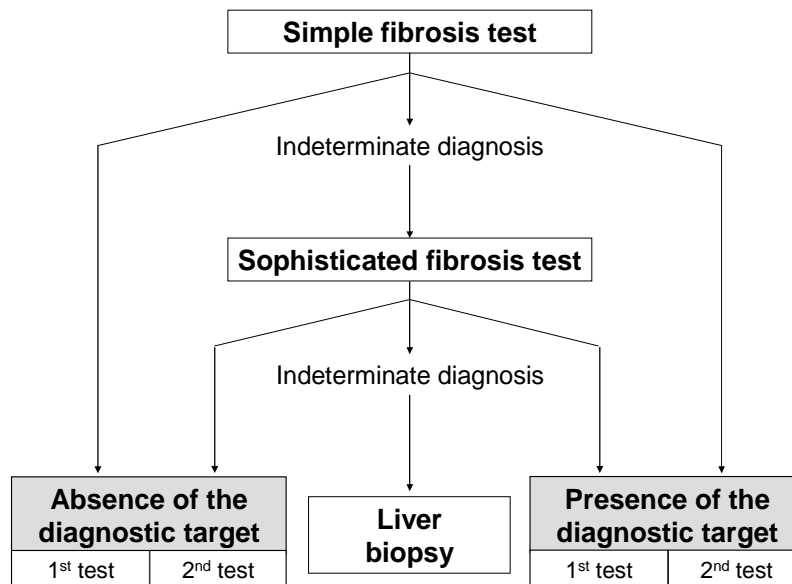


Figure S1: Sequential algorithm.

Intervals of $\geq 90\%$ predictive values for a diagnostic target (**Figure S2**): the diagnostic accuracy of a fibrosis test is maximum at its extremes values (48). Thus, two thresholds of fibrosis test values can be distinguished. Usually, those thresholds are fixed at 90% predictive values. These thresholds of 90% negative predictive value and 90% positive predictive value defined two intervals of blood tests values:

- A lower interval, defined by a blood test value \leq the 90% negative predictive value threshold, where patients have a $\geq 90\%$ chance of not having the diagnostic target.
- A higher interval, defined by a blood test value \geq the 90% positive predictive value threshold, where patients have a $\geq 90\%$ risk of having the diagnostic target.

If the fibrosis test value lies within the interval between the two 90% predictive value thresholds, the diagnosis remains indeterminate and liver biopsy might be secondly required.

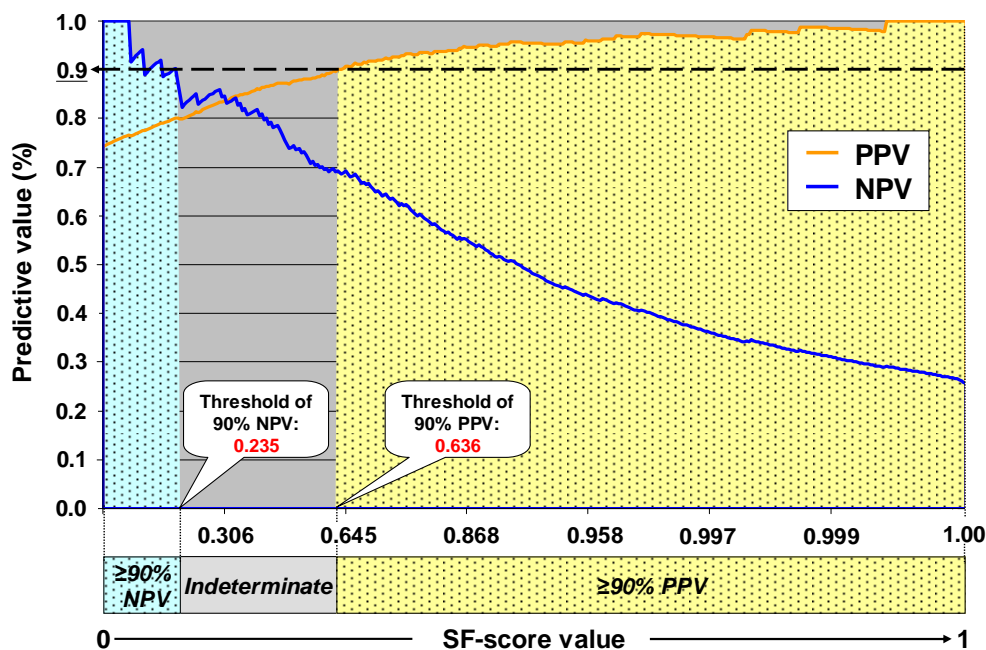


Figure S2: Intervals of $\geq 90\%$ negative (NPV) and positive (PPV) predictive values of SF-score, for the diagnosis of significant fibrosis (Metavir $F \geq 2$). NPV is $\geq 90\%$ in the 8.1% of patient with a score ≤ 0.235 ; PPV is $\geq 90\%$ in the 73.6% of patients with a score ≥ 0.636 .

Table S1: AUROCs of blood tests and liver stiffness evaluation (LSE) as a function of diagnostic target

	Significant fibrosis	Cirrhosis
AUROC:		
FibroMeter (FM)	0.836 ± 0.02	0.834 ± 0.02
Fibrotest (FT)	0.826 ± 0.03	0.813 ± 0.03
Hepascore (HS)	0.799 ± 0.03	0.806 ± 0.03
FIB-4	0.787 ± 0.03	0.793 ± 0.03
APRI	0.762 ± 0.03	0.691 ± 0.03
LSE	0.858 ± 0.02	0.915 ± 0.02
Comparison (p) ^a:		
FM vs FT	0.622	0.326
FM vs HS	0.074	0.101
FM vs FIB-4	0.030	0.078
FM vs APRI	0.004	<10 ⁻³
FM vs LSE	0.417	<10 ⁻³
FT vs HS	0.195	0.786
FT vs FIB-4	0.119	0.416
FT vs APRI	0.022	<10 ⁻³
FT vs LSE	0.257	<10 ⁻³
HS vs FIB-4	0.700	0.663
HS vs APRI	0.264	<10 ⁻³
HS vs LSE	0.046	<10 ⁻³
FIB-4 vs APRI	0.302	<10 ⁻³
FIB-4 vs LSE	0.016	<10 ⁻³
APRI vs LSE	0.003	<10 ⁻³

^a By Delong test.

Table S2: Combination by logistic regression. Fibrosis tests independently associated with significant fibrosis or cirrhosis, as a function of patient group.

Patient Group	Significant fibrosis			Cirrhosis		
	Independent variables ^a	p	Diagnostic accuracy (%) ^b	Independent variables ^a	p	Diagnostic accuracy (%) ^b
A	1. FibroMeter	<10 ⁻³	82.0	1. LSE	<10 ⁻³	89.7
	2. LSE	<10 ⁻³	87.6	2. FibroMeter	0.031	88.7
B	1. FibroMeter	<10 ⁻³	78.2	1. LSE	<10 ⁻³	82.4
	2. LSE	0.012	80.3	2. FibroMeter	0.017	83.0
All	1. FibroMeter	<10 ⁻³	80.6	1. LSE	<10 ⁻³	85.1
	2. LSE	<10 ⁻³	85.3	2. FibroMeter	10 ⁻³	86.1

LSE: liver stiffness evaluation

^a Variables independently associated with significant fibrosis or cirrhosis with increasing order of step (the first step is the most accurate variable); ^b Cumulative diagnostic accuracy for the second step.

Table S3: Discordances. Impact of FM+LSE index on discordances between FibroMeter (FM) and liver stiffness evaluation (LSE) for the diagnosis of significant fibrosis (F \geq 2) or cirrhosis (F4) in the whole population.

Classification by fibrosis tests ^a		Impact of FM+LSE index on classification by FM and LSE	Patients (n) according to diagnostic target studied	
FM+LSE index ^b	FM and LSE ^c		F \geq 2	F4
Correct	Both incorrect	Favorable	0	0
	Discordant		68	54
	Both correct	Neutral	248	275
Both incorrect	31		28	
Incorrect	Discordant		Unfavorable	35
	Both correct	0		0
		Net improvement	33 ^d (8.6%)	29 ^e (7.6%)

^a Respective diagnostic cut-off values used for significant fibrosis or cirrhosis, according to the highest Youden index: FibroMeter: 0.538 and 0.873; LSE: 6.9 and 13.0 kiloPascal; FibroMeter+LSE index: 0.753 (SF-index) and 0.216 (C-index)

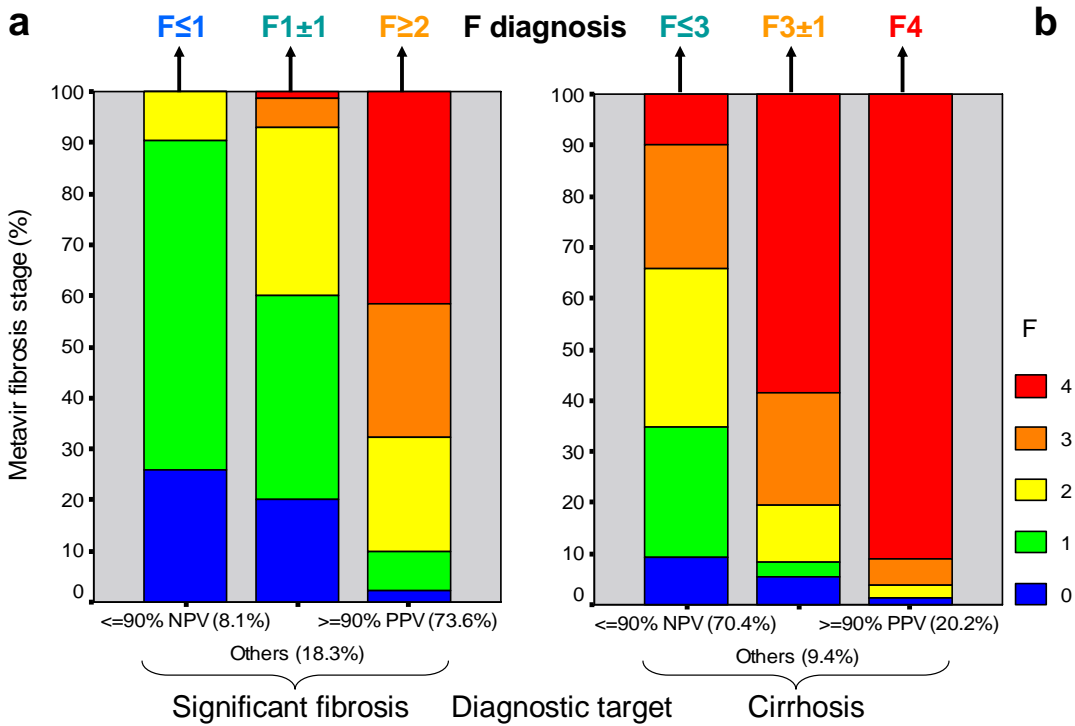
^b Classification by SF-index for significant fibrosis or C-index for cirrhosis expressed as correct or incorrect according to liver biopsy.

^c Classification of both tests based on liver biopsy. “Discordant” means than one test is correct and the other one is incorrect.

^d Favorable (68) - unfavorable (35) effect = improvement (33)

^e Favorable (54) - unfavorable (25) effect = improvement (29)

Figure S3: Diagnosis based on high predictive values for significant fibrosis (3a) or cirrhosis (3b) and intermediate fibrosis stages in the remaining patients. The X axis distinguishes the patients into 3 groups as a function of thresholds of 90% negative (NPV) and positive (PPV) predictive values of SF-index (3a) or C-index (3b); the rates of patients (%) included in each group are depicted in parentheses. The Y axis indicates the proportion of Metavir fibrosis (F) stages in the 3 patient groups; the stages including $\geq 90\%$ of patients provided the final F diagnosis indicated in the upper line.



Annexe 9 : Etude VINDIAG 7

Boursier Jérôme, de Ledinghen Victor, Zarski Jean-Pierre, Rousselet Marie-Christine, Sturm Nathalie, Foucher Juliette, Leroy Vincent, Fouchard-Hubert Isabelle, Bertrais Sandrine, Gallois Yves, Oberti Frédéric, Dib Nina, Cales Paul, and multicentric group ANRS HC EP 23 Fibrostar. A new combination of blood test and Fibroscan for accurate non-invasive diagnosis of liver fibrosis stages in chronic hepatitis C. Am J Gastroenterol, in press.

A New Combination of Blood Test and Fibroscan for Accurate Non-Invasive Diagnosis of Liver Fibrosis Stages in Chronic Hepatitis C

Jérôme Boursier, MD^{1,2,3}, Victor de Ledinghen, MD, PhD^{4,5}, Jean-Pierre Zarski, MD, PhD^{6,7}, Marie-Christine Rousselet, MD, PhD^{2,3,8}, Nathalie Sturm, MD⁹, Juliette Foucher, MD⁴, Vincent Leroy, MD, PhD⁶, Isabelle Fouchard-Hubert, MD^{1,2,3}, Sandrine Bertrais, PhD^{2,3}, Yves Gallois, PharmD, PhD^{2,3,10}, Frédéric Oberti, MD, PhD^{1,2,3}, Nina Dib, MD^{1,2,3}, Paul Calès, MD^{1,2,3} and the Multicentric Group ANRS HC EP 23 Fibrostar

OBJECTIVES: Precise evaluation of the level of liver fibrosis is recommended in patients with chronic hepatitis C (CHC). Blood fibrosis tests and Fibroscan are now widely used for the non-invasive diagnosis of liver fibrosis. Detailed fibrosis stage classifications have been developed to provide an estimation of the liver fibrosis stage from the results of these non-invasive tests. Our aim was to develop a new and more accurate fibrosis stage classification by using new scores combining non-invasive fibrosis tests.

METHODS: In all, 729 patients with CHC (exploratory set: 349; validation set: 380) had liver biopsy for Metavir fibrosis (F) staging, and 6 fibrosis tests: Fibroscan, Fibrotest, FibroMeter, Hepascore, FIB-4, APRI.

RESULTS: *Exploratory set:* Fibroscan and FibroMeter were the independent predictors of different diagnostic targets of liver fibrosis. New fibrosis indexes combining FibroMeter and Fibroscan were thus developed for the diagnosis of clinically significant fibrosis (CSF-index) or severe fibrosis (SF-index). The association of CSF- and SF-indexes provided a new fibrosis stage classification (CSF/SF classification): F0/1, F1/2, F2±1, F2/3, F3±1, F4. *Validation set:* CSF/SF classification had a high diagnostic accuracy (85.8% well-classified patients), significantly higher than the diagnostic accuracies of FibroMeter (69.7%, $P < 0.001$), Fibroscan (63.3%, $P < 0.001$), or Fibrotest (43.9%, $P < 0.001$) classifications.

CONCLUSIONS: The association of new fibrosis indexes combining FibroMeter and Fibroscan provides a new fibrosis stage classification. This classification is significantly more accurate than Fibrotest, FibroMeter, or Fibroscan classifications, and improves the accuracy of the non-invasive diagnosis of liver fibrosis stages to 86% without any liver biopsy.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/ajg>

Am J Gastroenterol advance online publication, 5 April 2011; doi:10.1038/ajg.2011.100

INTRODUCTION

The prognosis of patients with chronic hepatitis C (CHC) is related to the level of fibrosis in the liver (1). Usually, liver fibrosis is evaluated on a biopsy by using semi-quantitative scores that depict the progression of fibrosis according to histological fibrosis stages. Because it has a good prognosis value in patients with CHC (2), histological fibrosis staging is recommended for the clinical

management of patients (1). However, liver biopsy is an invasive procedure hampered by sample bias, and histological fibrosis staging has a poor interobserver reproducibility (3).

Several studies have shown that blood fibrosis tests and liver stiffness evaluation by transient elastography (e.g., Fibroscan) are accurate for the non-invasive diagnosis of liver fibrosis in patients with CHC (4–7). However, these studies have mainly evaluated the

¹Service d'Hépatogastroentérologie, CHU, Angers, France; ²Laboratoire HIFIH, UPRES 3859, IFR 132, Université, Angers, France; ³PRES UNAM, France; ⁴Service d'Hépatogastroentérologie, Hôpital Haut-Lévêque, Pessac, CHU, Bordeaux, France; ⁵INSERM U889, Université Victor Segalen, Bordeaux, France; ⁶Clinique d'Hépatogastroentérologie, Pôle digestif-DUNE, CHU, Grenoble, France; ⁷INSERM/UJF U823, IAPC, IAB, Grenoble, France; ⁸Département de Pathologie Cellulaire et Tissulaire, CHU, Angers, France; ⁹Département d'Anatomie et de Cytologie Pathologiques, Pôle de Biologie, CHU, Grenoble, France; ¹⁰Laboratoire de Biochimie et Biologie Moléculaire, CHU, Angers, France. **Correspondence:** Jérôme Boursier, MD, Service d'Hépatogastroentérologie, CHU, Angers Cedex 9, 49933 France. E-mail: jboursier@chu-angers.fr

Received 23 October 2010; accepted 16 February 2011

fibrosis tests only for a binary diagnosis of significant fibrosis or cirrhosis. Manufacturers or authors of Fibrotest, FibroMeter, and Fibroscan have developed *fibrosis stage classifications* derived from the results of the fibrosis test for a more precise estimation of the histological fibrosis stage (see **Supplementary Figure S1** online). In a recent study, we have shown that the diagnostic accuracy of those *fibrosis stage classifications* is perfectible (8).

We have recently proposed several methods to improve the non-invasive diagnosis of liver fibrosis. First, we showed that by adapting the marker coefficients of the original test specifically for the diagnosis of cirrhosis, the diagnostic accuracy was significantly increased as compared with the original test (9). Second, we proposed a new statistical method based on “*reliable diagnosis intervals*” (RDIs) that provides a more precise estimation of histological fibrosis stage(s), as compared with the original binary diagnosis. With this method, accuracy was $\geq 90\%$ (10) and $\geq 95\%$ (11) depending on the diagnostic target, thus eliminating the need for liver biopsy. However, these RDIs were less precise than Metavir F staging because they included fewer classes; in particular, no/mild fibrosis (F0/1), significant fibrosis (F ≥ 2), and cirrhosis (F4) were not grouped in the same classification. Finally, we showed that a synchronous combination of FibroMeter and Fibroscan results increased the diagnostic accuracy for significant fibrosis as well as for the RDI (12).

The main aim of this study was to improve the non-invasive diagnosis of histological fibrosis stages by creating a new *fibrosis stage classification* that uses new fibrosis indexes combining blood fibrosis tests and Fibroscan results. The secondary aim was to compare the diagnostic accuracy of this new *fibrosis stage classification* with those of Fibrotest, FibroMeter, and Fibroscan classifications.

METHODS

Patients

Exploratory set. Patients with CHC hospitalized for a percutaneous liver biopsy were prospectively enrolled from March 2004 to September 2008 in three tertiary centers in France (Angers, Bordeaux, and Grenoble). Patients with cirrhosis complications (ascites, variceal bleeding, systemic infection, hepatocellular carcinoma) were not included. Blood fibrosis tests and Fibroscan were performed in the week preceding biopsy. All patients gave their informed consent. The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and received approval from the local Ethics committee.

Validation set. The validation set corresponded to the multicenter population of the FIBROSTAR study promoted by the French National Agency for research in AIDS and hepatitis (13). This study prospectively included 512 patients with CHC. All patients had liver biopsy, blood fibrosis tests, and Fibroscan. Patients included in both the exploratory set and the FIBROSTAR study were excluded from the validation set.

Methods

Histological assessment. Liver fibrosis was evaluated according to Metavir staging. Significant fibrosis was defined as Metavir stages

F ≥ 2 , severe fibrosis (SF) as Metavir F ≥ 3 , and cirrhosis as F4. In the exploratory set, liver fibrosis was evaluated by two senior experts with a consensus reading at Angers, and by a senior expert at Bordeaux and Grenoble. In the FIBROSTAR study, liver fibrosis was centrally evaluated by two senior experts with a consensus reading in cases of discordance. Fibrosis staging was considered as reliable when liver specimen length was ≥ 15 mm and/or portal tract number ≥ 8 (14). Liver biopsy was used as the reference for the liver fibrosis evaluations by non-invasive tests.

Fibrosis blood tests. The following blood tests were calculated according to published or patented formulas: Fibrotest (15), FibroMeter (16), Hepascore (17), FIB-4 (18), and APRI (19). All blood assays were performed in the same laboratories of each center, or centralized in the FIBROSTAR study.

Liver stiffness evaluation. Fibroscan (EchoSens, Paris, France) examination was performed by an experienced observer (> 50 examinations before the study), blinded for patient data. Examination conditions were those recommended by the manufacturer (5). Fibroscan examination was stopped when 10 valid measurements were recorded. Results (kilopascals) were expressed as the median and the interquartile range (IQR) of all valid measurements. Fibroscan results were considered as reliable when the ratio IQR/result (IQR/median) was < 0.21 (20).

Statistical analysis

Quantitative variables were expressed as mean \pm s.d. The diagnostic cutoffs of fibrosis tests were calculated according to the highest Youden index (sensitivity + specificity – 1), unless otherwise specified.

Fibrosis stage classifications

We evaluated the accuracy of Fibrotest, FibroMeter, and Fibroscan *fibrosis stage classifications* (see **Supplementary Figure S1** online). Fibrotest, Fibroscan, and FibroMeter classifications were those previously published (16,21,22). Fibrotest classification includes eight classes (F0, F0/1, F1, F1/2, F2, F3, F3/4, F4), Fibroscan classification: six classes (F0/1, F1/2, F2, F3, F3/4, F4), and FibroMeter classification: six classes (F0/1, F1, F1/2, F2/3, F3/4, F4).

New fibrosis stage classification

The three-step procedure used to implement the new *fibrosis stage classification* (see **Supplementary Figure S2** online) and precise definitions (see Glossary in the **Supplementary Material** online) are detailed in the **Supplementary Material** online.

First step: new combined fibrosis indexes. To identify the best combination of *single fibrosis tests* for the diagnosis of significant fibrosis, we performed a stepwise binary logistic regression repeated on 1,000 bootstrap samples in the exploratory set. Independent variables tested were the five blood fibrosis tests and Fibroscan. The bootstrap method consists of a repeated sampling (with replacement) from the original entire data set, followed by a stepwise logistic regression procedure in each subsample (1,000 subsamples here). The most frequently ($> 50\%$) selected *single*

Table 1. Patient characteristics at inclusion

	All patients	Set		P
		Exploratory	Validation	
Patients (n)	729	349	380	—
Males (%)	61.3	60.2	62.4	0.531
Age (years)	51.7±11.2	52.1±11.2	51.3±11.2	0.347
Metavir (%)				<0.001
F0	4.0	1.4	6.3	
F1	37.7	30.7	44.2	
F2	25.8	35.5	16.8	
F3	17.6	20.6	14.7	
F4	15.0	11.7	17.9	0.020
Significant fibrosis (%)	58.3	67.9	49.5	<0.001
Reliable biopsy (%)	93.5	92.6	94.2	0.391
Fibroscan result (kPa)	10.0±7.9	9.9±8.1	10.1±7.7	0.755
IQR/median <0.21 (%)	66.9	66.2	67.6	0.700

IQR, interquartile range; kPa, kilopascal.

fibrosis tests among the 1,000 analyses were then included in a single binary logistic regression performed in the whole population of the exploratory set. Using the regression score of this multivariate analysis, we constructed a new *combined fibrosis index* for clinically significant fibrosis called “CSF-index”, ranging from 0 to 1. We also constructed *combined fibrosis indexes* for the diagnosis of SF-index and cirrhosis (C-index) using the same process.

Second step: RDIs. RDIs correspond to the intervals of fibrosis test values where the individual diagnostic accuracy is considered sufficiently reliable for clinical practice. This method has been previously described (10) and is detailed in the Glossary in the **Supplementary Material** online. Briefly, we first calculated the 90% negative and positive predictive value thresholds for significant fibrosis of the CSF-index. These two thresholds determined three intervals of CSF-index values: a low interval (from 0 to the 90% negative predictive value threshold) where the non-invasive diagnosis was consequently “F0/1”; a high interval (from the 90% positive predictive value threshold to 1) where the diagnosis was “F≥2”; and an intermediate interval between the two thresholds. The intermediate interval was then divided into two new intervals according to the diagnostic cutoff corresponding to the highest Youden index. In each of these two new intermediate intervals, the non-invasive diagnosis corresponded to the combined Metavir F stages having ≥90% prevalence (e.g. F1/2 for the interval between the 90% negative predictive value threshold and the highest Youden index cutoff). Finally, the four RDI that were obtained provided ≥90% diagnostic accuracy by definition.

We also calculated the RDIs of SF-index and C-index in the same way. Because SF-index was developed for the diagnosis of SF, its 90% negative and positive predictive value thresholds and its highest Youden index cutoff were determined for this diagnostic target.

For C-index, we calculated the thresholds for cirrhosis according to the 95% predictive values due to the clinical importance of cirrhosis diagnosis.

Third step: new fibrosis stage classifications. A new *fibrosis stage classification* was derived by associating RDIs for CSF- and SF-indexes. For example, if CSF-index provided a reliable diagnosis of “F≥2” and SF-index a reliable diagnosis of “F2±1”, the ensuing diagnosis of the new *fibrosis stage classification* was “F2/3”. Another *fibrosis stage classification* was derived by associating RDIs for CSF- and C-indexes.

Statistical softwares were SPSS, version 17.0 (SPSS, Chicago, IL) and SAS 9.1 (SAS Institute, Cary, NC).

RESULTS

Patients

The exploratory and validation sets included 349 and 380 patients, respectively. The characteristics of both sets are detailed in **Table 1**. Among the two sets, 93.5% of liver biopsies were considered as reliable.

Development of new fibrosis stage classifications

Main results are presented here. For more details, see **Supplementary Results** online.

First step: new combined fibrosis indexes. For each diagnostic target of liver fibrosis, Fibroscan and FibroMeter were *single fibrosis tests* the most frequently selected by the stepwise binary logistic regression repeated on the 1000 bootstrap samples. These two fibrosis tests were independent variables in logistic models that ran in the exploratory set and thus provided three new *combined*

Table 2. Accuracy (AUROC±s.d., bold values) of FibroMeter, Fibroscan, and their synchronous combination in new combined fibrosis indexes, as a function of diagnostic target and patient group

Diagnostic target	Fibrosis test	Set			All
		Exploratory	Validation	P	
Significant fibrosis (F≥2)	FibroMeter	0.806±0.026	0.839±0.022	0.333	0.813±0.017
	Fibroscan	0.785±0.026	0.828±0.022	0.207	0.791±0.017
	CSF-index	0.835±0.023	0.875±0.019	0.180	0.846±0.015
	Comparison (P)	—	—	—	—
	FibroMeter vs. Fibroscan	0.513	0.685	—	0.301
	FibroMeter vs. CSF-index	0.027	0.002	—	<0.001
	Fibroscan vs. CSF-index	0.024	0.009	—	<0.001
	SF (F≥3)	FibroMeter	0.776±0.025	0.880±0.020	0.001
Fibroscan	0.816±0.025	0.881±0.019	0.038	0.847±0.016	
SF-index	0.830±0.022	0.918±0.017	0.002	0.875±0.014	
Comparison (P)	—	—	—	—	
FibroMeter vs. Fibroscan	0.163	0.993	—	0.324	
FibroMeter vs. SF-index	<0.001	<0.001	—	<0.001	
Fibroscan vs. SF-index	0.458	0.014	—	0.019	
Cirrhosis (F4)	FibroMeter	0.814±0.031	0.897±0.021	0.027	0.861±0.018
	Fibroscan	0.878±0.032	0.927±0.017	0.176	0.905±0.017
	C-index	0.890±0.028	0.947±0.014	0.069	0.921±0.015
	Comparison (P)	—	—	—	—
	FibroMeter vs. Fibroscan	0.059	0.193	—	0.026
	FibroMeter vs. C-index	<0.001	<0.001	—	<0.001
	Fibroscan vs. C-index	0.511	0.120	—	0.133

AUROC, area under the receiver operating characteristic curve; C, cirrhosis; CSF, clinically significant fibrosis; SF, severe fibrosis.

fibrosis indexes for three diagnostic targets: CSF-index for significant fibrosis, SF-index for SF, and C-index for cirrhosis. CSF-index had a significantly higher area under the receiver operating characteristic curve than its composite tests, that is, FibroMeter or Fibroscan, in the exploratory set (Table 2). SF-index and C-index also had higher area under the receiver operating characteristic curves than FibroMeter or Fibroscan in the exploratory set, but the difference was significant only with FibroMeter.

Second step: RDIs. By using the thresholds of 90% predictive values for significant fibrosis and the diagnostic cutoff corresponding to the maximum Youden index, CSF-index provided four RDIs (F0/1, F1/2, F2±1, F≥2), which provided 90.3% diagnostic accuracy. By using the same method for SF, SF-index provided four RDIs (F1±1, F2±1, F3±1, F≥3) with 92.0% diagnostic accuracy. Finally, by using the thresholds of 95% predictive values for cirrhosis, C-index provided three RDIs (F3, F3±1, F4) with 95.1% diagnostic accuracy.

Third step: new fibrosis stage classifications. The first classification (CSF/SF classification) was derived from the association of CSF- and SF-index RDIs (Table 3). CSF/SF classification included

six classes (F0/1, F1/2, F2±1, F2/3, F3±1, F4) and provided 87.7% diagnostic accuracy in the exploratory set (Figure 1a).

The second classification (CSF/C classification) was derived from CSF- and C-index RDIs (Table 3). CSF/C classification also included six classes (F0/1, F1/2, F2±1, F2/3, F3±1, F4) and provided 86.5% diagnostic accuracy ($P=0.503$ vs. CSF/SF classification; Table 4).

Validation of the new fibrosis stage classifications

The diagnostic accuracies of CSF-index, SF-index, and C-index RDIs were not significantly different between the exploratory and the validation sets, with respectively: 90.3% vs. 86.7% ($P=0.142$), 92.0% vs. 91.5% ($P=0.827$), and 95.1% vs. 94.5% ($P=0.731$). Similarly, diagnostic accuracies of CSF/SF and CSF/C classifications were not significantly different between the two sets (Table 4).

In the validation set, CSF/SF classification provided a significantly higher diagnostic accuracy (85.8%) than CSF/C classification and those of single fibrosis tests ($P<0.008$; Table 4). Figure 1b shows the proportion of Metavir fibrosis stages as a function of CSF/SF classification. According to diagnostic accuracy in the validation set, classification ranking was: CSF/SF > CSF/C > FibroMeter > Fibroscan > Fibrotest (Table 4).

Figure 2 shows the diagnostic accuracy of each *fibrosis stage classification* as a function of Metavir fibrosis stage in the validation set. Among *single fibrosis tests*, FibroMeter provided the most homogeneous profile with no significant differences among histological fibrosis stages ($P=0.352$). The new CSF/SF and CSF/C classifications provided better profiles than those of *single fibrosis tests*. However, the rate of well-classified patients among cirrhotic patients was significantly higher with CSF/SF classification (94.5%) than with CSF/C classification (67.3%, $P<0.001$).

Table 3. Development in the exploratory set of new *fibrosis stage classifications* derived from the association of the *reliable diagnosis intervals (RDIs)* of *combined fibrosis indexes* (CSF- and SF-indexes, CSF- and C-indexes)

		RDI of CSF-index				
		F0/1	F1/2	F2±1	F≥2	
RDI of SF-index	F1±1	F0/1 (29/32)	F1/2 (84/96)	F1/2 (21/28)	— (0/0)	CSF/SF classification
	F2±1	— (0/0)	— (0/0)	F2±1 (32/32)	F2/3 (32/40)	
	F3±1	— (0/0)	— (0/0)	— (0/0)	F3±1 (94/103)	
	F≥3	— (0/0)	— (0/0)	— (0/0)	F4 (14/18)	
RDI of C-index	F≤3	F0/1 (29/32)	F1/2 (84/96)	F2±1 (57/60)	F2/3 (90/118)	CSF/C classification
	F3±1	— (0/0)	— (0/0)	— (0/0)	F3±1 (35/36)	
	F4	— (0/0)	— (0/0)	— (0/0)	F4 (7/7)	

C, cirrhosis; CSF, clinically significant fibrosis; SF, severe fibrosis. Gray cells indicate the RDIs of the combined fibrosis indexes. Colored cells indicate the F stages provided by the new fibrosis stage classifications. Figures into brackets are the rates of correctly classified patients in each class of the new fibrosis stage classifications according to liver biopsy results.

Influencing factors

In the whole study population, we performed a stepwise binary logistic regression including age, sex, biopsy length, Metavir F, and IQR/median as independent variables. The rate of well-classified patients by CSF/SF classification was independently associated with the ratio IQR/median (first step, $\exp(\beta)=0.322$), Metavir F (second step, $\exp(\beta)=1.370$), and age (third step, $\exp(\beta)=0.976$).

In the validation set, CSF/SF classification provided 89.5% diagnostic accuracy in patients with IQR/median <0.21 vs. 78.1% in patients with IQR/median ≥ 0.21 ($P=0.006$). In the subgroup of patients with IQR/median <0.21 , CSF/SF classification had the highest diagnostic accuracy ($P=0.006$ vs. other classifications, see **Supplementary Figure S4** online).

DISCUSSION

Because of its important role in guiding patient management, fibrosis staging in chronic viral hepatitis is clinically relevant (2). Blood fibrosis tests and liver stiffness evaluation are now used for the non-invasive assessment of liver fibrosis in CHC. In clinical practice, their results are interpreted by using *fibrosis stage classifications*, which provide an estimation of the histological fibrosis stage. We developed a new and accurate *fibrosis stage classification* for the non-invasive diagnosis of histological fibrosis stages by using an original method that includes several statistical techniques we have previously described: combinations of fibrosis tests (12), scores of blood tests adapted to specific diagnostic target (9,12), and RDI (10,11). This new *fibrosis stage classification* had a very significantly higher diagnostic accuracy than those previously published for *single fibrosis tests*.

Our study has several noteworthy aspects. First, both the exploratory and validation sets had a multicenter design (especially for the validation set). Second, histological examination was performed by expert pathologists and liver specimens were considered as reliable in 94% of the cases. Finally, fibrosis stage prevalence in our validation set was quite similar to the reference prevalence reported in a large cohort of around 33,000 patients with CHC (23). This

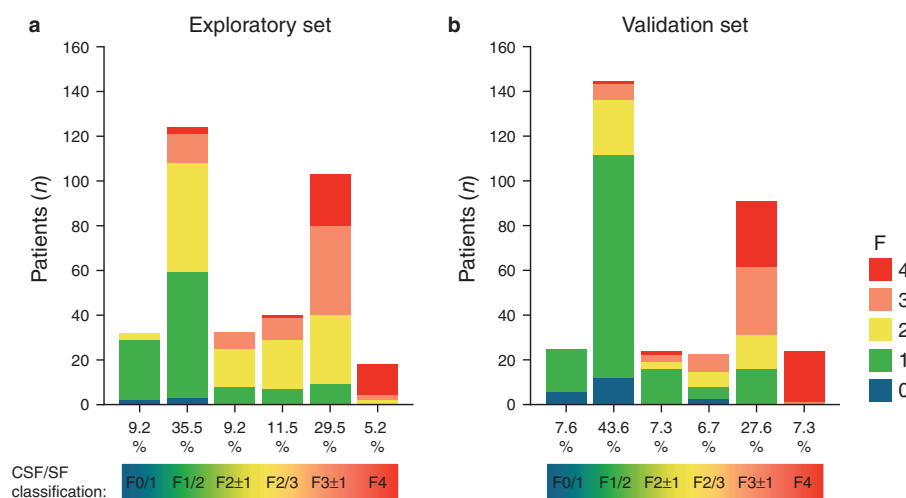


Figure 1. Proportion of Metavir fibrosis (F) stages as a function of clinically significant fibrosis/severe fibrosis (CSF/SF) classification (x axis with the rate of patients included in each class in italics), in the exploratory (a) and validation (b) sets. The bottom line indicates the *fibrosis stage classification*.

Table 4. Diagnostic accuracy (% of correctly classified patients) of fibrosis stage classifications as a function of patient group

Classification	Set			P ^a
	All	Exploratory	Validation	
CSF/SF	86.7	87.7	85.8	0.461
CSF/C	84.4	86.5	82.1	0.113
FibroMeter	68.7	67.6	69.7	0.550
Fibroscan	58.7	54.4	63.3	0.020
Fibrotest	38.8	33.5	43.9	0.005

C, cirrhosis; CSF, clinically significant fibrosis; SF, severe fibrosis.

^aComparison between sets.

Comparison between fibrosis stage classifications.

All patients: CSF/SF vs. CSF/C, $P=0.014$; others comparisons, $P<0.001$.

Exploratory set: CSF/SF vs. CSF/C, $P=0.503$; other comparisons, $P<0.001$.

Validation set: CSF/SF vs. CSF/C, $P=0.008$; FibroMeter vs. Fibroscan, $P=0.029$; other comparisons, $P<0.001$.

aspect is important as diagnostic accuracy relies on fibrosis stage prevalence (24).

Why a combination of fibrosis tests?

Best candidates for combination. We have previously shown that the best combination among several fibrosis tests evaluated was FibroMeter + Fibroscan (12). However, this study included various causes of chronic liver diseases. The present study confirms these findings in a homogeneous population of patients having CHC. Indeed, the bootstrap method identified FibroMeter and Fibroscan as the best independent predictors for each diagnostic target studied in the exploratory set (see **Supplementary Table S1** online).

The contribution of fibrosis test combination. Our results clearly show several advantages provided by combined fibrosis indexes. Combined fibrosis indexes increased the accuracy of fibrosis diagnosis: in the validation set and in the whole population, they provided significantly higher area under the receiver operating characteristic curves for significant fibrosis and SF than did single fibrosis tests (Table 2). Combined fibrosis indexes also increased the reliability of fibrosis diagnosis: they provided the highest rates of patients included in the intervals of $\geq 90\%$ predictive values (see **Supplementary Table S2** online).

RDI of combined fibrosis indexes had similar diagnostic accuracy than those of single fibrosis tests (see **Supplementary Results** online). There were thus two options: to associate RDIs of combined fibrosis indexes or, more simply, to associate RDIs of single fibrosis tests. The association of single fibrosis tests RDIs provided discordant results and a more detailed but less accurate fibrosis stage classification than the association of combined fibrosis indexes RDIs. Thus, combined fibrosis indexes improved diagnostic accuracy and erase discrepancy between single fibrosis tests.

Combined fibrosis indexes provided more robust fibrosis stage classifications. Indeed, despite the significantly different prevalence of histological fibrosis stages between the exploratory and validation sets, diagnostic accuracies of CSF/SF and CSF/C classifica-

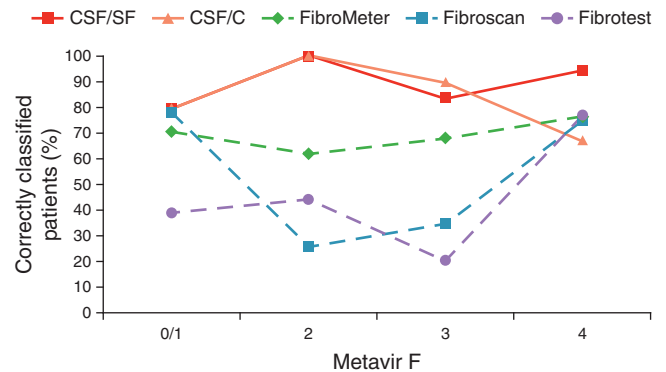


Figure 2. Rates of correctly classified patients by fibrosis stage classifications as a function of Metavir fibrosis stages in the validation set. Hatched lines: single fibrosis tests, continuous lines: new fibrosis stage classifications derived from new combined fibrosis indexes. Because of the few number of F0 patients, F0 and F1 were pooled together.

tions were not significantly different between the two patient sets (Table 4). By contrast, the fibrosis stage classification derived from the association of the single fibrosis tests RDIs showed a dramatic decrease in diagnostic accuracy, from 83% in the exploratory set to 69% in the validation set.

Finally, the new CSF/SF and CSF/C classifications provided significantly higher diagnostic accuracies than the fibrosis stage classifications of single fibrosis tests (Table 4).

Best-performing new fibrosis stage classification

CSF/SF classification provided significantly higher diagnostic accuracy (85.8%) than did the other fibrosis stage classifications (Table 4). This high diagnostic accuracy was very close to what would be expected for a perfect diagnostic test as evaluated against the imperfect gold standard that is liver biopsy (25). Moreover, the rate of correctly classified cirrhotic patients was significantly higher with CSF/SF classification: 95% vs. 67% with CSF/C classification. Overall, these results suggest that CSF/SF classification is more relevant for use in clinical practice than CSF/C or single fibrosis tests classifications.

How does it run in clinical practice?

Determination of CSF/SF classification. The use of CSF/SF classification in clinical practice may seem complex because it requires several calculations. However, after computerization of all steps, the user has only to provide two results (Fibroscan and blood test) to obtain an accurate estimation of the histological fibrosis stage (Figure 3).

CSF/SF classification vs. previously published algorithms. Several combinations of fibrosis tests such as SAFE (26) or Bordeaux algorithm (15) have been proposed to improve the non-invasive diagnosis of liver fibrosis in clinical practice. SAFE is accurate (27,28) but has several limits. Indeed, SAFE requires a sequential use of blood fibrosis tests and thus needs to repeat blood samples in patients. Moreover, SAFE includes two algorithms, each one

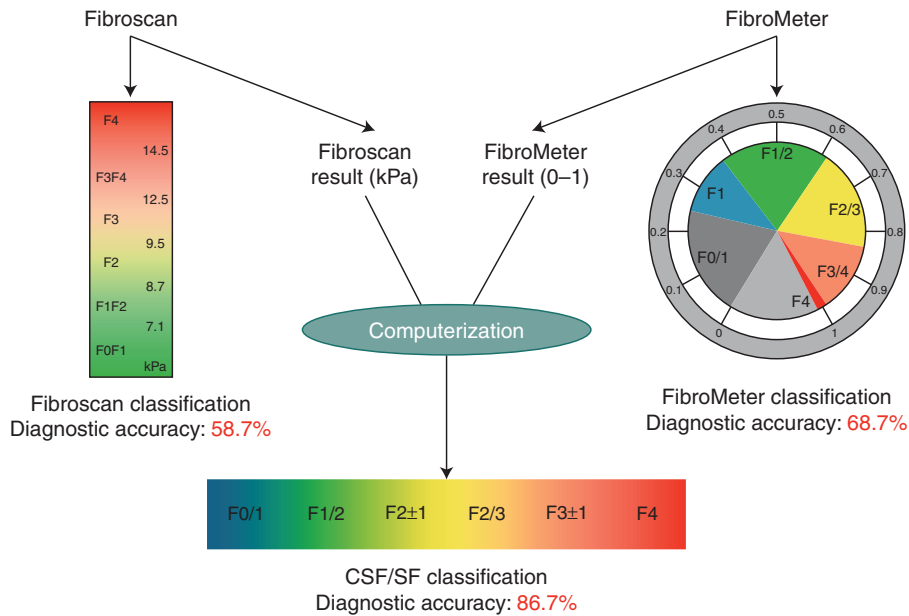


Figure 3. Non-invasive determination of the liver fibrosis stage in clinical practice. Fibroscan or FibroMeter could be used alone: results are interpreted according to their respective fibrosis stage classifications that provide fair/moderate diagnostic accuracy. Otherwise, Fibroscan and FibroMeter results could be introduced in a computerized algorithm to determine the clinically significant fibrosis/severe fibrosis (CSF/SF) classification that provides a high diagnostic accuracy.

providing only a binary diagnosis of significant fibrosis or cirrhosis. Thus, physician must first complete the SAFE for significant fibrosis and then, in case of $F \geq 2$ diagnosis, the SAFE for cirrhosis. This finally induces a decrease in the diagnostic accuracy by accumulation of each algorithm errors. Finally, liver fibrosis diagnosis remains undetermined in a large proportion of patients ($>50\%$ for the diagnosis of significant fibrosis and 20–25% for cirrhosis), with thus a high rate of liver biopsy required (27,28). The six-class CSF/SF classification circumvents all these limits by providing an accurate estimation of the fibrosis stage in a one-step procedure, without any liver biopsy required.

Castera *et al.* (15) have proposed the Bordeaux algorithm combining blood test (Fibrotest) and Fibroscan. However, because of disagreement between fibrosis tests, it requires liver biopsy in 30% of cases for the diagnosis of significant fibrosis and 20% for cirrhosis (27). The synchronous combination of blood test with Fibroscan into new *combined fibrosis indexes* allows CSF/SF classification to circumvent this limit.

Could CSF/SF classification replace liver biopsy in clinical practice? Liver biopsy is currently the reference procedure for liver fibrosis evaluation in clinical practice (1). However, it is now well established that liver biopsy cannot be considered a “gold standard” because of sampling bias and poor interobserver reproducibility of histological semi-quantitative scores (29). Thus, when a liver biopsy and a non-invasive test give discordant results, it is difficult to know which one is right. Several studies have evaluated discordances between liver biopsy and either blood fibrosis tests (30–32) or Fibroscan results (33). In these studies, discordances attributed to misdiagnoses of non-invasive tests were as frequent

as those attributed to liver biopsy errors, suggesting that non-invasive tests were as accurate as liver biopsy for liver fibrosis evaluation in clinical practice. In our study, liver biopsy and CSF/SF classification results were discordant in only 13.3% of cases and we may suppose that a significant part of these discordances could be attributed to liver biopsy errors. The subsequent decrease in diagnostic accuracy of liver biopsy (i.e., $<100\%$) and increase in accuracy of CSF/SF classification (i.e., $>86.7\%$) suggest that both methods most likely provide similar accuracy for the evaluation of liver fibrosis in clinical practice. Other studies with different methodology (e.g., surgical biopsy) will address that issue more completely. Nevertheless, the CSF/SF classification has the advantage of being totally non-invasive; it can thus be repeated at regular intervals, unlike liver biopsy. Finally, although liver biopsy remains the best standard (29), non-invasive tests have other indications, for example, a closer follow-up of patients. In this setting, cirrhotic patients initially misclassified by CSF/SF classification due to compensated cirrhosis will most likely be correctly classified later thanks to a slight progression of liver disease.

Beyond diagnostic accuracy, another crucial point is to compare the cost-effectiveness of liver biopsy and CSF/SF classification for liver fibrosis evaluation. Such a comparison would need to take into account not only numerous cost factors (liver biopsy, blood markers, Fibroscan, resulting antiviral prescriptions, treatments for initially misdiagnosed cirrhosis complications), but also morbidity and mortality due to liver biopsy and liver disease decompensation. The question of cost-effectiveness thus requires further longitudinal studies using clinical events as study end points.

Finally, we have previously shown that blood fibrosis tests adapted to the cause of liver disease provide higher diagnostic accuracy (34).

Continuing in this direction, we have developed the CSF/SF classification in a homogeneous population of patients with a single cause of chronic liver disease, that is, CHC. Future studies may evaluate if combinations of non-invasive tests according to our methodology could replace liver biopsy in other causes of liver disease.

In conclusion, the association of *combined fibrosis indexes* including FibroMeter and Fibroscan provides a new *fibrosis stage classification* for the non-invasive diagnosis of liver fibrosis stages in patients with CHC. This classification has a high and robust diagnostic accuracy, and is significantly more accurate than the previously published *fibrosis stage classifications* of Fibrotest, FibroMeter, and Fibroscan. It now seems possible to implement fully non-invasive management of patients with CHC.

ACKNOWLEDGMENTS

We thank other investigators from the Fibrostar study: *Hepatologists*: R. Poupon, A. Poulou, Saint-Antoine, Paris; A. Abergel, Clermont-Ferrand; J.P. Bronowicki, Nancy; J.P. Vinel, S. Metivier, Toulouse; V. De Ledinghen, Bordeaux; O. Gorla, Rouen; M. Maynard-Muet, C. Trepo, Lyon; Ph. Mathurin, Lille; D. Guyader, H. Danielou, Rennes; O. Rogeaux, Chambéry; S. Pol, Ph. Sogni, Cochin, Paris; A. Tran, Nice; P. Calès, Angers; P. Marcellin, T. Asselah, Clichy; M. Bourliere, V. Oulès, Saint Joseph, Marseille; D. Larrey, Montpellier; F. Habersetzer, Strasbourg; M. Beaugrand, Bondy; V. Leroy, M.N. Hilleret, Grenoble. *Biologists*: R.-C. Boisson, Lyon Sud; M.-C. Gelineau, B. Poggi, Hôtel Dieu, Lyon; J.-C. Renversez, Candice Trocmé, Grenoble; J. Guéchet, R. Lasnier, M. Vaubourdoles, Paris; H. Voitot, Beaujon, Paris; A. Vassault, Necker, Paris; A. Rosenthal-Allieri, Nice; A. Lavoinne, F. Ziegler, Rouen; M. Bartoli, C. Lebrun, Chambéry; A. Myara, Paris Saint-Joseph; F. Guerber, A. Pottier, Elibio, Vizille. *Pathologists*: E.-S. Zafrani, Créteil; N. Sturm, Grenoble. *Methodologists*: A. Bechet, J.-L. Bosson, A. Paris, S. Royannais, CIC, Grenoble; A. Plages, Grenoble. We also thank the following contributors: Juliette Foucher, Laurent Castera, Pascal Veillon, Gwénaëlle Soulard; and Kevin L. Erwin (for English proofreading).

CONFLICT OF INTEREST

Guarantor of the article: Paul Calès, MD.

Specific author contributions: Planning and conducting the study, collecting and interpreting data, drafting the manuscript, and approving the final draft submitted: Jérôme Boursier; planning and conducting the study, collecting data, drafting the manuscript, and approving the final draft submitted: Victor de Ledinghen and Jean-Pierre Zarski; collecting data: Marie-Christine Rousselet, Nathalie Sturm, Juliette Foucher, Isabelle Fouchard-Hubert, Yves Gallois, Frédéric Oberti, and Nina Dib; planning and conducting the study and collecting data: Vincent Leroy; collecting and interpreting data: Sandrine Bertrais; planning and conducting the study, collecting and interpreting data, drafting the manuscript, approving the final draft submitted, accepts full responsibility for the conduct of the study, has access to the data, and has control of the decision to publish: Paul Calès.

Financial support: This study was supported by ANRS (French national agency for research on AIDS and viral hepatitis) for HC EP 23 Fibrostar study.

Potential competing interests: Paul Calès, Isabelle Fouchard Hubert, and Frédéric Oberti have stock ownership in BioLiveScale. BioLiveScale has a license for FibroMeter from Angers University.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Precise evaluation of the level of liver fibrosis is recommended in patients with chronic hepatitis C.
- ✓ Liver fibrosis stages are usually determined on a biopsy by using histological semi-quantitative scores of fibrosis.
- ✓ Fibrosis stages classifications derived from blood fibrosis tests or Fibroscan results allow for a non-invasive determination of liver fibrosis stages.

WHAT IS NEW HERE

- ✓ The association of new fibrosis indexes combining FibroMeter and Fibroscan provides a new fibrosis stage classification.
- ✓ This classification is highly accurate for the non-invasive diagnosis of histological fibrosis stages in chronic hepatitis C.
- ✓ The new classification has a significantly higher diagnostic accuracy than those previously published for blood fibrosis tests or Fibroscan.

REFERENCES

1. Ghany MC, Strader DB, Thomas DL *et al.* American Association for the Study of Liver Diseases. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009;49:1335–74.
2. Everhart JE, Wright EC, Goodman ZD *et al.* Prognostic value of Ishak fibrosis stage: findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. *Hepatology* 2010;51:585–94.
3. Rousselet MC, Michalak S, Dupre F *et al.* Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005;41:257–64.
4. Cales P, Boursier J, Oberti F *et al.* FibroMeters: a family of blood tests for liver fibrosis. *Gastroenterol Clin Biol* 2008;32:40–51.
5. Castera L, Fornis X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008;48:835–47.
6. Halfon P, Munteanu M, Poynard T. FibroTest-ActiTest as a non-invasive marker of liver fibrosis. *Gastroenterol Clin Biol* 2008;32:22–39.
7. Leroy V. Other non-invasive markers of liver fibrosis. *Gastroenterol Clin Biol* 2008;32:52–7.
8. Calès P, Boursier J, Bertrais S *et al.* Accuracy of liver fibrosis classifications provided by non-invasive tests. *J Hepatol* 2010;52 (Suppl 1): S406.
9. Boursier J, Bacq Y, Halfon P *et al.* Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009;21:28–38.
10. Cales P, de Ledinghen V, Halfon P *et al.* Evaluating the accuracy and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C. *Liver Int* 2008;28:1352–62.
11. Cales P, Boursier J, de Ledinghen V *et al.* Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests. *Gastroenterol Clin Biol* 2008;32:1050–60.
12. Boursier J, Vergniol J, Sawadogo A *et al.* The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. *Liver Int* 2009;29:1507–15.
13. Zarski JP, Sturm N, Guechet J *et al.* Prospective independent validation and comparison of 10 biomarkers and transient elastography (Fibroscan™) for liver fibrosis assessment in patients with chronic hepatitis C. *J Hepatol* 2010;52:S175.
14. Noursbaum JB, Cadranet JF, Bonnemaïson G *et al.* Clinical practice guidelines on the use of liver biopsy. *Gastroenterol Clin Biol* 2002;26:848–78.
15. Castera L, Vergniol J, Foucher J *et al.* Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343–50.

16. Leroy V, Halfon P, Bacq Y *et al.* Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008;41:1368–76.
17. Adams LA, Bulsara M, Rossi E *et al.* Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* 2005;51:1867–73.
18. Sterling RK, Lissen E, Clumeck N *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;43:1317–25.
19. Wai CT, Greenon JK, Fontana RJ *et al.* A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003;38:518–26.
20. Lucidarme D, Foucher J, Le Bail B *et al.* Factors of accuracy of transient elastography (fibrosan) for the diagnosis of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1083–9.
21. de Ledinghen V, Vergniol J. Transient elastography (FibroScan). *Gastroenterol Clin Biol* 2008;32:58–67.
22. Poynard T, Imbert-Bismut F, Munteanu M *et al.* Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004;3:8.
23. Thein HH, Yi Q, Dore GJ *et al.* Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression. *Hepatology* 2008;48:418–31.
24. Poynard T, Halfon P, Castera L *et al.* Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007;53:1615–22.
25. Mehta SH, Lau B, Afdhal NH *et al.* Exceeding the limits of liver histology markers. *J Hepatol* 2009;50:36–41.
26. Sebastiani G, Vario A, Guido M *et al.* Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol* 2006;44:686–93.
27. Castera L, Sebastiani G, Le Bail B *et al.* Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. *J Hepatol* 2010;52:191–8.
28. Sebastiani G, Halfon P, Castera L *et al.* SAFE biopsy: a validated method for large-scale staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1821–7.
29. Bedossa P, Carrat F. Liver biopsy: the best, not the gold standard. *J Hepatol* 2009;50:1–3.
30. Bourliere M, Penaranda G, Renou C *et al.* Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies. *J Viral Hepat* 2006;13:659–70.
31. Halfon P, Bourliere M, Deydier R *et al.* Independent prospective multicenter validation of biochemical markers (fibrotest-actitest) for the prediction of liver fibrosis and activity in patients with chronic hepatitis C: the fibropaca study. *Am J Gastroenterol* 2006; 101:547–55.
32. Poynard T, Munteanu M, Imbert-Bismut F *et al.* Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. *Clin Chem* 2004;50:1344–55.
33. Myers RP, Crotty P, Pomier-Layrargues G *et al.* Prevalence, risk factors and causes of discordances in fibrosis staging by transient elastography and liver biopsy. *Liver Int* 2010;30:1471–80.
34. Calès P, Oberti F, Michalak S *et al.* A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005;42:1373–81.

SUPPLEMENTARY MATERIAL

GLOSSARY

Single fibrosis test: Corresponds to already published blood fibrosis test obtained by a biomarker combination (Fibrotest, FibroMeter, Hepascore, APRI, FIB-4), or Fibroscan.

Fibrosis stage classification: Classification giving an estimation of the histological Metavir fibrosis stage, as derived from the fibrosis test result. These classifications are available for Fibrotest, FibroMeter, and Fibroscan (**Figure s1**).

Combined fibrosis index: New fibrosis test combining *single fibrosis tests*. 3 combined fibrosis indexes combining FibroMeter and Fibroscan results were implemented in this study:

- CSF-index for the diagnosis of clinically significant fibrosis, i.e. Metavir $F \geq 2$
- SF-index for the diagnosis of severe fibrosis, i.e. Metavir $F \geq 3$
- C-index for the diagnosis of cirrhosis

Reliable diagnosis interval (RDI): RDIs correspond to the intervals of fibrosis test values where the individual diagnostic accuracy is considered sufficiently reliable for clinical practice.

In practice, the range of the fibrosis test value is divided into several consecutive intervals. First, the thresholds of 90% predictive values for the diagnostic target are calculated (Figure A). They define two intervals of blood tests values:

- A lower interval, defined by a blood test value \leq the 90% negative predictive value threshold, where patients have a $\geq 90\%$ chance of not having the diagnostic target;
- And a higher interval, defined by a blood test value \geq the 90% positive predictive value threshold, where patients have a $\geq 90\%$ risk of having the diagnostic target.

Second, the remaining intermediate interval between the thresholds of the 90% predictive values is divided according to the fibrosis test value providing the maximum Youden index (Figure B). These two intermediate intervals correspond to a class of fibrosis stages different from the initial diagnostic target but with a combined prevalence of $\geq 90\%$. Thus, the interest

of this recently described method (10) is to offer a reliable diagnosis (high predictive values) in 100% of patients without the need for liver biopsy.

Figure A: Intervals of $\geq 90\%$ negative (NPV) and positive (PPV) predictive values for the diagnosis of significant fibrosis (Metavir $F \geq 2$). NPV for significant fibrosis is $\geq 90\%$ in patients with a score ≤ 0.235 ; PPV is $\geq 90\%$ in patients with a score ≥ 0.636 (these figures are examples).

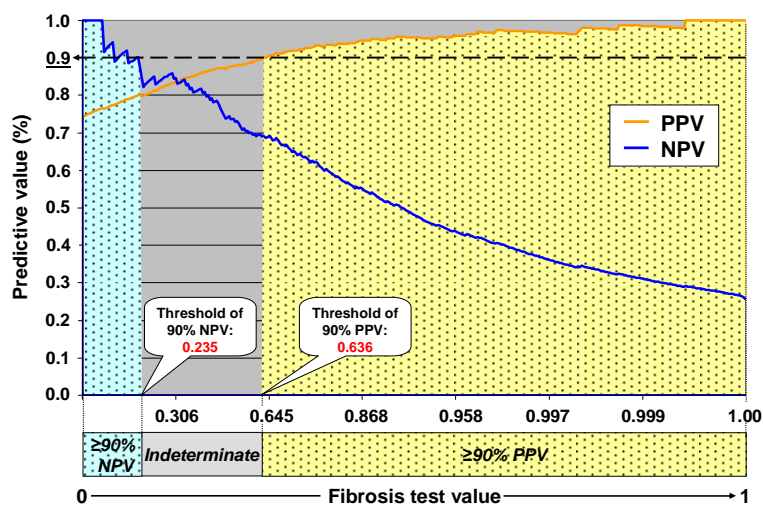
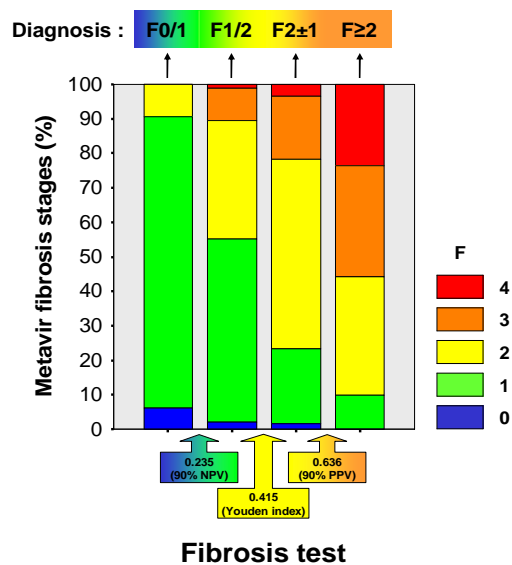


Figure B: Intervals of reliable diagnosis. Diagnostic accuracy is $\geq 90\%$ without performing liver biopsy.



SUPPLEMENTARY RESULTS

Development of new fibrosis stage classifications

1st step: New combined fibrosis indexes

For each diagnostic target, Fibroscan and FibroMeter were the *single fibrosis tests* most frequently selected by the stepwise binary logistic regression repeated on the 1000 bootstrap samples (**Table s1**). These 2 fibrosis tests were independent variables in logistic models ran in the exploratory set and thus provided 3 new *combined fibrosis indexes* for different diagnostic targets: CSF-index for significant fibrosis, SF-index for severe fibrosis, and C-index for cirrhosis.

CSF-index had a significantly higher AUROC than its composite tests, i.e. FibroMeter or Fibroscan, in the exploratory set (**Table 2**). SF-index and C-index also had higher AUROCs than FibroMeter or Fibroscan in the exploratory set, but the difference was significant only with FibroMeter.

2nd step: Reliable diagnosis intervals

CSF-index (diagnostic target: significant fibrosis) – CSF-index was divided into 4 reliable diagnosis intervals. The extreme intervals were the traditional intervals of $\geq 90\%$ negative (NPV) or positive (PPV) predictive values for significant fibrosis. CSF-index included 9.2% of patients in the $\geq 90\%$ NPV interval (CSF-index value ≥ 0 and ≤ 0.248) and 46.1% in the $\geq 90\%$ PPV interval (CSF-index value ≥ 0.784 and ≤ 1). Thus, CSF-index displayed a reliable diagnosis of significant fibrosis with $\geq 90\%$ accuracy in 55.3% of patients versus 33.8% with Fibroscan ($p < 0.001$) and 55.6% with FibroMeter ($p = 1.00$, **Table s2**).

The indeterminate interval (between CSF-index values > 0.248 and < 0.784) was then divided into 2 new intervals according to the diagnostic cut-off corresponding to the maximum Youden index (0.615). 87.5% of the patients included in the lower new interval (> 0.248 - < 0.615) had F1/2 stages according to liver biopsy results, and 95.0% of patients included in the higher new interval (≥ 0.615 and < 0.784) had F1/2/3 stages (**Figure s3a**). Finally, CSF-index provided 4 RDIs whose F classification was: F0/1, F1/2, F2 \pm 1, and F \geq 2. The diagnostic accuracy of these RDIs was 90.3% (**Figure s3a**).

FibroMeter provided the same 4 RDIs with 89.4% diagnostic accuracy ($p = 0.664$ vs CSF-index).

SF-index (diagnostic target: severe fibrosis) – SF-index was also divided into 4 RDIs. The extreme intervals were the traditional intervals of $\geq 90\%$ negative or positive predictive values for severe fibrosis. SF-index included 44.7% of patients in the $\geq 90\%$ NPV interval (SF-index value ≥ 0 and ≤ 0.220) and 5.2% in the $\geq 90\%$ PPV interval (SF-index value ≥ 0.870 and ≤ 1). Thus, SF-index displayed a reliable diagnosis of significant fibrosis with $\geq 90\%$ accuracy in 49.9% of patients (**Table s2**) versus 41.8% with FibroMeter ($p < 0.001$) and 46.4% with Fibroscan ($p = 0.235$). By dividing the indeterminate interval of SF-index according to the diagnostic cut-off (maximum Youden index: 0.364), SF-index provided 4 RDI ($F1 \pm 1$, $F2 \pm 1$, $F3 \pm 1$, $F \geq 3$; **Figure s3b**) with 92.0% diagnostic accuracy.

Fibroscan provided the same 4 RDIs with 91.1% diagnostic accuracy ($p = 0.728$ vs SF-index).

C-index (diagnostic target: cirrhosis) – C-index included 87.7% of patients in the $\geq 95\%$ NPV interval for cirrhosis (C-index value ≥ 0 and ≤ 0.244), and 2.0% in the $\geq 95\%$ PPV interval for cirrhosis (C-index value ≥ 0.896 and ≤ 1). Thus, C-index displayed a reliable diagnosis of cirrhosis with $\geq 95\%$ accuracy in 89.7% of patients (**Table s2**) versus 65.9% with FibroMeter ($p < 0.001$) and 87.4% with Fibroscan ($p = 0.096$). Dividing the indeterminate interval according to the diagnostic cut-off did not distinguish two different groups. Finally, C-index provided 3 RDIs ($F \leq 3$, $F3 \pm 1$, $F4$) with 95.1% diagnostic accuracy (**Figure s3c**).

3rd step: New fibrosis stage classifications

The first classification (CSF/SF classification) was derived from the association of CSF- and SF-index RDIs (**Table 3**). CSF/SF classification included 6 classes: $F0/1$, $F1/2$, $F2 \pm 1$, $F2/3$, $F3 \pm 1$, and $F4$. It provided 87.7% diagnostic accuracy in the exploratory set (**Figure 1a** and **Table 4**).

The second classification (CSF/C classification) was derived from CSF- and C-index RDIs (**Table 3**). CSF/C classification also included 6 classes ($F0/1$, $F1/2$, $F2 \pm 1$, $F2/3$, $F3 \pm 1$, $F4$) and provided 86.5% diagnostic accuracy ($p = 0.503$ vs CSF/SF classification, **Table 4**).

Association of combined fibrosis indexes RDIs or single fibrosis tests RDIs?

As previously shown, the accuracies of RDIs from *combined fibrosis indexes* and their composite *single fibrosis tests* were not significantly different (i.e., the FibroMeter RDIs for significant fibrosis vs that of CSF-index, and the Fibroscan RDIs for severe fibrosis vs that of SF-index). Therefore, we implemented a third classification (FM/FS classification) that was derived from the FibroMeter RDIs for significant fibrosis and the Fibroscan RDIs for severe

fibrosis. Results of FibroMeter and Fibroscan RDIs were discordant in 2 patients, who thus had indeterminate diagnoses. FM/FS classification ultimately included 7 classes (F0/1, F1, F1/2, F2, F2/3, F3±1, F4) and provided 82.8% diagnostic accuracy ($p=0.006$ vs CSF/SF classification). However, diagnostic accuracy of FM/FS classification dramatically decreased to 69.4% in the validation set ($p<0.001$ vs CSF/SF and CSF/C classifications).

Table s1: Selection of candidate predictors at bootstrapped stepwise binary logistic regressions, as a function of diagnostic target in the exploratory set. Stepwise binary logistic regressions were performed on 1000 bootstrap samples of 349 subjects from the exploratory set. The table depicts the number of times any *single fibrosis test* was selected across the 1000 multivariate analyses. For each diagnostic target, Fibroscan and FibroMeter were the most frequently selected variables.

Fibrosis tests	Significant fibrosis (Metavir F\geq2)	Severe fibrosis (Metavir F\geq3)	Cirrhosis (Metavir F4)
FibroMeter	944	925	779
Fibrotest	120	152	108
Hepascore	187	77	171
FIB-4	57	78	84
APRI	346	515	66
Fibroscan	961	1000	999

Table s2: Rate of patients included in the intervals of reliable diagnosis defined by the $\geq 90\%$ negative (NPV) and positive (PPV) predictive values for significant fibrosis (Metavir $F \geq 2$) or severe fibrosis (Metavir $F \geq 3$), and the $\geq 95\%$ predictive values for cirrhosis (Metavir F4), as a function of diagnostic target and fibrosis test, and according to patient group.

Set	Fibrosis test	Metavir $F \geq 2$		Metavir $F \geq 3$		Metavir F4	
		Patients ^a	Correctly classified ^b	Patients ^a	Correctly classified ^b	Patients ^a	Correctly classified ^b
Exploratory	FibroMeter	55.6	89.7	41.8	89.7	65.9	94.8
	Fibroscan	33.8	90.7	46.4	90.1	87.4	94.8
	Combined index ^c	55.3	90.2	49.9	89.7	89.7	94.9
Validation	FibroMeter	48.8	72.7	47.0	94.2	64.2	97.2
	Fibroscan	38.2	77.0	46.7	93.5	85.2	93.2
	Combined index ^c	49.1	85.2	58.5	95.9	87.3	93.8
All	FibroMeter	52.3	82.0	44.3	92.0	65.1	95.9
	Fibroscan	35.9	83.6	46.5	91.8	86.3	94.0
	Combined index ^c	52.3	87.9	54.1	92.9	88.5	94.3

^a Rate of patients (%) included in both intervals of $\geq 90/95\%$ negative and $\geq 90/95\%$ positive predictive values; ^b Rate of correctly classified patients (%) among those included in the NPV+PPV intervals; ^c CSF-index for significant fibrosis, SF-index for severe fibrosis, and C-index for cirrhosis

Cut-offs for NPV $\geq 90/95\%$ and PPV $\geq 90/95\%$ were calculated in the exploratory set and tested in the validation set and the whole population. **Significant fibrosis.** Cut-offs for NPV $\geq 90\%$: FibroMeter: ≤ 0.110 , Fibroscan: ≤ 3.2 , CSF-index: ≤ 0.248 ; cut-offs for PPV $\geq 90\%$: FibroMeter: ≥ 0.608 , Fibroscan: ≥ 9.2 , CSF-index: ≥ 0.784 . **Severe fibrosis.** Cut-offs for NPV $\geq 90\%$: FibroMeter: ≤ 0.554 , Fibroscan: ≤ 6.8 , SF-index: ≤ 0.220 ; cut-offs for PPV $\geq 90\%$: Fibroscan: ≥ 32.3 , SF-index: ≥ 0.870 . **Cirrhosis.** Cut-offs for NPV $\geq 95\%$: FibroMeter: ≤ 0.757 , Fibroscan: ≤ 14.5 , C-index: ≤ 0.244 ; cut-offs for PPV $\geq 95\%$: Fibroscan: ≥ 35.6 , C-index: ≥ 0.896 .

Figure s1: Various fibrosis stage classifications. **A:** Histological Metavir fibrosis stages. **B-E:** Fibrosis stage classifications by non invasive tests; **B:** Fibrotest classification (22); **C:** FibroMeter classification (16); **D:** Fibroscan classification (21); **E:** new CSF/SF classification derived from the association of new fibrosis indexes combining FibroMeter and Fibroscan.

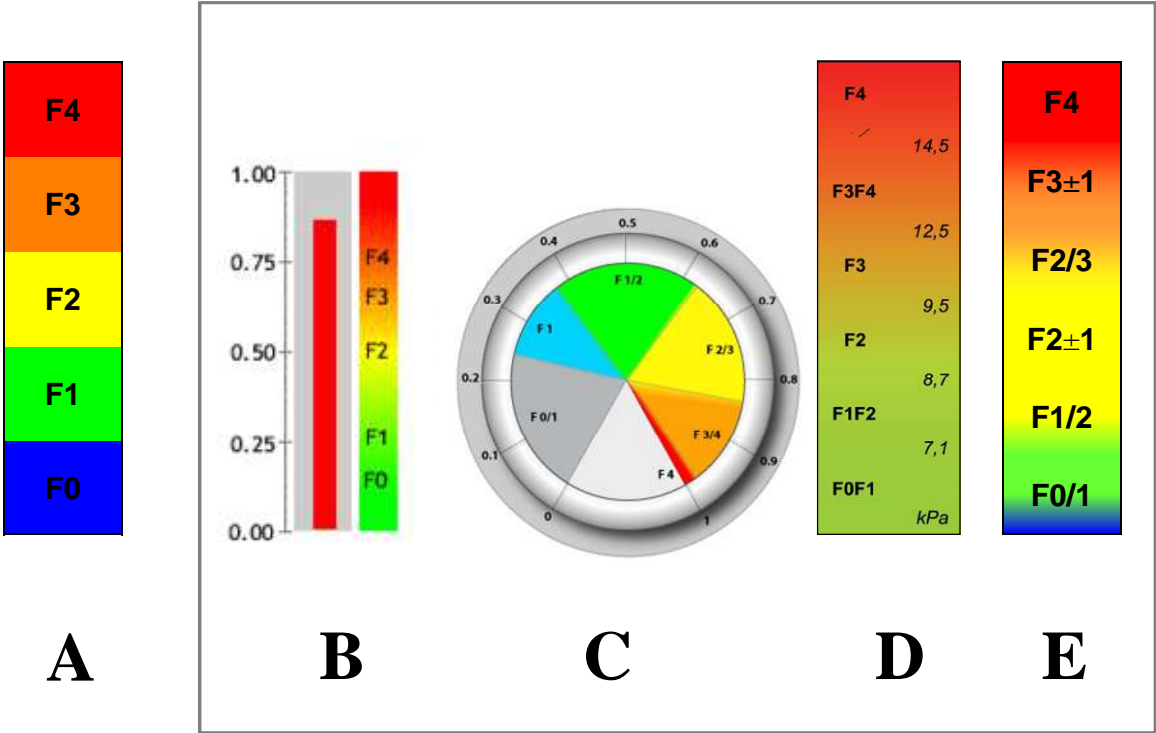


Figure s2: Study methodology. Implementation of new *fibrosis stage classifications* from new *combined fibrosis indexes* (exploratory set). BLR: binary logistic regression, RDI: reliable diagnosis intervals.

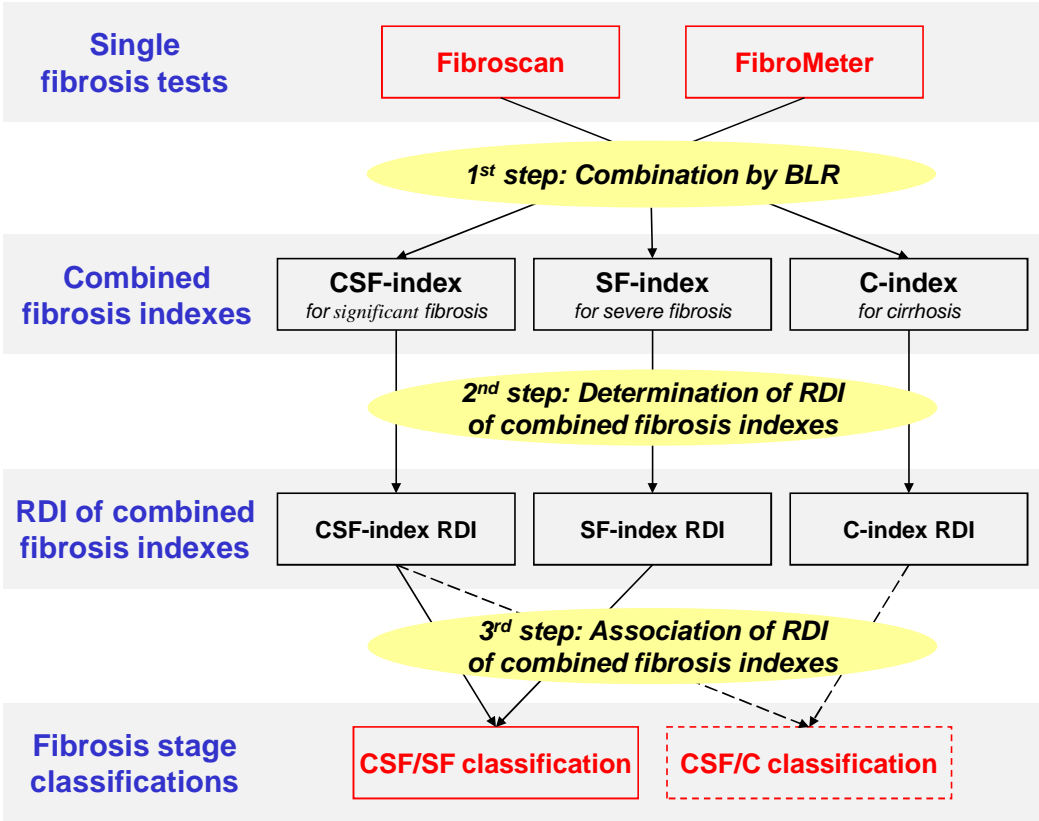


Figure s3: Reliable diagnosis intervals of CSF-, SF- and C-indexes in the exploratory set.

Panel s3a: Proportion of Metavir fibrosis (F) stages according to the maximum Youden index cut-off and the thresholds of 90% negative and positive predictive values for significant fibrosis with CSF-index. **Panel s3b:** Proportion of Metavir F stages according to the maximum Youden index cut-off and the thresholds of 90% negative and positive predictive values for severe fibrosis with SF-index. **Panel s3c:** Proportion of Metavir F stages according to the thresholds of 95% predictive values for cirrhosis with C-index.

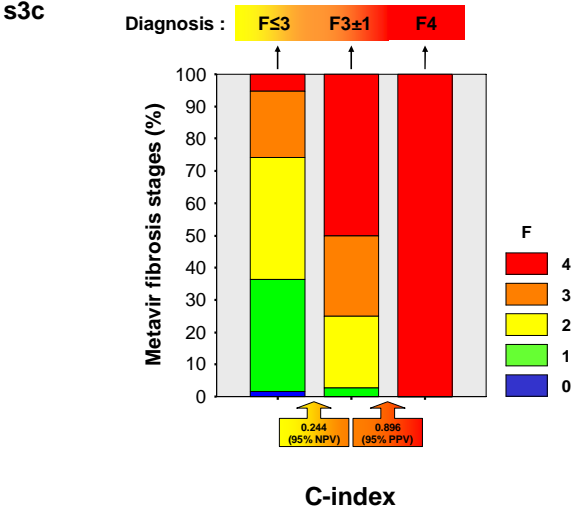
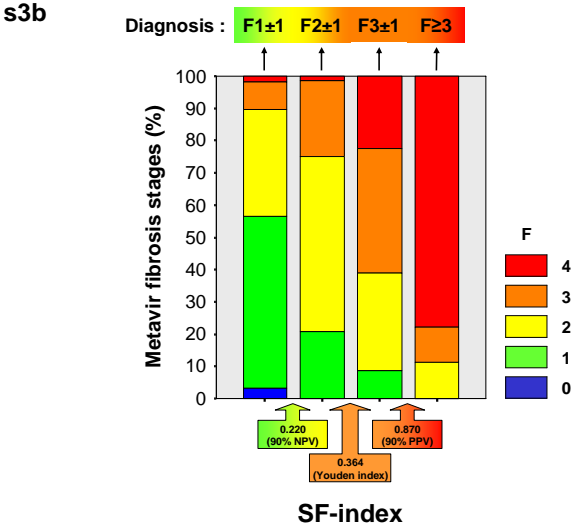
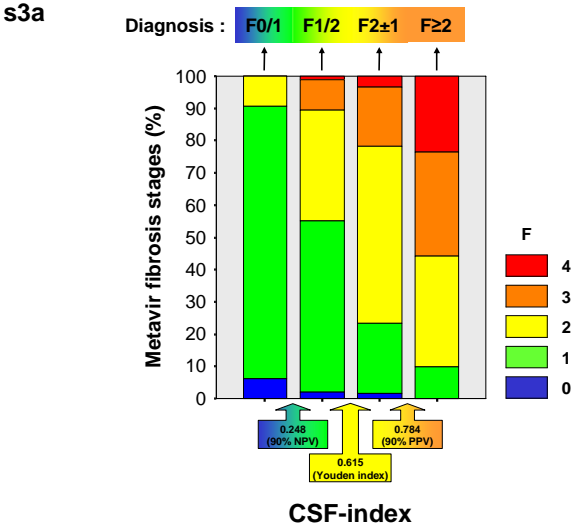
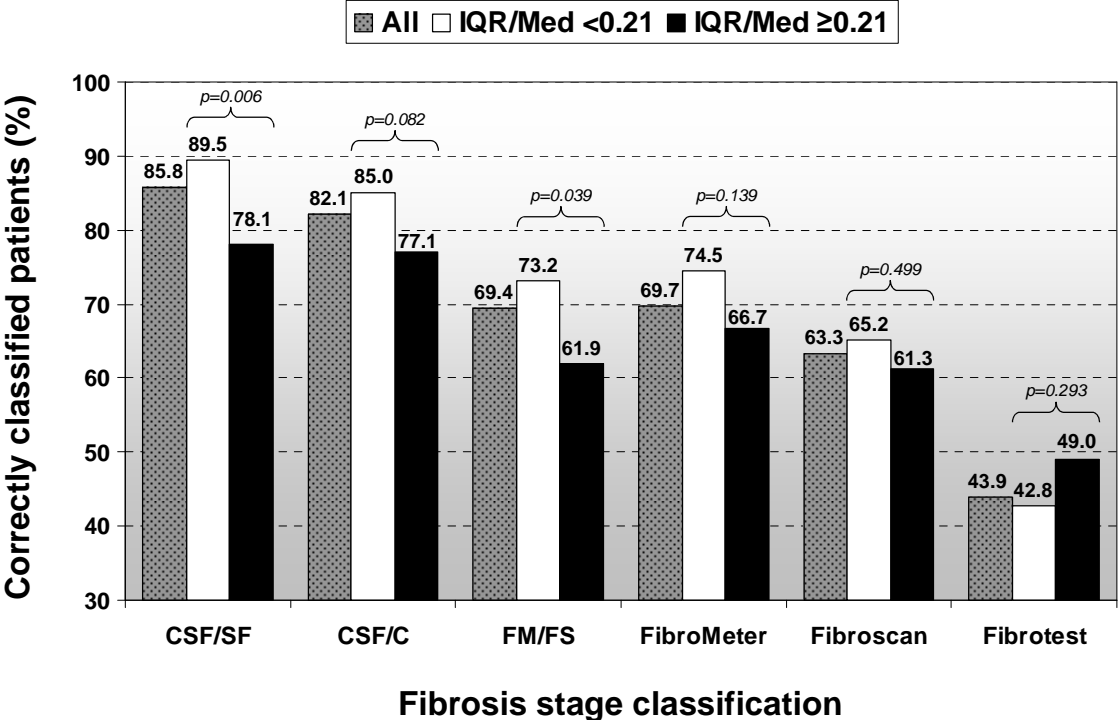


Figure s4: Rate of correctly classified patients by *fibrosis stage classifications* as a function of IQR/median ratio in the validation set.



Annexe 10 : Avantages (cases blanches) et inconvénients (cases grises) des recommandations de la Haute Autorité de Santé française, et des combinaisons de tests non invasifs de fibrose pour le diagnostic non invasif de fibrose hépatique au cours de l'hépatite chronique C.

		Recommandations	Combinaisons de tests non invasifs de fibrose		
		HAS (déc 2008)	SAFE	Algorithme de Bordeaux	Classification CSF/SF
Tests non invasifs	1 ^{ère} intention	FT ou FM ou HS ou FS	APRI	FT et FS	1 ^{ère} intention : FM et FS
	2 ^e intention	Un autre test ou la PBH	FT	PBH	-
	3 ^e intention	-	PBH	-	-
Etapes		2	3	2	1
Seuils diagnostiques		Valeurs non précisées	Valeurs précisées	Valeurs précisées	Non nécessaire
Diagnostic		Non précisé	2 classes (F0/1 ou F \geq 2 ; F \leq 3 ou F4)	3 classes (F0/1, F2/3, F4)	6 classes (F0/1, F1/2, F2 \pm 1, F2/3, F3 \pm 1, F4)
Performance diagnostique		Compteurs de fibrose : performance médiocre	Excellente	Excellente	Excellente
Taux de PBH		Inconnu	>50%	20-30%	0%

MÉTHODOLOGIE POUR L'AMÉLIORATION DE LA PERFORMANCE DU DIAGNOSTIC NON INVASIF DE LA FIBROSE HÉPATIQUE AU COURS DE L'HÉPATITE CHRONIQUE C

Le pronostic et la prise en charge des patients avec une hépatite C dépend du degré de fibrose hépatique qui est classiquement évalué par l'examen histologique d'une ponction-biopsie hépatique. Cependant, la biopsie est une procédure invasive et l'examen histologique est limité par le biais d'échantillonnage et une reproductibilité perfectible. Récemment, des tests permettant d'évaluer de façon non invasive la fibrose du foie ont été développés : les tests sanguins de fibrose et l'élastométrie impulsionnelle (Fibroscan). Depuis décembre 2008, la Haute Autorité de Santé française recommande ces tests en première intention pour l'évaluation de ma fibrose/cirrhose chez les patients avec une hépatite chronique C. Néanmoins, ces recommandations ne précisent pas les modalités pratiques d'utilisation de ces tests. Par ailleurs, la performance des tests non-invasifs de fibrose a surtout été évaluée de façon globale et les quelques travaux ayant évalué leur utilisation en pratique clinique ont montré des performances diagnostiques moyennes. Ce travail de thèse avait pour but de développer une méthodologie pour améliorer la performance du diagnostique non invasif de la fibrose hépatique au cours de l'hépatite C. Nous avons mis au point plusieurs concepts méthodologiques : tests adaptés à la cible diagnostique, combinaisons synchrones de tests, association des intervalles de diagnostic fiable. Ces nouvelles méthodes ont permis d'élaborer une nouvelle classification non invasive de fibrose, très simple d'utilisation et très significativement plus performante que celles utilisées jusqu'alors en pratique clinique.

Methodology to improve the accuracy of the non-invasive diagnosis of liver fibrosis in chronic hepatitis C

Prognosis and management of patients with chronic hepatitis C depends on the degree of liver fibrosis which is classically assessed by histological examination of a liver biopsy. However, liver biopsy is an invasive procedure and histological examination is impaired by sampling bias and fair reproductibility. Recently, new tests allow for a non-invasive evaluation of liver fibrosis : fibrosis blood tests and transient elastography (Fibroscan). Since December 2008, the French Health Authority has recommended these tests as first-line exams for the assessment of fibrosis/cirrhosis in patients with chronic hepatitis C. However, these recommendations do not specify the practical use of these tests. In addition, the accuracy of non-invasive fibrosis tests has mainly been globally evaluated and the implement a methodology to improve the accuracy of the non-invasive diagnosis of liver fibrosis in hepatitis C. We developed several methodological concepts : fibrosis tests adapted to the diagnostic target, synchronous combinations of tests, association of intervals of reliable diagnosis. These new concepts were finally used to develop a new non-invasive classification of liver fibrosis, easy to use and very significantly more accurate than those actually used in clinical practice.